Comparing sensitivity of Radial Basis Function method with Multilayer Perceptron Network and Cox Proportional Hazard Model in Survival Data

Mohammad Salehi Veisi[1], Sadegh Rezaei[2*] and Khadejeh Salehivaysi[3]
[1]Department of Statistics, Faculty of Mathematics and Computer Science
Amirkabir University of Technology (Tehran Polytechnic), Tehran, IRAN
[2]Department of Statistics, Faculty of Mathematics and Computer Science
Amirkabir University of Technology (Tehran Polytechnic), Tehran, IRAN
[3]Department of Computer, Mahshahr branch, Islamic Azad University, Mahshahr, Iran

Summary
Cox proportional hazard model is broadly deployed in the analysis of survival data and reliability of systems and its application is contingent upon accepting some assumptions such as appropriateness of the risk. Neural network model, obviating the need of making any specific assumption, is an appropriate substitute in predicting survival data. To compare the sensitivity of neural network models with the Cox proportional hazard model, the present study investigates the sensitivity and specificity of radial basis function method, multilayer perceptron network, and cox proportional hazard model in the survival analysis of patients with myocardial infarction. The results of the study revealed that, compared to other models, neural network models performed better and were more precise in the survival analysis of patients with myocardial infarction. Moreover, compared to the other two methods, the radial basis function method is more sensitive, specific, and precise in the survival prediction of the patients under investigation in the present study and, accordingly, it is more reliable.
Key-words:
Radial Basis Function, network, Cox Proportional Hazard, model efficacy, survival analysis

1. Introduction

Most of the statistical models are deployed with the aim of determining the relationship between variables, selecting the effective variables, estimating, and predicting (19, 23, 28, 29, and 32). These models are exploited in reliability discussions and survival analysis, which are widely used in industry and healthcare systems. Some of the studies involve data whose measuring is dependent on factors which cause risk for the patients under investigation. However, the recovery condition of these patients after undergoing the prescribed treatments and their survival probability has attracted researchers' attention. Selecting an appropriate model for analyzing such data and survival prediction of these patients is subject to some conditions which are recognized as the assumptions of the model. However, if the model's assumptions are not satisfied, it cannot be applied (23, 28, and 32) in that it would lead to incorrect estimation and unwarranted predictions. Accordingly, models with less error should be opted for data analysis. One of the solutions adopted to circumvent such problems is deploying artificial neural network models which have recently received unprecedented attention (44). Despite the non-interpretability of artificial neural network models, compared to other similar statistical models, they are widely used in different fields of medical sciences, including prediction and disease detection systems (3, 31), cancer (1, 8, 12, 24, and 27), medical image analysis, survival prediction, etc. The wide use of these methods can be attributed to the fact that these models are not dependent on the distribution of variables and do not assume any assumption or postulate.

Diseases resulting from heart failure are among common diseases in Iran and in the world (22, 41). A bulk of research, drawing on classic methods such as cox regression, logistic regression and parametric proportional hazards regression, have been conducted to investigate the effect of age, gender, blood pressure, cholesterol level, etc. on these diseases and the to examine survival analysis of patients suffering from these diseases. As the amount of information and the number of effective parameters in detecting diseases increase, the classic methods fail to take all of them into account and; accordingly, lose their efficiency. Hence, some of the studies conducted in this regard adopted the artificial neural network models and introduced them into heart disease-related studies. In this regard, we can refer to a paper entitled "comparing artificial neural networks with logistic regression in determining predictors of in-hospital morality after open heart surgery". In this study, an artificial neural network with 18 input neuron, 4 hidden neurons, and 2 output neurons is used and the results revealed that, in the training group, sensitivity and specificity are 100%; while, in the testing group, they were 99.33% and 100%, respectively. However, in the proposed logistic regression model, sensitivity and specificity were 99% and 90%,

Manuscript received July 5, 2017
Manuscript revised July 20, 2017

respectively. Thus, comparing these findings with the sensitivity and specificity of the artificial neural networks reveals that the artificial neural networks outperform the logistic regression model (4). Another related study is devoted to comparing artificial neural networks and other statistical models such as logistic regression in predicting coronary artery disease. This comparison was based on the Roc curve, and the results demonstrated that the area under the Roc curve was more in the artificial neural networks, compared to the logistic regression model. This is inductive of the superiority of artificial neural networks to the logistic regression in solving such problems (26). In the same line of research, another study was conducted to compare artificial neural networks with Cox regression models in survival prediction of gastric cancer patients. The results suggested that the artificial neural networks, compared to the Cox regression model, offer better predictions regarding the survival of patients and the artificial neural networks were recommended to be used for survival prediction studies (5). Furthermore, in another study devoted to determining prognostic factors in the survival of gastric cancer patients, artificial neural networks were used. The results revealed that the correct prediction of artificial neural networks was 82.65%; while, for the Weiball model, this value was 75.7%. Accordingly, it goes without saying that, this study also substantiated the superiority of the artificial neural networks to the Weiball regression model (7). Given that no study has been conducted to investigate the sensitivity of Cox proportional hazard model, multilayer neural network with radial basis function method simultaneously, the present study aims at comparing the performance of MLP, RBF and Cox PH model using Heart attack data.

2. Materials and Methods

2.1 Cox Proportional Hazard Model

In survival analysis, the deployed statistical methods for data analysis are mostly based on the time required for the occurrence of the event. If survival data analysis aims at describing the survival time, without taking the covariates into account, non-parametric methods such as life tables and Kaplan-Meier method are used (24). However, if the survival analysis pursues modeling, determining the relationship between effective variables, and prediction, survival regression models (which are mostly based on hazard function) such as Cox proportional hazard model, as a semi-parametric model, and parametric regression models are used (28). The equation of Cox regression model is as follows:

$$h_i(\mathrm{t}) = \mathrm{h}_0(\mathrm{t})\exp(\beta_1 \mathrm{x}_{i1} + \beta_2 \mathrm{x}_{i2} + \ldots + \beta_k \mathrm{x}_{ik}) \quad (1)$$

Where $h0\ (t)$ is the hazard function of time only, $\exp(\beta_1 \mathrm{x}_{i1} + \beta_2 \mathrm{x}_{i2} + \ldots + \beta_k \mathrm{x}_{ik})$ is the proportional hazard, and β_i represents regression parameters (21). Despite having some limitations, the Cox regression model is the most commonly-used method in survival data modeling. In Cox proportional hazard models, mostly the effect of covariates on the hazard function (in which it is assumed that the hazard ratio of one group to another group is constant over time) is investigated. In this condition, if the hazard baseline is indeterminate and lacks any specific distributional form, the semi-parametric model of Cox proportional hazard is obtained; while, if a parametric distribution is considered for the baseline hazard function, exponential parametric models, Weibull, Gompertz, etc. are obtained. Proportionality of hazard for all the covariates in the final model and the independency of the occurrence time of each event are among the basic assumptions underlying the Cox model. If the data do not satisfy the basic assumptions of the Cox model, this model cannot be used for data analysis.

2.2 Artificial Neural Networks

Artificial neural network models are related to simulating human learning in the mind and implementing them as a computer algorithm. An artificial neural network should be capable of categorizing patterns, it should be small enough to be physically realistic, it should be programmable (by applying training) and, finally, it should be capable of being generalized by the examples provided during the training. In these networks, after learning, by inserting a specific input, a specific output is achieved. Networks are coordinated according to compliance and contrast between the input and purpose so that the output of the network and the purpose are matched and coordinated with each other. Most often, a large number of input-output pairs are deployed so that, in this process, which is dubbed supervised learning, the network is taught (2). Neural networks are composed of a number of neurons which are interconnected by weights or synapses, each having a specific weight. It is worth mentioning that the weight of each synapse can be changed, commensurate with the condition.

Each neuron transforms its activities to an output activity and transmits it to other neurons which receive synapse from it (39). The composition of weights, biases, distribution function, and performed multiplication and summation operations are referred to as a layer of the network (13). Neural networks are performed in two learning and testing stages. In the learning stage, the network receives a random sample of the input and output data, which mostly exceed 50% of the observations, and modifies the weight values of the network. Finally, after repeating this process several times, weights are matched

in such a way that, by having access to the information of each pattern, the network is capable of assessing it. When the learning stage is completed, the new inputs, which were not used in the previous stage, are inserted into the network and their output is registered. By comparing the outputs of the second stage and real values, network efficiency is evaluated (20).
One of the widely-used architectures of artificial neural networks is the multi-layer Perceptron neural network (MLP) which uses a back propagation learning algorithm (17, 30, and 46). One of the other neural networks which has attracted researchers' attention and will be expounded in the following is the Radial Basis Function Neural Network (RBFNN) which is widely deployed to estimate non-parametricity in multi-dimensional functions, using limited learning information sets. These neural networks have received unprecedented attention due to their fast and pervasive learning. Networks with radial basis function are powerful approximators, such that, by having enough neurons in their hidden layer, they can approximate every continuous function with every degree of precision. These networks have these characteristics merely by having one hidden layer (20, 35). Networks with radial basis function mostly use the statistical techniques for classifying patterns. The most significant advantage these networks offer is classifying patterns with non-linear space. The reason behind subsuming these techniques under the artificial neural networks is that, although the number of these techniques is small, they are extensively used. These networks are often compared with the progressive neural network.

2.3 Preparing and Classifying the Data

Given that the numerical values of the inputs and outputs have different units and values, the learning process of artificial neural networks is not performed properly and satisfactorily. The reason can be attributed to the fact that the difference in the value of numbers affects the modification of the weights of the network considerably and, accordingly, the number of experimental data should be great to normalize the weights and obtain the desired result. Thus, before learning the neural network, a normalizing function is required to modify the weights and refute the mentioned problem. One of the normalizing techniques is using the equation (2) in which the data are divided to their maximum values. In this way, not only the effect of different units is eliminated but also we deal with values in the interval of zero to one.

$$T_n = \frac{T}{T_{\max}} \tag{2}$$

Where Tn is the value of the normalized variable and Tmax is the maximum value of the variable which is under investigation.

2.4 Architecture of the Multilayer Perceptron Neural Network

The structure of the multilayer Perceptron neural network is represented in figure 1. As can be seen, this network involves three layers: input layer (which depends on the number of inputs, in this figure, there are 4 input neurons), the hidden layer (whose neurons are selected after correct learning of the network), and output layer (which has one neuron in this model).

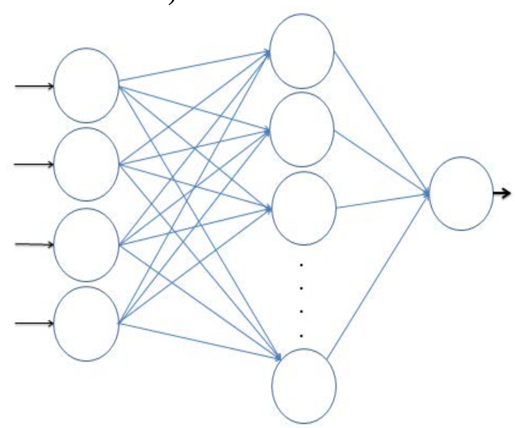

Fig. 1 The structure of multilayer Perceptron neural network

The first step in designing this neural network is creating the architecture of the network. The first input is a R*2 matrix as minimum and maximum values of each input with the R number of input vector. The second input, is a combination of the size of each layer, determining the neuron number of different layers. The third input, is a combination including transmission functions used in each layer, and the fourth input is the name of the applied learning function. Transmission functions include Tansig ، *Logsig* and Purelin which are defined in 3 to 5 equations, respectively:

$$Logsig\,(\mathrm{n}) = 1/(1+\exp(-\mathrm{n})) \tag{3}$$

$$\mathrm{Tansig(n)} = 2/(1+\exp(-2\times \mathrm{n}))-1 \tag{4}$$

$$\mathrm{Purelin(n)} = n \tag{5}$$

Where n is the input of the function. The Logsig function generates an output between zero and one for the $(-\infty,+\infty)$ input. The Tansing function generates an output between 1 and -1 for the $(-\infty,+\infty)$ input. Finally, for the $(-\infty,+\infty)$ input, the Purelin function generates the same input data.

2.5 The Architecture of Radial Basis Function Network

The neural network with radial basis function, depicted in figure 2, is composed of input, hidden, and output layers such that each of the neurons of the output layer is a linear combination of the outputs of the hidden layer neurons. The input layer is merely an input layer in which no processing is performed. The second layer, or otherwise construed as the hidden layer, creates a non-linearity between the input space and another space with larger dimensions and plays a crucial role in transforming non-linear patterns to linearly separable patterns. Finally, the third layer generates the weighted sum along with a linear output.

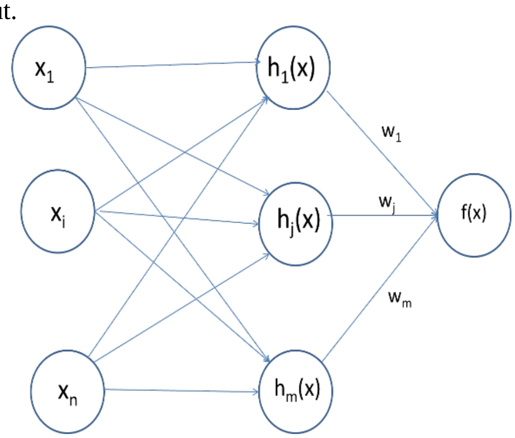

Fig. 2 The structure of neural networks with radial basis function

If the RBF is used to approximate the function, such output would be useful; however, if the patterns are not required to be classified, a hard-limiter or a sigmoid function can be placed on the output neurons to generate zero or one output values. The processing performed in the hidden layer can be conceived of as the distinctive characteristic of this network. The function of the hidden layer can be obtained by the equation 6:

$$f\ (\mathrm{x}) = \sum_{j=1}^{p} w_j \phi(\|x - u_j\|) \tag{6}$$

This equation suggests that, to approximate the f function, p radial basis functions with the center of uj, are used. The $\|.\|$ symbol represents the distance function in Rn space and often the Euclidean distance is selected. Given that the function curve of radial basis functions is symmetrical radial, the hidden layer neurons are recognized as the radial basis neurons. Different radial basis neural functions have been proposed. In the following, some of them are sketched in passing:

Cubic Splines $\phi : \Re \to \Re : r \to |r|^3$

Thin Plate Splines

$$(\mathrm{x}, \mathrm{y}) \to \|(\mathrm{x}, \mathrm{y})\|^2 \log \|(\mathrm{x}, \mathrm{y})\|$$

$$\phi_{MQ}(\mathrm{r}) = \sqrt{a^2 + r^2}$$ Multiquardics

The recognized and common function in radial basis neural function is the Gaussian or exponential function, which is represented in 7:

$$\phi(\|x - u_j\|) = \mathrm{e}^{-\frac{(\|x - u_j\|)}{\sigma_j}} \tag{7}$$

Where σ_j is the j Kernel. The reason behind choosing the exponential Gaussian function as neurons' response function, in networks with radial basis function, is that exponential functions are among functions having the best characteristics for approximation. This confirms that, there is a set of weights which approximate the relationship between inputs and purpose vectors better that other weight sets. The Sigmoid function, used in designing error back propagation, does not have these characteristics.

According to the above-mentioned equation, for an input value equal to the average x=u, neurons' response function reaches its maximum value. As distance of x from the average value increases, the response value decreases considerably and, the more the distance increases, the more, the values of neuron's response function decreases. Hence, neuron's output value can be investigated in a specific field of x's values. This specific field is dubbed the "receptive field". The size and range of this field is determined by the σ parameter.

Compared to the standard normal distribution, which is like the neuron's response fnction curve with one or two variable (s), u can be construed as the median and σ as the standard deviation of neurons' response curve. Although spatial visualization of the display curves of hidden neurons' response function is not possible in spaces with more than three dimensions, in this condition, likewise, each hidden layer has a response which can be represented by the equation in 8:

$$\phi(\|\mathbf{x} - \mathbf{u}_j\|) = \mathrm{e}^{-\frac{(\|\mathbf{x} - \mathbf{u}_j\|)}{\sigma_j}} = e^{-\frac{(\mathbf{x} - \mathbf{u}_j)^T (\mathbf{x} - \mathbf{u}_j)}{\sigma_j}} \tag{8}$$

Where, T is the transpose of the vector, x is the input column vector, and uj is the center vector of the jth neuron which is equal to one learning input vector.

For describing, modeling, and comparing the data, SPSS20 and MATLAB (version R2015) were used. Probability values below 0.05 are considered as being statistically significant. To evaluate and validate the performance of networks in multilayer Perceptron neural network, mean square error is defined as presented in equation 9:

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(\mathrm{e}_i)^2 = \frac{1}{N}\sum_{i=1}^{N}(t_i - \mathrm{a}_i)^2$$

Where N is the number of all learning samples, ti is the value of real output and ai can be network's output value. In learning multilayer Perceptron neural network in MATLAB, the Levenberg-Marquardt algorithm is used.

To compare different models, the sensitivity, specificity, and accuracy criteria are deployed. Sensitivity can be defined as the possibility of the accurate detection of a patient who is actually sick. Moreover, specificity refers to the possibility of the accurate detection of the health of a person who is healthy, and accuracy is possibility of the accurate detection of a patient who is sick and detecting the health of healthy people to all of the participants of the study. The more the value of a method for these three criteria, the more reliable the method is.

2.6 Dataset

The data were obtained from the following websites, 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 were 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 were used.

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 model 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 Software.

2.7 Discussion

Table1: Statistical equivalent phrases in neural networks

Neural network	Analysis
Network	Model
Learning	Estimation
Regression	Regression
Generalization	Interpolation
Learning Collections	Observation
Synapse	Parameters
Inputs	Independent Variables
Outputs	Dependent Variables

MLP and RBFNN are the comprehensively-used Feed forward neural networks. Both differ basically in the way how the hidden units combine values approaching from the inputs. The MLP uses inner products and the RBF uses the Euclidian distance. We used both RBF and MLP algorithms for the prediction of heart attack data. Previously-conducted studies demonstrate that artificial neural networks outperform the classic models.

Many researchers have compared the efficiency of the RBF and MLP models and most of them have contended that RBF network was better than MLP, and some of them doubt their prediction efficiency. In a study conducted to predict the survival of patients suffering from breast cancer, findings revealed that compared to regression models, the artificial neural network models can predict patients' survival more precisely (43). Thus, the better performance and precision of artificial neural networks in predicting the survival of patients with cancer, in comparison with other statistical models is confirmed in the literature (11). In investigating the performance of artificial neural network in detecting electromagnetic parameters, the findings revealed that the RBFNN is characterized by a more rapid detection, compared to the MLP method (18). Moreover, it is demonstrated that, in non-linear systems, the RBF neural networks perform better than the MLP methods and are more specific and precise (15). Furthermore, in a study conducted to predict the survival of patients with breast cancer, it was revealed that the RBF method was more precise in detection and required less time for detection and, compared to the MLP method, it was more reliable (33). Moreover, in diabetes diagnosis, it was likewise revealed that the RBFNN method is more precise, compared to the MLP and logistic regression method (42). Another study carried out to compare the precision of the MLP and RBFNN methods in predicting the weight and length of fetuses, it was demonstrated that, although the findings are rather similar, the RBF method was more precise and efficient.

3. Results

WHAS data set with 481 records is used for the purpose of the present study. The Cox PH model was fitted using SPSS. The covariates AGE, SHO and CHF are significantly connected with the time to event data. The

Cox PH regression model fitted to the data gave a sensitivity of 85%, specificity of 82%, and overall accurate prediction of 83%. The MLP architecture had six input variables and one hidden layer with three hidden nodes and one output node. The best MLP model was obtained at lowest Root Mean Square Error (RMSE) of 0.2125. MLP sensitivity was 92%, specificity was 91%, and the percentage of accurate prediction was 91%. RBFNN was executed best at ten centers and maximum number of tried centers was 18. Root Mean Square Error (RMSE), using the best centers, was 0.3212. Sensitivity of the RBFNN model was 98.5%, specificity was 98.5% and the percentage of accurate prediction was 98.5%. Execution time of RBFNN is less than the MLP and compared Cox PH model.

Table 2: The models and their specifications

Model	Sensitivity (%)	Specificity (%)	Accuracy (%)
Cox	85	82	83
MLP	92	91	91
RBF	98.5	98.5	98.5

4. 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 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 have showed better outcomes and more applications would prove the efficiency of this model over other models. These findings likewise reveal that the RBF method can be an appropriate and reliable substitute for the PH model in the predicting and analyzing survival data.

References

[1] Amiri, Zohreh, et al. (2008). "Assessment of gastric cancer survival: using an artificial hierarchical neural network."Pak J Biol Sci 11.8. 1076-84.

[2] Anderson A. (1995). An introduction to neural networks, Cambridge, MA: MIT: 795-851.

[3] Baxt, William G., and Joyce S. (1996). "Prospective validation of artificial neural network trained to identify acute myocardial infarction." The Lancet 347.8993: 12-15.

[4] Biglarian A, Babaee GR, and Azmie R. (2004). Application of Artificial Neural Network Model in Determining Important Predictors of InHospital Mortality After Coronary Artery Bypass Graft Surgery, and it's Comparison with Logistic Regression Model. Modarres J Med Sci. 7(1): 23-30.

[5] Biglarian A, Hajizadeh E, Kazemnejad A. (2010). Comparison of Artificial Neural Network and COX Regression Models in Survival Prediction of Gastric Cancer Patients. Koomesh. 11(3): 211-205.

[6] Biglarian, A., Hajizadeh E., and Kazemnejad, A. (2010) "Comparison of artificial neural network and Cox regression models in survival prediction of gastric cancer patients." Koomesh 11.3: 215-220.

[7] Biglarian A, Hajizadeh E, Kazemnejad A, and Zayeri F. (2010). Determining of prognostic factors in gastric cancer patients using artificial neural networks. Asian Pac J Cancer Prev. 11(2): 533-536

[8] Bollschweiler, E. H., et al. (2004). "Artificial neural network for prediction of lymph node metastases in gastric cancer: a phase II diagnostic study." Annals of surgical oncology 11.5: 506-511.

[9] Borhani, D. A., and Fatehi, M. A. (2008). "Application of artificial neural network in stream flow forecasting using climatic indices (Case study: Nazloochay River Basin)." 25-36.

[10] Borglund, A. (2014). Statistical Pattern Recognition. International Journal of Computer (IJC), 7(1).

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

[12] Chien, Ching-Wen, et al. (2007). "The application of artificial neural networks and decision tree model in predicting post-operative complication for gastric cancer patients." Hepato-gastroenterology 55.84: 1140-1145.

[13] Jae-ho, Ch. M. Adams, T. and Bahia, H. (2004). "Pavement Roughness Modeling Using Back-Propagation Neural Networks." Computer-Aided Civil and Infrastructure Engineering 19.4: 295-303.

[14] Collett, D. (2015). Modelling survival data in medical research. CRC press.

[15] Ardelean, D., Kloetzer, M. and Pastravanu, O. (2001). "RBF neural networks in nonlinear system identification" Proc 7th International symposium on Automatic Control and Computer Science. 82-86.

[16] Dickman, P. W., Coviello, E., and Hills, M. (2015). Estimating and modeling relative survival. Stata J, 15(1), 186-215.

[17] Fakhri M. Seedalhosseini E. (2012). Apply Artificial Neural Network in Roughness. Available from: http://www.compcivil.com.

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

[19] Harrell, F. (2015). Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis. Springer.

[20] Hartman, Eric J., James D. Keeler, and Jacek M. Kowalski. (1990). "Layered neural networks with Gaussian hidden units as universal approximations." Neural computation 2.2: 210-215.

[21] Hougaard, Philip. (2012). Analysis of multivariate survival data. Springer Science & Business Media.

[22] Hunt, S. A., Abraham, W. T., Chin, M. H., Feldman, A. M., Francis, G. S., and Ganiats T. G. et al. (2005). ACC/AHA 2005 guideline update for the diagnosis and management of chronic heart failure in the adult. Circulation; 112:154-235.

[23] Jobson, J. (2012). Applied multivariate data analysis: volume II: Categorical and Multivariate Methods. Springer Science & Business Media.
[24] Karakitsos, P. et al. (2000). "Neural network application in the discrimination of benign from malignant gastric cells." Analytical and quantitative cytology and histology 22.1: 63-69.
[25] Kohl, M., Plischke, M., Leffondré, K., & Heinze, G. (2015). PSHREG: A SAS macro for proportional and nonproportional subdistribution hazards regression. Computer methods and programs in biomedicine, 118(2), 218-233.
[26] Kurt, I., Ture, M., Kurum, A. T. (2008). Comparing performances of logistic regression, classification and regression tree, and neural networks for predicting coronary artery disease. Expert Syst Appl 2008; 34(1): 366-374
[27] Lai, Kuang-Chi, et al. (2007). "Artificial neural network-based study can predict gastric cancer staging." Hepato-gastroenterology 55.86-87: 1859-1863.
[28] Lee, E. T., and Wang, J. (2003). Statistical methods for survival data analysis. Vol. 476. John Wiley & Sons.
[29] Lindsey, J. K.., and Jones, B. (1998). "Choosing among generalized linear models applied to medical data." Statistics in medicine 17.1: 59-68.
[30] Menhaj, M. B. (2008). "Artificial Neural Network Basics." Tehran: Amirkabir Publications. 93-121.
[31] Mobley BA, Schecheer E, Moore WE. (2000). Prediction of coronary artery enosis by artificial networks artificial intelligence in medicine, 18, 187-203.
[32] Neter, J. et al. (1996). Applied linear statistical models. Vol. 4. Chicago: Irwin.
[33] Padmavathi J. A. (2001). Comparative study on breast cancer prediction using RBF and MLP, International Journal of Scientific and Engineering Research, 2(1): 1-5.
[34] Pérez-Godoy, M. D., Rivera, A. J., Carmona, C. J., and Del Jesús, M. J. (2014). Training algorithms for radial basis function networks to tackle learning processes with imbalanced data-sets. Applied Soft Computing, 25, 26-39.
[35] Tomaso, P., and Girosi, F. (1990). "Networks for approximation and learning." Proceedings of the IEEE 78.9, 1481-1497.
[36] Rapsomaniki, E., Timmis, A., George, J., Pujades-Rodriguez, M., Shah, A. D., Enaxas, S., and Williams, B. (2014). Blood pressure and incidence of twelve cardiovascular diseases: lifetime risks, healthy life-years lost, and age-specific associations in 1· 25 million eople. The Lancet, 383(9932), 1899-1911.
[37] Ross, S. M. (2014). Introduction to probability models. Academic press.
[38] Marques, S. and Bernardes, De. S. (2000). 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.
[39] Sahoo, G. B., Ray, C. and De Carlo E. H. (2006). "Use of neural network to predict flash flood and attendant water qualities of a mountainous stream on Oahu, Hawaii." Journal of Hydrology 327.3: 525-538.
[40] Sprent, P., and Smeeton, N. C. (2016). Applied nonparametric statistical methods. CRC Press.
[41] Swedberg K, Cleland J, Dargie H, Drexler H, Follath, F., Komajda, M. et al. (2005). Guidelines for diagnosis and treatment of chronic heart failure. Eur Heart J, 26(11):1115-40.
[42] Venkatesan P and Anitha S, (2006). Application of Radial basis function neural network for diagnosis of diabetes mellitus, Current Sceience, 91(9), 1195-1199.
[43] Venkatesan P. and Suresh M. L. (2009). Brest Cancer survival prediction using Artificial Neural Network, IJCSNS International Journal of Computer Science and Network Security, 9 (5), 169-174.
[44] Warner, B. and Manavendra M. (1996). "Understanding neural networks as statistical tools."The American statistician 50.4: 284-293.
[45] Yadav, P., Steinbach, M., Kumar, V., and Simon, G. (2015). Mining Electronic Health Records (EHR): A Survey.
[46] Zoqi, M. J., and Ghavidel, A. (2009). "Neural network modeling and prediction of methane fraction in biogas from landfill bioreactors." Iranian Journal of Health and Environment 2.2: 140-149


Mohammad Salehi Veisi


Sadegh Rezaei


Khadejeh Salehivaysi

Appendix1. Data set

AME: Heart Disease Survival (WHAS1.DAT)
SIZE: 481 observations, 14 variables

SOURCE: Hosmer D.W. and Lemeshow, S. (1998) Applied Survival Analysis:
Regression Modeling of Time to Event Data,
John Wiley and Sons Inc., New York, NY

DESCRIPTIVE ABSTRACT:

This data set is described in Table 1.4 of the source text.

DISCLAIMER: This data is also available at the publisher's FTP site:
ftp//ftp.wiley.com/public/sci_tech_med/survival

LIST OF VARIABLES:

Variable	Description	Codes / Units
ID	Identification Code	1 - 481
AGE	Age (per chart)	years
SEX	Gender	0 = Male 1 = Female
CPK	Peak Cardiac Enzyme	International Units (iu)
SHO	Cardiogenic Shock Complications	0 = No 1 = Yes
CHF	Left Heart Failure Complications	0 = No 1 = Yes
MIORD	MI Order	0 = First 1 = Recurrent
MITYPE	MI Type	1 = Q-wave 2 = Not Q-wave 3 = Indeterminate
YEAR	Cohort Year	1 = 1975 2 = 1978 3 = 1981 4 = 1984 5 = 1986 6 = 1988
YRGRP	Grouped Cohort Year	1 = 1975 & 1978 2 = 1981 & 1984 3 = 1986 & 1988
LENSTAY	Length of Hospital Stay Days in Hospital	Days
DSTAT	Discharge Status from Hospital	0 = Alive 1 = Dead
LENFOL	Total Length of Follow-up from Hospital Admission	Days
FSTAT	Status as of Last Follow-up	0 = Alive 1 = Dead