

Multilayer Feed-Forward Neural Network Approach to Lymphoma Cancer Data

Norizan Mohamed¹, Nor Azlida Aleng², Wan Muhamad Amir W Ahmad³

and Maizah Hura Ahmad⁴

^{1,2,3} Jabatan Matematik, Fakulti Sains dan Teknologi, Malaysia, Universiti Malaysia Terengganu (UMT), 21030 Kuala Terengganu, Terengganu, Malaysia

⁴Jabatan Matematik, Fakulti Sains, Universiti Teknologi Malaysia, 81030 UTM Skudai, Johor, Malaysia.

¹e-mail: norizan@umt.edu.my

Abstract

Lymphoma is a cancer in the lymphatic cells of the immune system, called lymphocytes. One of the most important organ systems of the human body is the lymphatic or lymphoid system, which is a network of node-like structures located throughout the body. This system helps filter out bacteria and plays an important role in fighting diseases. However, just like any other organ system in the body, it is also exposed to developing cancers. The objective of the current study is to study from the view of MLFF neural network model the most significant factor that influences Hodgkin's lymphoma. The input variables of MLFF neural network model are selected based on the significant variables of logistic regression (LR) model. The view from (MLFF) neural network model showed that two variables which were diffuse large B-Cell lymphoma (DLBCL) and hypertension status (HPT) were most influence of Hodgkin's lymphoma.

Keywords: Multilayer Feed-Forward neural network, multiple logistic regression and Lymphoma.

INTRODUCTION

Lymphoma is a general term for cancer in the lymph system. The lymph system is

made up of many cells and organs, including the lymph nodes, thymus gland, spleen, and liver. This system produces B-Cells and T-Cells, which make up your body's immune system. Since these cells travel between the lymphatic and circulatory system while fighting infections and viruses, lymphomas are blood-related cancers. It is a type of cell that forms part of the immune system. Typically, lymphomas present as a solid tumor of lymphoid cells. Generally, there are more than 35 types of lymphoma. The two main types are Hodgkin lymphoma and non-Hodgkin lymphoma (NHL). In 2011, about 662,789 people are living with lymphoma or are in remission (no sign of the disease). This number includes about 159,846 people with Hodgkin lymphoma and 502,943 people with NHL [13]. The lymph system can be illustrated in Figure 1.

Hodgkin lymphoma also known as Hodgkin disease. It has characteristics that distinguish it from other diseases classified as lymphoma, including the presence of Reed-Sternberg cells. These are large, cancerous cells found in Hodgkin lymphoma tissues, named for the scientists who first identified them. Hodgkin lymphoma is one of the most curable forms of cancer. It was named after Thomas Hodgkin (most prominent British physician in 1798 - 1866), who first described abnormalities in the lymph system in 1832 (Hellman).

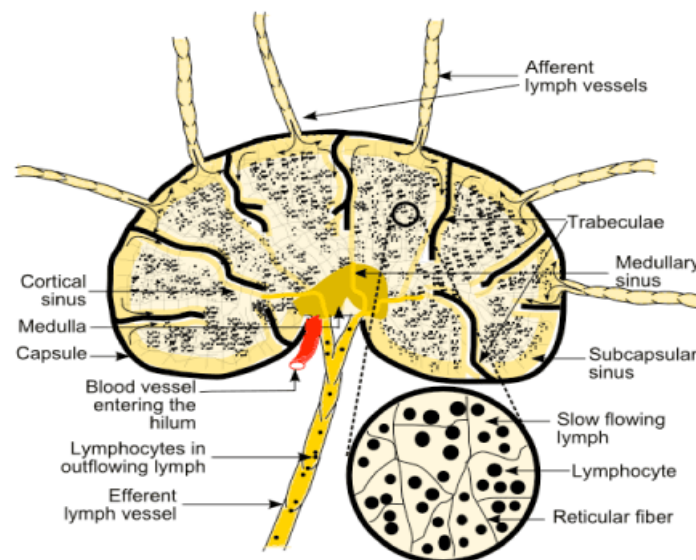


Figure 1: Lymph node schematic [11]

NHL represents a diverse group of diseases distinguished by the characteristics of the cancer cells associated with each disease type. Most people with NHL have a B-cell type of NHL (about 85 percent). For patients with slow-growing NHL, treatment may

keep the disease in check for many years. Hodgkin's lymphoma may be treated with Radiation therapy, chemotherapy, Surgery and High-dose chemotherapy and radiation therapy with stem cell transplant [1].

NHL is a cancer of the fifth highest in the world today. In 2009, over 66,000 adults and children have been diagnosed with the NHL disease. 95 percent of cases involving adults aged 60 and above suffer from this disease. Men at higher risk of women suffering from this disease though the number of female patients who were in NHL were diagnosed increases. Among the 74,030 new cases of lymphoma in 2010, the diseases will affect 40,050 males and 33,980 females. Hodgkin lymphoma will account for 8,490 cases (4,670 males and 3,820 females), and 65,540 cases will be NHL (35,380 males and 30,160 females). Many NHL disease existed by white people against people of African and Asian Americans [1].

Based on studies by Becker, Deeg & Nieters [5,7] of factors for the disease of cancer has proved effective use of multiple logistic regression analysis in the medical field. Among the factors studied were gender, age, ethnicity, medical history and smoking. Age, gender, medical history and smoking are factors that are absolute and can be categorized with the value as 0 and 1. In general, the risk for lymphoma cancer face starting at the age of 35 years to 50 years and almost 50% of those who reach the age of 65 years of age. In these studies, he also found that more men at high risk for lymphoma cancer are more likely to develop heart disease. Simple logistic regression analysis was also used as a guide to continue the study with logistic regression analysis [3-4, 6].

In 2011, multilayer feed-forward (MLFF) neural network was applied to medical data [10]. Norizan Mohamed et al. [10] develop a multiple linear regression (MLR) and two multilayer feed-forward (MLFF) neural network models of body mass index (BMI). Model 1 was developed based on all the independent variables which are systolic blood pressure (SBP), diastolic blood pressure (DBP), age, serum cholesterol (CHOLEST), gender, pulse, height, weight, arm, leg, waist and wrist. Model 2 was developed based on four significant independent variables of multiple linear regression. The result shown that Model 2 was more accurate than Model 1.

Recently, Nor Azlida Aleng et al. [8] applied MLFF neural network model on tuberculosis patients. The combinations of significant variables from multiple linear regression (MLR) model were used as input variables of MLFF neural network model. The MLR of tuberculosis showed that gender, married status, diseases genital and descendant were significant. Hence, four variables were used to develop the best (MLFF) neural network model of tuberculosis.

The objective of the current study is to study from the view of MLFF neural network model the most significant factor that influences Hodgkin's lymphoma. The input variables of MLFF neural network model are selected based on the significant variables of logistic regression (LR) model. Material of this study is a hypothetical sample which is composed of six variables. Data of 326 respondents (patients) were collected and selected variables are as in Table 1.

Table 1. Explanation of the Variables

Variables	Code	Explanation of the variables	Categorical
LYMPHOMA	Y	Type Of Hodgkin Lymphoma Cancer	0 = No 1 = Yes
SEX	X1	Gender	0 = Male 1 = Female
DLBCL	X2	Diffuse Large B-Cell Lymphoma	0 = No 1 = Yes
HPT	X3	Hypertension status	0 = No 1 = Yes
AGE	X4	Age Of Patients In Year	-
STATUS	X5	Patients Status	0 = alive 1 = dead

METHOD

Multilayer Feed-forward Neural Network (MLFF)

Multilayer Feed-forward Neural Network (MLFF) consists of an input layer, one or several hidden layers and an output layer. The neurons in the feed-forward neural network are generally grouped into layers. Signals flow in one direction from the input layer to the next, but not within the same layer [12]. An essential factor of successes of the neural networks depends on the training network. Among the several learning algorithms available, back-propagation has been the most popular and most widely implemented [2]. Basically, the BP training algorithm with three-layer feed-forward architecture means that, the network has an input layer, one hidden layer and an output layer. In this research the output node is fixed at one since there is only one independent variable. Thus, for the feed-forward network with N input nodes, H hidden nodes and one output node, the values \hat{Y} are given by:

$$\hat{Y} = g \left(\sum_{j=1}^H w_j h_j + w_0 \right) \tag{1}$$

where w_j is an output weight from hidden node j to output node, w_0 is the bias for output node, and g is an activation function. The values of the hidden nodes h_j , $j = 1, \dots, H$ are given by:

$$h_j = k \left(\sum_{i=1}^N v_{ji} X_i + v_{j0} \right), \quad j = 1, \dots, H \tag{2}$$

Here, v_{ji} is the input weight from input node i to hidden node j , v_{j0} is the bias for hidden node j , X_i are the independent variables where $i = 1, \dots, N$ and k is an activation function. The architecture of the multilayer feed-forward neural network model is illustrated in Figure 2.

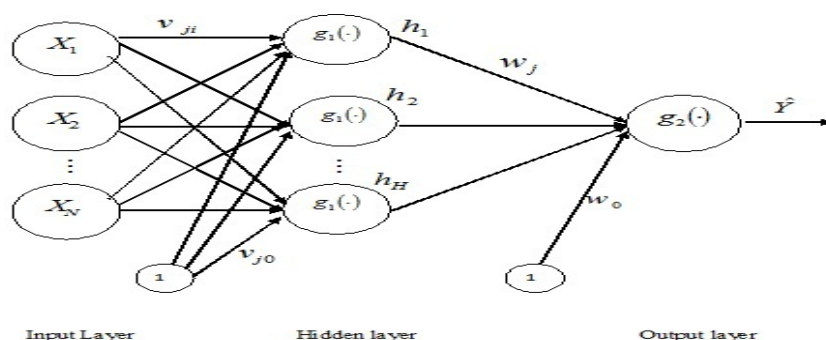


Fig. 2: The architecture of the multilayer feed-forward neural network model with N input nodes, H hidden nodes and one output node.

RESULT AND DISCUSSION

Previous study, Norizan Mohamed et al. [10] used the significant variables from MLR model as input nodes of MLFF neural network model. Instead of significant variables from MLR model, the significant variables from LR model are used in this current study. Previously, Wan Muhammad Amir et al. [14] found that three independent variables which were sex, DLBCL and HPT significantly influence the logistic regression model of lymphoma cancer. Hence these three independent variables (sex, DLBCL and HPT) are used as the input variables of MLFF model.

The output node in this study is one node since we have one dependent variable which is the type of Hodgkin lymphoma cancer. To find the appropriate number of

hidden nodes and best combination of input variables, the model selection strategies which proposed by Norizan Mohamed [9] is followed. The best number of hidden nodes is one node. Using difference combination of significant variables, result shown that the best combination is two input variables as illustrated in Table 2. The architecture of MLFF neural network model of Hodgkin lymphoma is composed of two input variables, one hidden node and one output node as presented in Figure 3. Thus, MLFF neural network model with two input variables, one hidden node and one output node, the values \hat{Y} are given by:

$$\hat{Y} = g(w_1 h_j + w_0) \quad j = 1 \quad (3)$$

where w_1 is an output *weight* from hidden node j to output node, w_0 is the bias for output node, and g is linear function. The values of the hidden nodes h_j , $j = 1$ are given by:

$$h_j = k \left(\sum_{i=2}^3 v_{ji} X_i + v_{j0} \right) \quad j = 1 \quad (4)$$

Here, v_{ji} is the input *weight* from input node i to hidden node j , v_{j0} is the bias for hidden node j , X_i are the independent variables where $i = 2, 3$ and k is a sigmoid function. Equations 3 and 4 can also be represented as follows:

$$\hat{Y} = w_0 + w_1 h_1 \quad (5)$$

$$h_j = \left[1 + \exp \left[- \left(v_{j0} + v_{j2} X_2 + v_{j3} X_3 \right) \right] \right]^{-1}, \quad j = 1 \quad (6)$$

The architecture of MLFF neural network model is illustrated in Figure 3.

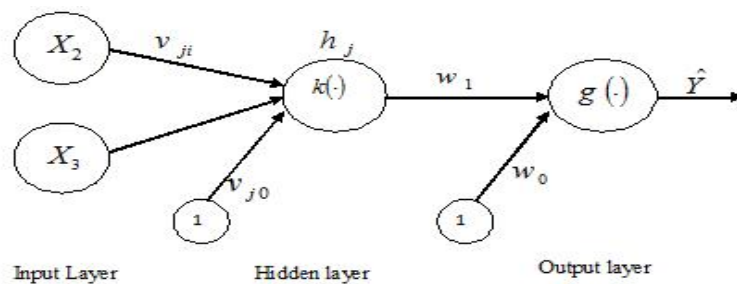


Figure 3: The architecture of MLFF neural network model with two input variables, one hidden node and one output node.

Table 2: The results of MSE training and MSE testing.

Input Variables	MSE Training	MSE Testing
X ₁ , X ₂ , X ₃	0.1883	0.2143
X ₂ , X ₃	0.1912	0.2119
X ₃	0.1935	0.2144

X₁, X₂, and X₃ are sex, DLBCL and HPT respectively

CONCLUSION

The purpose of the current study is to examine from the view of MLFF neural network model the most significant factor that influences Hodgkin’s lymphoma or Hodgkin’s disease. The combinations of significant variables from logistic regression (LR) model are used as input variable of MLFF neural network model. The LR of Hodgkin’s lymphoma showed that sex, DLBCL and HPT were significant. Using the significant variables of LR model, the performance of MLFF neural network model for difference combinations are tested. The performance of MLFF was evaluated MSE of testing/out-sample. The combination of two input variables outperformance other combinations. Hence, two variables which were DLBCL and HPT were most influence of Hodgkin’s lymphoma from the view of (MLFF) neural network model.

REFERENCES

- [1] American Cancer Society: Cancer facts & figures, 2010.
- [2] G.A. Darbellay and M. Slama, Forecasting the short-term demand for electricity. Do neural networks stand a better chance? *International Journal of Forecasting*, 16 (2000), 71-83.
- [3] G.C.C. Lim, H. Yahaya, and T.O. Lim, The first report of the national cancer registry cancer incidence in Malaysia. National Cancer Registry, Kuala Lumpur, National Cancer Institute, 2002.
- [4] Long and Freese, Review of regression model for categorical dependent variable using stata, second edition. *The Stata Journal*, 6 (2006), 273-278.
- [5] N. Becker, E. Deeg and A. Nieters, Population-based study of lymphoma in germany: Rationale, study design and first results. *Leukemia Research* 28, (2004), 713–724.

[6] National Cancer Institute, Physician Data Query, Adult Non-Hodgkin Lymphoma Treatment, 2009.

[7] National Comprehensive Cancer Network, Practice guidelines in oncology: non-hodgkin's lymphoma, 2009.

[8] Nor Azlida Aleng, Norizan Mohamed and Wan Muhamad Amir W Ahmad, Modelling of tuberculosis patients using multilayer feed forward neural network. *Applied Mathematical Sciences*, 6 (50) (2012), 2481 – 2488.

[9] Norizan Mohamed, Parametric and artificial intelligence based methods for forecasting short term electricity load demand, Unpublished PhD thesis, 2011.

[10] Norizan Mohamed, Wan Muhamad Amir W. Ahmad, Nor Azlida Aleng and Maizah Hura Ahmad, Assessing the efficiency of multilayer feed-forward neural network model: Application to body mass index data, *World Applied Sciences Journal*, 15 (5) (2011), 677-682.

[11] Schematic of lymph node, <http://commons.wikimedia.org>, 13 February 2012.

[12] T. D. Pham and X. Liu, *Neural Networks for Identification, Prediction and Control*, Great Britain. 1995.

[13] The leukemia & lymphoma society, lymphoma, <http://www.lls.org/diseaseinformation/lymphoma>, 14 February 2012.

[14] Wan Muhamad Amir W. Ahmad, Nor Azlida Aleng Zalila Ali and Siti Aisyah Abdullah, Logistic regression approach to lymphoma cancer data, *European Journal of Scientific Research*, 51(4) (2011), 448-456.

Received: March, 2012