RESEARCH ARTICLE

Use of an Artificial Neural Network to Construct a Model of Predicting Deep Fungal Infection in Lung Cancer Patients

Jian Chen[1&], Jie Chen[2&], Hong-Yan Ding[1], Qin-Shi Pan[1], Wan-Dong Hong[3], Gang Xu[1], Fang-You Yu[1]*, Yu-Min Wang[1]*

Abstract

<u>Background</u>: The statistical methods to analyze and predict the related dangerous factors of deep fungal infection in lung cancer patients were several, such as logic regression analysis, meta-analysis, multivariate Cox proportional hazards model analysis, retrospective analysis, and so on, but the results are inconsistent. <u>Materials and Methods</u>: A total of 696 patients with lung cancer were enrolled. The factors were compared employing Student's t-test or the Mann-Whitney test or the Chi-square test and variables that were significantly related to the presence of deep fungal infection selected as candidates for input into the final artificial neural network analysis (ANN) model. The receiver operating characteristic (ROC) and area under curve (AUC) were used to evaluate the performance of the artificial neural network (ANN) model and logistic regression (LR) model. <u>Results</u>: The prevalence of deep fungal infection from lung cancer in this entire study population was 32.04% (223/696), deep fungal infections occur in sputum specimens 44.05% (200/454). The ratio of candida albicans was 86.99% (194/223) in the total fungi. It was demonstrated that older (≥65 years), use of antibiotics, low serum albumin concentrations (≤37.18g /L), radiotherapy, surgery, low hemoglobin hyperlipidemia (≤93.67g /L), long time of hospitalization (≥14days) were apt to deep fungal infection and the ANN model consisted of the seven factors. The AUC of ANN model(0.829±0.019)was higher than that of LR model (0.756±0.021). <u>Conclusions</u>: The artificial neural network model with variables consisting of age, use of antibiotics, serum albumin concentrations, received radiotherapy, received surgery, hemoglobin, time of hospitalization should be useful for predicting the deep fungal infection in lung cancer.

Keywords: Artificial neural network (ANN) - predictors - lung cancer - deep fungal infection

Introduction

The mortality of lung cancer mortality is the highest in all tumors and its incidence is gradually growing (Moore et al., 2010). Although surgical resection, chemotherapy and radiation therapy technology are continuously improving, patients with lung cancer remains extremely apt to relapse and fatal (Salim et al., 2011). The cure rate of lung cancer is extremely low and the average 5-year survival of patients with lung cancer is under 15% (Gridelli et al., 2003; Jemal et al., 2005; Ogawa et al., 2008; Chen et al., 2009).

In recent years, the deep fungal infections increased and had a large proportion of hospital infections, with extensive use of broad-spectrum antibiotics, immune-suppressants and anticancer drugs. Especially the fungal infection of elderly patients had significantly increasing trend with malignant tumors for their decreased immune function, severe malnutrition, surgery, radiotherapy, chemotherapy and prolonged use of broad-spectrum

antibiotics, as well as the original basis of disease. The deep fungal infection rate of patients with lung cancer showed high trend (Bereket et al., 2012). The deep fungal infection not only affected the treatment and rehabilitation, prolonged time of hospitalization, increased health care costs, but also significantly led to prognosis, even life-threatening (Bereket et al., 2012). To analyze the characteristics and dangerous factors of deep fungal infections in lung cancer patients, that will be used for clinicians adopt effective prevention and control of deep fungal infection for improving patient outcomes and survival in lung cancer patients.

The statistical methods to analyze and predict the related dangerous factors of deep fungal infection in lung cancer patients were several, such as logic regression analysis, meta-analysis, multivariate Cox proportional hazards model analysis, retrospective analysis , and so on , but the results are inconsistent. As one of the clinical prediction rules (Adams and Leveson, 2012), an artificial neural network (ANN) is consisted of a

[1]Department of Laboratory Medicine, [2]Department of Intensive Care Unit, [3]Department of Gastroenterology and Hepatology, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou, China &Equal contributors *For correspondence: wzjxyfy@163.com, wym0577@163.com

series of interconnecting parallel nonlinear processing elements (nodes) with limited numbers of inputs and outputs (Hong et al., 2011). A systematic review suggested that ANN is potentially more useful than conventional statistical techniques at predicting clinical outcomes when the relationship between the variables that determine the prognosis is complex, multidimensional and non-linear (Bartosch-Harlid et al., 2008). Chen J found that hospitalization, age, clinical type, hormone use is predictive dangerous factors of nosocomial infection in patients with lung cancer by ANN model and LR model (Chen et al., 2014). It was shown that the ANN can accurately predict the hepatitis B surface antigen negative (Zheng et al., 2014). There are no reports concerning the ANN model was used to predict fungal infections and relative dangerous factors of patients with lung cancer at home and abroad. The aim of this work was to develop an ANN to predict deep fungal infection in lung cancer.

Materials and Methods

Subjects

The 696 patients with lung cancer came from the First Affiliated Hospital of Wenzhou Medical University, China, from January April 2005 to December 2014. The above cases were confirmed by the histopathological results. The lung cancers consisted of 536 male and 160 women, aged between 31 to 90 years and including 54 small cell lung cancer, 642 non-small cell lung cancer (consisting of 400 adenocarcinoma, 240 squamous cell carcinoma, 2 large cell carcinoma). The criterion for the histopathologic diagnosis of lung cancer was the World Health Organization (WHO)/International Study of Lung Cancer (IASLC) lung cancer histological classification standards. TNM clinical stage was in the light of TNM clinical stage from American Joint Committee on Cancer (AJCC) and Union for International Cancer Control (UICC) in 2002.The following information was collected for each patient on admission: age, sex, clinical stage, histological classification, invasive procedures, mechanical ventilation, surgery, radiotherapy, chemotherapy, hemoglobin, serum albumin, white blood cell count, use of antibiotics, use of hormone, non-neoplastic lung disease, concurrent diabetes or renal insufficiency, smoking (smoking index=number of cigarettes smoked per day× smoking years), time of hospitalization. This examine was approved by the Institutional Ethics Review Board of the First Affiliated Hospital of Wenzhou Medical University and all patients provided written informed consent to this examine.

The culture and identification of fungi

Specimens include 454 cases sputum, 121 cases blood, and 121 urine sample. The culture and identification of fungi were used with Kemaijia chromogenic medium (France, Meraux Corporation).

Diagnostic criteria of deep fungi infection

The fungi infection diagnostic criteria and incidence statistics with reference to the Chinese Ministry of Health nosocomial infection diagnostic criteria in 2001. Confirmed information according to clinical manifestations, laboratory tests and/or identification of fungi.

Statistical analysis

Continuous values were expressed as the means±SD or medians and compared employing Student's t-test or the Mann-Whitney test. Categorical values were described by counts and proportions and compared applying the X^2 test. Variables that were significantly related to the presence of deep fungal infection were selected as candidates for input into the final ANN model. Independent variable importance analysis was performed to determine the optimum variables for construction of the final ANN model (Hong et al., 2011). An exploratory three-layer multiplayer perceptron (MLP) ANN model with a back propagation algorithm was constructed for independent variable importance analysis. The data were randomly divided into a training sample (557 cases, 80%) and a test sample (139 cases, 20%) in the exploratory ANN model. Sigmoid transfer functions were performed in the hidden and output layers. Gradient descent was performed to estimate the synaptic weights. In light of the results of the univariate and sensitivity analyses, a final three-layer feed-forward ANN model with a back propagation algorithm was constructed for all 696 patients. The ANN model was trained with a maximum of 500 iterations and 10 tours. The overfit penalty was assigned as 0.001, and the convergence criterion was 0.00001 (Hong et al., 2011). Fivefold cross-validation was performed (Balak et al., 2006). The output of the ANN model was transformed to range from 0-1. Deep fungal infection was predicted if the output was greater than or equal to 0.5 (Hong et al., 2011). The sensitivity, specificity, negative predictive value, positive predictive value and diagnostic accuracy of the ANN model are reported herein. A Fagan plot allows the reader to estimate the post-test probability of the target condition in an individual patient in light of a selected pretest probability (Whiting et al., 2008). Forward conditional step-wise logistic regression analysis was performed to develop a logistic regression function (LR) for comparison. The conditional probabilities for stepwise entry and removal of a factor were 0.05 and 0.06, respectively. The receiver operating characteristic (ROC) and area under curve (AUC) were performed to evaluate the performance of the ANN model and LR model. Differences were taken into account statistically significant if the two-tailed P value was less than 0.05. SPSS 18.0 (SPSS Inc., Chicago, IL, USA) was performed for ANN analysis.

Results

The distribution of deep fungal infection in patients with lung cancer

The prevalence of deep fungal infection from lung cancer in this entire study population was 32.04% (223/696), deep fungal infections occur in sputum specimens 44.05% (200/454), followed by urine sample17.36% (21/121), and blood1.65% (2/121). The ratio of Candida albicans was 86.99% (194/223) in the total deep fungal, Candida glabrata accounted for 7.62%

Use of an Artificial Neural Network to Construct a Model of Predicting Deep Fungal Infection in Lung Cancer Patients

(17/223), that of Candida tropicalis was 4.94% (9/223) and that of krusei was 1.35% (3/223).

Chi-Square and t-test/Mann-Whitney test results for deep fungal infection and non-infection from lung cancer patients

Our work found that the sex (X^2=6.257, p=0.012), invasive procedures (X^2= 6.313, p=0.012), surgery (X^2=39.634, p=0.000), radiotherapy (X^2=51.937, p=0.000), use of antibiotics (X^2=98.874, p=0.000), age (t=-5.135, p=0.000), hemoglobin (t=2.284, p=0.023), serum albumin (t=3.436, p=0.001), and time of hospitalization (Mann-Whitney test=47492.000, p=0.047) were statistically different between deep fungal infection and non-infection from lung cancer patients. The data was manifested in Table 1-2.

Univariate and multivariate analysis

Nine variables considered relevant to the presence of deep fungal infections were tested applying univariate and multivariate analyses. Multivariate analysis by logistic regression identified the following three independent variables as predictive of persistent deep fungal infections in lung cancer: sex (p=0.016), age (p=0.000), serum albumin (p=0.026), use of antibiotics (p=0.000), radiotherapy (p=0.011), surgery (p=0.000). A logistic regression function (LR model) was developed to predict nosocomial infections in lung cancer as follows: -0.24-0.585sex+ 0.041age (years)-0.038 serum albumin -1.947 use of antibiotics +0.051 radiotherapy +1.206 surgery.

ANN analysis

As manifested in Figure 1, age, and use of antibiotics, serum albumin, radiotherapy, surgery, hemoglobin and time of hospitalization were the most crucial predictors of nosocomial infections by sensitivity analysis (the exploratory ANN model constructed for the sensitivity analysis is not manifested). The final three-layer 5-5-1-feed-forward back propagation ANN model with variables consisting of age, use of antibiotics, serum albumin, radiotherapy, surgery, hemoglobin and time of hospitalization was developed and trained in 696 patients (Figure 2). The sensitivity, specificity, positive likelihood ratio, negative likelihood ratio of the ANN was 58.56%, 80.39%, 2.99 and 0.52, respectively. The ROC curves for the ANN model and LR model for predicting nosocomial infections in lung cancer patients are manifested in Figure 3. The AUC of the ANN model (0.829±0.019) was statistically higher than the AUC of the LR model (0.756±0.021, p=0.0041), Table3. We fatherly analyze

Table 1. The Demographics Characteristics of 696 NSCLC Patients and Chi-Square Results

Variable name	cases		X^2	P
	infection	non-infection		
sex			6.257	0.012
male	185	346		
female	46	119		
Clinical stage			3.267	0.514
Ia+Ib	1	4		
II	5	7		
IIIa	3	11		
IIIb	15	23		
IV	63	85		
Histologicaldifferentiation			1.705	0.426
Poorly differentiated	49	71		
Moderately differetiation	14	12		
Well-differentiation	58	40		
invasive procedures			6.313	0.012
yes	87	139		
no	136	333		
chemotherapy			19.328	0.199
yes	102	237		
no	120	232		
surgery			39.634	0.000
yes	54	34		
no	169	438		
radiotherapy			51.937	0.000
yes	68	86		
no	149	379		
mechanical ventilation			0.262	0.608
yes	17	31		
no	109	441		
use of antibiotics			98.874	0.000
yes	129	426		
no	94	46		
use of hormone			0.292	0.589
yes	22	53		
no	201	419		
benign lung disease			1.006	0.212
yes	118	363		
no	105	118		
diabetes			1.063	0.302
yes	25	66		
no	198	404		
renal insufficiency			2.505	0.113
yes	7	28		
no	216	442		

**Table 2. T-test/Mann-Whitney Test Results for Deep Fungal Infection and Non-Infection from Lung Cancer Patients**

| Variable name	Cases		t-test/Mann-Whitney test	P
	infection	non-infection		
age	68.55±9.93	64.23±11.08	-5.135	0
hemoglobin(g/L)	106.93±22.54	110.86±20.48	2.284	0.023
serum albumin(g/L)	32.97±6.47	34.58±5.24	3.436	0.001
white blood cell count(*109)	8.87(5.80-11.87)	8.00(5.80-12.10)	50266.000*	0.591
time of hospitalization(days)	17.00(10.00-33.00)	20.00(0.00-33.00)	47492.000*	0.047
Smoking index	500.00(0.00-1000.00)	400.00(0.00-860.00)	49438.000*	0.237

* Mann-Whitney test

Table 3. The Sensitivity, Specificity, and AUC of ANN Model and LR Model

|	Sensitivity (%)	Specificity (%)	AUC	LR+	LR-	PPV	NPV
ANN model	58.6	80.4	0.829±0.019	2.99	0.52	75.7	69.9
LR model	48.2	74.2	0.756±0.021	2.81	0.41	71.3	60.4

LR+: positive likelihood ratio, LR- : negative likelihood ratio, PPV: positive predictive value, NPV: negative predictive value
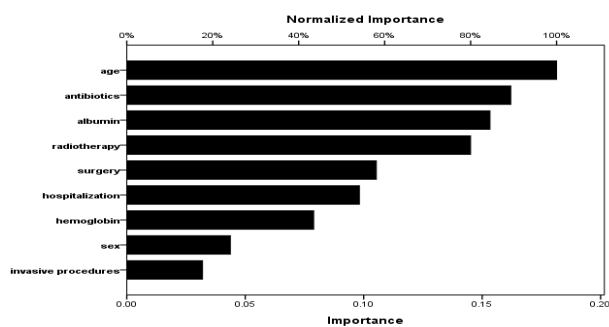
Figure 1. Sensitivity Analysis of the input Variables. The value demonstrated for each input variable is a measure of its relative importance
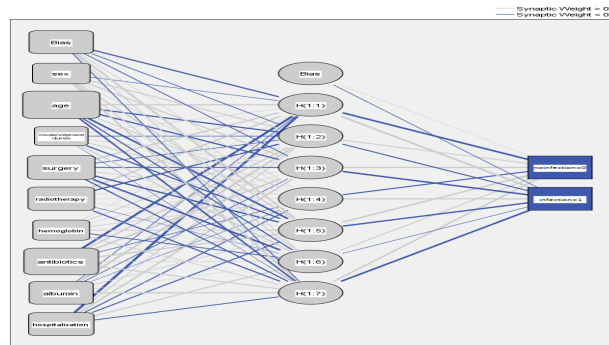
Figure 2. A Neural Network for the Prediction of Infection in Patients with Lung Cancer Consisting of Nine Input Variables, a Hidden Layer with Seven Nodes, and One Output Variable. The 1 is infection and 0 is non-infection
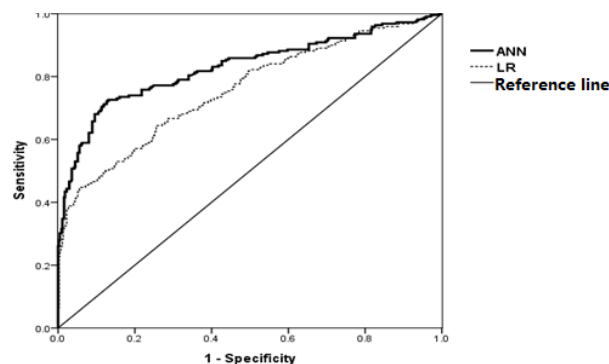
Figure 3. Receiver Operating Characteristic Curves for the ANN ?odel and LR ?odel. The AUC of the ANN model (0.829±0.019) was statistically higher than the AUC of the LR model (0.756±0.021)

to show that older (≥65 years of age, *p*=0.023), use of antibiotics (*p*=0.000), low serum albumin concentrations (less than 37.18g/L, *p*=0.003), radiotherapy (*p*=0.013), surgery (*p*=0.000), low hemoglobin hyperlipidemia (less than 93.67g/L, *p*=0.025), long time of hospitalization (≥14days, *p*=0.000) were apt to deep fungal infection and the ANN model consisted of the seven factors.

Discussion

The results of this work demonstrate that older, use of antibiotics, low serum albumin concentrations, received radiotherapy, received surgery, low hemoglobin hyperlipidemia; long time of hospitalization was the most crucial predictors of nosocomial infections. According to ROC analysis, the diagnostic performance of the ANN model was superior to that of the LR model.

Of the standard ANNs, the MLP is perhaps the most popular network architecture currently in use (Saftoiu et al., 2010). MLP models include an input layer, a hidden layer and an output layer. The artificial neurons are arranged in a layered feed-forward topology. Our ANN model was developed by applying the SPSS neural networks program and JMP software, which can both perform the MLP model (Balak et al., 2006; Hong et al., 2011). ANNs are nonlinear statistical data modeling tools. They can take into account outliers and nonlinear interactions among variables and can reveal previously unrecognized and/or weak relationships between given input variables and an outcome (Balak et al., 2006). Thus, ANNs often consist of parameters that may not reach significance making use of conventional statistics, as evidenced by the fact that hemoglobin and time of hospitalization included in our ANN model was not significant in logistic regression analysis.

We further analyze to imply that older (≥65 years of age, *p*=0.023), use of antibiotics (*p*=0.000), low serum albumin concentrations (less than 37.18g/L, *p*=0.003), radiotherapy (*p*=0.013), surgery (*p*=0.000), low hemoglobin hyperlipidemia (less than 93.67g/L, *p*=0.025), longer time of hospitalization (≥14days, *p*=0.000) were apt to deep fungal infection and the ANN model consisted of the seven factors. The searchers (Jiang et al., 2004) found that pulmonary fungal infection rate was 6.35% (78/1229). The major fungus was Candida albicans (68.18%). The main risk factors were age of > or =50 years (*p*<0.005), primary site (lung cancer, *p*<0.001), pulmonary radiotherapy (*p*<0.001), chemotherapy (*p*<0.001), and long-term hospitalization (>2 weeks, *p*<0.005). Our results were mostly consistent to Jiang Y. Improved hospitalization increased the chances of opportunistic infections. Older were prone to infection for decreased immunity. The employ of antibiotics to suppress the normal flora and is conducive to fungal growth. The immune of patients declined for receiving radiotherapy and surgery. The serum hypoalbuminemia and hyperlipidemia could reflect the malnutrition of the patients. These factors fatherly triggered fungal infection.

In conclusion, an artificial neural network model with variables consisting of age, employ of antibiotics, serum albumin concentrations, radiotherapy, surgery, hemoglobin, time of hospitalization may be useful for predicting the fungal infection in patients with lung cancer.

Acknowledgements

References

Adams ST, Leveson SH (2012). Clinical prediction rules. *BMJ*, **344**, 8312.

Balak MN, Gong Y, Riely GJ, et al (2006). Novel D761Y and common secondary T790M mutations in epidermal growth factor receptor-mutant lung adenocarcinomas with acquired resistance to kinase inhibitors. *Clin Cancer Res*, **12**, 6494-501.

Bartosch-Harlid A, Andersson B, Aho U, et al (2008). Artificial neural networks in pancreatic disease. *Br J Surg*, **95**, 817-26.

Bereket W, Hemalatha K, Getenet B, et al (2012). Update on bacterial nosocomial infections. *Eur Rev Med Pharmacol Sci*, **16**, 1039-44.

Chen CH, Lai JM, Chou TY, et al (2009). VEGFA upregulates FLJ10540 and modulates migration and invasion of lung cancer via PI3K/AKT pathway. *PloS One*, **4**, 5052.

Chen J, Pan QS, Hong WD, et al (2014). Use of an artificial neural network to predict risk factors of nosocomial infection in lung cancer patients. *Asian Pac J Cancer Prev*, **15**, 5349-53.

Gridelli C, Rossi A, Maione P (2003). Treatment of non-small-cell lung cancer: state of the art and development of new biologic agents. *Oncogene*, **22**, 6629-38.

Hong WD, Ji YF, Wang D, et al (2011). Use of artificial neural network to predict esophageal varices in patients with HBV related cirrhosis. *Hepat Mon*, **11**, 544-7.

Jemal A, Murray T, Ward E, et al (2005). Cancer statistics, 2005. *CA Cancer J Clin*, **55**, 10-30.

Jiang Y, Li JY, Li M, et al (2004). Clinical analysis of nosocomial pulmonary fungal infection in patients with cancer. *Ai Zheng*, **23**, 1707-9.

Moore MA, Ariyaratne Y, Badar F, et al (2010). Cancer epidemiology in South Asia-past, present and future. *Asian Pac J Cancer Prev*, **11**, 49-66.

Ogawa E, Takenaka K, Katakura H, et al (2008). Perimembrane Aurora-A expression is a significant prognostic factor in correlation with proliferative activity in non-small-cell lung cancer (NSCLC). *Ann Surg Oncol*, **15**, 547-54.

Saftoiu A, Vilmann P, Gorunescu F, et al (2012). Efficacy of an artificial neural network-based approach to endoscopic ultrasound elastography in diagnosis of focal pancreatic masses. *Clin Gastroenterol Hepatol*, **10**, 84-90.

Salim EI, Jazieh AR, Moore MA (2011). Lung cancer incidence in the arab league countries: risk factors and control. *Asian Pac J Cancer Prev*, **12**, 17-34.

Whiting PF, Sterne JA, Westwood ME, et al (2008). Graphical presentation of diagnostic information. *BMC Med Res Methodol*, **8**, 20.

Zheng MH, Seto WK, Shi KQ, et al (2014). Artificial neural network accurately predicts hepatitis B surface antigen seroclearance. *PloS one*, **9**, 99422.