Application of BP and RBF Neural Network in Classification Prognosis of Hepatitis B Virus Reactivation

Wu Guan-peng[1], Wang Shuai[1], Huang Wei[2], Liu Tong-hai[2], Yin Yong[2], Liu Yi-hui[1]

[1]School of Information, Qilu University of Technology, Jinan, China

[2]Department of Radiation Oncology, Shandong Cancer Hospital, Shandong Academy of Medical Sciences, Jinan, China

Email address:
zbxywgp@163.com (Wu Guan-peng)

Abstract: This study aims at finding the risk factors (the key feature subset) and building the classification prognosis model of hepatitis B virus (HBV) reactivation after precise radiotherapy (RT) in patients with primary liver carcinoma. We find out that the outer margin of RT, TNM of tumor stage and the HBV DNA levels are the risk factors ($P<0.05$) of HBV reactivation by feature extraction method of logistic regression analysis in this article. The feature extraction method reduced the dimension and improved the classification accuracy. Establish the classification prognosis model of BP and RBF neural network for original data set and the key feature subset. The experimental results show that BP and RBF neural network have good performance in classification of HBV reactivation.

Keywords: Primary Liver Carcinoma, HBV Reactivation, Feature Extraction, BP, RBF

1. Introduction

Primary liver carcinoma (PLC) is extremely prevalent malignant tumors in South China and mostly patients were infected with hepatitis B virus (HBV). In recent years, precise radiotherapy (RT) is a good method for treating primary liver carcinoma. Precise radiotherapy method can cause HBV [1] reactivation for PLC patients and the rate of HBV reactivation approximately reached to 25%. The mortality reached to 25% for patients after HBV reactivation. Finding the risk factors of HBV reactivation has important signification for PLC patients with HBV reactivation. In 2007, Kim [2] researched 32 PLC patients who underwent three-dimensional conformal radiotherapy (3D-CRT). Their study showed that the HBV reactivation may be associated with the HBV DNA levels. Wu [3] [4] speculated that HBV reactivation has closely associated with the HBV DNA replication. In 2013, Huang [5] studied 69 PLC patients who underwent RT. They adopted logistic regression analysis affection on HBV reactivation of clinical features. Their results showed that the HBV DNA levels were the independent risk factors which lead to HBV reactivation. In 2014, they found dosimetric parameters also

were high risk factors which lead to HBV reactivation [6]. We still need to research the risk factors of HBV reactivation. So far, there is no the classification prognosis model of HBV reactivation. This study aims at finding the risk factors and building the classification prognosis model of HBV reactivation after precise radiotherapy in PLC patients.

Feature extraction has been widely used in complex data analysis in biomedical field. Such as Feature extraction [7, 8] applied in mass spectrometry data analysis. Artificial neural network [9] [10] [11] [12] is deemed to important intelligent algorithm in pattern recognition. It has been widely used in a variety of biomedical respects. For example, used in magnetic resonance spectroscopy in hepatocellular carcinoma based on neural network [13] [14] [15]. We find out that the outer margin of RT, TNM of tumor stage and the HBV DNA levels are the risk factors ($P<0.05$) of HBV reactivation by feature extraction method of logistic regression analysis in this article. The feature extraction method reduced the dimension and improved the classification accuracy. Establish the classification prognosis model of BP and RBF neural network for original data set and the key feature subset. The experimental results show that BP and RBF neural network have good performance in classification of HBV reactivation.

2. Data and Feature Extraction Methods

2.1. Data

The research data comes from Shandong cancer hospital treated 90 PLC patients after precise radiotherapy. 20 PLC patients were occurred HBV reactivation, the rest of patients didn't occurred HBV reactivation. Each patient is considered as a research sample. Each sample includes 30 features.

2.2. Feature Extraction Methods

We adopt two independent samples t-test, χ^2 test and rank sum test three methods to extract significant factors which are affecting on HBV reactivation. Put significant factors into logistic regression analysis for finding the risk factors. The risk factors are considered as the key feature subset that has good performance for distinguishing classification.

2.2.1. Two Independent Samples t-test Method

Two independent samples t-test [16] is suitable for normal distribution or approximating normal distribution of sample with enumeration data. There is no connection between two samples. Mathematical formula is:

$$t = \left(\left| \overline{X}_1 - \overline{X}_2 \right| \right) \Big/ S_{\overline{X}_1 - \overline{X}_2} \qquad (1)$$

$\overline{X}_1$, $\overline{X}_2$ represent the mean of sample 1 and sample 2 respectively. $S_{\overline{X}_1 - \overline{X}_2}$ represent the standard error of mean differences for $\overline{X}_1$ and $\overline{X}_2$.

2.2.2. χ^2 Test Method

χ^2 test [17] is suitable for qualitative data analysis of sample. It can use to check the connection of feature. Mathematical formula is as follows.

$$\chi^2 = \sum \frac{(A - T)^2}{T} \qquad (2)$$

The A is the actual value, T is theoretical value.

2.2.3. Rank Sum Test

Rank sum test [18] is nonparametric hypothesis test. It doesn't consider the distribution of sample. It is suitable for the data with abnormal distribution, ranked data and heterogeneity of variance.

2.2.4. Logistic Regression Analysis

The outer margin of RT, TNM of tumor stage and the HBV DNA levels are the risk factors of HBV reactivation by feature extraction method with logistic regression analysis [19]. Mathematical formula as follows:

$$y = b_0 + b_1 X_1 + \cdots + b_k X_k \qquad (3)$$

b_k is partial regression coefficient, X_k is independent variable, y is true or false value.

3. BP and RBF Neural Network

3.1. BP Neural Network

BP neural network [20] is able to classify and recognize the complex data. There are three layer structure of classic BP neural network, including input layer, hidden layer output layer. As shown in figure 1.
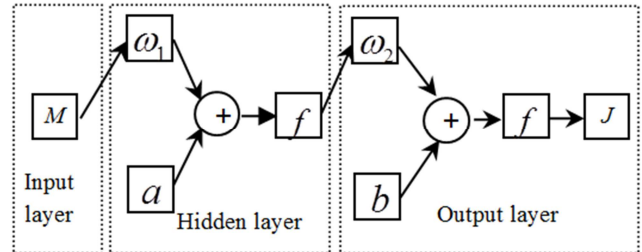

Figure 1. The structure of BP neural network.

Input layer: M represent the neurons of input. Hidden layer I consists of the weight ω_1 between input layer and hidden layer, threshold value a, accumulator and function f. Output layer J consists of the weight ω_2 between hidden layer and output layer, threshold value b, accumulator and function f. In this article, $M=30$, $I=5$, $J=2$ for original data; $M=3$, $I=2$, $J=2$ for feature subset. Function f adopted Sigmoid. Training times 1000, learning rate 0.05, training precision 0.001.

3.2. RBF Neural Network

RBF neural network [21] has a good performance in dealing with local optimum. There are three layers structure similar to BP, including input layer, hidden layer and output layer. As shown in figure 2.
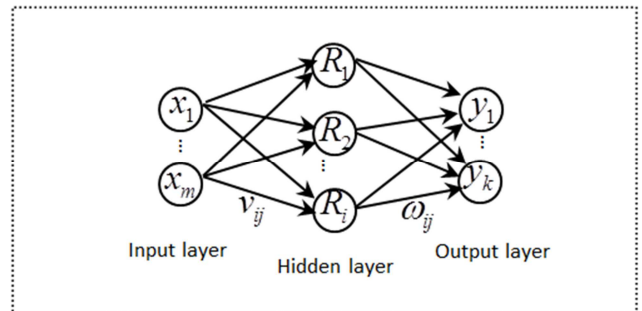

Figure 2. The structure of RBF neural network.

RBF neural network finished nonlinear mapping from input layer x_m to hidden layer R_i by connectional weight v_{ij} and linear mapping from hidden layer R_i to output layer y_k by connectional weight ω_{ij}. Hidden layer R_i is given by Gauss function. Gauss function is defined as follows:

$$R_i(x) = \exp\left[-\frac{\|x - \omega_i\|}{2\sigma^2} \right] \qquad (4)$$

Input layer x_m is the number of feature. Output layer y_k represent the number of output result that HBV reactivation and non-reactivation. Spread of radial basis functions adopted default: Spread = 1.0.

4. Experimental Results of Feature Extraction

4.1. Experimental Results of Two Independent Samples *t* -test

The enumeration data adopted two independent samples *t* -test method. The experimental results show that the outer margin of RT is correlated with HBV reactivation (P<0.05), and HBV reactivation had no correlation with other parameter. As shown in table 1.

Table 1. *Two independent samples t -test results of enumeration data.*

Parameter	Mean	Standard error	p-value
Age, years	56.14	10.63	0.866
AFP, ng/ml	630.98	1022.58	0.644
Radiotherapy dose, Gy	57.93	7.19	0.917
Equivalent biometric, Gy	69.969	8.198	0.891
Radiotherapy times, times	28.48	5.484	0.871
GTV, cm3	179.59	228.76	0.570
PTV, cm3	392.07	318.93	0.915
outer margin of RT(mm)	11.04	2.76	0.012
V5(%)	51.645	17.776	0.751
V10(%)	16.357	1.724	0.722
V15(%)	37.216	14.638	0.977
V20(%)	31.299	13.262	0.862
V25(%)	25.643	11.448	0.782
V30(%)	21.205	10.304	0.635
V35(%)	17.015	8.513	0.786
V40(%)	13.352	7.099	0.982
V45(%)	10.169	6.308	0.393
Dmax, Gy	6902.6	1160.4	0.941
Dmean, Gy	1597.1	623.8	0.733

AFP, α -fetoprotein; GTV, gross tumor volume; PTV, planning target volume; Dmax, maximum dose; Dmean , Mean dose.

4.2. Experimental Results of χ^2 Test

The measurement data adopted χ^2 test method. The experimental results show that the measurement data had no correlation with HBV reactivation. As shown in table 2.

Table 2. χ^2 *test results of measurement data.*

Parameter	Minimum expected value T	χ^2 value	P-value
Gender	8.44	0.81	0.775
KPS grade	7.56	1.634	0.201
HBeAg	7.56	0.54	0.816
PVTT	7.56	0.54	0.816
Child-Pugh	7.56	1.634	0.201
Split method	2.44	0.02	0.966

HBeAg, hepatitis Be antigen; PVTT, portal vein tumor thrombus;

4.3. Experimental Results of Rank Sum Test

The HBV DNA levels and code of outer margin of RT had correlation with HBV reactivation. The P value of TNM tumor stage is so close to 0.05 that is deemed correlation with HBV reactivation. As shown in table 3.

Table 3. *Rank sum test results of measurement data.*

Parameter	χ^2 value	P-value
TNM tumor stage	3.406	0.065
HBV DNA levels	8.631	0.03
code of outer margin of RT	6.729	0.09
TACE times before RT	0.16	0.899

4.4. Experimental Results of Logistic Regression Analysis

The outer margin of RT, TNM of tumor stage and the HBV DNA levels are the risk factors of HBV reactivation. As shown in table 4.

Table 4. *Logistic regression analysis results.*

Parameter	(B)	SE	Exp(B)	P-value	95% C.I for Exp(B)	
					Ceiling	Inferior
outer margin of RT	0.558	0.160	1.747	0.001	2.392	1.276
TNM tumor stage	1.566	0.592	4.785	0.008	15.281	1.498
HBV DNA levels	1.630	0.463	5.103	0.000	12.663	2.061
code of outer margin of RT	-0.876	1.195	0.416	0.463	4.328	0.040

B, coefficient; SE, standard error; Exp(B), coefficient logarithm.

5. Experimental Evaluation Methods and Results of BP and RBF Neural Network

5.1. Experimental Methods of K-fold Cross Validation and Performance Evaluation Standard of Classifier

Experiment took the k-fold cross validation in order to get the accuracy results. The k-fold cross validation defined as follows:

$$\hat{u}_k = \frac{1}{k} \sum_1^k u_k \ .$$

The $\hat{u}_k$ is the mean value of sum of u_k by k-fold cross validation. The k is the number of sample. The k value selected 3, 5, 10.

Define five formulas for evaluating the performance of classifier, including accuracy, sensitivity, specificity, balance accuracy (BACC), positive predictive value (PPV). The accuracy is the most important evaluation standard for classifier.

5.2. Experimental Results of BP and RBF Neural Network

Build the classification prognosis model of BP and RBF neural network for original data set (table 5) and the key feature subset (table 6) of HBV reactivation. Experimental results show that BP and RBF neural network have a good classification performance of HBV reactivation. As shown in following table.

Table 5. BP and RBF neural network experimental results of original data.

Neural network	k value	Accuracy	Sensitivity	Specificity	BACC	PPV
BP	3	0.7333	0.7857	0.5500	0.6678	0.8286
	5	0.6889	0.7429	0.5000	0.6214	0.8087
	10	0.7000	0.7857	0.4000	0.5927	0.8209
RBF	3	0.7556	0.8143	0.5500	0.6821	0.8636
	5	0.6778	0.6571	0.7500	0.7036	0.9020
	10	0.6555	0.6714	0.6000	0.6357	0.8545

Table 6. BP and RBF neural network experimental results of the key feature subset.

Neural network	k value	Accuracy	Sensitivity	Specificity	BACC	PPV
BP	3	0.7889	0.8143	0.7000	0.7571	0.9048
	5	0.7444	0.7714	0.6500	0.7107	0.8852
	10	0.7111	0.7571	0.5500	0.6535	0.8548
RBF	3	0.8000	0.8857	0.5000	0.6929	0.8611
	5	0.7556	0.8286	0.5000	0.6643	0.8529
	10	0.7222	0.7714	0.5500	0.6607	0.8571

The accuracy was worst when the k value adopted 10 whatever the model is BP or RBF neural network. These results due to samples of training were very less. The accuracy of results increased with the samples of training increasing. The obvious result showed by compared to table 5 and table 6. The classification performance was optimal when the experiment adopted 3 fold cross validation. The key feature subset has better classification performance than original data whatever the model is BP or RBF neural network. The classification accuracy of BP increased from 73.33% to 78.89%. Sensitivity: 81.43%; Specificity: 70.00%; BACC: 75.71%; PPV: 90.48%. The classification accuracy of RBF increased from 75.56% to 80.00%. Because of RBF has better learning ability and approximate consistency than BP. BP is easily fall into local optimum, especially for small sample data. So, the classification accuracy of RBF is superior to BP.

The experimental results show that BP and RBF neural network have good performance in classification of HBV reactivation. The classification performance of the key feature subset is superior to original data set. The feature extraction method reduced the dimension and improved the classification accuracy.

6. Conclusion

The classification prognosis model of BP and RBF has higher application value for HBV reactivation after precise radiotherapy in PLC patients. We can take antiviral treatment and liver protective measures for had infected HBV in PLC patients in order to improve their quality of survival and prolong survival time. In the future, we will continue to study other intelligent algorithms for application in HBV reactivation, in order to improve the classification performance.

Acknowledgements

References

[1] Yeo W, Lam KC, Zee B, et al. Hepatitis B reactivation in patients with hepatocellular carcinoma undergoing systemic chemotherapy [J]. Annals of Oncology, 2004, 15(7): 1661-1666.

[2] Kim JH, Park J W, Kim TH, et al. Hepatitis B virus reactivation after three-dimensional conformal radiotherapy in patients with hepatitis B virus-related hepatocellular carcinoma [J]. International Journal of Radiation Oncology Biology Physics, 2007, 69(3): 813–819.

[3] Xiao-An W U, Zhang Z Y, Hong J Y. Therapeutic Effect of Three-Dimensional Conformal Radiotherapy in the Treatment of Primary Hepatoellular Carcinoma [J]. Practical Journal of Cancer, 2007.

[4] Xiao-An W U, Zhang Z Y, Yan Z B, et al. Clinical study of three-dimensional conformal radiation therapy for primary hepatoellular carcinoma [J]. Chinese Clinical Oncology, 2008.

[5] Wei H, Yanda L, Min F, et al. Analysis of hepatitis B virus reactivation inducedby precise radiotherapy in patients with primary liver cancer [J]. Chinese Journal of Radiation Oncology, 2013, 22(3): 193-197.

[6] Wei H, Wei Z, Min F, et al. Risk factors for hepatitis B virus reactivation after conformal radiotherapy in patients with hepatocellular carcinoma. [J]. Cancer Science, 2014, 105(6): 697-703.

[7] Liu Y H. Feature extraction and dimensionality reduction for mass spectrometry data [A]. Computers in Biology and Medicine, 2009, 39: 818-823.

[8] Liu Y, Bai L. Find Key m/z Values in Predication of Mass Spectrometry Cancer Data [C]// Advanced Intelligent Computing Theories and Applications. With Aspects of Theoretical and Methodological Issues, 4th International Conference on Intelligent Computing, ICIC 2008, Shanghai, China, September 15-18, 2008, Proceedings. 2008:196-203.

[9] Agarwal M, Jain N, Kumar M, H Agrawal. Face Recognition Using Eigen Faces and Artificial Neural Network [C]// IEEE Computer Society Conference on Computer Vision & Pattern Recognition, Cvpr. 2010:624-629.

[10] Scanzio S, Cumani S, Gemello R, F Mana, P Laface. Parallel implementation of Artificial Neural Network training for speech recognition [J]. Pattern Recognition Letters, 2010, 31(11):1302-1309.

[11] Yang Z, Zhang H. Prediction of melt rate of vibrating-electrode Electroslag Remelting process using artificial neural network [C]// Information Science and Technology (ICIST), 2015 5th International Conference on. IEEE, 2015.

[12] Liu M, Tsai H P. USB3.1 silicon and channel design optimization using artificial neural network modeling [C]// Electromagnetic Compatibility and Signal Integrity, 2015 IEEE Symposium on. IEEE, 2015.

[13] Wang S Q, Liu Y H, Wang L J, et al. ~(31)P-MRS data analysis of liver based on back-propagation neural networks [J]. Chinese Journal of Medical Imaging Technology, 2009, 25(10):1875-1878.

[14] Qiang L, Liu Y H, Wang L J, JY Cheng. ~(31)P-MRS data analysis of liver based on self-organizing map neural networks [J]. Journal of Medical Imaging, 2009, 2:151-153.

[15] Liu Q, Wang N, Liu Y H, et al. 31P MRS Data Analysis of Liver Cancer Based on Neural Networks [J]. Applied Mechanics & Materials, 2010, 20:630-635.

[16] Farrell-Singleton P A. Critical values for the two independent samples Winsorized t test[J]. Dissertations & Theses - Gradworks, 2010.

[17] Mchugh M L. The chi-square test of independence. [J]. Biochemia Medica, 2013, 23(2): 143-9.

[18] Murakami H. The power of the modified Wilcoxon rank-sum test for the one-sided alternative [J]. Statistics A Journal of Theoretical & Applied Statistics, 2014, 49(4):1-14.

[19] Yao H, Gong J, Li L I, Do Radiotherapy, SS Hospital. Risk Factors of Hepatitis B Virus Reactivation Induced by Precise Radiotherapy in Patients with Hepatic Carcinoma [J]. Practical Journal of Cancer, 2014.

[20] Li X, Liu S, Hui L, T Zhang. Research on Prediction Model of Bending Force Based on BP Neural Network with LM Algorithm [C]// the 25th Chinese Control and Decision Conference. 2013:832-836.

[21] Han H G, Qiao J F. Adaptive Computation Algorithm for RBF Neural Network [J]. IEEE Transactions on Neural Networks & Learning Systems, 2012, 23(2):342-347.