SciencePG
Science Publishing Group

The Predictive Model of Hepatitis B Virus Reactivation Induced by Precise Radiotherapy in Primary Liver Cancer

Wang Shuai[1], Wu Guan-peng[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:
932203027@qq.com (Wang Shuai)

Abstract: In this paper, to build a predictive model of hepatitis B virus (HBV) reactivation in primary liver cancer (PLC) patients after precise radiotherapy (RT). Logistic regression analysis was adopted to extract the optimal feature subset, TNM, HBV DNA level and outer margin of RT were risk factors for HBV reactivation ($P < 0.05$). A predictive model of support vector machine (SVM) was established for the optimal feature subset and all of PLC data sets. The experimental results proved that the former obviously improves the classification accuracy, which increased from 74.44% to 78.89%. In this paper, it is concluded that TNM, HBV DNA levels and outer boundary are the risk factor for HBV reactivation ($P < 0.05$).

Keywords: Primary Liver Cancer, Data Set, Feature Extraction, Support Vector Machine (SVM)

1. Introduction

Primary liver cancer (PLC) is one of the most common malignant tumors in China.The clinical discovery of patients with advanced, only about 10% can be resected, and most patients need conservative treatment. In recent year. Conformal radiotherapy three-dimensional (3D-CRT) and modulated radiotherapy intensity (IMRT) have been widely used in clinical practice for the treatment of advanced PLC patients [1-3]. However, the HBV of patients with the primary liver cancer are activated easily after precise radiotherapy [4-6]. The influence factors of HBV reactivation remains to be research, and related prediction model needs to be established.

In 2007 Wu etc [7] reported 86 cases of PLC patients with 3-DCRT, among of them 1 case combined hepatitis b patients with cirrhosis appears liver atrophy in radiotherapy after 6 months, the amount of HBV DNA exponentially increases continuously and the liver was damaged to death. The author concluded that the cause of death is closely related to HBV reactivation. Huang etc [8] retrospective analysis of clinical characteristics of 69 cases of HBsAg positive PLC patients with HBV reactivation after precise radiotherapy. They used the statistics of logistic method to evaluate the effects of each index of HBV reactivation, and the result show that the baseline blood HBV DNA levels was the risk factor of HBV reactivation.

At present the intelligent computing technology has been widely used in biomedical field to analysis the complex data. Such as Zhang etc [9] used the support vector machine (SVM) to predict esophageal squamous carcinoma postoperative survival to provide guidance for clinical diagnosis of thyroid nodules; Gao [10] described the application of SVM in speech recognition. We use the logistic regression analysis to extract the features of the feature subset in this paper. Firstly we acquire the tumor staging TNM, HBV DNA level and outer boundary were the risk factors for HBV virus activation ($P < 0.05$). Then we use the SVM to establish the classification prediction model of the optimal feature subset after feature extraction and all primary liver cancer data sets.

2. Data and Feature Extraction

2.1. Data

We select Shandong province tumor hospital treated 90

liver cancer patients after precise radiotherapy as the research object, namely 90 samples, and each sample has the 30 characteristics. The matrix is the 90*30. All cases had complete record by clinical examination, B ultrasonic examination, abdominal CT or pathological examination, and the diagnosis is the primary liver after that. There are 52 males and 38 females. The range of the age is 30 to 74 years old, the average age is 56.1 years old. In the 90 patiens, there are 20 patiens of HBV reactivation and 70 patiens of HBV without activation.

2.2. Feature Extraction

We use the SPSS 17.0 software to screen the risk factors by the single factor logistic method, then use the Binarys logistic regression method to analyze many factors. Putting the statistically significant single factor into logistic regression analysis [11].

2.2.1. Independent Sample T-test
Independent samples t-test was performed on the various influencing factors of primary liver cancer data set. Independent sample means there is no connection between two samples each other. Two independent samples accept the same measurements respectively.

2.2.2. Rank Sum Test
Rank sum test is a nonparametric statistics. It is needn't to consider the distribution of the data. We use the rank-sum test with comparing the multiple samples in this article.

2.2.3. Chi Square Analysis
Chi-square analysis is a kind of widely used hypothesis test method. It is suitable for data classification of variables (measurement data).

2.2.4. Multiple Factors Analysis
Each variable turns into logistic regression analysis follow the effect from high to low. Every time introducing an independent variable, we will take a significance test on the role of each independent variable in the regression equation. We just select significant meaning of the independent variables.

3. Support Vector Machine

The SVM is a linear separable cases surface theory which is mainly based on the optimal classification. optimal classification is able to rightly separate the two different categories (that is, to achieve the training error rate is 0) while the classification interval is enough far. SVM classification is ensure the both sides of the blank area (margin) reach maximum [12]. Figure 1 for the optimal classification plane.

In the figure, black and white points represent different types of samples. H is the optimal classification on the surface of the straight line. H1 and H2 are respectively classification line H recent samples in a straight line.
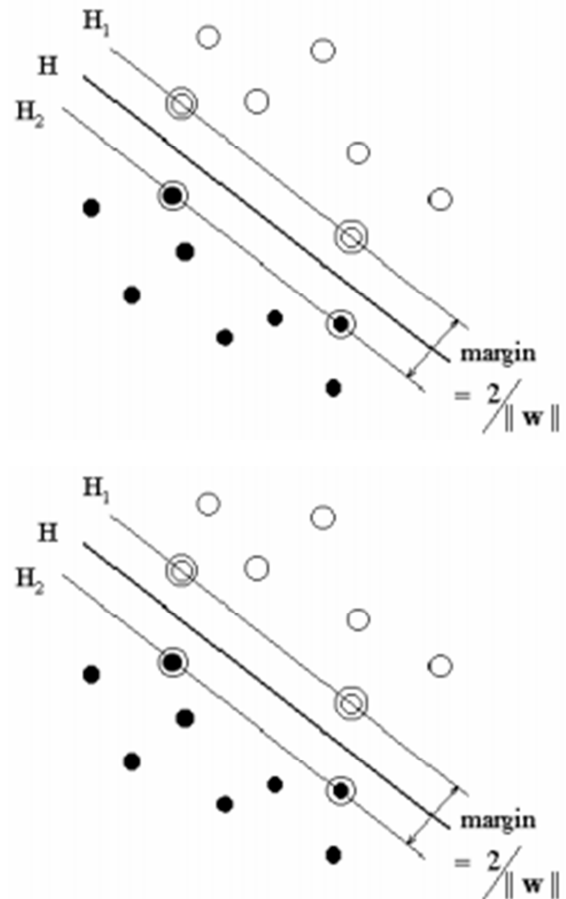

Figure 1. The optimal classification plane.

4. Experimental Analysis

We use SVM classification to classify the original data and the feature data after statistical analysis respectively. The matrix size of the original data set is 90*30. The matrix size of feature data set is 90*3.

4.1. Cross Validation

Cross validation is a commonly used machine learning method in data sampling, it has many different forms. We uses the K fold cross validation (k-fold cross-validation) in this article. The basic process is that the samples were divided into K sets equally. The k-1 sets is used to training sample and leaving set is used to test samples. This process is repeated K times, and putting average value of the test error in the K times process as generalization error [13].

4.2. Feature Extraction

4.2.1. Independent Samples T-test
The table 1 data takes independent sample T test method, the result as follows. We concluded that there is a significant relationship between external boundary and HBV reactivation ($P < 0.05$). While to the age, there is no significant relationship between liver with such factors as the maximum dose and HBV virus reactivation.

Table 1. *Effect of HBV virus reactivation count data independent sample T test analysis.*

Factors (units)	average	standard deviation	p-values
Age (years)	56.14	10.626	0.866
AFP (ng/ml)	630.977	1022.58	0.565
Total dose radiation (Gy)	57.929	7.1907	0.917
Equivalent biometric (Gy)	69.969	8.1987	0.891
Number of radiotherapy (time)	28.48	5.484	0.871
GTV volume (cm^3)	179.59	228.768	0.441
PTV volume (cm^3)	392.0676	318.933	0.891
MPTV (mm)	11.04	2.764	0.012
V5 (%)	51.645	17.77625	0.723
V10 (%)	16.35714	1.72419	0.696
V15 (%)	37.216	14.63808	0.975
V20 (%)	31.2992	13.26243	0.859
V25 (%)	25.6433	11.44842	0.782
V30 (%)	21.2053	10.30448	0.635
V35 (%)	17.0147	8.51392	0.786
V40 (%)	13.3516	7.09971	0.977
V45 (%)	10.1686	6.30801	0.287
Maximum dose (Gy)	6902.56	1160.37	0.562
The average dose (Gy)	1597.09	623.795	0.689

4.2.2. Rank Sum Test

The table 2 data takes rank sum test, the result as follows. We concluded that HBV DNA levels, put a boundary encoding, outside the two classification of encoding has significant relationship with HBV reactivation ($P < 0.05$).The P value of TNM is close to 0.05. So we think that TNM is related to the HBV reactivation.

Table 2. *Rank sum test analysis of all the influencing factors of primary liver cancer data set.*

	Mann - Whitney U	Wilcoxon W	Z	p-values
TNM	527.500	3012.500	-1.845	0.065
HBV baseline three categories	416.000	2901.000	-2.938	0.003
Outside the boundary encoding	446.500	2931.500	-2.594	0.009
The outer boundary are two classification and coding	480.000	2965.000	-2.475	0.013

4.2.3. Chi-square Analysis

The table 3 data takes chi-square analysis method, the result as follows. We conclude that there is no significant relationship between the table 3 data with HBV reactivation.

Table 3. *Affect HBV reactivation of measuring parameters of the chi-square analysis.*

factor	case number	reactivation	chi-square value	p-values
sex				
man	52	11	0.081	0.486
woman	38	9		
HbeAg				
masculine	34	8	0.054	0.507
feminine	56	12		
PVTT				
exist	56	12	0.054	0.507
nothing	34	8		

factor	case number	reactivation	chi-square value	p-values
split method				
routine	79	17	0.02	0.966
Big	11	3		
TACE times before radiotherapy				
one	12	2		
two	37	9	0.323	0.851
three	40	9		

4.2.4. Multivariate Analysis

The table 4 data takes multivariate analysis method, the result as follows. We found that the baseline serum HBV DNA levels, outer boundary, TNM is the risk factor for HBV reactivation.

Table 4. *The dangerous factors of HBV reactivation occurred multiariable Logistic regression analysis results.*

factor	B	S.E,	p.	Exp (B)
HbeAg	-.059	.706	.934	.943
TNM	1.626	.615	.008	5.085
HBV baseline	1.710	.479	.000	5.530
MPTV (mm)	.744	.352	.035	2.104
Outside the boundary encoding	-1.206	1.308	.356	0.299
The outer boundary are two classification and coding	1.702	1.463	.245	5.488

4.3. Support Vector Machine (SVM) Results

We use SVM classification to classify the original data and the feature data after statistical analysis, we use K fold cross validation to get the classification. Results are shown in the following table.

Table 5. *The classification of the original data under different cross validation results.*

k-Fold	accuracy	sensitivity	specificity
3	0.7444	0.7887	0.5789
5	0.7111	0.7606	0.5263
10	0.7222	0.7606	0.5789

Table 6. *The feature extraction of the classification of the related data under different cross validation results.*

k-Fold	accuracy	sensitivity	specificity
3	0.7889	0.8028	0.6842
5	0.7222	0.7606	0.5789
10	0.7222	0.7746	0.5263

The matrix size of the original data set is 90*30. The SVM use 10 fold cross-validation, we concluded that the accuracy is 72.22%, sensitivity is 76.06%, specificity is 57.89%. The SVM use 5 fold cross-validation, we concluded that the accuracy is 71.11%, sensitivity is 76.06%, specificity is 52.63%. The SVM use 3 fold cross-validation, we concluded that the accuracy is 71.11%, sensitivity is 76.06%, specificity is 52.63%.

The matrix size of feature data set is 90*30. The SVM use 10 fold cross-validation, we concluded that the accuracy is

72.22%, sensitivity is 77.46%, specificity is 52.63%. The SVM use 5 fold cross-validation, we concluded that the accuracy is 72.22%, sensitivity is 76.06%, specificity is 57.89%. The SVM use 3 fold cross-validation, we concluded that the accuracy is 78.89%, sensitivity is 80.28%, specificity is 68.42%.

5. Conclusion

In this paper, we concluded that TNM, HBV DNA levels and outer boundary are the risk factor for HBV reactivation ($P < 0.05$). The data set after feature extraction had the highest accuracy, sensitivity and specificity by using the support vector machine (SVM) and 3 fold cross-validation. The accuracy is 78.89%, sensitivity is 80.28%, specificity is 68.42%.

Acknowledgements

References

[1] Lok ASF, Lai CL, Wu PC, et al. Hepatitis B virus infection in Chinese families in Hong Kong. Am J Epidemiol, 1987, 126(3): 492-499.

[2] Yao Hui Gong Jinlan, lily, tinna. Accurate liver cancer patients after radiotherapy of HBV virus reactivation risk factor analysis [J]. Journal of cancer, 2014, 29 (6): 675-677.

[3] Yang Binghui. Primary liver cancer / / China anti-cancer association. New standard of diagnosis and treatment of common malignant tumors. Beijing: Beijing union medical university press, 1999: 389-479.

[4] Tamori A, Nishiguchi S, Tanaka M, et al. Lamivudine therapy for hepatitis B virus reactivation in a patient receiving intra-arterial chemotherapy for advanced hepatocellular carcinoma. Hepatol Res, 2003, 26(1): 77-80.

[5] Jang JW, Choi JY, Bae SH, et al. Transarterial chemolipiodolization can reactivate hepatitis B virus replication in patients with hepatocellular carcinoma. J Hepatol, 2004, 41(3): 427-435.

[6] Jang JW, Choi JY, Bae SH, et al. A randomized controlled study of preemptive lamivudine in patients receiving transarterial chemo-lipiodolization. Hepatology, 2006, 43(2): 233-240.

[7] Xiao-an wu Zhang Zhi Yong, hong, saving, etc. The three dimensional conformal radiotherapy in the treatment of 86 cases of liver cancer clinical efficacy analysis. Journal of cancer, 2007, 22(4): 373-375.

[8] Huangwei, Zhangwei, Fanmin et al. Risk factors for hepatitis B virus reactivation after conformal radiotherapy in patients with hepatocellular carcinoma. Cancer Science, 2014, 193-197.

[9] Zhangtian Zhao Yungang, wang Ming, etc Support vector machine (SVM) to predict esophageal squamous carcinoma postoperative survival. Cancer prevention and control research. 2015, 765-771.

[10] Gao jiabao. The application of the support vector machine (SVM) in speech recognition. Software Tribune. 2015, 39-40.

[11] Ma Binrong SPSS17.0 application in medical statistics Science press. 2010.

[12] BIAN Zhaoqi, ZHANG Xuegong. Pattern recognition(the second edition) [M]. TSINGHUA University Press, 2003: 284-300.

[13] Christel R, Anuradha B, Herbert I. H, Andrew B. N, Herbert P. A leave-one-out cross-validation SAS macro for the identification of markers associated with survival [J]. Computers in Biology and Medicine, 2014, 57: 123-129.