

吴冠朋, 崔晓笛, 刘文强. 基于粗糙集和 SVM 的 HBV 再激活危险因素分析与预测[J]. 智能计算机与应用, 2024, 14(12): 46-50. DOI:10.20169/j.issn.2095-2163.24031802

基于粗糙集和 SVM 的 HBV 再激活危险因素分析与预测

吴冠朋, 崔晓笛, 刘文强

(山东省立第三医院, 济南 250031)

摘要: 文章主要分析接受精确放疗后的原发性肝癌患者的乙型肝炎病毒(HBV)再激活的危险因素。通过收集多维度的临床数据,包括 HBV DNA 水平、KPS 评分、甲胎蛋白(AFP)、TNM 分期以及外放边界等,并利用粗糙集算法来评估各个因素对 HBV 病毒再激活的影响。研究发现 HBV DNA 水平是引起 HBV 病毒再激活的独立危险因素。建立了支持向量机(SVM)预测模型,用于预测 HBV 病毒再激活的可能性。经过验证,该模型分类准确率达到 74.46%,表明该模型在预测 HBV 病毒再激活方面具有一定的可靠性和有效性。

关键词: HBV 再激活; 危险因素; 粗糙集; SVM 预测模型

中图分类号: TP18 文献标志码: A 文章编号: 2095-2163(2024)12-0046-05

Analysis and prediction of risk factors for HBV reactivation based on Rough Set theory and SVM

WU Guanpeng, CUI Xiaodi, LIU Wenqiang

(Shandong Provincial Third Hospital, Jinan 250031, China)

Abstract: The article primarily delves into the risk factors for Hepatitis B Virus (HBV) reactivation among patients with primary liver cancer after receiving precision radiotherapy. By collecting multi-dimensional clinical data, including HBV DNA levels, KPS scores, AFP (Alpha-FetoProtein), TNM staging, and external radiation boundaries, and utilizing the rough set algorithm, the article assesses the impact of various factors on HBV reactivation. Through analysis, the study finds that HBV DNA level is an independent risk factor for HBV reactivation. Based on this risk factor, a Support Vector Machine (SVM) prediction model is established to predict the possibility of HBV reactivation. After verification, the classification accuracy of this model reaches 74.46%, indicating that the model has a certain degree of reliability and effectiveness in predicting HBV reactivation.

Key words: HBV reaction; risk factory; Rough Set; SVM prediction model

0 引言

原发性肝癌(Primary Liver Cancer, PLC)在中国属于高发的恶性肿瘤,其中乙肝病毒相关性肝癌占比较大。放疗是 PLC 治疗中常用的物理疗法,但放疗后易引起 HBV 病毒再激活,这对患者的治疗效果和生活质量都产生了不良影响。因此,识别与 HBV 再激活相关的关键临床因素,对于提高 PLC 患者的治疗效果至关重要^[1-4]。文献[1]综述了原发性肝癌相关危险因素的研究进展,以及分析了危险因素和如何干预方法,未建立预测分析模型。文献[2]通过 Logistic 方法分析 HBV 病毒再激活的危险因素,判断 HBV DNA 载量是 HBV 再激活的危险因素。文献[3]

研究 56 例接受 3D-CRT 治疗的 PLC 患者,发现 HBV 再激活和 HBV 未激活与 Child-Pugh 等级、HBV DNA 水平具有统计学意义。文献[4]中采用 Logistic 回归分析方法,发现 HBV 再激活与 HBV DNA 水平、外放边界、肿瘤分期 TNM 有密切关系,研究建立了基于反向传输(Back Propagation, BP)神经网络预测模型。文献[5]用遗传算法(Genetic Algorithm, GA)的特征选择方法选出致 HBV 再激活的危险因素特征子集。以上研究进行了部分特征提取和预测分类的算法研发,但对于 HBV 再激活的实用方法有待进一步探究,对预测模型的使用也要有更多选择,本文使用粗糙集理论对临床因素如 HBV DNA 水平、KPS 评分、AFP、TNM、外放边界、DVH 剂量体积参数等组成的高维数

作者简介: 吴冠朋(1989—),男,工程师,主要研究方向:人工智能,图像处理技术; 刘文强(1995—),男,工程师,主要研究方向:自然语言处理。

通信作者: 崔晓笛(1995—),女,工程师,主要研究方向:自然语言处理。Email: zbyxwgp@163.com。

收稿日期: 2024-03-18

哈尔滨工业大学主办 ◆ 学术研究与应用

据向量进行简约分析,以找到关键的特征向量集。粗糙集是一种处理不确定性和模糊性的数学工具,能够有效地去除数据中的冗余信息,提取出对决策最有影响的特征。通过粗糙集简约分析,可以从众多临床因素中筛选出与 HBV 再激活最相关的特征向量集。这些关键特征不仅可以有助于更好地理解 HBV 再激活的机制,还可以为肝癌的临床诊断和治疗提供重要参考。

传统的支持向量机 (Support Vector Machine, SVM) 机器学习算法针对高维数据空间的小样本数据识别率高。基于上面提取的关键临床特征训练 SVM^[6] 预测模型,实现对 HBV 再激活的准确预测,从而为临床决策提供有力支持。

1 粗糙集理论

1.1 粗糙集理论概述

粗糙集理论 (Rough Set Theory) 由 Pawlak 于 1982 年提出,通过对不完整、不精确和不一致的数据集进行处理,寻找规律与规则,将不精确或不确定的知识用已知知识库中的知识来近似刻画,能够有效地识别关键特征,降低数据维度,提高预测精度。本文针对原发性肝癌 (PLC) 患者放疗过程中产生的多种临床因素,如 HBV DNA 水平、KPS 评分、AFP、TNM 分期、外放边界、DVH 剂量体积参数等,利用粗糙集理论进行约简,分析筛选出对 HBV 再激活具有显著影响的主要因素,为后续的诊断和治疗提供更为精确的依据。

1.2 粗糙集的基本概念

(1) 知识表示系统^[7]。通过描述指定对象的基本特征和特征值,在海量数据中挖掘可用的知识或者决策规则。

定义知识表达公式:

$$S = (U, \mathbf{A}, V, f) \quad (1)$$

其中, U 是论域,表示非空有限集合; \mathbf{A} 表示非空条件属性; $V = \bigcup_{r \in R} V_r$, 表示属性值集合; V_r 表示属性 $r \in R$ 的范围; $f: U \times \mathbf{A} \rightarrow V$ 表示信息函数。

系统数据用矩阵 \mathbf{A} 来表示。其中,行为研究对象,列为对象属性,对象属性值即为对象信息。

(2) 决策表。信息系统 $S = (U, \mathbf{A}, V, f)$, 如果 $\mathbf{A} = C \cup D, C \cap D = \phi$, 其中 C 表示条件属性集, D 表示决策属性集,称为决策表^[8]。

(3) 约简。就是在保持知识库中初等范畴的条件下,去除知识库中冗余非必要知识,仅保留核心信息,减少数据集复杂性,提取出数据中关键特征和规律。

1.3 连续属性离散化

连续属性的离散化作为知识获取中数据预处理的一项重要内容,对于粗糙集而言,所处理的属性值必须是离散化的数据,若输出是连续数据时,在应用时需处理为离散化数据。通常有等频率划分算法、Rough 集理论和布尔逻辑相结合的离散化算法、等距离划分算法等。本文使用等距离划分算法,具体步骤如下:

(1) 确定每组属性值范围,找出属性最大值 x_{\max} 和属性最小值 x_{\min} ;

(2) 确定可分割的区间数目 m_0 。

(3) 计算区间增量 $d = (x_{\max} - x_{\min}) / m_0$ 。

(4) 确定等分点值 $A_0 = x_{\min}, A_1 = A_0 + d, \dots, A_n = x_{\max}$ 。

进一步地,判断各组属性的每个属性值 x 位于哪 2 个等分点之间,确定离散化^[9]后的决策表属性值 p , 当 $A_{k-1} \leq x < A_k$ 时, $p = k - 1$, 这里 $k = 1, 2, \dots, m_0$ 。

1.4 属性约简

决策表中决策属性 D 和条件属性 C 的依赖度^[10] 计算关系为:

$$r(C, D) = |POS(C, D)| / |X| \quad (2)$$

其中, $|POS(C, D)|$ 表示 $POS(C, D)$ 元素个数; $|X|$ 表示全部对象集合的个数。 $r(C, D)$ 性质如下。若在已知条件 C 下,则有以下情况:

(1) $r = 0$, 不能分类到决策属性 D 的类别中。

(2) $0 < r < 1$, 只能将 X 上那些属于正域的个体分类到决策属性 D 的类别中。

(3) $r = 1$, 将 X 上所有个体分类到决策属性 D 的类别中。

$C, D \subset A$, 这里 C 表示条件属性集, D 表示决策属性集, $a \in C$, 属性 a 关于 D 的重要度定义为:

$$SGF(a, C, D) = r(C, D) - r(C - \{a\}, D) \quad (3)$$

其中, $r(C - \{a\}, D)$ 表示在 C 中缺少属性 a 后的条件属性与决策属性依赖程度^[11]; $SGF(a, C, D)$ 表示 C 中缺少属性 a 后导致不能被准确分类的对象在系统中的占比。

$SGF(a, C, D)$ 性质如下:

(1) $SGF(a, C, D) \in [0, 1]$ 。

(2) $SGF(a, C, D) = 0$ 表示属性 a 关于 D 是不可约简的。

(3) $SGF(a, C, D) \neq 0$, 表示属性 a 关于 D 是不可约简的。

属性约简^[12-13]的步骤具体如下:

(1) 生成区分矩阵,矩阵中的单个属性列入约简集合,作为核属性。

(2) 区分矩阵中找出不包含该核属性且 $m_{ij} \neq 0$ 的属性组合 M 。

(3) 将所有不包含该核属性的条件属性组合表示为合取范式形式, 即 $T = \prod_{ij=1,2,3,\dots,n} \sum m_{ij}, m_{ij} \in M$ 。

(4) 将 T 转换为析取范式形式, 进行属性约简。

2 支持向量机

支持向量机^[14-16] (Support Vector Machine, SVM) 主要是基于统计学线性可分情况下的最优分类面理论, 属于模式识别方法范畴, 且得到广泛应用。在最优分类面当中, 分类线既能够把 2 个不同的类别正确分开 (即达到训练的误差率是 0), SVM 公式主要涉及到分类超平面的定义以及间隔最大化的目标。对于线性可分的情况, SVM 的核心公式可以表示为:

$$f(\mathbf{x}) = \text{sign}(\mathbf{w}\mathbf{x} + b) \quad (4)$$

其中, $f(\mathbf{x})$ 表示预测结果的符号; \mathbf{x} 表示输入样本的特征向量; \mathbf{w} 表示权重向量; b 表示偏置项。式 (4) 表示通过计算特征向量与权重向量的内积, 再加上偏置项, 得到预测结果的符号。

在 SVM 中, 边距 (margin) 定义为超平面到样本点的最大距离, 而 SVM 的目标就是最大化这个边距。设超平面为:

$$f(x) = \boldsymbol{\omega}^T \mathbf{x} + b \quad (5)$$

其中, $\boldsymbol{\omega}$ 为法向量, b 为截距。SVM 的目标就是找到一个超平面, 使得样本点到该平面的距离最大,

即实现最大间隔。同时, 这个超平面还需要满足样本点的约束条件, 即样本点应在超平面的同侧。

求解 SVM 的过程中, 还会涉及到核函数的选择和应用。核函数可以将低维特征空间中的样本映射到高维特征空间, 使得在高维空间中存在一个线性超平面能够将不同类别的样本分隔开来。

3 基于粗糙集的 HBV 再激活的影响因素分析与 SVM 预测模型

3.1 基于粗糙集的 HBV 再激活的影响因素分析

3.1.1 数据来源与准备

收集了 90 例原发性肝癌患者的完整记录。这些记录包括临床体检、胸腹部 CT、B 超、血清甲胎蛋白测定或病理学检查的结果。所有患者放疗后的相关指标均被观察与记录。

3.1.2 信息系统表的构建

根据可能影响放疗后 HBV 病毒再激活的因素 (如年龄、性别、HBV baseline、Child-Pugh 分级、甲胎蛋白水平、肿瘤大小及放疗剂量等), 构建了信息系统 (见表 1)。信息系统: $S = (U, \mathbf{A}, V, f)$, 在这个表中决策属性是 HBV 是否再激活, 而其他因素 (HBV baseline、Child-Pugh 分级、甲胎蛋白水平、肿瘤大小及放疗剂量等) 共计 30 个组成条件属性集 $C = \{c_1, c_2, \dots, c_{30}\}$, 属性集合是 $\mathbf{A} = C \cup D$ 。

3.1.3 数据离散化

对信息系统 (见表 1) 进行了等距离离散化, 得到了决策表 (见表 2)。

表 1 部分数据的 HBV 再激活及其影响因素

Table 1 Partial data on HBV reactivation and its influencing factors

1	2	3	4	5	6	7	8	9	10	11	12
'HBV 再激活'	'性别'	'年龄'	'KPS 评分'	'HbeAg'	'门脉癌栓有...'	'TNM'	'Child-Pugh'	'AFP'	'HBV basel...'	'放疗总剂量...'	'等效生物计...'
2	1	54	1	1	2	4	1	120.500 0	3	56.000 0	63.600 0
2	1	54	1	1	1	2	2	580.000 0	3	58.000 0	69.600 0
2	2	59	2	2	2	3	1	625.800 0	1	51.000 0	66.300 0
2	2	36	2	2	1	4	2	2.800 0	3	45.000 0	58.500 0
2	1	61	2	1	2	3	1	52.500 0	2	52.000 0	62.400 0
2	1	63	1	2	2	4	2	1 132.000 0	1	60.000 0	72.000 0
2	1	74	2	1	1	4	2	556.000 0	2	66.000 0	79.200 0
2	2	48	1	2	1	3	1	46.500 0	3	64.000 0	76.800 0
2	1	48	2	1	1	3	1	8.400 0	2	50.000 0	62.500 0
2	2	47	2	1	1	4	2	1 820.000 0	1	60.000 0	72.000 0
2	1	61	1	2	2	3	2	63.200 0	3	64.000 0	76.800 0
2	1	71	1	1	2	3	2	922.500 0	3	39.600 0	46.700 0
2	2	59	1	1	1	3	2	37.800 0	2	56.000 0	67.200 0
2	1	72	2	1	1	3	1	2 650.000 0	3	58.000 0	69.600 0
2	2	60	2	1	1	3	1	1 250.000 0	2	66.000 0	79.200 0
2	1	45	1	1	1	4	2	42.000 0	3	62.000 0	74.400 0
2	2	63	1	1	2	4	2	6.000 0	3	61.200 0	73.700 0

表 2 离散化后的决策表
Table 2 Discretized decision table

1	2	3	4	5	6	7	8	9	10	11	12
'HBV 再激活'	'性别'	'年龄'	'KPS 评分'	'HbeAg'	'门脉癌栓有...'	'TNNM'	'Child-Pugh'	'AFP'	'HBV basel...'	'放疗总剂量...'	'等效生物计...'
2	1	1	1	1	2	4	1	2	3	1	1
2	1	1	1	1	1	2	2	2	3	1	1
2	2	1	2	2	2	3	1	2	1	1	1
2	2	1	2	2	1	4	2	1	3	1	1
2	1	2	2	1	2	3	1	1	2	1	1
2	1	2	1	2	2	4	2	2	1	1	1
2	1	2	2	1	1	4	2	2	2	2	2
2	2	1	1	2	1	3	1	1	3	2	2
2	1	1	2	1	1	3	1	1	2	1	1
2	2	1	2	1	1	4	2	2	1	1	1
2	1	2	1	2	2	3	2	1	3	2	2
2	1	2	1	1	2	3	2	2	3	0	0
2	2	1	1	1	1	3	2	1	2	1	1
2	1	2	2	1	1	3	1	2	3	1	1
2	2	2	2	1	1	3	1	2	2	2	2

3.1.4 属性约简与核属性的确定

通过对表 2 构造出区分矩阵,进而进行约简处理,最终确定核属性为 HBV baseline,而且是单个属性。未发现不包含核属性且 $m_{ij} \neq 0$ 的属性组合,这也进一步验证了基线血清 HBV DNA 水平作为独立危险因素的重要性。

3.1.5 结论与验证

得出结论:基线血清 HBV DNA 水平是 PLC 患者精确放疗后 HBV 再激活的独立危险因素。这个结论与已有研究的报道结果相符,验证了研究的准确性和可靠性。同时相比其他特征提取发现的多种危险因素,能够更进一步地降低数据计算量,消除原始数据中的噪声和冗余信息,提高了数据的质量。

3.2 SVM 预测模型结果分析

3.2.1 试验方法

利用原数据与通过粗糙集得到的 HBV DNA 水平数据分别建立支持向量机(SVM)预测模型,并通过准确率、特异性和敏感性来评价模型性能,同时采

用 10 折交叉验证来确保结果的稳定性和可靠性,是一种科学且有效的方法。这里用到的各指标,具体含义如下。

(1)准确率 (Accuracy):模型正确分类的样本数与总样本数之比,反映了模型的整体分类性能。

(2)特异性 (Specificity):模型正确识别负样本(即未发生 HBV 再激活的样本)的能力,即真正负例 (True Negative, TN) 与所有负例 (TN + FP) 之比。

(3)敏感性 (Sensitivity):模型正确识别正样本(即发生 HBV 再激活的样本)的能力,即真正正例 (True Positive, TP) 与所有正例 (TP + FN) 之比。

10 折交叉验证 ($k = 10$) 意味着数据集被分成 10 份,每次使用其中的 9 份作为训练集,剩下的 1 份作为测试集,重复这个过程 10 次,确保每份数据都被用作测试集一次。这样得到的实验结果更加稳定和可靠,能够更好地评估模型的泛化能力。

构建的模拟实验结果见表 3。

表 3 原数据与 HBV DNA 水平数据在 10 折交叉验证下的分类结果

Table 3 Classification results of original data and HBV DNA level data under 10 fold cross validation

实验数据	K-fold	准确率/%	灵敏性/%	特异性/%
原数据 90×30 的二维矩阵	10	72.22	76.06	57.89
粗糙集处理后 HBV DNA 水平数据矩阵	10	74.46	78.48	64.68

3.2.2 实验结果

(1)准确率提升。从 72.22% 提升至 74.46%,说明经过粗糙集处理后的数据在 SVM 模型中的整体分类性能有所提高。

意味着模型在识别发生 HBV 再激活的正样本方面的能力有所增强。

(3)特异性提升。从 57.89% 提升至 64.68%,说明模型在识别未发生 HBV 再激活的负样本方面的能力也得到了提升,尽管提升幅度不如灵敏性显著。

(2)灵敏性提升。从 76.06% 提升至 78.48%,

3.2.3 原因分析

(1)特征选择的有效性。粗糙集方法通过知识约简,有效地从原始数据中提取了关键的预测因子(即 HBV DNA 水平),使得模型能够更专注于这个重要特征,从而提高了分类性能。

(2)数据质量的提升。通过粗糙集处理,消除了原始数据中的噪声和冗余信息,提高了数据的质量,使得 SVM 模型能够更好地学习和泛化。

(3)模型稳定性。SVM 作为一种强大的分类器,在处理简化后的数据集时仍能保持较好的性能,说明其具有较强的鲁棒性和稳定性。

3.2.4 试验结论

综合以上分析,可以得出结论:使用粗糙集处理后的 HBV DNA 水平数据作为 SVM 模型的输入特征,能够提升模型的性能,包括准确率、灵敏性和特异性。这进一步验证了 HBV DNA 水平是预测 PLC 患者精确放疗后 HBV 再激活的重要因素^[17-19],与文献[17-19]相比,粗糙集在选择关键因素时更准确,同时在利用 HBV DNA 水平独立因素进行分类的模型中已经具有相当优秀的分类能力,SVM 建立的预测具有更强的鲁棒性,准确率达到 74.46%。因此,粗糙集方法在处理此类问题时具有实际应用价值,能够有效地提取关键信息,过滤掉冗余数据,从而提高模型的性能。

4 结束语

本文研究了原发性肝癌患者在接受精确放疗后 HBV 再激活的危险因素。通过收集多维度的临床数据,包括 HBV DNA 水平、KPS 评分、AFP(甲胎蛋白)、TNM 分期以及外放边界等,并利用粗糙集算法来评估各个因素对 HBV 病毒再激活的影响。经过分析,得到 HBV DNA 水平是致使 HBV 病毒再激活的独立危险因素。

进一步地,文章基于 HBV DNA 水平这一危险因素,建立了支持向量机(SVM)预测模型,用于预测 HBV 病毒再激活的可能性。经过验证,该模型的性能分类准确率达到 74.46%,表明该模型在预测 HBV 病毒再激活方面具有一定的可靠性和有效性。

粗糙集算法的应用在本研究中起到了关键作用。通过粗糙集算法,可以有效地降低数据的计算量,提高计算效率,从而更快速、更准确地分析数据并找出关键因素。这种算法的优势在于其能够处理不完整和不确定的数据,并能够在保持分类能力的同时简化数据表示。

这项研究不仅揭示了原发性肝癌患者精确放疗后 HBV 病毒再激活的危险因素,还通过建立 SVM 预测模型为临床提供了实用的预测工具。这些结果和发现有助于医生制定更精准、更有效的治疗方案,从而改善患者的生活质量和预后。同时,粗糙集算法的应用也展示了其在医学数据处理和分析中的潜力和价值,为未来的研究提供了新的思路和方法。

参考文献

- [1] 乌日古门勒,其其格. 原发性肝癌相关危险因素的研究进展[J]. 临床医学进展, 2024, 14(2):4479-4485.
- [2] 孟金霞,卢喜玲,别文倩,等. 乙型肝炎肝硬化并发原发性肝癌的相关危险因素分析[J]. 实用癌症杂志, 2022,37(5): 777-779.
- [3] 张晶晶,曲颂,余建荣,等. 原发性肝癌三维适形放疗致乙型肝炎病毒再激活相关研究[J]. 癌症进展, 2015, 13(2): 183-187.
- [4] 吴冠朋,王帅,黄伟,等. 基于 BP 神经网络的肝癌放疗致乙型肝炎病毒再激活分类预测模型[J]. 智能计算机与应用, 2016, 6(2):43-47.
- [5] WU Guanpeng, LIU Yihui, WANG Shuai, et al. The classification prognosis models of hepatitis b virus reactivation based on Bayes and support vector machine after feature extraction of genetic algorithm[C]// The 2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD 2016). Piscataway, NJ: IEEE,2016: 572-577.
- [6] 王帅. 基于数据分析的肝癌放疗后乙型肝炎病毒再激活模型[D]. 济南:齐鲁工业大学, 2016.
- [7] 高爽,冬雷,高阳,等. 基于粗糙集理论的中长期风速预测[J]. 中国电机工程学报, 2012, 32(1):32-37.
- [8] 黄旭. 无线传感器网络性能测试与智能故障诊断技术研究[D]. 济南:山东大学, 2014.
- [9] 李卫国,运泽健,张师,等. 基于粗糙集理论的 t 检验广域保护区划分方法[J]. 电力系统保护与控制, 2019, 47(19):122-130.
- [10] 任婷婷,郭建胜,万江川,等. 基于粗糙集与 ID3 的航空维修数据分析方法研究[J]. 航空维修与工程, 2010(1):48-49.
- [11] 彭慧伶. 基于最佳分类数和粗糙集的故障诊断方法[J]. 计算机与数字工程, 2010,38(5):31-34.
- [12] 周黎莎,李晨,余顺坤. 基于粗糙集的电网企业低碳管理型员工绩效评价研究[J]. 水电能源科学, 2011,29(12):184-187.
- [13] 张岩,石林. 基于粗糙集约简技术的排水管路运行风险评估[J]. 常州大学学报(自然科学版), 2011,23(4):56-58.
- [14] 林泰安,罗亚梅,黄志伟,等. 基于支持向量机的肝外胆管癌术前淋巴结状态预测模型构建[J]. 西南医科大学学报, 2023, 46(4):303-307.
- [15] 孟霖宜,刘屿鸿. 基于 PSO 优化 SVM 算法的癌症诊断方法研究[J]. 计算机仿真, 2023, 40(8):279-283.
- [16] 文洁,王猛,刘周,等. 基于 MRI 放射组学模型预测乳腺癌腋窝淋巴结转移状态的初步研究[J]. 现代肿瘤医学, 2023, 31(3):506-512.
- [17] 吴冠朋. 基于智能计算的 HBV 病毒再激活分类预测模型研究[D]. 济南:齐鲁工业大学, 2017.
- [18] 王会娜. 基于特征提取的肝癌患者乙型肝炎病毒再激活的分类预测[D]. 济南:齐鲁工业大学, 2018.
- [19] 赵咏旺. 原发性肝癌精确放疗后乙型肝炎病毒(HBV)再激活分类预测[D]. 济南:齐鲁工业大学,2019.