

病理图像纹理分析在胃癌 MSI 预测中的应用研究

安卫超¹, 阎 婷², 张 楠¹, 张 杉¹, 相 洁¹, 曹 锐¹, 王 彬¹

1. 太原理工大学 信息与计算机学院, 山西 晋中 030600

2. 山西医科大学 转化医学研究中心, 山西 晋中 030600

摘 要:微卫星是遍布于人类基因组中的短串重复序列, 肿瘤组织的微卫星由于重复单位的插入或缺失而导致微卫星长度的改变的现象叫做微卫星不稳定性(Microsatellite Instability, MSI)。MSI 型胃癌往往拥有独特的分子表型以及临床病理特征, 且微卫星的不稳定性决定了胃癌患者对免疫疗法的反应是否良好, 因此 MSI 状态的术前检测对于胃癌患者治疗方案的制定具有重要意义。传统的 MSI 检测方法需要进行免疫组化及基因分析, 不仅需要增加额外的成本, 而且在临床实践中难以推广至每一个患者。应用图像特征提取技术和机器学习算法对胃癌患者的高分辨率组织病理图像进行定量分析, 实现对胃癌患者 MSI 状态的预测。从 TCGA 数据库获取 279 例原始数据, 经预处理和上采样后得到 442 个样本, 从每例样本的组织病理图像中提取出 445 个定量图像特征, 包括图像的一阶统计量, 纹理特征以及小波特征。应用 Lasso 回归进行特征筛选并构造胃癌 MSI 状态的预测标签(Risk-score), 并通过 logistics 分类模型对预测标签分类性能进行验证, 进而结合每例患者的临床特征进行多变量分析, 构建个性化的列线图进行 MSI 状态预测。实验结果显示, 基于组织病理图像纹理特征的预测标签的预测性能 AUC 值为 0.74, 现有的基于组织病理图像纹理特征的 MSI 预测模型 AUC 值为 0.73; 基于全部样本, 结合临床特征与 Risk-score 构建的 MSI 预测模型 AUC 值为 0.802, 而现有的结合临床特征和图像特征的 MSI 预测模型的 AUC 值仅为 0.752, 相较于现有方法, 提出的 MSI 预测模型具有更优的预测性能, 可以为胃癌患者的临床决策提供更有价值的参考信息。

关键词:机器学习; 组织病理图像; MSI 预测; 纹理特征; 胃癌

文献标志码:A **中图分类号:**TP391 **doi:**10.3778/j.issn.1002-8331.2008-0420

Application of Pathological Image Texture Analysis in MSI Prediction of Gastric Cancer

AN Weichao¹, YAN Ting², ZHANG Nan¹, ZHANG Shan¹, XIANG Jie¹, CAO Rui¹, WANG Bin¹

1. College of Information and Computer, Taiyuan University of Technology, Jinzhong, Shanxi 030600, China

2. Translational Medicine Research Center, Shanxi Medical University, Jinzhong, Shanxi 030600, China

Abstract: Microsatellites are short strings of repeated sequences scattered throughout the human genome. Microsatellite Instability (MSI) is a phenomenon in which the length of microsatellites changes due to the insertion or deletion of repeated units in tumor tissues. MSI type gastric cancer often has unique molecular phenotypes and clinicopathological characteristics, and the instability of microsatellites determines whether gastric cancer patients respond well to immunotherapy. Therefore, preoperative detection of MSI status is of great significance for the formulation of treatment plans for gastric cancer patients. Traditional MSI detection methods require immunohistochemistry and genetic analysis, which not only require additional costs, but also are difficult to be extended to every patient in clinical practice. In this paper, image feature extraction technology and machine learning algorithm are applied to quantitative analysis of high-resolution histopathological images of gastric cancer patients to predict the MSI status of gastric cancer patients. The original data of 279 cases are obtained from the TCGA database. After pre-processing and up-sampling, 442 samples are obtained, and 445 quantitative image features are extracted from the histopathological images of each sample, including the first-order statistics, texture features and small wave characteristics of the images. Lasso regression is used to screen features and construct predic-

基金项目:国家重点研发计划(2018AAA0102604); 国家自然科学基金(81702449)。

作者简介:安卫超(1994—), 男, 硕士研究生, 研究领域为医学数字图像处理, E-mail: anweichao0343@link.tyut.edu.cn; 阎婷(1983—), 女, 博士, 副教授, 研究领域为肿瘤病因学; 张楠(1996—), 女, 硕士研究生, 研究领域为脑影像学; 张杉(1998—), 女, 硕士研究生, 研究领域为胶质瘤分类; 相洁(1970—), 女, 博士, 教授, 研究领域为大数据管理与分析; 曹锐(1982—), 男, 博士, 副教授, 研究领域为医疗大数据处理; 王彬(1983—), 男, 博士, 副教授, 研究领域为医学影像数据分析。

收稿日期:2020-08-28 **修回日期:**2021-02-26 **文章编号:**1002-8331(2021)24-0205-07

tive labels(Risk-score) of gastric cancer MSI status, and the performance of predictive labels is verified through logistics classification model. Then, multivariate analysis is carried out in combination with the clinical characteristics of each patient, and personalized train diagram is constructed for MSI status prediction. The experimental results show that the prediction performance AUC value of the prediction label based on histone image texture features is 0.74, and the AUC value of the existing MSI prediction model based on histone image texture features is 0.73. Based on all samples, the AUC value of the MSI prediction model constructed by combining clinical features and Risk-score is 0.802, while the AUC value of the existing MSI prediction model combining clinical features and image features is only 0.752, compared with the existing methods, the MSI prediction model proposed in this paper has better prediction performance and can provide more valuable reference information for the clinical decision-making of gastric cancer patients.

Key words: machine learning; histopathological images; MSI prediction; texture features; gastric cancer

胃癌是全球最常见的恶性肿瘤之一,2018年全球胃癌新发病例1 033 701例,占全球癌症新发病例的5.7%,死亡782 685例,占全球癌症死亡病例的8.2%,在全球癌症发病率中位居第五,死亡率位居第三,且死亡率和发病率均无下降趋势^[1]。由于癌症的异质性,胃癌的表型复杂多样,使得癌症的诊断和治疗更为困难。

微卫星不稳定性是由DNA错配修复受损而导致的一种特殊癌症表型,是基因组中短重复序列超变异性的形式,其特征是由于DNA聚合酶滑移和单核苷酸变异(SNVs)频率的升高而导致微卫星重复的广泛长度多态性^[2],研究表明,MSI型胃癌约占胃癌患者的15%,这些患者更容易从免疫疗法中获益^[3]。MSI型胃癌患者有其独特的临床特征,例如:弥散型癌组织基因组更不稳定、发病部位常见于肿瘤组织远端、肿瘤类型多为Borrmann3型等;MSI型胃癌患者通常具有良好的总体长期预后,与同期的MSS型胃癌患者相比,MSI型胃癌患者生存率较高;且从癌前到发病,MSI逐渐累积增加^[4-7]。因此,MSI检测对胃癌的早期诊断和筛查,延长胃癌患者的预后,辅助胃癌治疗的临床决策有重要意义。主流的MSI检测方法主要有两种:免疫组织化学(Immunohistochemistry, IHC)和聚合酶链式反应(PCR)^[8];IHC是通过检测错配修复基因的表达情况来反应MSI状态,PCR则是通过特定的单核苷酸位点的基因标记进行遗传分析;但无论是IHC还是PCR检测手段均需要在大容量的三级医疗中心进行,且需要较高的经济和时间成本,在临床实践中难以推广至每一个患者^[9]。因此,无法为大量潜在的免疫疗法敏感者提供及时的免疫检查点抑制剂治疗,从而丧失了控制疾病的机会。

病理组织学是癌症诊断和预测的重要工具,它的表型信息反应了分子改变对癌细胞行为的综合影响,并为评估疾病的进展情况提供了直接的可视化的工具。组织病理学家可以通过评估细胞密度、组织结构、有丝分裂状态等组织学特征来对病变进行分类和分级。随着显微镜成像技术和计算机技术的进步,基于病理图片的辅助诊断模型发展迅速。其中,图像纹理分析用于病理

图像的纹理特征提取,用于癌症的分级^[10-11]、分类^[12-15]和预测。例如,Belsare等人^[16]通过提取乳腺癌患者组织病理图像的灰度共生矩阵(GLCM)、图游程长度矩阵(GRLM)和欧拉数等纹理特征,使用线性判别分类器(LDA)对图像组织学图像进行分类,对恶性和非恶性如下组织病理图像的分类准确率分别为80%和100%;Hamidreza等人^[17]通过提取软组织肉瘤的三组纹理特征:灰度共生矩阵(GLCM)、灰度游程长度矩阵(GLRLM)和局部二值模式(LBP)进行纹理分析,实现了软组织肉瘤的转移和坏死预测。Nicolas等人^[18]训练了一个深层卷积神经网络,可以通过组织病理图像准确地区分肺癌的两种亚型:腺癌(LUAD)和鳞状细胞癌(LUSC),还能确定细胞中与肺癌相关的六种基因的突变情况。Andrew等人^[19]建立了融合多个残差网络的统计模型,可以从标准苏木精和曙红染色后的组织病理图像中精准预测前列腺癌患者Speckle-type POZ基因的突变状态。

本文提出一种基于组织病理图像纹理特征的胃癌MSI预测方法,针对肿瘤的异质性,从胃癌组织病理图像中提取定量的图像特征,使用Lasso回归构建预测标签,将预测标签作为独立的预测因子,结合患者的临床特征,通过logistics回归进行多变量分析构建预测模型,绘制个性的列线图,为胃癌患者MSI预测提供了有力工具。方法流程如图1所示。

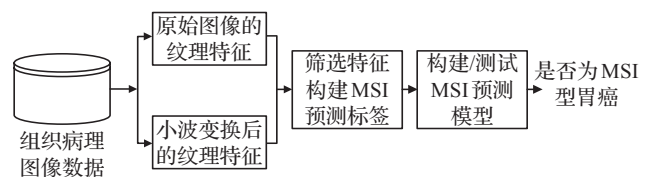


图1 胃癌MSI预测模型构建过程

Fig.1 Construction process of MSI prediction model for gastric cancer

1 数据和方法

1.1 病人数据

本文所用胃癌组织病理图像均来自TCGA数据库。为有效利用所得数据对胃癌患者的MSI状态进行

预测,本研究对收集到的数据制定了三个纳入标准:(1)病理图像染色均匀、成像清晰、无组织粘连;(2)具有统一完整的个人基本信息及临床特征;(3)具有明确的 MSI 状态信息。经筛选,共有 277 个病例样本符合入组标准。

1.2 数据预处理

为保证实验的有效性,得到有价值的结果,需要解决样本不均衡问题。采用上采样的方式扩充少数类样本:对于 MSI 型的病例,针对每例患者的组织病理图像选取多个 ROI,每个 ROI 作为独立的样本,上采样后的数据集共有 442 例样本。

将所得样本按照 3:1 的比例随机划分为训练集和验证集:训练集共有样本 313 例,其中 MSI 型 156 例, MSS 型 157 例;验证集共有 129 例样本,其中 MSI 型 64 例, MSS 型 65 例。

1.3 图像分割

为了确保所得图像特征的准确性并降低计算复杂度,在进行图像特征提取之前需要对组织病理图像进行分割。为得到最具代表性的病灶区域,在一名具有丰富组织病理图像检测经验的主任医师指导下对肿瘤区域进行标注,并通过另一名专家对标记的病灶区域进行验证,最终分割得到全部组织病理图像的 ROI。

1.4 特征提取

本研究从分割所得 ROI 的原始图像和经过 wavelet 滤波后的图像中共提取到 445 个图像特征,可以分为两类,每类 6 组:一阶统计量(First order statistics)、灰度共生矩阵(GLCM)、灰度大小区域矩阵(GLSZM)、灰度游程长度矩阵(GLRLM)、相邻灰度色调差异矩阵(NGTDM)、灰度依赖矩阵(GLDM)。

First order statistics 通过常用统计指标来描述感兴趣区域内的像素强度分布。GLCM 是描述图像灰度的空间相关特性的二阶联合概率函数,通过计算 GLCM 得到矩阵的部分特征值,用以表示图像的纹理特征,能够反应图像灰度关于方向、相邻间隔、变化幅度的综合信息。GLSZM 用于量化图像中的灰度级区域,灰度级区域定义为共享相同灰度级强度的已连接像素的数量。GLRLM 用于量化灰度游程,灰度游程被定义为具有相同灰度值的连续像素的长度。NGTDM 通过灰度等级的绝对差之和反应相邻像素的平均灰度值之间的差异。GLDM 可以量化图像中的灰度依赖性,灰度依赖性定义为距离 δ 内依赖于中心像素的连通体像素的数量。

本研究从 First order statistics 中提取了 18 种特征,主要包括熵、总能量、平均绝对偏差、偏度等;从 GLCM 中提取了 22 种特征,主要包括自相关、联合海损、群集阴影、群集倾向等;从 GLSZM 中提取了 16 种特征,主要包括灰度不均匀归一化、区域大小不均匀、区域百分比、

尺寸区域非均匀性标准化等;从 GLRLM 中提取了 16 种特征,主要包括游程熵、游程差异、灰度方差、游程非均匀标准化等;从 NGTDM 中提取了 5 种特征,主要包括粗糙度、对比度、复杂度、强度;从 GLDM 中提取了 14 种特征,主要包括依赖熵、依赖非均匀性、依赖非均匀性标准化、依赖方差等。

1.5 特征选择

为降低模型复杂度,防止过拟合,在建模之前本文使用 Lasso 方法对特征进行选择。Lasso 为改进传统的线性回归方法提供了新的视角,在普通线性回归算法的基础上加入 L1 惩罚项,使线性回归参数具有稀疏性,从而使得到的模型具有良好的预估性,所选择的特征与预测标签更相关。

对于给定样本的特征向量 $x_i(i=1,2,\dots,N), x_i \in \mathbb{R}^n$, Lasso 回归的目标函数为:

$$L(\alpha, \lambda) = \sum_{i=1}^N (y_i - \sum_j x_{ij} a_j)^2 + \lambda \sum_{j=1}^P |a_j| \quad (1)$$

其中, y 是样本的标签, $\alpha = \{a_j\}$ 是回归参数。为获得最佳回归参数,将目标函数最小化问题转化为以下子问题:

$$a_j^{(k+1)} = \arg \min \frac{L}{2} \|a - z_j\|_2^2 + \lambda \|a\|_1 \quad (2)$$

其中:

$$z_j = a_j^{(k)} - \frac{1}{L} \nabla f(a_j^{(k)}) \quad (3)$$

$$\nabla f(a_j^{(k)}) = 2 \sum_{i=1}^N \left[x_{ij} \left(\sum_{s=1}^P x_{is} a_s^{(k)} - y_i \right) \right] \quad (4)$$

使用近端梯度下降(Proximal Gradient Descent)^[20] 算法对式(3)进行迭代求解;使用软域函数对式(2)求解,最终求解方法如下:

$$a_j^{(k+1)} = \begin{cases} z_j - \frac{\lambda}{L}, & z_j > \frac{\lambda}{L} \\ 0, & |z_j| \leq \frac{\lambda}{L} \\ z_j + \frac{\lambda}{L}, & z_j < -\frac{\lambda}{L} \end{cases} \quad (5)$$

通过上述算法,最终获得稀疏后的特征矩阵,用于构建分类模型。

1.6 预测标签构建

本研究应用稀疏后的特征值及其回归系数构建样本的预测标签,公式如下:

$$Risk-score = \sum_{i=1}^n Feature_i \times \alpha_i \quad (6)$$

其中, $Feature_i$ 为样本特征向量的第 i 个特征值, α_i 为该特征值对应的回归系数。

将 Risk-score 作为独立的预测指标,与样本的临床特征相结合,构建 logistics 回归模型,绘制个性化列线图。并通过 C 指数、AUC 值、校准曲线和决策曲线评价模型的预测性能。

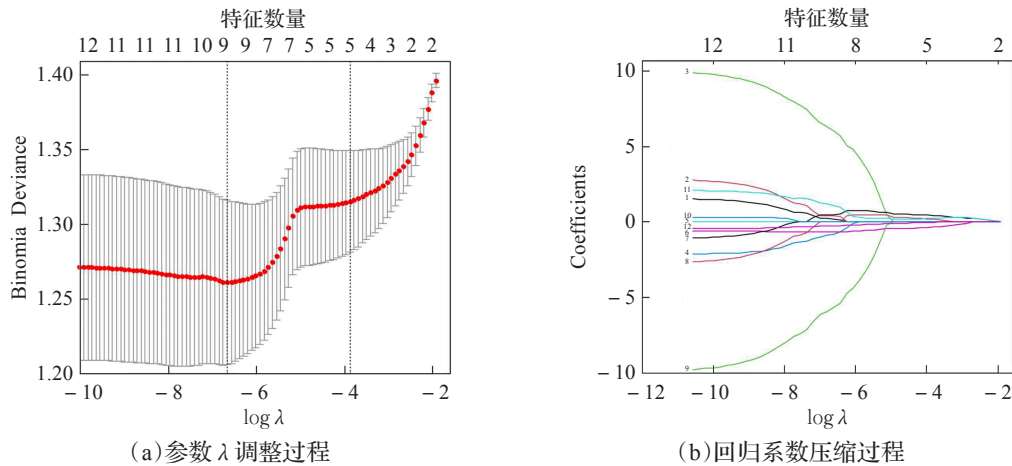


图2 Lasso回归过程

Fig.2 Lasso regression process

2 实验结果分析

2.1 临床特征

本研究所用组织病理图像来自277例胃癌患者,包括55例MSI型胃癌患者和222例MSS型胃癌患者,其中,男性患者188例,女性患者89例,中位年龄67.64岁(33~90岁),MSI型患病率为19.85%(55/277)。根据MSI状态将胃癌患者分为两组,经过T检验,MSI型患者与MSS型患者在性别、年龄、TNM分期方面的差异有显著统计学意义,患者的临床特征如表1所示。

表1 患者的临床特征

Table 1 Clinical features of patient

特征项	分类	MSI ($n=55$)	MSS ($n=222$)	P 值
Age	Mean	70.81	64.74	<0.001***
	Range	44~90	34~90	
Gender	Male	30(54.5%)	158(71.2%)	<0.001***
	Female	25(45.5%)	64(28.8%)	
TNM stage	I	12(21.8%)	25(11.3%)	<0.001***
	II	20(36.4%)	66(29.7%)	
	III	19(34.5%)	105(47.3%)	
	IV	4(7.3%)	24(10.8%)	

2.2 图像特征筛选和预测标签构建

基于MSI状态,应用Lasso回归在训练集上对图像特征进行筛选,图2(a)显示了二项式误差分类点与 $\log \lambda$ 的关系,其中最小二项式误差分类点表示保留的最佳特征数量适合模型。基于最小准则和1个标准误差标准,通过10倍交叉验证,以最佳 λ 值绘制虚线垂直线。图2(b)显示了图像特征的Lasso系数曲线。

Lasso回归的结果如表2所示,最终筛选得到9个系数不为零的特征,包括基于原始图像4个图像特征,和基于小波滤波后的5个图像特征。通过式(6)计算出样本的Risk-score。将9个图像特征与MSI状态进行单因素方差分析, P 值均小于0.001,表明筛选所得特征与胃癌患者的MSI状态显著相关。

表2 Lasso回归结果

Table 2 Result of Lasso regression

特征名称	回归系数	P 值
original_firstorder_10Percentile	0.212 204	<0.001***
original_firstorder_90Percentile	0.404 922	<0.001***
original_firstorder_Median	6.118 815	<0.001***
original_firstorder_Skewness	-0.817 240	<0.001***
wavelet-HL_glcmlmc2	-0.650 800	<0.001***
wavelet-LL_firstorder_10Percentile	0.490 395	<0.001***
wavelet-LL_firstorder_Median	-5.750 580	<0.001***
wavelet-LL_glcmlmc2	1.133 542	<0.001***
wavelet-LL_glcmlmc2_GrayLevelEmphasis	-0.254 150	<0.001***

2.3 预测准确性验证

基于所选图像纹理特征和Logistics回归,使用训练集构建了MSI预测性分类模型。如图3所示,在ROC曲线分析中,AUC值为0.75。然后将该模型应用于验证集,可以有效预测MSI状态,在ROC曲线分析中,AUC值为0.74。因此,构成模型的9个特征被证明是与胃癌患者MSI状态相关的组织病理图像特征。表3给出了分类模型各项评价指标的结果。

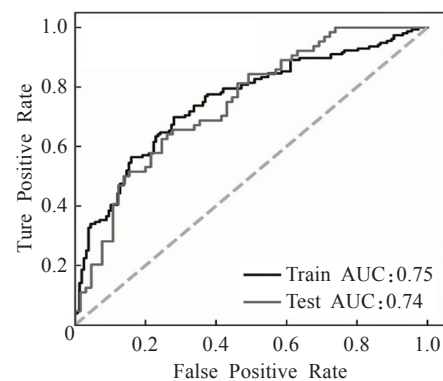


图3 训练集和测试集的ROC曲线

Fig.3 ROC curves for training sets and test sets

2.4 列线图的构建和评价

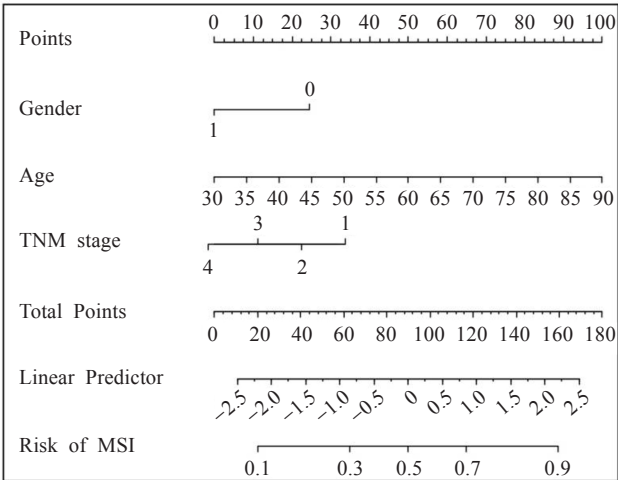
为体现预测模型的临床价值,本研究在全部数据集

表3 模型评价结果

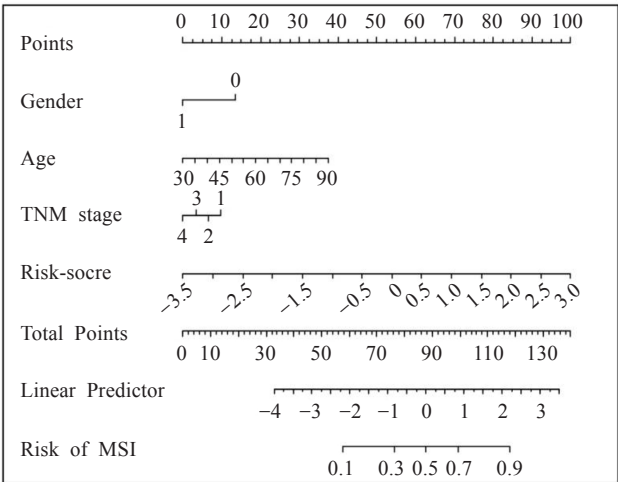
Table 3 Evaluation results of model

数据集	查准率	查全率	F1 值	AUC 值
训练集	0.68	0.72	0.70	0.75
验证集	0.65	0.67	0.66	0.74

上分别构建了基于临床特征的列线图和加入Risk-score后的列线图用于预测胃癌患者的MSI状态。如图4所示。



(a)加入 Risk-score 前



(b)加入 Risk-score 后

图4 列线图对比

Fig.4 Nomogram comparison

该列线图包括性别、年龄、TNM分期和Risk-score,允许用户获得对应于患者协变量组合的MSI状态预测概率。例如,定位患者的TNM分期轴,在该轴上画一条垂线,以确定与该TNM分期对应的预测分数。对每个变量重复此过程,并将每个协变量的得分相加,将总分对应得到预测概率,以此来实现胃癌患者的MSI状态预测。

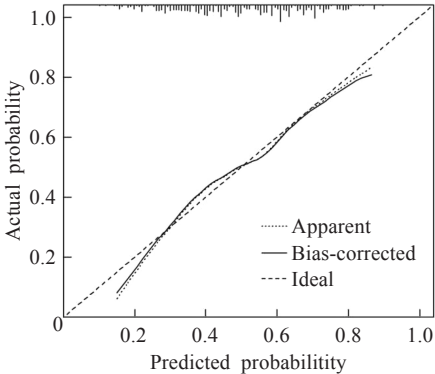
分别应用一致性指数(index of concordance,C-index),AUC和校准曲线来评估列线图的预测性能。AUC值在加入Risk-score前后分别为0.696和0.802;一致性指数如表4所示,在加入Risk-score后,C-index的值由0.69提

升至0.80;校准曲线如图5所示,虚线表示理想的预测状态,结果表明加入本研究构建的预测标签后校准曲线拟合更优。

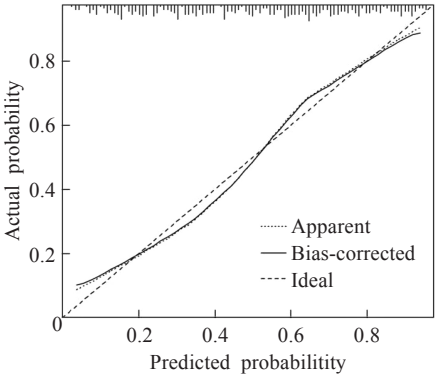
表4 预测模型C-index 评价

Table 4 C-index evaluation of prediction model

预测模型	C-index	95%CI
加入 Risk-score 前	0.69	0.64~0.74
加入 Risk-score 后	0.80	0.76~0.84



(a)加入 Risk-score 前



(b)加入 Risk-score 后

图5 校准曲线对比

Fig.5 Calibration curve comparison

为进一步验证预测模型的临床实用性,通过决策曲线分析,量化净收益来评估基于病理图像纹理特征的列线图的临床应用价值。如图6所示,在整个风险阈值区间内,加入Risk-score后的预测模型获得了更大的净收

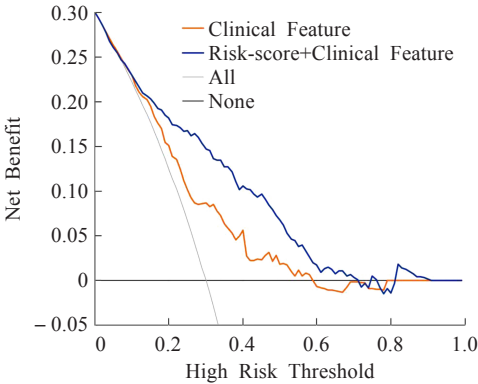


图6 决策曲线对比

Fig.6 Decision curve comparison

益。该结果表明加入 Risk-score 的列线图具有更大的临床应用潜力。

2.5 与其他 MSI 预测研究的对比

为了进一步验证模型的性能,对比了其他关于 MSI 预测的研究,对比结果如表 5 所示。Fan^[21]通过提取 CT 图像的形态、纹理、Gabor 小波等放射组学特征,结合临床特征,使用 Lasso 和朴素贝叶斯分类器开发了三个预测模型用于 MSI 的预测,单独使用临床特征的模型 AUC 值为 0.598,单独使用放射组学特征的模型 AUC 值为 0.688,联合放射组学和临床特征的模型 AUC 值为 0.752,与提出的 MSI 预测模型的性能有较大差距。Wang^[22]通过组织病理图像的切片训练了一个 ResNet-18 网络,得到患者 MSI 状态的似然分布,生成斑块似然直方图特征,使用 XGboost 分类器对患者的 MSI 状态进行预测,得到的模型在训练集上 AUC 值为 0.93,测试集为 0.73,存在较明显的过拟合现象。

表 5 MSI 预测模型性能对比

Table 5 Performance comparison of MSI prediction models

方法	数据类型	图像特征	临床特征	联合模型
Wang	组织病理图像	0.730	—	—
Fan	CT 图像	0.688	0.598	0.752
本文方法	组织病理图像	0.740	0.696	0.802

3 讨论

本文提出了一种基于胃癌组织病理图像纹理特征的 MSI 预测方法,提取原始图像的 GLCM、GLSZM、GLRLM 等纹理特征,以及经过 wavelet 变换后的纹理特征。使用 Lasso 回归进行特征选择,得到 9 个与胃癌患者 MSI 状态最相关的纹理特征,基于这些纹理特征构建了胃癌的 MSI 预测标签,并在训练集和验证集上对预测标签分类性能进行验证,AUC 值分别为 0.75 和 0.74,该结果表明提出的预测标签对胃癌患者的 MSI 状态有较好的区分度。

结合临床特征与本文提出的预测标签构建了胃癌 MSI 预测模型,与基于临床特征的预测模型相比,在加入本文提出的预测标签后,模型的 AUC 值从 0.696 提升至 0.802。为了进一步验证本文提出的预测标签的有效性和预测模型的临床价值,分别对加入预测标签前后的预测模型进行校准曲线、C-index 值和决策曲线评价,结果显示在加入本文提出的预测标签后,C-index 值和校准曲线表现有明显提升,决策曲线分析也显示出更大的净收益。

与传统的 MSI 检测方法相比,利用机器学习技术基于容易获得的组织病理图像直接预测胃癌患者的 MSI 状态,无需额外的实验室进行基因检测和免疫组织化学分析,可以以较低的成本实现 MSI 状态的预测。

与基于 CT 图像的计算机辅助 MSI 预测方法相比,

考虑到不同扫描仪和成像方案在放射学特征可重复性方面的潜在差异^[23],而 H&E 染色的组织病理图像的成像更为稳定^[24],这可能是本文提出的 MSI 预测模型性能优于 Fan^[21]的原因。与 Wang^[22]提出的方法相比,本文方法在基于组织病理图像纹理特征的预测模型上展现出了相近的性能,但 Wang^[22]在模型构建过程存在较为明显的过拟合,且没有充分利用患者的临床数据,而将患者的组织病理图像特征与临床特征相结合构建的列线图预测模型可以实现个性化的 MSI 状态预测。

4 结束语

本研究提出并验证了基于组织病理图像的胃癌 MSI 预测方法,能够有效预测胃癌患者的 MSI 状态,为 MSI 的普遍筛查提供了可能,有潜力使更多的胃肿瘤患者在免疫治疗中获益。

参考文献:

- [1] FREDDIE B, JACQUES F, REBECCA L, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries[J]. CA: A Cancer Journal for Clinicians, 2018, 68(6): 394-424.
- [2] ISIDRO C, SEJIOON L, PETER J, et al. A molecular portrait of microsatellite instability across multiple cancers[J]. Nature Communications, 2017, 8(8): 15180.
- [3] RUSSELL B, MELANIE A K, ESKO A K, et al. Landscape of microsatellite instability across 39 cancer types[J]. JCO Precision Oncology, 2017(1): 1-15.
- [4] POLOM K, MARANO L, MARRELLI R, et al. Meta-analysis of microsatellite instability in relation to clinicopathological characteristics and overall survival in gastric cancer[J]. British Journal of Surgery, 2018, 105(3): 159-167.
- [5] NICK B, JENNA L S, SHAHID A, et al. Gastric biomarkers: a global review[J]. BioMed Central, 2016, 14(1): 212.
- [6] LI B, LIU H Y, GUO S H, et al. Microsatellite instability of gastric cancer and precancerous lesions[J]. International Journal of Clinical and Experimental Medicine, 2015, 8(11): 21138-21144.
- [7] JOONHONG P, HAN M Y, WOORI J, et al. Distribution of somatic mutations of cancer-related genes according to microsatellite instability status in Korean gastric cancer[J]. Medicine, 2017, 96(25): e7224.
- [8] MARIA L, MAYUR A, JACKY Z, SHIVANI M, et al. Epidemiology of Microsatellite Instability High (MSI-H) and Deficient Mismatch Repair (dMMR) in solid tumors: a structured literature review[J]. Journal of Oncology, 2020(22): 1-17.
- [9] ARON J F, MAIRA G, MICHAEL F, et al. Results from an institutional universal reflex testing program for MSI/MMR in colorectal cancer[J]. Journal of Clinical Oncology, 2018, 36(4): 796.

- [10] NIAZI M K K, YAO K, ZYNGER D, et al. Visually meaningful histopathological features for automatic grading of prostate cancer[J]. *IEEE Journal of Biomedical and Health Informatics*, 2016, 21(4): 1.
- [11] REN J, SADIMIN E T, WANG D, et al. Computer aided analysis of prostate histopathology images Gleason grading especially for Gleason score 7[C]//Annual International Conference of the Engineering in Medicine & Biology Society, 2015: 3013-3016.
- [12] HOU L, SAMARAS D, KURC T, et al. Patch-based convolutional neural network for whole slide tissue image classification[C]//Proceedings IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016: 2424-2433.
- [13] FAUZI M F A, PENNEL M S B, CHEN W, et al. Classification of follicular lymphoma: the effect of computer aid on pathologists grading[J]. *BMC Medical Informatics and Decision Making*, 2015, 15(1): 1-10.
- [14] KOTHARI S, PHAN J H, YOUNG A N, et al. Histological image classification using biologically interpretable shape-based features[J]. *BioMed Central*, 2013, 13(1): 9-25.
- [15] JUN K, COOPER L A D, FUSHENG W, et al. Machine-based morphologic analysis of glioblastoma using whole-slide pathology images uncovers clinically relevant molecular correlates[J]. *Plos One*, 2013, 8(11): e81049.
- [16] BELSARE A D, MUSHRIF M M, PANGARKAR M A, et al. Classification of breast cancer histopathology images using texture feature analysis[C]//TENCON 2015-2015 IEEE Region 10 Conference, 2016: 1-5.
- [17] HAMIDREZA F, DMITRY B G, LAWRENCE O, et al. Texture feature analysis to predict metastatic and necrotic soft tissue sarcomas[C]//2015 IEEE International Conference on Systems, Man, and Cybernetics, 2015: 2798-2802.
- [18] NICOLAS C, SANTIAGO O P, THEODORE S, et al. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning[J]. *Nature Medicine*, 2018, 24(10): 1559-1567.
- [19] SCHAUMBERG A J, RUBIN M A, FUCHS T J, et al. H&E-stained whole slide image deep learning predicts SPOP mutation state in prostate cancer[J]. *BioRxiv*, 064279, 2018.
- [20] ZHANG L, YANG T, JIN R, et al. Stochastic proximal gradient descent for nuclear norm regularization[J]. *arXiv*: 1511.01664, 2015.
- [21] FAN S, LI X, CUI X, et al. Computed tomography-based radiomic features could potentially predict microsatellite instability status in stage II colorectal cancer: a preliminary study[J]. *Academic Radiology*, 2019, 26(12).
- [22] WANG T, LU W, YANG F, et al. Microsatellite instability prediction of uterine corpus endometrial carcinoma based on H&E histology whole-slide imaging[C]//IEEE International Symposium on Biomedical Imaging (ISBI 2020), 2020: 1289-1292.
- [23] PERRIN T, MIDYA A, YAMASHITA R, et al. Short-term reproducibility of radiomic features in liver parenchyma and liver malignancies on contrast-enhanced CT imaging[J]. *Abdominal Radiology*, 2018, 43(12).
- [24] LIAO H, XIONG T, XU L, et al. Classification and prognosis prediction from histopathological images of hepatocellular carcinoma by a fully automated pipeline based on machine learning[J]. *Annals of Surgical Oncology*, 2020, 27(7): 2359-2369.