

肝细胞癌MR图像的纹理分类研究

邱甲军¹, 吴跃¹, 惠宇^{2*}, 刘彦伯²

(1. 电子科技大学计算机科学与工程学院 成都 611731; 2. 电子科技大学信息与软件工程学院 成都 610054)

【摘要】结合小波多分辨率分析方法与统计分析方法提出了一种复合纹理分类模型, 评估其基于核磁共振(MR)图像辅助诊断肝细胞癌(HCC)与正常肝脏组织的价值。首先, 训练样本按类别分成两组, 在每组中执行小波系数统计; 其次, 对新样本的小波系数基于两组统计结果执行两次离散化, 以直方图、共生矩阵、游程长度矩阵等方法提取到两组特征; 最后, 基于两组特征执行两次分类以计算新样本的类别属性概率并决策。实验结果显示, 该模型获得了比传统方法更好的分类性能, 表明采用该模型对基于MR图像进行HCC与正常肝脏组织的计算机辅助诊断是有益的。

关键词 计算机辅助诊断; 肝细胞癌; 核磁共振图像; 纹理分析

中图分类号 TP391.4

文献标志码 A

doi:10.3969/j.issn.1001-0548.2019.04.021

Texture Classification Study of MR Images for Hepatocellular Carcinoma

QIU Jia-jun¹, WU Yue¹, HUI Bei^{2*}, and LIU Yan-bo²

(1. School of Computer Science and Engineering, University of Electronic Science and Technology of China Chengdu 611731;

2. School of Information and Software Engineering, University of Electronic Science and Technology of China Chengdu 610054)

Abstract Combining wavelet multi-resolution analysis method and statistical analysis method, a composite texture classification model is proposed to evaluate its value in computer-aided diagnosis of hepatocellular carcinoma (HCC) and normal liver tissue based on magnetic resonance (MR) images. First, training samples are divided into two groups by two categories, statistics of wavelet coefficients are calculated in each group. Second, two discretizations are performed on wavelet coefficients of a new sample based on the two sets of statistical results, and two groups of features can be extracted by histogram, co-occurrence matrix, and run-length matrix, etc. Finally, classification is performed twice based on the two groups of features to calculate the category attribute probabilities, then a decision is conducted. The experimental results demonstrate that the proposed model can obtain better classification performance than routine methods, it is rewarding for the computer-aided diagnosis of HCC and normal liver tissue based on MR images.

Key words computer-aided diagnosis; hepatocellular carcinoma; magnetic resonance image; texture analysis

肝细胞癌(HCC)是肝脏病灶中最常见的恶性肿瘤, 其早期的诊断对于后续的治疗至关重要。高场强核磁共振(MR)的动态增强扫描提高了HCC特别是小HCC的检出敏感性, 但当影像表现不典型时仍难以鉴别, 需穿刺活检证实^[1-2]。随着医学影像技术的发展, 大量医学图像在协助医师进行疾病诊断的同时, 也加重了医师的工作负担, 从而增加了经验依赖性诊断出现纰漏的可能。因此, 基于医学图像的计算机辅助诊断具有重要的研究意义与临床应用价值, 其中基于医学图像的定量数据分析, 又被称

为放射组学或影像组学(radiomics)^[3-4]。纹理, 作为数学上的一种描述属性, 可以很容易地提供病灶特征的定量测量, 基于纹理分类的计算机辅助诊断是影像组学研究中常见的临床应用: 文献[5]基于灰度直方图、灰度共生矩阵等方法, 提取CT图像中肝脏肿瘤区域的纹理特征, 采用概率神经网络对HCC与肝血管瘤进行分类; 文献[6]基于灰度共生矩阵、Gabor变换等方法, 提取CT图像中肝纤维化区域的纹理特征, 采用K近邻与支持向量机对肝纤维化分期进行分类; 文献[7]基于灰度直方图、灰度共生矩阵、

收稿日期: 2018-11-21; 修回日期: 2019-03-15

基金项目: 中央高校基本科研业务费专项(ZYGX2016J092)

作者简介: 邱甲军(1984-), 男, 博士生, 主要从事医学图像处理、影像组学的方面的研究。

通信作者: 惠宇, 副教授, E-mail: bhui@uestc.edu.cn

灰度游程长度矩阵、Gabor变换等方法提取MR-T2期图像中肝脏肿瘤区域的纹理特征,采用K近邻、支持向量机等对肝血管瘤、肝转移瘤、以及HCC进行分类;文献[8]基于灰度直方图、灰度游程长度矩阵、灰度共生矩阵等方法,提取CT图像中肝脏肿瘤区域的纹理特征,采用K近邻对HCC与肝血管瘤进行分类。然而,基于MR图像的纹理特征对HCC与正常肝脏组织进行计算机辅助诊断的影像组学研究并不多见。因此,本文提出了一种结合小波多分辨率分析与统计分析融合的复合纹理分类模型,旨在探索其在计算机辅助诊断HCC与正常肝脏组织中对低年资医师的指导作用,观察其是否可作为诊断时的参考。

1 方法

近年来,深度学习方法在图像模式识别问题中取得显著进步,但该算法需大量已标记数据(即使是基于小样本、基于生成对抗网络、或基于无标记数据训练的深度学习方法也需较多样本数据),且对微小目标的检测与分类存在巨大挑战^[9-11]。然而,在基于医学图像的辅助诊断研究中,受纳入标准限制而收集到的病例通常很少,且图像感兴趣区(region of interest, ROI)也非常小。此外,在疗效评价或预后中涉及定量测量时,定量特征需分别与临床表现、病理组织学特性进行关联分析。因此,目前基于定量特征提取与特征挖掘的影像组学方法,仍然是医学图像计算机辅助诊断研究的主要方向^[4]。

医学图像中丰富的纹理代表了大量的非可视的数据。在影像组学研究中,统计分析类方法如灰度直方图、灰度游程长度矩阵、灰度共生矩阵等,以及信号分析类方法如Gabor变换、小波变换等,被广泛用于纹理特征提取;机器学习方法如K近邻、概率神经网络、支持向量机等,常被用于基于纹理特征的预测与分类中^[4,12]。本文基于MR图像的纹理特征对HCC与正常肝脏组织进行影像组学研究,其内容包括:1) 图像预处理,主要是为了获取ROI; 2) 参考文献[5-8]提取ROI的纹理特征,建立常规的纹理分类模型; 3) 提出一种基于小波多分辨率分析与统计分析融合的纹理分析方法,提取纹理特征并构建复合分类模型。

1.1 图像预处理

从每个病例中选择1张T2WI(T2-weighted imaging)序列的图像,并导出为8位的BMP位图图像存储,被选择的图像其肿瘤截面积相对较大。然后在图像的肿瘤区域内截取一个面积尽可能大的矩

形,以形成HCC的ROI;在肿瘤区域以外的正常肝脏组织区域截取一个面积与HCC的ROI接近的矩形,以形成正常肝脏组织的ROI。本文所采用的数据集被预处理后,有少量样本ROI的像素尺寸接近 24×24 ,为处理方便,对小于该尺寸的ROI执行最近邻插值法。

1.2 常规纹理分类模型

常规纹理分类模型主要包括特征提取、特征选择、分类,本文参考相近文献[5-8]中的方法构建常规纹理分类模型。

1) 纹理特征提取。参考文献[5-8]中的方法,采用4种纹理分析方法提取ROI的纹理特征:灰度直方图(histogram, HIS)、灰度共生矩阵(co-occurrence matrix, COM)、灰度游程长度矩阵(gray level run-length matrix, RLM)以及Gabor变换(Gabor transform, GTF)。

2) 特征选择。鉴于本文所使用的数据集的异构性较大,采用了文献[13]提出的无限潜在特征选择算法(infinite latent feature selection, ILFS)。该算法构造概率潜在图形成排序阶梯,将所有可能的特征子集作为图上的路径,再通过分析绕开组合问题,对于异构数据的特征选择非常有效。此外,考虑到样本数量较少,为了减少过拟合,在ILFS特征选择基础上进一步限制了选择的特征数目。

3) 训练与测试:参考文献[5-8]中的方法,分别采用K近邻(K-nearest neighbors, KNN)、概率神经网络(probabilistic neural networks, PNN)、与支持向量机(support vector machine, SVM)构建分类模型,因纹理分析方法采用了4种,因此实验中共有12个常规纹理分类模型被使用。为减少过拟合,采取留一法进行交叉训练与测试。

1.3 复合纹理分类模型

在基于MR图像进行HCC辅助诊断的影像组学的处理中,MR成像存在伪影,ROI区域空间分辨率又小,病灶区域的细节表现复杂,因此对ROI的定量解释比较困难。然而,MR成像对软组织分辨率较高,因此在纹理分析中捕获细节可能非常重要。此外,病患存在一些个体差异、ROI区域的内部结构复杂,采用合适的统计方法描述纹理应该有较大帮助。基于小波变换的多分辨率分析对捕获图像局部区域的细节具有优势。一级二维离散小波变换^[14]将图像分解为4个子带:1个代表低频的近似分量;3个分别代表水平、垂直以及对角线方向的高频的细节分量(灰度变化较多、较快的部分)。一个分量是一

个系数矩阵。鉴于反向双正交小波 (reverse biorthogonal wavelet, rbio)在对称、紧凑支持以及精确重建上的特性, 本文通过实验选择rbio3.1小波对图像进行两级分解。

在使用小波变换捕获到灰度变化较多、较快的细节分量之后, 如何从细节分量中获得数学上的描述符(即纹理特征), 也是纹理分析方法的重要组成部分

分, 目的是尽可能使得所提取的纹理特征在同类样本间具有更大的相似性, 而在不同类的样本间具有更大的差异性。对于HCC与正常肝脏组织的二分类问题, 本文提出一种基于两组纹理特征的复合纹理分类模型(composite classification model based on bi-groups of texture features, CCM-BGTF), 其结构如图1所示, 划分为3个部分。

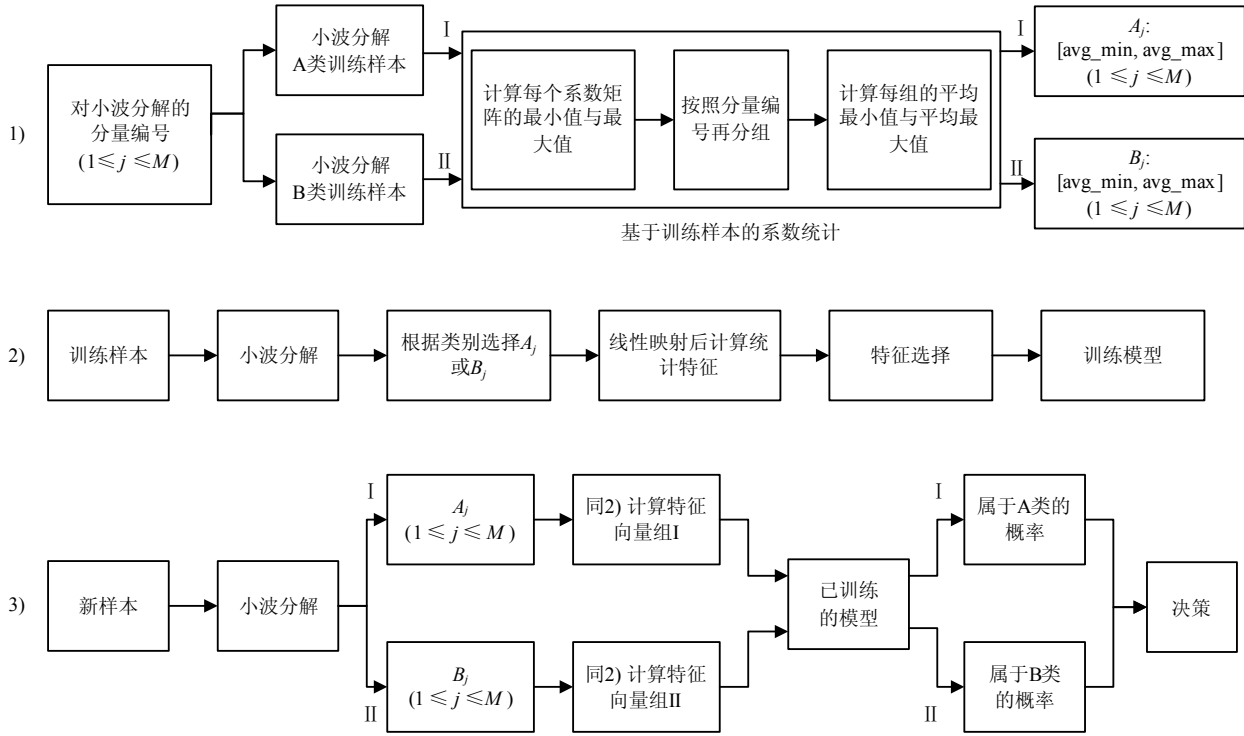


图1 CCM-BGTF模型的结构

1) 基于训练样本的系数统计

步骤1: 以字母A、B表示两个类别, 可以令A类表示HCC、B类为正常肝脏组织; 再令对ROI的两级分解可以得到的1~M个子带分量, 并以字母j表示第j号子带分量。

步骤2: 训练样本的系数矩阵按照类别分为两个组(命名为第I组与第II组), 依次输入到“基于训练样本的系数统计”过程, 参考步骤3。

步骤3: 对第I组系数矩阵, 令其中第j号子带分量对应的系数矩阵数量为K, 计算这K个矩阵中的每个矩阵的最大值与最小值, 如式(1)与式(2)所示, 第i个训练样本的第j号子带分量的系数矩阵为 C_{ij} , 其元素的最小值为 l_{ij} 、最大值为 h_{ij} , 对K个系数矩阵处理后将获得K个最小值与K个最大值。计算这K个最小值的平均值与K个最大值的平均值, 输出 A_j : [avg_min, avg_max], 其计算如式(3)与式(4)所示。

步骤4: 对于第II组系数矩阵也按照步骤3进行处

理, 输出 B_j : [avg_min, avg_max]。

$$l_{ij} = \min(C_{ij}) = \min \begin{pmatrix} c_{ij11} & c_{ij12} & \cdots & c_{ij1n} \\ c_{ij21} & c_{ij22} & \cdots & c_{ij2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{ijm1} & c_{ijm2} & \cdots & c_{ijmn} \end{pmatrix} \quad (1)$$

$$i = 1, 2, \dots, K$$

$$h_{ij} = \max(C_{ij}) = \max \begin{pmatrix} c_{ij11} & c_{ij12} & \cdots & c_{ij1n} \\ c_{ij21} & c_{ij22} & \cdots & c_{ij2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{ijm1} & c_{ijm2} & \cdots & c_{ijmn} \end{pmatrix} \quad (2)$$

$$i = 1, 2, \dots, K$$

$$A_j : \text{avg_min} = \text{avg}(\{l_{ij} \mid i = 1, 2, \dots, K\}) = \text{avg}(l_{1j}, l_{2j}, \dots, l_{Kj}) \quad (3)$$

$$A_j : \text{avg_max} = \text{avg}(\{h_{ij} \mid i = 1, 2, \dots, K\}) = \text{avg}(h_{1j}, h_{2j}, \dots, h_{Kj}) \quad (4)$$

2) 训练模型

步骤1: 对每个训练样本采取相同的小波分解。

步骤2: 根据样本类别选择 A_j 或 B_j 将第 j 号子带分量的系数矩阵离散化为 $[0, N-1]$ 的整数, 其中 N 是该系数矩阵所代表的子带图像的灰度级, 如图2所示。

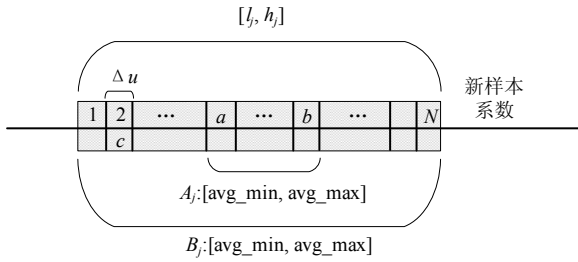


图2 离散化方法

离散化方法: ① 对第 j 号子带分量按照式(5)与式(6)计算区间 $[l_j, h_j]$, 并将这个区间划分为 N 个等宽区间, 一个系数值落入某个区间即被离散化为对应区间的整数值, 图2中的系数值 c 被离散化为2; ② 区间 A_j 或 B_j 显然是区间 $[l_j, h_j]$ 的子区间, 且不存在一个类别的最小值均值大于另一个类别的最大值均值的情况, 图2中 A_j 或 B_j 只是表达它们均属于 $[l_j, h_j]$ 的子区间的情况, 实际区间根据 A_j 或 B_j 中的最小值均值与最大值均值的真实大小决定; ③ 如果第 j 号子带分量对应的系数矩阵根据 A_j 离散化, 在图2的示例中, A_j 在 $[l_j, h_j]$ 中对应的离散化整数范围为 $[a, b]$, 系数大于或等于 A_j 中的 avg_max 时映射为 b , 小于或等于 A_j 中的 avg_min 时映射为 a ; ④ 如果第 j 号子带分量对应的系数矩阵根据 B_j 离散化, 其离散化处理方法与第③步类似, 不再赘述。

$$l_j = \min(A_j : avg_min, B_j : avg_min) \quad (5)$$

$$h_j = \max(A_j : avg_max, B_j : avg_max) \quad (6)$$

步骤3: 对已离散化的系数矩阵采用直方图、共生矩阵、游程长度矩阵提取纹理特征, 特征相关的统计量的计算参考文献[15-17]。

步骤4: 采用1.2节中所讨论的ILFS方法对训练样本的纹理特征进行特征选择。

步骤5: 基于已选择的特征训练SVM模型。

3) 对新样本的类别属性计算与决策

步骤1: 对新样本采取与第1)部分相同的小波分解。

步骤2: 对新样本的第 j 号子带分量对应的系数矩阵分别基于 A_j 与 B_j 进行离散化并提取特征, 将得到两组特征(分别为第I组特征与第II组特征), 离散化方法参考第2)部分所述方法; 训练样本由于其类别已知, 直接选择 A_j 或 B_j 离散化即可, 因此仅获得一组特征。

步骤3: 采用第2)部分的特征选择结果对新样本特征做选择。

步骤4: 基于第2)部分的训练模型计算类别属性概率, 使用SVM后验概率作为类别属性评分。SVM将线性可分数据的训练样本映射到空间中的点, 并在空间中构建一个或一组超平面, 以将各类别的样本划分到各区域, 并使各区域间有尽可能宽的、清晰的间隙, 然后将新样本映射到同一空间, 并根据它们更靠近哪个类别区域进行类别预测。对于线性不可分的数据, 可以使用核函数有效地执行非线性分类, 隐式地将输入映射到高维特征空间。越靠近某个类别区域中心的点, 其属于该类别的概率值就越大^[18]。以二维平面为例, SVM类别属性概率的概念如图3所示。

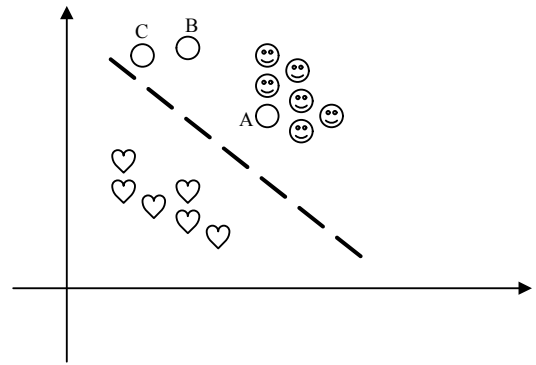


图3 二维平面中SVM超平面划分类别区域

图3所示为笑脸类与心形类的训练模型, 虚线为超平面。对于圆形表示的A、B、C三个点, 可按超平面被划分为笑脸类, 但它们的类别属性概率显然不同: A距离笑脸区域最近, 概率最大; B其次; 概率最小的是C点, 处于超平面边缘, C点属于心形类的概率也比较大。这种输出概率实际上是新样本属于某个类别的一种测量。

步骤5: 比较类别属性值大小, 根据如下伪代码进行类别决策, 决策算法遵循类别属性概率大则优先选取的原则: 采用第I组特征分类时, 第I组特征本来就是根据A类训练集统计计算出的特征, 因此新样本属于A类的概率表达出了新样本与A类训练样本的相似程度; 采用第II组特征分类时, 第II组特征本来就是根据B类训练集统计计算出的特征, 因此新样本属于B类的概率表达出了新样本与B类训练样本的相似程度。

输入:

α : 第I组特征分类时新样本属于A类的概率为 α

β : 第II组特征分类时新样本属于B类的概率为 β

输出: 类别A或B, $score_A$ 与 $score_B$ 是新样本分别属于A类与B类的评分

if ($\alpha \geq \beta$) return A; //第I组属于A的概率 \geq 第II

组属于B的概率

```

else return B;
score_A=α/(α+β); //计算新样本属于A的评分
score_B=1-score_A;

```

2 实验结果

本文实验所使用数据集来源于TCGA Research Networks公开的97个HCC病例, 选择其中属于MR病例(共40例)的T2WI(T2-Weighted Imaging)序列图像作为研究, 正常肝脏组织区域也从这些病例的图像中选择, 数据集参考网址 <http://cancer.genome.nih.gov/>^[19-20]。然而, 在这40个病例中: 1例无T2WI图像、3例未能从T2WI图像中找到病灶、2例图像质量极差、还有1例无术前T2WI图像。因此总共有66个ROI被分析, HCC与正常肝脏组织各33个。这些图像的扫描层厚为5 mm~10 mm, 窗位为88~712, 窗宽241~1 237, 窗位窗宽比大约在0.3~0.6之间, 图像分辨率256×208~640×640; 所有T2WI图像以DICOM格式存储。

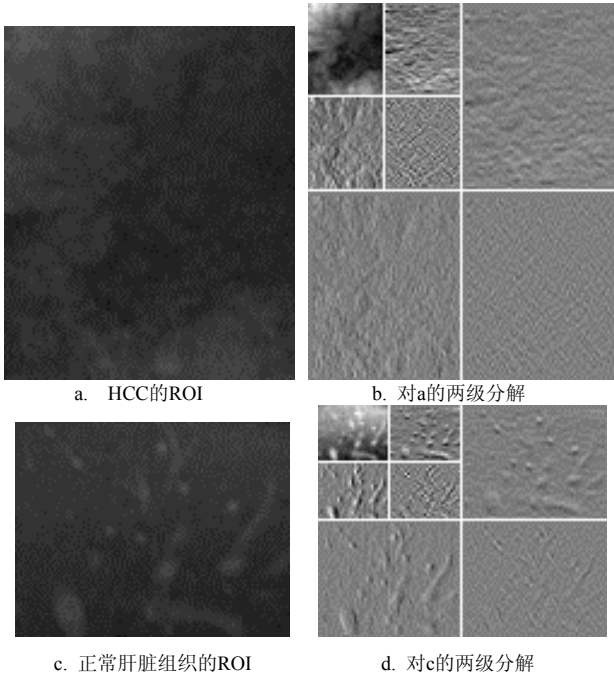


图4 rbio3.1小波对ROI的两级分解

数据集图像的预处理参考1.1节, 实验采用两级rbio3.1小波变换, 其示例如图4所示, 图4b与4d的左上角为近似分量。CCM-BGTF模型在提取纹理特征时分别采用了直方图、共生矩阵以及游程长度矩阵, 为表达方便, 它们对应的复合分类模型分别简称为HIS-CCM、COM-CCM、RLM-CCM。实验中共采用15个纹理分类模型, 表1所示为4种纹理分析方法提取的特征。

表1中: 1) COM方法的1个共生矩阵提取5个常用特征, RLM方法1个游程矩阵提取11个特征, GTF方法中1个分量提取均值与平均能量^[15-17]; 2) 在CCM-MGTF模型中, 多分辨率分析中的近似分量是通过低通滤波器得到的, 系数离散化后其值大多相近(接近常量), 这对于计算共生矩阵特征的一些特征是不可能的, 因此COM-CCM方法仅使用了细节分量, 而HIS-CCM与RLM-CCM方法则可以使用所有分量; 3) 在RLM-CCM模型中, 由于提取的特征数量过多, 通过计算某个特征的4个方向的均值以降低特征向量维度。

表1 提取的纹理特征

方法	特征
HIS	mean, standard deviation, smoothness, cubic moment, uniformity, entropy, fourth moment
COM	contrast, correlation, energy, uniformity, entropy
RLM	short run emphasis, long run emphasis, gray-level non-uniformity, run length non-uniformity, run percentage, low gray-level run emphasis, high gray-level run emphasis, short run low gray-level emphasis, short run high gray-level emphasis, long run low gray-level emphasis, long run high gray-level emphasis
GTF	mean, mean energy

分类性能指标: 二分类的问题通常以混淆矩阵表达结果, 以HCC为正样本类(positive), 以正常肝脏组织为负样本类(negative), TP表示正类被正确分类为正类的样本数, TN表示负类被正确分类为负类的样本数, FP表示负类被错误地分类为正类的样本数, FN表示正类被错误地分类为负类的样本数。性能指标准确度(Accuracy)、灵敏度(Sensitivity)以及特异性(Specificity)的定义如式(7)~式(9)所示:

$$Accuracy: ACC = \frac{TP+TN}{TP+TN+FP+FN} \tag{7}$$

$$Sensitivity: SEN = \frac{TP}{TP+FN} \tag{8}$$

$$Specificity: SPE = \frac{TN}{TN+FP} \tag{9}$$

分类结果如表2所示, “Train”列与“Test”列分别表示各模型的训练时间与测试时间, 它们是10次完整的留一法实验后的均值, 其中加粗显示的是所消耗的最小时间值; “Ref.”列是指模型所参考的文献编号; 第6~9列加粗显示的是各模型中某个性能指标相对突出的值; AUC是指ROC曲线(receiver operating characteristic curve)下的面积, ROC曲线如图5所示。

由表2可知, RLM与GTF结合KNN与PNN分类

器时特异性高,但准确率不突出、敏感度较低,且图5表明仅RLM+KNN的ROC曲线较快地靠近左上角(较好的ROC曲线在图5中以椭圆标记);复合分类模型HIS-CCM与COM-CCM的AUC、准确率、敏感度与特异性均比较高,特别是HIS-CCM模型的性能最好,其ROC曲线在图5中也很快接近左上角;关于训练时间与测试时间,PNN快于其他模型,而CCM-MGTF的最长,但在可接受范围内,且

CCM-MGTF相比于常规SVM分类器,训练时间与测试时间并无显著增加。

此外,CCM-MGTF中的3个模型(基于SVM的复合纹理分类模型)比基于SVM的4种常规纹理分类模型的性能均有提高,其中HIS-CCM的分类性能比HIS-SVM有显著的性能提升,COM-CCM的分类性能比起COM-SVM也有所提升,而RLM-CCM的分类性能只是比RLM-SVM的略有提高。

表2 CCM-MGTF与常规纹理分类模型的分分类结果

模型	TP	FN	TN	FP	ACC/%	SEN/%	SPE/%	AUC	Train	Test	Ref.
HIS-KNN	28	5	30	3	87.88	84.85	90.91	0.934 3	0.311 6	0.052 7	[7-8]
COM-KNN	28	5	29	4	86.36	84.85	87.88	0.918 3	0.311 8	0.062 4	[6-8]
RLM-KNN	26	7	32	1	87.87	78.79	96.97	0.953 2	0.301 3	0.059 7	[7-8]
GTF-KNN	27	6	32	1	89.39	81.82	96.97	0.897 6	0.300 2	0.059 8	[6-7]
HIS-PNN	27	6	29	4	84.85	81.82	87.88	0.792 5	0.005 7	0.005 3	[5]
COM-PNN	28	5	28	5	84.85	84.85	84.85	0.874 7	0.005 1	0.004 3	[5]
RLM-PNN	27	6	33	0	90.91	81.82	100	0.844 4	0.005 9	0.005 2	--
GTF-PNN	26	7	32	1	87.88	78.79	96.97	0.819 6	0.005 8	0.004 9	--
HIS-SVM	28	5	30	3	87.88	84.85	90.91	0.909 1	6.083 2	0.036 3	[7]
COM-SVM	28	5	28	5	84.85	84.85	84.85	0.871 9	6.337 1	0.037 4	[6-7]
RLM-SVM	29	4	29	4	87.88	87.88	87.88	0.927 0	6.507 9	0.038 3	[7]
GTF-SVM	27	6	29	4	84.85	81.82	87.88	0.877 0	5.835 2	0.035 8	[6-7]
HIS-CCM	30	3	31	2	92.42	90.91	93.94	0.941 7	6.580 7	0.074 4	--
COM-CCM	30	3	29	4	89.39	90.91	87.88	0.947 7	5.815 0	0.064 1	--
RLM-CCM	27	6	31	2	87.88	81.82	93.94	0.882 9	5.814 2	0.066 8	--

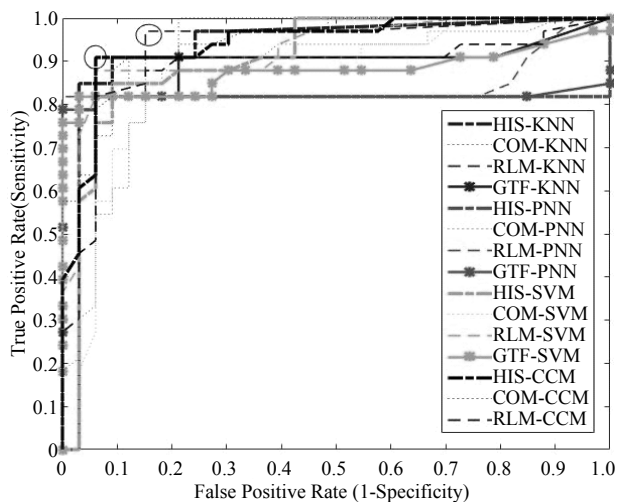


图5 ROC曲线图

3 讨论

MR成像对软组织的分辨率较高,正常肝脏组织在灰度纹理表现上总体平滑且包含血管分布等软组织的细节。HCC的细胞结构比正常肝脏组织的更紧

凑,其血管分布更丰富,因此其在MR图像的灰度纹理表现上既呈现出局部有规律的统计模式,又因细节较多而呈现出较大的异构性^[21]。另外,因成像原理、设备、操作等各种原因,MR成像容易产生伪影等噪声^[22],加剧了纹理模式上的异构性,这对传统纹理分析方法的性能有一定影响。小波变换在信号的时频分析上具有多分辨率分析的优势,统计方法可以表达强度和空间分布的多维统计模式。因此组合小波变换和统计方法是能够提高其所提取的特征的辨别能力的。文献[23]在分析自然图像的纹理特征时指出,如何选择 N 个值将小波系数矩阵离散化,对基于这些小波系数所提取的纹理特征的代表能力起到重要作用。这个基本原理对于分析MR图像的纹理特征仍具有参考意义,本文的解决思路是:1)按两个类别对训练集样本的小波系数进行统计,对于小波分解后的一个在某个分辨率下的某个方向的细节分量(编号为 j),生成两个限制区间: A_j 与 B_j ;2)其他样本的对应的第 j 号细节分量基于这两个区间执行两次离散化;3)提取离散化系数矩阵的统计特

征, 得到两组特征(训练样本因其类别已知, 只使用其所属类别对应的区间离散化, 得到一组特征)。

采用不同的两个限制区间离散化 N 个值时的表现参考图2。 A_j 与 B_j 分别是基于两类训练集统计出来的限制区间, 可以分析: 1) 如新样本属于A类, 则它的第 j 个系数矩阵所提取的第I组特征与A类训练样本对应的特征应该更相似; 2) 类似地, 如新样本属于B类, 则它的第 j 个系数矩阵所提取的第II组特征

与B类训练样本的对应特征应该更加相似。因此本文接着提出了一种类别决策算法, 基于SVM分类的评分(类别属性的后验概率)作为类别相似性程度的判别依据。

对所有样本进行小波系数统计, 得到的各个小波分量的限制区间如表3所示: H、V、D分别表示水平、垂直、对角线的高频分量, “1”与“2”表示第1级分解与第2级分解。

表3 各小波分量的限制区间

		H	V	D	A
1	HCC(A类)	[-56.693 2, 55.162 9]	[-62.958 3, 62.492 4]	[-51.685 6, 50.242 4]	[122.016 1, 311.156 3]
	正常组织(B类)	[-47.119 3, 41.706 4]	[-46.842 8, 47.081 4]	[-42.087 1, 44.003 8]	[85.447 0, 187.107 0]
2	HCC(A类)	[-99.250 2, 105.555 6]	[-110.704 5, 109.804 4]	[-89.725 1, 87.336 1]	[273.288 7, 572.995 0]
	正常组织(B类)	[-69.510 2, 66.927 6]	[-60.492 8, 59.222 7]	[-67.792 4, 70.955 1]	[185.739 9, 323.329 1]

从表3可知, 在不同的子带分量中, A_j 与 B_j 还是存在较大的差异性, 特别是第2级分解中的水平分量与垂直分量, A_j 与 B_j 的差异非常明显。因此可以推断, 在第 j 号子带分量的离散化总区间 $[l_j, h_j]$ 不变的情况下, 根据 A_j 与 B_j 离散化的两个系数矩阵之间存在差异。此外, 在离散化时, 基于 A_j 或 B_j 的离散化将被限制在 A_j 与 B_j 对应的离散化区间, 如果新样本是A类样本, 其第 j 号子带分量基于 A_j 离散化后得到的矩阵特征, 应该与训练样本中A类样本的第 j 号子带分量离散化后得到的矩阵特征更相似(A_j 本来就是基于A类训练样本的第 j 号子带分量的系数矩阵统计而来); 如果新样本是B类时, 可作类似的分析。

本实验也有一些局限性: 1) 图像的窗位窗宽比范围较大, 造成的干扰对纹理分析性能有一定的影响; 2) 肿瘤与正常肝脏组织的ROI由人工选取, 可能存在选择性偏差; 3) 为评估CCM-MGTF的分类性能, 并未组合所有纹理特征再评估分类性能, 而是对每种纹理分析方法进行了比较。下一步将增加样本、在纹理分类前进行降噪与肿瘤分割处理, 并组合各类方法提取的纹理特征, 进行计算机辅助诊断相关的应用研究。

4 结束语

本文首先基于灰度直方图、灰度共生矩阵、灰度游程长度矩阵、以及Gabor变换进行了HCC病例的MR T2WI图像的纹理分析, 并采用了相近的影像组学研究中常用的KNN、PNN、SVM构建了常规纹理分类模型进行分类。然后基于小波变换的多分辨率分析, 结合统计方法在强度和空间分布模式上的统

计特性进行了纹理分析, 并针对本文所讨论的二分类问题构建了一种复合分类模型, 获得了比常规分类模型更好的分类性能, 表明采用该模型对基于MR图像进行HCC与正常肝脏组织的计算机辅助诊断是有益的, 对低年资医师具有较好的指导作用, 作为诊断时的参考可减少医师的工作量, 且对将来基于MR图像的肝脏病灶相关的影像组学研究具有一定的参考作用。

参 考 文 献

- [1] WARD J, ROBINSON P J. How to detect hepatocellular carcinoma in cirrhosis[J]. European Radiology, 2002, 12(9): 2258-2272.
- [2] WILLATT J M, HUSSAIN H K, ADUSUMILLI S, et al. MR imaging of hepatocellular carcinoma in the cirrhotic liver: Challenges and controversies[J]. Radiology, 2008, 247(2): 311-330.
- [3] 郑光远, 刘峡壁, 韩光辉. 医学影像计算机辅助检测与诊断系统综述[J]. 软件学报, 2018, 29(5): 1471-1514. ZHENG Guang-yuan, LIU Xia-bi, HAN Guang-hui. Survey on medical image computer aided detection and diagnosis systems[J]. Journal of Software, 2018, 29(5): 1471-1514.
- [4] GILLIES R J, KINAHAN P E, HRICAK H. Radiomics: Images are more than pictures, they are data[J]. Radiology, 2015, 278(2): 563-577.
- [5] KUMAR S S, MONI R S, RAJEESH J. An automatic computer-aided diagnosis system for liver tumours on computed tomography images[J]. Computers & Electrical Engineering, 2013, 39(5): 1516-1526.
- [6] KAYAALTI O, AKSEBZECI B H, KARAHAN I O, et al. Liver fibrosis staging using CT image texture analysis and soft computing[J]. Applied Soft Computing, 2014, 25: 399-413.
- [7] LI Z, MAO Y, HUANG W, et al. Texture-based

- classification of different single liver lesion based on SPAIR T2W MRI images[J]. *BMC Medical Imaging*, 2017, 17(1): 42.
- [8] 王永芹, 黄子星, 袁放, 等. CT平扫图像纹理分析对肝癌与肝血管瘤鉴别诊断的初步研究[J]. *中国普外基础与临床杂志*, 2017, 24(2): 254-258.
- WANG Yong-qin, HUANG Zi-xing, YUAN Fang, et al. Preliminary study on differential diagnosis of liver cancer and hepatic hemangioma by texture analysis of non-enhanced CT images[J]. *Chinese Journal of Bases and Clinics in General Surgery*, 2017, 24(2): 254-258.
- [9] ZHOU Z, SHIN J, ZHANG L, et al. Fine-tuning convolutional neural networks for biomedical image analysis: Actively and incrementally[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Hawaii: IEEE, 2017: 4761-4772.
- [10] LITJENS G, KOOI T, BEJNORDI B E, et al. A survey on deep learning in medical image analysis[J]. *Medical Image Analysis*, 2017, 42: 60-88.
- [11] ERHAN D, SZEGEDY C, TOSHEV A, et al. Scalable object detection using deep neural networks[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Columbus: IEEE, 2014, 2147-2154.
- [12] CASTELLANO G, BONILHA L, LI L M, et al. Texture analysis of medical images[J]. *Clinical Radiology*, 2004, 59(12): 1061-1069.
- [13] ROFFO G, MELZI S, CASTELLANI U, et al. Infinite latent feature selection: A probabilistic latent graph-based ranking approach[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Hawaii: IEEE, 2017: 1398-1406.
- [14] MALLAT S G. A theory for multiresolution signal decomposition: The wavelet representation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1989, 11(7): 674-693.
- [15] GONZALEZ R C, WOODS R E, EDDINS S L. *Digital image processing using MATLAB*[M]. USA: Gatesmark Publishing, 2009.
- [16] WEI X. Gray level run length matrix toolbox v1.0[EB/OL]. [2018-07-01]. <http://www.mathworks.com/matlabcentral/fileexchange/17482-gray-level-run-length-matrix-toolbox>.
- [17] HARALICK R M. Statistical and structural approaches to texture[J]. *Proceedings of the IEEE*, 1979, 67(5): 786-804.
- [18] CORTES C, VAPNIK V. Support-vector networks[J]. *Machine Learning*, 1995, 20(3): 273-297.
- [19] ERICKSON B J, KIRK S, LEE Y, et al. Radiology data from the cancer genome atlas liver hepatocellular carcinoma [TCGA-LIHC] collection[EB/OL]. [2018-07-01]. <http://doi.org/10.7937/K9/TCIA>.
- [19] CLARK K, VENDT B, SMITH K, et al. The cancer imaging archive (TCIA): Maintaining and operating a public information repository[J]. *Journal of Digital Imaging*, 2013, 26(6): 1045-1057.
- [20] NGUYEN B N, FLEJOU J F, TERRIS B, et al. Focal nodular hyperplasia of the liver: A comprehensive pathologic study of 305 lesions and recognition of new histologic forms[J]. *The American Journal of Surgical Pathology*, 1999, 23(12): 1441.
- [21] 徐子森, 唐立民, 王燕. 影响MR成像质量的因素分析[J]. *中国医疗设备*, 2003, 18(9): 31-34.
- XU Zi-sen, TANG Li-min, WANG Yan. Analysis of the influence factors for MR imaging quality[J]. *Information of Medical Equipment*, 2003, 18(9): 31-34.
- [22] VAN D W G, SCHEUNDERS P, VAN D D. Statistical texture characterization from discrete wavelet representations[J]. *IEEE Transactions on Image Processing*, 1999, 8(4): 592-598.

编辑 蒋晓