

基于¹⁸F-FDG PET/CT 多模纹理特征的自身免疫性胰腺炎与胰腺导管腺癌鉴别方法



张玉全^{1,2}, 程超³, 刘兆邦², 潘桂霞³, 孙高峰³, 杨晓冬², 左长京³

1. 中国科学技术大学(合肥 230026)
2. 中国科学院苏州生物医学工程技术研究所(江苏苏州 215163)
3. 第二军医大学附属长海医院(上海 200433)

【摘要】 自身免疫性胰腺炎(AIP)是慢性胰腺炎中的一种独特亚型,其临床表现与胰腺导管腺癌(PDA)非常相似,故AIP患者经常被误诊为PDA,承受不必要的手术。¹⁸F-FDG正电子发射断层显像/X线计算机断层成像(PET/CT)检查可以同时提供胰腺形态、密度和功能代谢的综合信息,有助于对AIP和PDA进行鉴别。然而目前临床上缺乏对PET/CT图像纹理特征的分析,基于现有的诊断手段对二者进行准确鉴别依然存在困难。因此,本文基于多模纹理特征研究AIP与PDA的鉴别。本文首先采用多种特征提取算法来提取CT和PET图像内的纹理特征,然后采用Fisher准则和与支持向量机(SVM)相结合的序列前向浮动选择算法(SFFS)选择鉴别性能最优的多模特征子集,最后采用SVM分类器实现AIP与PDA的鉴别。结果表明,对病灶的纹理分析有助于实现AIP与PDA的准确鉴别。

【关键词】 自身免疫性胰腺炎; 胰腺导管腺癌; ¹⁸F-FDG正电子发射断层显像/X线计算机断层成像; 多模纹理特征; 支持向量机

Differentiation of autoimmune pancreatitis and pancreatic ductal adenocarcinoma based on multi-modality texture features in ¹⁸F-FDG PET/CT

ZHANG Yuquan^{1,2}, CHENG Chao³, LIU Zhaobang², PAN Guixia³, SUN Gaofeng³, YANG Xiaodong², ZUO Changjing³

1. University of Science and Technology of China, Hefei 230026, P.R.China
 2. Suzhou Institute of Biomedical Engineering and Technology, Chinese Academy of Sciences, Suzhou, Jiangsu 215163, P.R.China
 3. Changhai Hospital Affiliated to Second Military Medical University, Shanghai 200433, P.R.China
- Corresponding author: ZUO Changjing, Email: changjing.zuo@qq.com

【Abstract】 Autoimmune pancreatitis (AIP) is a unique subtype of chronic pancreatitis, which shares many clinical presentations with pancreatic ductal adenocarcinoma (PDA). The misdiagnosis of AIP often leads to unnecessary pancreatic resection. ¹⁸F-FDG positron emission tomography/ computed tomography (PET/CT) could provide comprehensive information on the morphology, density, and functional metabolism of the pancreas at the same time. It has been proved to be a promising modality for noninvasive differentiation between AIP and PDA. However, there is a lack of clinical analysis of PET/CT image texture features. Difficulty still remains in differentiating AIP and PDA based on commonly used diagnostic methods. Therefore, this paper studied the differentiation of AIP and PDA based on multi-modality texture features. We utilized multiple feature extraction algorithms to extract the texture features from CT and PET images at first. Then, the Fisher criterion and sequence forward floating selection algorithm (SFFS) combined with support vector machine (SVM) was employed to select the optimal multi-modality feature subset. Finally, the SVM classifier was used to differentiate AIP from PDA. The results prove that texture analysis of lesions helps to achieve accurate differentiation of AIP and PDA.

【Key words】 autoimmune pancreatitis; pancreatic ductal adenocarcinoma; ¹⁸F-FDG positron emission tomography/ computed tomography; multi-modality texture features; support vector machine

DOI: 10.7507/1001-5515.201807012

基金项目: 江苏省自然科学基金(BK20170391); 江苏省社会发展面上项目(BE2017670)

通信作者: 左长京, Email: changjing.zuo@qq.com

引言

自身免疫性胰腺炎 (autoimmune pancreatitis, AIP) 是慢性胰腺炎中的一种独特亚型, 它与自身免疫机制相关, 类固醇激素治疗对其效果显著^[1]。胰腺导管腺癌 (pancreatic ductal adenocarcinoma, PDA) 是胰腺癌中最常见的一种形式, 需要根治性手术治疗, 且预后差^[2]。AIP 在临床表现上与 PDA 存在极大的相似性, 使得 AIP 患者常被误诊为 PDA 而接受不必要的胰腺切除手术。国内外研究表明, 在被诊断为 PDA 的患者中, 有 5% ~ 21% 最终被证实是 AIP^[3-4]。因此, 实现 AIP 与 PDA 的准确鉴别意义重大。

AIP 的诊断标准主要包括影像学检查、血清学检查、组织学检查、胰腺外器官受累情况和激素治疗后的反应等^[5-6]。血清学指标对 AIP 的敏感性和特异性仍有争议^[7], 而有创性的穿刺活检敏感性不高, 因此影像学检查在临床诊断中发挥了关键性的作用。传统影像学检查, 如超声、计算机断层显像 (computed tomography, CT) 和磁共振成像 (magnetic resonance imaging, MRI), 它们均反映胰腺组织的解剖学信息, 当 AIP 患者胰腺表现出局灶性增大和“双管征”时, 仅基于解剖学信息极易将其误诊为 PDA。¹⁸F-氟脱氧葡萄糖 (¹⁸F-fluorodeoxyglucose, ¹⁸F-FDG) 正电子发射断层显像/X 线计算机断层成像仪 (positron emission tomography/computed tomography, PET/CT) 集功能显像和解剖显像于一体, 可以同时提供胰腺形态、密度及功能代谢的综合信息, 尤其在胰腺外器官的检查方面拥有不可替代的优势, 对 AIP 和 PDA 的鉴别价值极大^[8]。然而, 目前临床上常用的定量分析指标, 如标准摄取值 (standardized uptake value, SUV) 无法挖掘病灶内的代谢变化信息, 而人眼对灰阶又不敏感, 当 AIP 患者胰腺表现出局灶性的¹⁸F-FDG 高摄取及肿大时, 容易导致误诊^[9]。针对这一情况, 本文旨在深入挖掘¹⁸F-FDG PET/CT 图像中的胰腺病灶内部特征^[10], 利用计算机辅助诊断技术实现 AIP 与 PDA 的鉴别诊断。

截止目前, 计算机辅助诊断技术在鉴别 AIP 与 PDA 方面的价值如何尚未见报道。在其他胰腺疾病分析场景中, Gazit 等^[11]基于增强 CT 图像对胰腺导管内乳头状黏液性肿瘤恶性程度进行评估, 构建了随机森林 (random forest, RF) 和 Adaboost 两个分类模型。在基于超声内镜图像对 PDA 与慢性胰腺炎^[12]和 AIP 与慢性胰腺炎^[13]进行鉴别的研究

中, 研究人员采用序列前向选择算法对提取到的特征进行筛选, 最后采用支持向量机 (support vector machine, SVM)^[14]进行分类。上述研究均基于单一模态的解剖学影像信息, 缺乏对功能代谢信息和多模态互补信息的探讨。鉴于¹⁸F-FDG PET/CT 在鉴别 AIP 与 PDA 中的价值, 本文基于该多模态图像研究 AIP 与 PDA 的鉴别诊断, 提出了一个基于多模纹理特征的 SVM 鉴别模型。

鉴于深度学习算法对数据量要求较高^[15], 本文采用统计、频域变换和高分辨率灰度直方图等特征提取算法来提取 CT 和 PET 图像内的纹理特征。经 Fisher 准则对特征进行预筛选后, 为充分考虑特征间的相关性, 本文采用序列前向浮动选择算法 (sequence forward floating selection algorithm, SFFS)^[16]选择最优多模特征子集。最后, 本文采用 SVM 分类器实现 AIP 与 PDA 的鉴别。

1 方法

本文首先基于胰腺在 PET/CT 成像中的显像特点及临床数据处理方式, 对收集到的 PET/CT 数据进行预处理, 然后采用多种特征提取算法分别提取 CT 和 PET 图像感兴趣区 (region of interest, ROI) 内的纹理特征。为降低特征维度, 我们先采用 Fisher 准则进行特征预筛选, 剔除 Fisher 比接近于 0 的特征, 之后, 我们选择在小样本上具有稳定表现的 SVM 分类器作为最终分类器, 并将它结合到 SFFS 内, 构建 SFFS-SVM 特征选择算法, 为 SVM 分类器选择最有利于其性能的多模特征子集。最后, SVM 分类器基于该多模特征子集实现 AIP 与 PDA 的鉴别。基于多模纹理特征的 AIP 与 PDA 鉴别的具体步骤如下所示。

1.1 数据预处理

对于 CT 数据, 根据胰腺组织 CT 值范围, 本文截取 $[-10, 100]$ HU 内的内容以增强图像细节; 对于 PET 数据, 本文先采用双三次内插使其获得与 CT 图像相同的空间分辨率 ($0.98 \text{ mm} \times 0.98 \text{ mm}$), 然后, 采用图像刚性配准算法将 PET 数据配准到与其对应的 CT 数据上, 最后将其灰度值变换到临床上常用的 SUV 值来进行后续特征计算。本文中, 我们以胰腺病灶为 ROI, 数据预处理过程如图 1 所示。

1.2 多模纹理特征提取

不同模态的图像因其成像原理不同而含有不同的信息。对于 CT 图像, 其灰度差异能够凸显病灶组织的粗糙度、规则性和方向性; 对于 PET 图

像,其灰度值可以反映病灶组织的代谢情况,而空间密度差异则可以反映病灶内部的代谢变化。本文首先基于采集到的原始数据计算临床上常用的4个诊断特征:CT图像ROI内的平均CT值和最大CT值(CT_{mean} 、 CT_{max})以及PET图像ROI内的平均SUV值和最大SUV值(SUV_{mean} 、 SUV_{max}),然后本文将CT和PET图像内的灰度值转换到0~255来进行纹理信息的提取。本文中,我们采用统计、频域变换和多分辨率直方图算法来提取多模态图像病灶内部的纹理特征。

对于灰度共生矩阵(gray level co-occurrence matrix, GLCM),为提高运算效率,本文在不影响纹理特征的前提下进行了灰度级压缩,将CT和PET图像压缩到16个灰度级。类似地,对于灰度

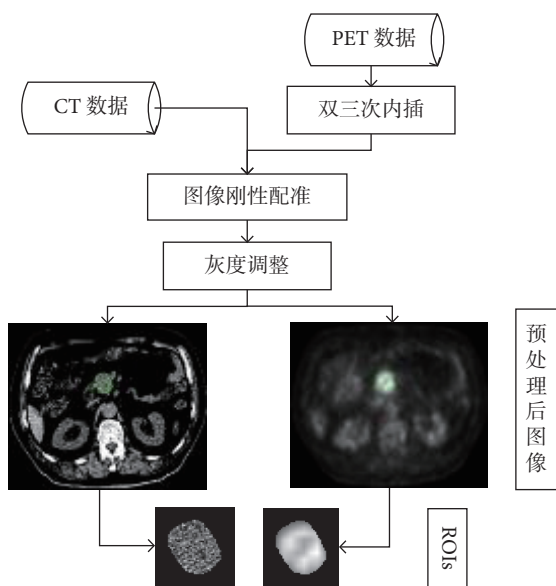


图1 数据预处理流程图

Fig.1 Flowchart of the data preprocessing method

差统计(gray level difference statistics, GLDS)、灰度游程矩阵(gray level run length matrix, GLRLM)、局部灰度差分矩阵(neighborhood gray tone difference matrix, NGTDM)和灰度区域尺寸矩阵(gray level zone size matrix, GLSZM)^[17],本文将CT和PET图像压缩至32个灰度级以进行相关统计特征的提取。

本文采用Gabor滤波器^[18]进行频域特征的提取。Gabor滤波器是经尺度伸缩和方向旋转产生的一组带通滤波器,在空域与频域均有较好的分辨率。本文中,我们采用以5个尺度(波长为2到6)、4个方向(0°, 45°, 90°, 135°)的Gabor函数为母波函数的小波变换来提取ROI在不同频域尺度和纹理方向的特征信息。

多分辨率灰度直方图特征^[19]具有平移和旋转不变性,同时对噪声和尺寸变化不敏感,且相较于高阶纹理统计特征,其在表达图像密度空间信息时计算更加简单。本文中,考虑到ROI的实际尺寸,我们先采用高斯滤波金字塔算法来获得3个不同分辨率的ROI图像,然后计算累计直方图来获得2个差分直方图,并以4为步长对灰度级为256的差分直方图进行欠采样,最后连接欠采样后的2个差分直方图形成128维的多分辨率灰度直方图特征向量。

本文最终从CT和PET图像中共提取了418个特征,特征分量集合见表1。

1.3 特征选择与建模

上述特征提取过程提取到的大量特征存在特征冗余,为了避免引发“维度灾难”,降低学习难度,需要选取合适的特征选择方法来剔除不相关和冗余特征。Fisher准则和SFFS^[16]算法是常见的两

表1 多模态特征分量统计表

Tab.1 Multi-modality feature component statistics

特征类别	特征分量
诊断特征	CT_{mean} 、 CT_{max} 、 SUV_{mean} 、 SUV_{max}
一阶统计特征	基于灰度直方图的4个统计特征:方差、熵、倾斜度、峰度
二阶统计特征	基于GLCM的20个统计特征:自相关、对比度、相关性、差异性、能量、熵、均匀性、最大概率、和均值、和方差、和熵、平方方差和、差方差、差熵、差分矩、逆差分矩、信息测度1、信息测度2、聚块阴影度、聚块突出性; 基于GLRLM的13个统计特征:短游程因子、长游程因子、灰度不均匀性、游程长度不均匀性、游程百分比、低灰度游程因子、高灰度游程因子、短游程低灰度因子、短游程高灰度因子、长游程低灰度因子、长游程高灰度因子、灰度方差、游程长度方差
高阶统计特征	基于GLDS的4个特征:均值、对比度、角度方向二阶距、熵; 基于NGTDM的5个特征:粗糙度、对比度、频度、复杂度、纹理强度; 基于GLSZM的13个特征:小区域因子、大区域因子、灰度不均匀性、区域尺寸不均匀性、区域百分比、低灰度区域因子、高灰度区域因子、小区域低灰度因子、小区域高灰度因子、大区域低灰度因子、大区域高灰度因子、灰度方差、区域尺寸方差
频域纹理特征	Gabor滤波器5个尺度4个方向子带图像的灰度均值
多分辨率灰度直方图特征	128个多分辨率灰度直方图特征

种特征选择算法,前者计算复杂度小且易于推广,后者直接把要使用的分类器的性能作为特征子集的评价准则,在分类器上获得的准确率一般较高。因此,本文先采用 Fisher 准则进行特征预筛选,剔除鉴别性能极差的特征,然后采用 SFFS 算法选出最优的多模特征子集。

Fisher 准则的主要思想是类内距离小、类间距离大的特征鉴别性能更强。本文采用单个特征的 Fisher 比作为筛选准则,将 Fisher 准则与后续的 SFFS 算法相结合对预筛选的 Fisher 阈值进行网格调优,最终从原有 418 个特征中筛选出 Fisher 比大于 0.01 的 293 个特征,从而节省后续 SFFS 特征选择中的计算开销。

本文中,我们采用在小样本上具有稳定表现的 SVM 分类器对 AIP 和 PDA 进行鉴别。SVM 分类器根据结构风险最小化准则,以最大化分类间隔构造最优分类超平面来提高分类器的泛化能力,它能较好地解决非线性、高维数和局部极小点等问题。为提高 SVM 分类器的鉴别能力,本文把在 SVM 分类器上的表现作为 SFFS 算法的评价准则,构建 SFFS-SVM 特征选择算法,为 SVM 分类器选择最有利于其鉴别性能的特征子集,该算法的特征子集搜索过程如表 2 所述。

在特征子集搜索过程中,对于搜索到的每个特征子集,我们在 SVM 分类器上进行 10 次 10 折交叉验证,以它们 F1 值的平均值作为该特征子集鉴别性能的评价函数。10 次 10 折交叉验证的具体过程如下:① 将数据集随机划分成 10 组,轮流取其中 9 组患者的 ROI 样本作为训练集,用于训练 SVM 分类器,剩余 1 组作为测试集,计算 SVM 分类器在当前测试集上的 F1 值,10 次实验后,计算 10 次实验 F1 值的平均值;② 将步骤① 重复 10 次,对 10 次随机划分得到的 F1 值的平均值求平均。步骤① 用于消除单一测试集上 F1 值具有的偶然性,步骤② 用于消除每次数据集随机划分带来的结果

的随机性。F1 值是精确率 (precision) 和召回率 (recall) 的调和平均,是对二者的综合评价。其中精确率可以反映 AIP 被错误预测的情况,而召回率则可以反映样本中所有 PDA 被正确预测的情况。F1 值、精确率及召回率的定义分别见公式 (1) ~ (3):

$$F1 = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (1)$$

其中:

$$\text{precision} = \frac{TP}{TP + FP} \quad (2)$$

$$\text{recall} = \frac{TP}{TP + FN} \quad (3)$$

其中,TP: 真阳性例数,是指 PDA 被识别为 PDA 的个数;FN: 假阴性例数,是指 PDA 被识别为 AIP 的个数;FP: 假阳性例数,是指 AIP 被识别为 PDA 的个数;TN: 真阴性例数,是指 AIP 被识别为 AIP 的个数。

当选中的特征总数连续增加 4 个而没有使整体分类性能获得提升时,停止搜索。最后,我们在 SFFS-SVM 特征选择算法搜索到的最优多模特征子集上构建 SVM 鉴别模型,实现 AIP 与 PDA 的鉴别。本文采用敏感性 (sensitivity)、特异性 (specificity)、阴性预测值 (negative predictive value, NPV) 和阳性预测值 (positive predictive value, PPV) 对模型的性能进行分析,其定义分别如式 (4) ~ (7) 所示。

$$\text{sensitivity} = \frac{TP}{TP + FN} \quad (4)$$

$$\text{specificity} = \frac{TN}{TN + FP} \quad (5)$$

$$\text{NPV} = \frac{TN}{TN + FN} \quad (6)$$

$$\text{PPV} = \frac{TP}{TP + FP} \quad (7)$$

表 2 SFFS-SVM 特征选择算法步骤

Tab.2 Steps of SFFS-SVM feature selection algorithm

步骤	算法
Step 1	起始于空集 $Y = \{\emptyset\}, k = 0$
Step 2	如果满足停止准则,退出; 否则,在未选择的特征中选择一个特征 x^+ , 使加入特征 x^+ 后评价函数 J 达到最优 $x^+ = \arg \max_{x \notin Y_k} J(Y_k + x), Y_k = Y_k + x^+; k = k + 1$
Step 3	在已选择的特征中选择特征 x^- , 使得剔除特征 x^- 后评价函数 J 达到最优 $x^- = \arg \max_{x \in Y_k} J(Y_k - x)$
Step 4	如果 $J(Y_k - x^-) > J(Y_k)$, 那么 $Y_k = Y_k - x^-; k = k + 1$, 跳至 Step 3; 否则, 跳至 Step 2

注: 评价函数 J 为特征子集在 SVM 分类器上进行 10 次 10 折交叉验证得到的 F1 值的平均值

2 结果与分析

2.1 实验环境

软件环境：Windows10 操作系统, MatlabR2017a;

硬件环境：Intel (R) Xeon (R) CPU E5-2960 v3 @ 2.60 GHz, 64 GB 内存。

2.2 数据来源

本研究为回顾性分析, 经上海市长海医院伦理委员会审查, 在确保病患个人隐私的基础上可以免除知情同意。PET/CT 原始数据集在长海医院核医学科采集, 其中 AIP 患者 45 例 (弥漫性 26 例, 局灶性 19 例), PDA 患者 66 例。所有患者均通过组织学检查或液基细胞学检查确诊。所有 PET/CT 数据均是在注射显像剂约 1 h 后由德国西门子公司 Biograph64 PET/CT 扫描仪扫描获得, CT 数据和 PET 数据的像素大小分别为 0.98 mm 和 4.07 mm, 层距为 3 mm。图2 为一例 AIP 患者的原始数据。本研究中的 ROI 是在经验丰富的核医学科医师指导下, 在 3D Slicer^[20] 软件上目测确认胰腺在 PET 图像中的代谢浓聚灶, 取横断面 SUV 值最大层面, 结合 CT 图像中的解剖学信息在 CT 图像上沿病灶边缘手动精确勾画得出。

2.3 实验分析

2.3.1 SFFS-SVM 特征选择性能分析 当 SFFS 结合不同的分类器时, 由于分类器的学习方式不同, 搜索到的最优特征子集及表现出的分类性能也不同。如图3 所示, 本文对比分析了几种常见的分类器与 SFFS 结合时的表现, 包括 Fisher 线性判别 (Fisher's linear discriminant analysis, LDA) 分类器、k-近邻 (k-nearest neighbors, kNN) 分类器、分类回归树 (classification and regression tree, CART) 和 SVM。需要指出的是, 由于 SFFS 算法在特征选择过程中的特性, 同一特征个数下可能对应不同时刻产生的多个特征子集, 因此, 本文对这些特征子集在 10 次 10 折交叉验证中产生的 F1 值平均值取最高值作为图3 中在当前特征个数下的 F1 值。

对于 4 种不同的分类器, 当选中的特征个数小于 5 时, F1 值随着选中的特征个数的增加而增加, 表明多个特征的组合更有利于 AIP 与 PDA 的鉴别。当选中的特征个数超过 5 时, LDA、kNN 和 CART 开始出现过拟合, 鉴别性能开始下降。由于 SVM 分类器在小样本上解决非线性、高维数问题的能力更强, 所以在其 F1 值达到最高值后, F1 值不会因选中的特征个数的增加而出现明显下降。在进一步的实验中我们发现, 当选中的特征个数超

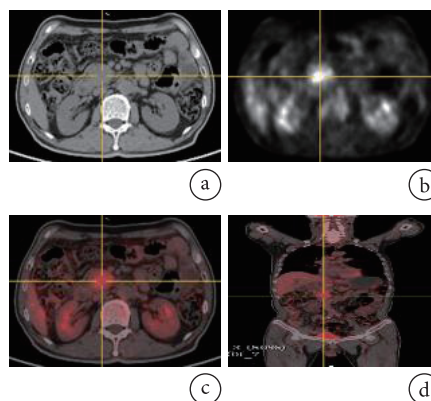


图2 PET/CT原始数据 a. 腹部横断面 CT 图像; b. 腹部横断面 PET 图像; c. 腹部横断面 PET/CT 图像; d. 全身冠状面 PET/CT 图像

Fig.2 PET/CT raw data a. abdominal cross section CT image; b. abdominal cross section PET image; c. abdominal cross section PET/CT image; d. whole body coronal PET/CT image

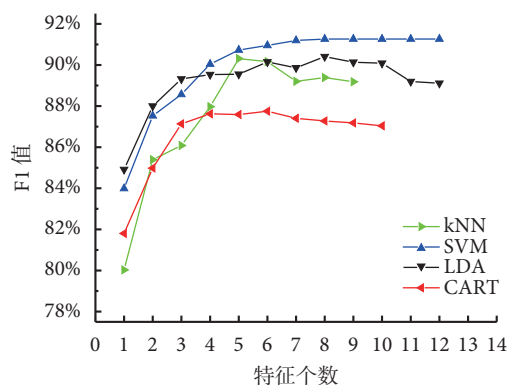


图3 结合不同分类器的SFFS算法

Fig.3 SFFS algorithm combined with different classifiers

过 19 个时, SVM 分类器才会出现过拟合现象。对比 4 种不同分类器我们可以发现, SVM 分类器优势较大, 其 F1 值最高可以达到 91.26%, 此时选中的最优特征子集共包含 8 个特征。

2.3.2 多模特征分析 SFFS-SVM 特征选择算法选中的 8 个多模特征如图4 所示。在选择的特征中, PET 图像中的基于 GLDS 的角度方向二阶距的 Fisher 比最大, 证明其所代表的纹理粗糙度对 AIP 与 PDA 的鉴别能力较强。对该特征进一步研究发现, 在 AIP 样本中该特征值较大, 表明 AIP 患者病灶内的功能代谢变化较大, 纹理更粗糙。当 SVM 分类器仅基于该特征进行鉴别时, 10 次 10 折交叉验证的 F1 值平均值可以达到 83.93%, 高于 SUV_{max} 的 80.37%, 表明该特征对 AIP 与 PDA 的鉴别能力高于临床上常用的 SUV_{max} 指标, 将这两个特征组合使用时, F1 值为 84.05%。选中的其余 6 个纹理特征单独使用时不具备鉴别能力, 将它们与上述两

个特征组合使用时,可以将 F1 值提升至 91.26%,表明胰腺病灶内部的多模纹理特征的组合使用有助于对 AIP 与 PDA 进行鉴别。

对比图 4 中不同模态的特征可以发现, PET 图像中的特征的 Fisher 比更大,这表明 PET 图像反映的病灶内部代谢的变化更有助于 AIP 与 PDA 的鉴别。为了进一步研究不同模态的特征的鉴别能力,本文基于 SFFS-SVM 特征选择算法分别对从 CT 图像中提取到的特征和从 PET 图像中提取到的特征进行筛选,得到的最优特征子集分别包括 9 个和 3 个特征, SVM 分类器在上述特征子集上的性能对比见图 5,其中 PET/CT 指的是上文中的将多模态特征组合后再进行特征选择获得的 8 个最优特征的性能。图 5 表明,相对于 CT 图像特征, PET 图像特征可以使 SVM 鉴别模型获得更高的特异性和准确率。同时我们可以发现,将两种模态的特征组合在一起进行筛选得到的最优多模特征子集可以使鉴别模型的特异性和准确率都得到大幅提升。

2.3.3 模型性能分析

(1) 不同鉴别模型的性能对比

本节中,我们将本文构建的 SVM 鉴别模型与其他胰腺医学图像分析任务^[12]中用到的 RF 和 Adaboost 模型进行性能对比。为了保证对比的公平性,本文同样以 10 次 10 折交叉验证得到的 F1 值的平均值作为评价指标,对 Fisher 准则预筛选的阈值、RF 中树的数量及 Adaboost 中的迭代次数进行网格调优。最终,对于 RF,其预筛选的 Fisher 阈值为 0.13,树的数量为 60 棵;对于 Adaboost,其对应的 Fisher 阈值为 0.11,迭代次数为 40 次。

为了在不同截断点对模型进行对比,我们进行受试者工作特征曲线(receiver operating characteristic curve, ROC)分析。对于每个模型,本文采用 bootstrap 自助采样的方法对数据集采样 5 000 次,计算模型在每次采样下得到的 ROC 曲线及曲线下面积(area under curve, AUC)。本文对 5 000 次采样得到的 ROC 曲线求平均,不同模型的平均 ROC 曲线如图 6 所示。本文对不同模型平均 ROC 曲线的 AUC 值进行计算,可以得到 SVM 鉴别模型的 AUC 值为 0.921 8,高于 RF 的 0.834 7 和 Adaboost 的 0.854 8。进一步地,本文基于不同模型 5 000 次采样得到的 AUC 值,采用成对 *t* 检验对模型的性能差异进行分析。结果表明,在 95% 的置信度下, SVM 鉴别模型相对于 Adaboost 和 RF 的差异有统计学意义($P < 0.001$, $P < 0.001$), Adaboost 与 RF 的识别性能差异也具有统计学意义($P < 0.001$)。

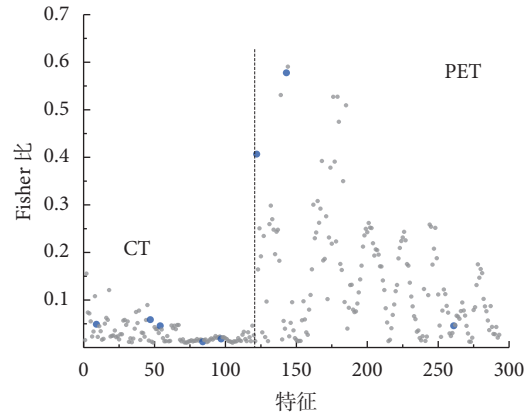


图 4 特征分析 灰点表示 Fisher 准则选出的 293 个特征;蓝点表示 SFFS-SVM 特征选择算法选出的 8 个特征,自左至右依次为: CT 图像的基于 GLCM 的聚块阴影度、基于 NGTDM 的纹理强度、基于 GLSZM 的小区域低灰度因子、多分辨率灰度直方图的第 26 和第 46 个特征,以及 PET 图像的 SUV_{max} 、基于 GLDS 的角度方向二阶距和多分辨率灰度直方图的第 90 个特征,相应的 Fisher 比依次为: 0.049 0, 0.058 7, 0.045 9, 0.012 2, 0.018 1, 0.406 8, 0.577 7, 0.045 9

Fig.4 Feature analysis gray points represents the 293 features selected by the Fisher criterion; blue points represents the 8 features selected by the SFFS-SVM algorithm, from left to right: GLCM-based cluster shade, NGTDM-based strength, GLSZM-based small zone low gray level emphasis, and multi-resolution gray histogram's 26th and 46th feature for CT image; SUV_{max} , GLDS based angular second-order moment and multi-resolution gray histogram's 90th feature for PET image; the corresponding Fisher ratios are 0.049 0, 0.058 7, 0.045 9, 0.012 2, 0.018 1, 0.406 8, 0.577 7, 0.045 9, respectively

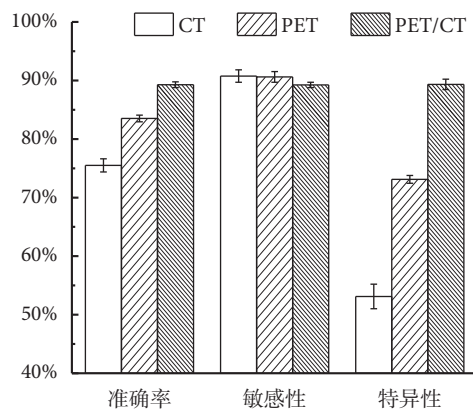


图 5 不同模态特征的性能对比

Fig.5 Performance comparison of different modalities' features

(2) 模型性能讨论

目前临床上基于 PET/CT 对 AIP 与 PDA 进行鉴别的研究大多基于 SUV 值。Ozaki 等^[21]对 15 例 AIP 和 26 例 PDA 的胰腺 SUV 值进行了定量分析,结果显示两组病例的 SUV_{max} 无显著差异。Zhang 等^[8]对 26 例 AIP 患者和 40 例 PDA 患者进行的一

项定量研究中, 胰腺 SUV_{max} 对 AIP 与 PDA 的鉴别准确率为 72.7%。因此, 由于肿瘤异质性的特点, SUV_{max} 不能完全表征病灶的代谢情况, 无法进行准确鉴别。

本文中, 我们采用 10 次 10 折交叉验证法和留一法两种验证方法对基于多模纹理特征的 SVM 模型的识别性能进行验证。留一法是指每次轮流留下一例患者的样本作为测试集, 其他全部用作训练集, 以提高样本利用率。留一法和 10 次 10 折交叉验证得到的模型性能统计见表 3。从表 3 可以发现, 本文方法在 10 次 10 折交叉验证中的平均鉴别准确率为 89.28%, 在留一法验证中的鉴别准确率为 89.19%。尽管与 Ozaki 等^[21]和 Zhang 等^[8]的研究基于不同的数据集, 但本文在更大的数据集上 (45 例 AIP, 66 例 PDA) 表现出的性能优势可以表明, 病灶内部的纹理特征有助于实现 AIP 与 PDA 的准确鉴别。此外, 本模型在两种验证方式中有相似的性能表现, 且在 10 次 10 折交叉验证中各项性能指标的标准差均小于 1%, 体现出本模型具有较高的稳定性; 同时, 在两种验证方式中, 本模型的敏感性与特异性都非常接近, 表明本模型可以无偏地对 AIP 与 PDA 进行鉴别。

本文中, 我们收集了 PET/CT 数据集对应的临床诊断结果, 该结果由上海市长海医院具有 10 年以上临床经验的两名核医学科医生共同讨论决定, 其中一人为副高以上人员。在临床诊断结果中, 敏感性为 95.45% (63 例/66 例), 特异性为 75.56% (34 例/45 例), 鉴别准确率为 87.39% (97 例/111 例)。相比于本文方法, 现有临床诊断手段敏感性高而特异性低, 容易将 AIP 误诊为 PDA, 在鉴别准确率方面, 本文方法略高于现有临床诊断手段。综上所述, 对病灶内部的纹理分析可以取得略高于现有临床诊断手段的结果, 由于其特异性较高, 将其作为现有临床诊断手段的补充将有助于提高 AIP 的诊断准确率。

3 总结与展望

针对临床上 AIP 与 PDA 准确鉴别存在的困难, 本文提出了一个基于¹⁸F-FDG PET/CT 多模纹理特征的 SVM 鉴别模型。本文首先以胰腺病灶为 ROI, 利用统计、频域变换和多分辨率灰度直方图算法来提取 CT 和 PET 图像内的纹理特征。然后, 利用 Fisher 准则对提取到的特征进行预筛选后, 本文采用 SFFS-SVM 特征选择算法来选择出鉴别性能最优的多模特征子集。最后, 本文基于该子集建立

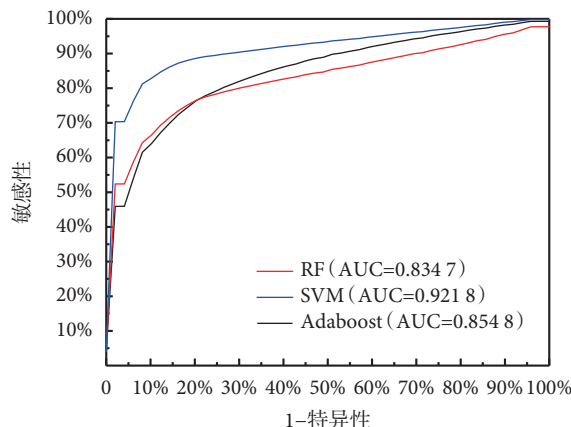


图 6 不同模型的ROC曲线图

Fig.6 ROC curves of different methods

表 3 模型性能统计 (%)

Tab.3 Model performance statistics (%)

指标	10次10折交叉验证	留一法
准确率	89.28 ± 0.49	89.19
敏感性	89.24 ± 0.45	89.39
特异性	89.33 ± 0.89	88.89
阳性预测值	92.47 ± 0.59	92.19
阴性预测值	84.99 ± 0.57	85.11

了一个 SVM 鉴别模型实现 AIP 与 PDA 的鉴别。bootstrap 抽样实验结果表明, 本文构建的 SVM 鉴别模型在识别性能上优于随机森林和 Adaboost 算法。10 次 10 折交叉验证和留一法实验结果表明, 该模型具有较高的鉴别准确率, 可以无偏地对 AIP 与 PDA 进行鉴别。对多模特征和模型性能的分析表明, 对胰腺 PET/CT 病灶内部的纹理特征进行分析有助于实现 AIP 与 PDA 的准确鉴别。

在后续工作中, 我们将进一步扩充数据集, 在大数据集上利用深度学习的方法提取多模态图像中的病灶内部特征。此外, 如何有效地实现 PET/CT 图像在像素级别上的数据融合, 在避免丢失原有信息的前提下引入更多具有区分性的信息, 也是我们后续的一个研究方向。

利益冲突声明: 本文全体作者均声明不存在利益冲突。

参考文献

- 1 Finkelberg D L, Sahani D, Deshpande V, et al. Autoimmune pancreatitis. *N Engl J Med*, 2009, 355(25): 2670-2676.
- 2 Ferrone C R, Pieretti-Vanmarcke R, Bloom J P, et al. Pancreatic ductal adenocarcinoma: Long-term survival does not equal cure. *Surgery*, 2012, 152(3): S43-S49.
- 3 Vijayakumar A, Vijayakumar A. Imaging of focal autoimmune pancreatitis and differentiating it from pancreatic cancer. *ISRN Radiology*, 2013(4): 569489.
- 4 Meng Qianqian, Xin Lei, Liu Wenyu, et al. Diagnosis and

- treatment of autoimmune pancreatitis in China: a systematic review. *PLoS One*, 2015, 10(6): e0130466.
- 5 王珏磊, 杨云生, 梁浩, 等. 自身免疫性胰腺炎不同诊断标准间诊断率的差异及其主要影响因素分析. *中华内科杂志*, 2013, 52(6): 498-502.
 - 6 O'Reilly D A, Malde D J, Duncan T, *et al.* Review of the diagnosis, classification and management of autoimmune pancreatitis. *World J Gastrointest Pathophysiol*, 2014, 5(2): 71-81.
 - 7 Shimosegawa T, Chari S T, Frulloni L, *et al.* International consensus diagnostic criteria for autoimmune pancreatitis: guidelines of the International Association of Pancreatology. *Pancreas*, 2011, 40(3): 352-358.
 - 8 Zhang J, Jia G, Zuo C, *et al.* 18 F-FDG PET/CT helps differentiate autoimmune pancreatitis from pancreatic cancer. *BMC Cancer*, 2017, 17(1): 695.
 - 9 王绍波, 吴湖炳, 王全师, 等. 自身免疫性胰腺炎的18F-FDG PET/CT特征. *临床放射学杂志*, 2016, 35(5): 736-739.
 - 10 朱凌云, 吴宝明, 曹长修. 医学数据挖掘的技术、方法及应用. *生物医学工程学杂志*, 2003, 20(3): 559-562.
 - 11 Gazit L, Chakraborty J, Attiyeh M, *et al.* Quantification of CT images for the classification of high- and low-risk pancreatic cysts// *SPIE Medical Imaging*, 2017. Orlando: SPIE, 2017: 101340X.
 - 12 Zhu Maoling, Xu Can, Yu J, *et al.* Differentiation of pancreatic cancer and chronic pancreatitis using computer-aided diagnosis of endoscopic ultrasound (EUS) images: a diagnostic test. *PLoS One*, 2013, 8(5): e63820.
 - 13 Zhu J, Wang L, Chu Y, *et al.* A new descriptor for computer-aided diagnosis of EUS imaging to distinguish autoimmune pancreatitis from chronic pancreatitis. *Gastrointest Endosc*, 2015, 82(5): 831-836.
 - 14 Schölkopf B, Smola A J. *Learning with kernels: support vector machines, regularization, optimization, and beyond*. Cambridge: MIT Press, 2002, 98(462): 489.
 - 15 Cho J, Lee K, Shin E, *et al.* How much data is needed to train a medical image deep learning system to achieve necessary high accuracy?. *arXiv: 1511.06348*, 2016.
 - 16 Pudil P, Novovičová J, Kittler J. Floating search methods in feature selection. *Pattern Recogn Lett*, 1994, 15(11): 1119-1125.
 - 17 Vallières M, Freeman C R, Skamene S R, *et al.* A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities. *Phys Med Biol*, 2015, 60(14): 5471-5496.
 - 18 Haghighat M, Zonouz S, Abdel-Mottaleb M. Identification using encrypted biometrics// Wilson R, Hancock E, Bors A, *et al.* *CAIP 2013: Computer Analysis of Images and Patterns*. Springer Berlin Heidelberg, 2013: 440-448.
 - 19 Hadjidemetriou E, Grossberg M D, Nayar S K. Multiresolution histograms and their use for recognition. *IEEE Trans Pattern Anal Mach Intell*, 2004, 26(7): 831-847.
 - 20 Pieper S, Halle M, Kikinis R. *3D Slicer*// 2004 2nd IEEE International Symposium on Biomedical Imaging: Nano To Macro. Arlington: IEEE, 2004, 1: 632-635.
 - 21 Ozaki Y, Oguchi K, Hamano H, *et al.* Differentiation of autoimmune pancreatitis from suspected pancreatic cancer by fluorine-18 fluorodeoxyglucose positron emission tomography. *J Gastroenterol*, 2008, 43(2): 144-151.

收稿日期: 2018-07-15 修回日期: 2018-11-28
本文编辑: 李蓓兰