

doi: 10.12068/j.issn.1005-3026.2020.02.008

基于活动轮廓模型和影像组学的 乳腺癌 LVI 状态预测

冯宝^{1,2}, 李昌林³, 李智², 刘壮盛³

(1. 中山大学 生物医学工程学院, 广东 广州 511400; 2. 桂林航天工业学院 电子信息与自动化学院, 广西 桂林 541004; 3. 中山大学附属江门市中心医院 放射科, 广东 江门 529000)

摘要: 针对乳腺癌患者术前 LVI 状态预测问题, 提出了活动轮廓模型和影像组学相结合的计算机辅助分析方法. 首先, 提出一种基于后验概率和模糊速度函数的活动轮廓模型方法来完成乳腺癌 DCE-MRI 图像分割. 通过在小波域下构建基于后验概率的活动轮廓模型的区域项, 同时利用模糊速度函数构建活动轮廓模型的边界项, 可以提高乳腺癌病灶分割的准确性. 其次, 提取形态、灰度、纹理等图像特征, 利用集成分类器随机森林方法构造 LVI 状态的预测模型. 实验结果表明, 所构建的模型对乳腺癌患者 LVI 状态具有较好的预测能力.

关键词: DCE-MRI; 乳腺癌; 活动轮廓模型; 影像组学; 图像分割; 图像分类

中图分类号: TP 391.41

文献标志码: A

文章编号: 1005-3026(2020)02-0193-07

Breast Cancer Related LVI Status Prediction Based on Active Contour Model and Radiomics

FENG Bao^{1,2}, LI Chang-lin³, LI Zhi², LIU Zhuang-sheng³

(1. School of Biomedical Engineering, Sun Yat-sen University, Guangzhou 511400, China; 2. School of Electronic Information and Automation, Guilin University of Aerospace Technology, Guilin 541004, China; 3. Department of Radiology, Affiliated Jiangmen Hospital of Sun Yat-sen University, Jiangmen 529000, China. Corresponding author: LI Zhi, E-mail: cclizhi@guet.edu.cn)

Abstract: To predict preoperative lymphovascular invasion (LVI) status of breast cancer patients, a computer-aided analysis method which combines an active contour model and radiomics is proposed. First, the image from dynamic contrast enhanced magnetic resonance imaging (DCE-MRI) of breast cancer is segmented by an active contour model (ACM) based on post probability and fuzzy velocity function. By constructing the region term of the active contour model based on post probability in the wavelet domain and the edge stop term of the active contour model by using the fuzzy velocity function, the accuracy of breast cancer lesion segmentation can be improved. Second, the image features such as morphology, grayscale and texture are extracted. Finally, a model for predicting LVI status is developed by the random forest classifier and its predictive ability is verified by experimental results.

Key words: DCE-MRI; breast cancer; active contour model; radiomics; image segmentation; image classification

乳腺癌是发病率最高的女性恶性肿瘤^[1-3]. 肿瘤对周围脉管(包括淋巴管和血管)的侵犯状态(lymphovascular invasion, LVI)是预测乳腺癌

患者局部复发和转移的重要指标^[4-6]. 目前, 对肿瘤切除术后的标本行病理检查是诊断乳腺癌 LVI 状态的金标准和常规方法, 但该方法的 LVI 信息

收稿日期: 2018-12-21

基金项目: 广西壮族自治区自然科学基金资助项目(2016GXNSFBA380160); 广西壮族自治区千名中青年骨干教师培育计划项目(2018GXQGF160); 国家自然科学基金地区科学基金资助项目(81960324).

作者简介: 冯宝(1986-), 男, 山西太原人, 中山大学博士后研究人员; 李智(1965-), 男, 广西桂林人, 桂林航天工业学院教授, 博士生导师.

是在肿瘤切除术后获得,这对乳腺癌病人术前新辅助化疗和手术方式优化无参考价值.医学影像技术术前无创提供整个肿瘤信息,在指导临床治疗上潜力巨大^[7-8].在乳腺癌患者的影像诊断中,动态对比增强磁共振成像(dynamic contrast enhanced magnetic resonance imaging, DCE-MRI)凭借高敏感性、高空间分辨率等特点成为乳腺癌影像检查的主要工具.因此,开展基于 DCE-MRI 的乳腺癌患者术前 LVI 状态预测,对指导和优化患者诊疗方案意义重大.

影像组学是当前医工交叉学科的研究热点之一^[9].它利用临床、基因和影像等信息,通过模式识别方法来提取病灶区的有效特征,构建描述肿瘤表型的定量辅助诊断模型,再进行计算机辅助诊断.在影像组学的研究中,肿瘤病灶的准确分割和定量辅助诊断模型的构建是两个关键技术.其中病灶的准确分割是一个重要的预处理步骤,病灶区分割的准确性将直接影响后续影像特征提取的准确性和稳定性,进而决定模型构建的有效性;而模型构建的有效性可以作为肿瘤病灶分割的评价标准之一.

近年来,基于影像组学的乳腺癌影像分析受到广泛关注,已取得较多成果.刘桐桐等^[10]利用动态活动轮廓模型对乳腺癌病灶分割,然后提取形态、纹理等特征,利用支持向量机建立乳腺癌雌激素受体表达情况的预测模型.该算法获得的 AUC 值为 0.738,准确率为 0.712,敏感度为 0.732,特异度为 0.758.吴佩琪等^[11]提出自动阈值方法获取病灶初始轮廓,同时结合医生手动分割调整来获取乳腺癌 DCE-MRI 图像的病灶,提取了灰度共生矩阵(GLCM)等纹理特征,利用逻辑回归方法建立乳腺导管癌病理分级模型.该方法获得的 AUC 值为 0.691,准确率为 0.676,敏感度为 0.905,特异度为 0.345. Huang 等^[12]提出基于模板匹配算法对乳腺癌 DCE-MRI 病灶图像分割,提取了基于 GLCM 的纹理特征,采用模糊 C 均值聚类方法对乳腺肿瘤良恶性分类. Fan 等^[13]利用医生手动分割方法来分割乳腺癌病灶,通过逻辑回归方法建立亚型分类模型(Luminal A 型和 non-Luminal A 型),获得 AUC 值为 0.718.此外,深度学习方法也受到关注.孙利雷等^[14]提出多尺度卷积神经网络对乳腺肿块良恶性分类,得到 AUC 值为 0.713.

上述方法一定程度上提高了乳腺癌计算机辅助诊断效率,但病灶区提取还是以半自动分割、医生手动分割为主,耗时耗力,重复性欠佳,不适合

大数据集上的影像分析.此外,大部分方法采用单分类器,应用在具有多样性的乳腺癌影像分析中存在一定局限性,临床应用很难达到满意的效果.

针对乳腺癌患者术前 LVI 状态预测问题,本文提出活动轮廓模型和影像组学结合的分析方法.首先对乳腺癌病灶区在 DCE-MRI 图像中存在的边界模糊、内部亮度不均匀等特点,提出一种基于后验概率和模糊速度函数的活动轮廓模型方法来分割病灶.通过在小波域下构建基于后验概率的活动轮廓模型的区域项,同时利用模糊速度函数改进活动轮廓模型的边界项,能提高分割的准确性.在构建定量辅助诊疗模型时,提取形态、灰度、纹理等影像特征来表征乳腺癌病灶,再用最小收缩和选择算子(least absolute shrinkage and selection operator, LASSO)方法选出有效特征.最终利用集成分类随机森林方法构建分类器.通过集成多棵决策树提高分类准确率,完成乳腺癌患者术前 LVI 状态预测任务.

1 LVI 状态预测方法

乳腺癌患者术前 LVI 状态预测问题有两个关键技术点:病灶的准确分割和分类预测模型的有效构建.病灶分割是重要的预处理步骤,分割准确度直接影响特征提取的准确性和稳定性,从而影响分类模型构建,因此病灶的准确分割是本文的关键;而分类模型的有效构建可以作为病灶分割的评价指标之一,更好地对乳腺癌患者进行术前 LVI 状态预测.

1.1 基于后验概率和模糊速度函数的乳腺癌病灶分割

由于部分乳腺癌病灶沿腺体分布,导致其形态极不规则,边界模糊,对比度低,准确分割困难极大^[15].针对上述问题,提出一种基于后验概率和模糊速度相结合的活动轮廓模型方法来分割病灶:①利用小波变换将图像灰度信息转变为小波能量,增强目标区域与背景差异,再在能量图中计算每个像素点的后验概率,建立基于后验概率的活动轮廓模型区域项;②结合 Gabor 纹理特征和灰度特征构建模糊速度函数,并引入到活动轮廓模型中作为边界检测项;③计算构建的活动轮廓模型的梯度下降流,驱动轮廓曲线演变,完成病灶的分割.

1.1.1 活动轮廓模型的能量函数定义

假定 C 是一条轮廓曲线,该曲线处于封闭状态并把曲线区域划分为内部和外部;因此,定义活

动轮廓模型的能量函数为

$$E(C, \phi) = \mu E_{\text{Region}}(\phi) + \beta E_{\text{Contour}}(C). \quad (1)$$

式中: $E_{\text{Region}}(\phi)$ 和 $E_{\text{Contour}}(C)$ 分别是控制轮廓曲线演变的区域项和边界检测项; μ 和 β 均是常数 (文中 μ 和 β 均取 1)^[16]; ϕ 是水平集函数, 在 C 内部取正数, 在外部取负数.

1.1.2 小波能量引导下基于后验概率的活动轮廓模型区域项

图像任意像素点 (x, y) 的小波能量定义为

$$W(x, y) = E_A(x, y) + E_H(x, y) + E_V(x, y) + E_D(x, y). \quad (2)$$

式中, $E_A(x, y)$, $E_H(x, y)$, $E_V(x, y)$ 和 $E_D(x, y)$ 分别是图像的近似细节、水平细节、垂直细节和对角细节小波能量.

将原始 DCE-MRI 图像分成病灶区和背景区, 假定这两个区域中像素的小波能量值都满足均值为 u_r 、方差为 σ_r ($r=1, 2$) 的高斯分布, 采用一个混合高斯模型^[17]来表达整张图像. 将图像像素定义为 $\{x_0, x_1, \dots, x_N\}$ (N 是像素个数), x_i 是第 i 个像素点小波能量值, 当区域 r 有 x_i 值时, 混合高斯分布定义为

$$P(x_i, r) = P(x_i | r) P_r. \quad (3)$$

式中: 系数 $P_r = \pi_r$ 且 $\sum_{r=1,2} \pi_r = 1, 0 \leq \pi_r \leq 1$; $P(x_i | r)$ 为第 r 个混合高斯模型的概率密度, 定义为

$$P(x_i | r) = \frac{1}{\sqrt{2\pi}\sigma_r} \exp\left[-\frac{(x_i - u_r)^2}{2\sigma_r^2}\right]. \quad (4)$$

通过最大期望算法^[18]求均值 u_r 、方差 σ_r 和 P_r , 得到区域 r 中有 x_i 值的后验概率 $P(r | W_{xy}, \theta^{(n)})$, $\theta^{(n)}$ 为通过最大期望算法求解最大似然函数获得的模型参数, W_{xy} 表示像素点 (x, y) 处的小波能量. 故活动轮廓模型区域项为

$$E_1(\phi) = \iint_{\Omega} P(1 | W_{xy}, \theta) H(\phi(x, y)) + P(2 | W_{xy}, \theta) [1 - H(\phi(x, y))] dx dy. \quad (5)$$

式中: $H(\cdot)$ 是 Heaviside 函数; Ω 为图像区域.

图 1 是乳腺癌病灶小波能量分布和后验概率分布图. 由图 1b 可看出病灶和正常腺体灰度值相近, 难以区分. 由图 1c 可看到正常腺体的小波能量没有病灶区强, 可据此区分病灶和正常腺体, 此外还可看到病灶区内部亮度不均匀. 由图 1d 可看出病灶区得到明显增强.

1.1.3 构建基于模糊速度的边界检测项

利用灰度特征和 Gabor 纹理特征相结合的模糊聚类算法, 计算模糊隶属度, 构建模糊速度函

数, 并将其引入活动轮廓模型中作为边界停止项.

模糊聚类算法需要多次迭代寻找最优的目标函数, 目标函数^[19]被定义为

$$J(U, v) = \sum_{l=1}^n \sum_{h=1}^2 (u_{lh})^m d^2(x_l, v_h). \quad (6)$$

式中: x_l 是二维特征向量; $U = (u_{lh})$ 是病灶区和背景区的模糊隶属度矩阵, u_{lh} 是 x_l 属于 h 类 ($h=1$ 表示背景, $h=2$ 表示病灶) 的隶属度大小; $v = [v_1, v_2]$ 是聚类中心向量; m 是模糊控制常数; v_h 是距离中心, $d^2(x_l, v_h)$ 是 x_l 和 v_h 的距离. 重复迭代计算 u_{lh} 和 v_h , 所有的二维向量被正确分类时, $J(U, v)$ 目标函数达到最小, 隶属病灶的模糊隶属度矩阵 Z 由 u_{lh} 得到.

为了更好分割病灶, 需设定较好的模糊速度函数: 离病灶边界处越远, 模糊速度越大; 边界处, 模糊速度为零.

将图像像素值 (x, y) 的模糊速度函数定义为

$$V(x, y) = e^{\alpha[Z(x, y) - 0.5]^2} - 1. \quad (7)$$

式中: α 是常数; $Z(x, y)$ 是 (x, y) 隶属病灶区的模糊隶属度矩阵.

把 $V(x, y)$ 加到活动轮廓模型中, 边界检测项 $E_2(C)$ 定义为

$$E_2(C) = \iint_{\Omega} V(x, y) \delta(\phi) |\nabla\phi| dx dy. \quad (8)$$

式中, $\delta(\phi) = \frac{\varepsilon}{\pi \varepsilon^2 + \phi^2}$ 为 Dirac delta 函数.

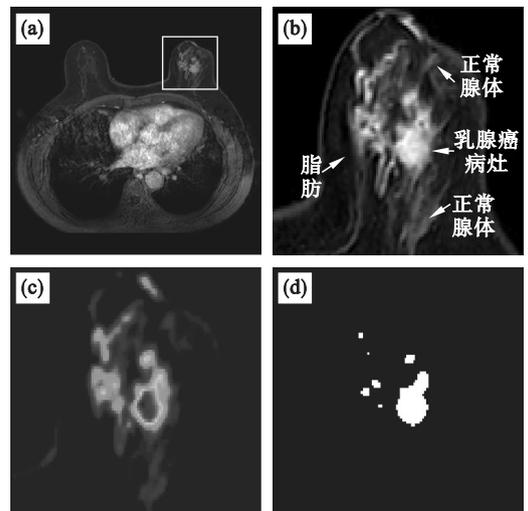


图 1 乳腺癌病灶的小波能量分布和后验概率分布
Fig. 1 Wavelet energy distribution and posterior probability distribution of breast cancer

(a)—乳腺 DCE-MRI 图像; (b)—病灶放大图;
(c)—小波能量分布图; (d)—后验概率分布图.

1.1.4 活动轮廓模型能量泛函演化

将式(5)区域项和式(8)边界检测项代入式(1)能量函数, 可得改进后的活动轮廓模型能量

泛函:

$$\begin{aligned}
 E(\phi) = & \beta \iint_{\Omega} p(1 | w_{xy}, \theta) H(\phi) dx dy + \\
 & \beta \iint_{\Omega} p(2 | w_{xy}, \theta) (1 - H(\phi)) dx dy + \\
 & \mu \iint_{\Omega} V(x, y) \delta(\phi) |\nabla \phi| dx dy + \\
 & \gamma \iint_{\Omega} \frac{1}{2} (|\nabla \phi| - 1)^2 dx dy. \quad (9)
 \end{aligned}$$

式中: β, μ 分别为控制区域项、边界检测项的权重系数; γ 用于控制 ϕ 偏离有符号距离函数的程度; $\iint_{\Omega} \frac{1}{2} (|\nabla \phi| - 1)^2 dx dy$ 是对 ϕ 的惩罚项, 防止水平集函数 ϕ 重新初始化, ϕ 要求 $|\nabla \phi| = 1$ (这里 $|\nabla \phi| = \sqrt{\phi_x^2 + \phi_y^2}$).

病灶分割是最小化能量函数 $E(\phi)$. 通过构建 $E(\phi)$ 关于水平集函数 ϕ 的 Euler-Lagrange 方程, 则

$$\begin{aligned}
 \beta \delta(\phi) (p_1 - p_2) - \mu \delta(\phi) \left(\operatorname{div} \left(V \frac{\nabla \phi}{|\nabla \phi|} \right) \right) - \\
 \gamma \left(\Delta \phi - \operatorname{div} \left(\frac{\nabla \phi}{|\nabla \phi|} \right) \right) = 0. \quad (10)
 \end{aligned}$$

根据变分原理, 求出 ϕ 的梯度下降流为

$$\begin{aligned}
 \frac{\partial \phi}{\partial t} = & -\beta \delta(\phi) (p_1 - p_2) + \\
 & \mu \delta(\phi) \left(\operatorname{div} \left(V \frac{\nabla \phi}{|\nabla \phi|} \right) \right) + \gamma \left(\Delta \phi - \operatorname{div} \left(\frac{\nabla \phi}{|\nabla \phi|} \right) \right). \quad (11)
 \end{aligned}$$

1.2 特征提取和特征选择

分割完病灶后, 提取其形态特征共 4 个^[20] (直径、体积、偏心率 and 密度), 灰度特征共 3 个^[20] (方差、偏度和峰度). 为提高特征的鲁棒性和可重复性, 对三维重建的病灶进行各向同性重采样^[21] 和灰度离散化^[22] 处理, 获得不同参数下的特征, 提取了基于灰度共生矩阵 (GLCM) 在不同重采样系数 (重采样系数 = [0.8, 1, 1.2, 1.5, 2]) 和不同灰度阶 (灰度阶 = [8, 16, 32, 64]) 下的纹理特征 (纹理能量、对比度、相关性、同质性、纹理方差、和均值、纹理熵、差异性、自相关)^[23-24], 共 $5 \times 4 \times 9 = 180$ 个.

由于特征间的相关性和冗余性会降低分类准确率, 过多无关特征易增加分类器复杂性, 出现过拟合; 故需对提取的特征进行优化选择^[25]. 本文选择“最小收缩和选择算子 (LASSO)”方法进行特征选择^[26]. 这是一种能稀疏特征的方法, 本质是在回归模型中引入 L_1 范数约束项, 将一些不显著特征对应的回归系数置 0, 实现特征的自动选择.

1.3 基于集成分类器随机森林的 LVI 状态预测模型

特征选择后, 采用集成分类器随机森林算法构建预测模型: 通过集成学习的思想将多棵决策树进行集成^[27]. 集成分类器随机森林算法能处理高维特征的输入样本, 训练速度快, 可有效避免过拟合.

集成分类器随机森林模型构建步骤如下:

① 选择乳腺癌 LVI 的训练集数据, 采用 Bootstrap 方法随机, 筛选出 k 个样本, 生成 m 个训练样本子集.

② 每个样本子集训练一棵分类决策树, 分别训练 m 个决策树模型. 每次训练过程中, 随机选取 n 个特征作为决策树每个节点的候选分裂特征, 然后通过计算基尼系数^[24] 来选取最好的特征.

③ 利用最优特征对决策树每个节点进行分裂, 划分决策树的左、右子数, 直到节点的所有训练样本属于同一类别. 决策树尽可能在不修剪的前提下生长, 直到达到最大迭代次数或者满足停止标准, 则停止该枝生长. 多个训练子集形成多棵决策树.

④ 将所有训练得到的决策树组成随机森林, 形成随机森林分类器.

⑤ 用训练好的随机森林分类器对测试集数据进行预测. 再根据多数投票准则, 将票数最多的分类结果作为最终预测结果.

1.4 分割和分类算法性能评价指标

1.4.1 分割性能评价指标

乳腺癌 DCE-MRI 的病灶分割以影像科医生的手动分割结果为金标准. 医生首先判定病灶区, 然后进行手动精准分割, 最后将分割的病灶区作为金标准. 分割指标采用阳性率 (true positive ratio, TPR)、假阳性率 (false positive ratio, FPR) 和相似度 (similarity degree, SD) 来评判:

$$\left. \begin{aligned}
 \text{TPR} &= \frac{|A_m \cap A_a|}{|A_m|}, \\
 \text{FPR} &= \frac{|A_m \cup A_a - A_m|}{|A_m|}, \\
 \text{SD} &= \frac{|A_m \cap A_a|}{|A_m \cup A_a|}.
 \end{aligned} \right\} \quad (12)$$

式中: A_m 是金标准区域; A_a 是算法分割区域. TPR 是 A_a 对 A_m 的覆盖率, TPR 越大, 分割效果越好. FPR 是 A_a 的背景区域对 A_m 的比值, FPR 值越小, A_a 包含的背景区域越小, 分割效果越好. SD 值是 A_a 和 A_m 的相似比值, SD 越大表明分割效果越好.

1.4.2 分类性能评价指标

为了评估分类算法性能,采用 3 个指标:准确率(Acc)、敏感度(Sen)、特异度(Spe),计算公式为

$$\left. \begin{aligned} \text{Acc} &= \frac{T_p + T_N}{T_p + T_N + F_p + F_N}, \\ \text{Sen} &= \frac{T_p}{T_p + F_N}, \\ \text{Spe} &= \frac{T_N}{T_N + F_p}. \end{aligned} \right\} \quad (13)$$

式中: T_p 和 F_p 分别表示将 LVI 状态正确和错误地划分为阳性的病例数; T_N 和 F_N 分别表示将 LVI 状态正确和错误地划分为阴性的病例数。

2 实验结果与分析

2.1 实验数据描述

本文选择中山大学附属江门市中心医院的 149 例乳腺癌患者 DCE-MRI 数据,90 例为训练集,59 例为测试集。根据术后病理结果,将患者 LVI 状态分为阳性(肿瘤侵犯周围脉管)和阴性(肿瘤没有侵犯周围脉管)。训练集包括 22 例 LVI 阳性和 68 例 LVI 阴性,测试集包括 22 例 LVI 阳性和 37 例 LVI 阴性。

DCE-MRI 图像数据的采集设备为西门子 Magnetom Espree 1.5T MR 扫描仪,配备 4 通道乳腺专用相控阵表面线圈。全部患者为俯卧位。选择轴向三维 T1 梯度回波法,利用脂肪饱和序列,最终采集 DCE-MRI 图像(TR: 4.6 ms; TE: 1.7 ms; 翻转角:7°;像素大小:1.0 mm × 1.0 mm; FOV:280 mm × 340 mm;矩阵大小:280 × 340;截面厚度:1.0 mm;差距:0 mm;采集时间:75 s/相)。所有患者均顺利完成 MRI 检查。对于多发乳腺病变的病人,仅选取最大的癌变病灶。

2.2 不同分割算法的分割性能比较

为了检验本文分割算法性能,选择 4 个形态不规则、边界模糊和对比度低的数据进行验证。将本文分割算法和基于边界的 ACM 方法^[28]、基于区域的 ACM 方法^[29]对比分析,图 2 是对比结果。

从图 2b 知,病灶区和周围正常腺体对比度较低,采用基于边界的 ACM 方法分割,发现正常腺体较多,部分边界易泄漏。图 2c 采用基于区域的 ACM 方法,分割效果也不佳,出现误分割。比较图 2a 和图 2d 知,本文算法接近医生分割结果。

基于区域的 ACM 方法和基于边界的 ACM 方法无法精准分割病灶的原因是:病灶和正常腺

体灰度值相近,在轮廓曲线演变时只用了图像梯度信息;而图像对比度低时,无法检测低对比度边界是否为病灶边界,导致边界泄漏,无法精准分割。

为了验证算法分割性能,选择 20 个临床样本。采用 TPR, FPR, SD 的平均值作为分割综合评价指标。从表 1 看出,基于区域的 ACM 方法和基于边界的 ACM 方法能获得较好的 TPR 值,但 FPR 值太大,SD 值较小,表明这两种分割方法尽管能把病灶分割出来,但出现大量背景,相似度低;而本文算法分割背景少,相似度高,分割性能更优。

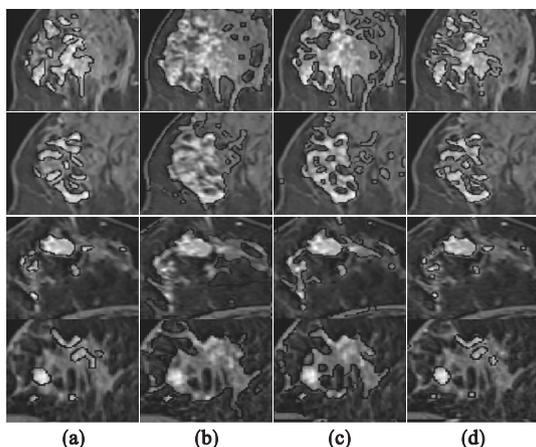


图 2 乳腺病灶分割算法对比

Fig. 2 Comparison of breast lesion segmentation algorithms

(a) — 医生手动分割; (b) — 基于边界的 ACM 方法; (c) — 基于区域的 ACM 方法; (d) — 本文分割算法。

表 1 20 个临床样本分割结果

Table 1 20 clinical sample segmentation results

算法	TPR	FPR	SD
基于边界的 ACM 方法	0.963 9	1.976 7	0.404 1
基于区域的 ACM 方法	0.942 1	1.174 2	0.499 4
本文分割算法	0.917 9	0.280 7	0.728 0

2.3 基于 LASSO 的特征选择

当全部特征经 LASSO 方法选择后,最终选出两个任务相关有效特征,分别是方差(灰度特征)和熵(纹理特征)。图 3 为 LASSO 特征选择可视化图。图 3a 是 10 重交叉验证结果,虚线①定义调整参数 λ 最优值,此时模型提供对数据的最佳拟合;图 3b 是全部特征 LASSO 回归系数分布图,虚线②表明当 λ 达到最优时非零回归系数对应的两个特征。

2.4 基于集成分类器随机森林的 LVI 状态预测

在训练集成分类器随机森林模型时,决策树的数量 N_1 对随机森林预测的准确性起到重要作

用. 表 2 列举了在选择不同决策树数量时得到的分类性能. 从表中看出, N_1 大于 100 或小于 100 时, 敏感度、特异性和准确率均不如 N_1 为 100 时的好, 只有当 $N_1 = 100$ 时分类性能最好.

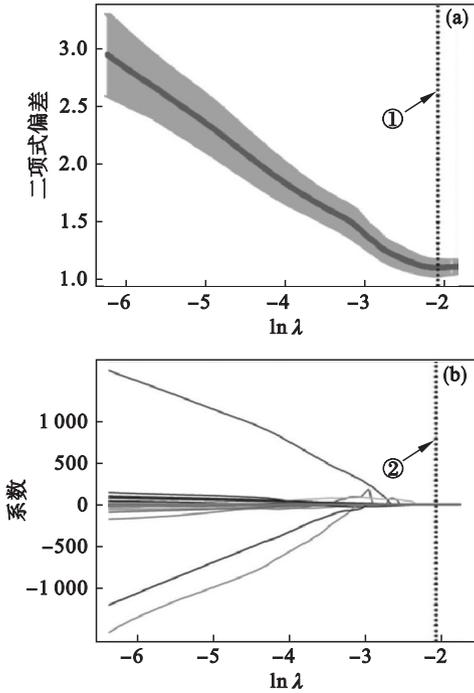


图 3 LASSO 特征选择可视化

Fig. 3 LASSO feature selection visualization

(a)—10 重交叉验证图;
(b)—全部特征 LASSO 系数分布图.

表 2 N_1 取不同值对应的分类性能比较

Table 2 Comparison of classification performance corresponding to different values of N_1

N_1	T_p	F_p	F_N	T_N	Sen	Spe	Acc
10	4	18	3	34	0.571 4	0.653 8	0.644 1
20	3	19	3	34	0.500 0	0.641 5	0.627 1
30	2	20	2	35	0.500 0	0.636 4	0.627 1
40	2	20	2	35	0.500 0	0.636 4	0.627 1
50	2	20	3	34	0.400 0	0.629 6	0.610 2
60	2	20	3	34	0.400 0	0.629 6	0.610 2
70	2	20	2	35	0.500 0	0.636 4	0.627 1
80	2	20	2	35	0.500 0	0.636 4	0.627 1
90	2	20	2	35	0.500 0	0.636 4	0.627 1
100	4	18	2	35	0.666 7	0.660 4	0.661 0
200	3	19	2	35	0.600 0	0.648 1	0.644 1
300	3	19	2	35	0.600 0	0.648 1	0.644 1
400	2	20	2	35	0.500 0	0.636 4	0.627 1
500	2	20	2	35	0.500 0	0.636 4	0.627 1
600	2	20	2	35	0.500 0	0.636 4	0.627 1

2.5 不同分割算法在随机森林模型的分类比较

采用三种分割算法进行乳腺癌病灶分割, 然后构建不同分割算法下的集成分类器随机森林模

型, 再进行乳腺癌患者 LVI 状态预测. 为了公平起见, 三种分割算法均采用 LASSO 选出的 2 个特征. 从表 3 可以看出, 本文提出的分割算法敏感度、特异度和准确率均比基于边界的 ACM 方法和基于区域的 ACM 方法好, 表明本文提出的分割算法能够提高分类器的分类性能.

表 3 不同分割算法下随机森林预测模型的性能比较
Table 3 Classification performance of random forest prediction models under different segmentation algorithms

算法	Sen	Spe	Acc
基于边界的 ACM 方法	0.500 0	0.636 4	0.627 1
基于区域的 ACM 方法	0.600 0	0.648 1	0.644 1
本文分割算法	0.625 0	0.666 7	0.661 0

2.6 不同分类器对测试集的分类性能比较

采用 6 种不同的分类器对测试集进行分类性能对比. 从表 4 看出, 集成分类器随机森林比其他分类器性能好, 尽管贝叶斯线性回归的特异度比集成分类器随机森林高, 但敏感度和准确率均较低. ResNet18 的敏感度高, 但特异度和准确率低.

表 4 不同分类器对测试集的分类性能
Table 4 Classification performance of test sets by different classifiers

算法	Sen	Spe	Acc
支持向量机	0.437 5	0.703 7	0.559 3
贝叶斯线性回归	0.500 0	0.810 8	0.694 9
K 邻近	0.666 7	0.642 9	0.644 1
ResNet18	0.950 9	0.489 0	0.578 8
VGG16	0.542 3	0.664 4	0.595 0
集成分类器随机森林	0.714 3	0.733 3	0.728 8

3 结 语

本文结合活动轮廓模型和影像组学方法, 完成乳腺癌患者术前 LVI 状态预测. 针对乳腺癌病灶在 DCE-MRI 图像中边界模糊、亮度不均匀的特点, 提出基于后验概率和模糊速度函数的活动轮廓模型方法, 完成病灶的分割, 然后进行影像学特征提取, 并利用 LASSO 方法选择有效特征, 最后利用集成分类器随机森林模型进行测试. 实验结果表明, 本文提出的分割算法可以提高乳腺癌病灶分割的准确性, 所构建的集成分类器随机森林模型能够对乳腺癌患者的术前 LVI 状态进行较准确的预测.

参考文献:

[1] Fitzmaurice C, Allen C, Barber R M, et al. Global, regional,

- and national cancer incidence, mortality, years of life lost, years lived with disability, and disability-adjusted life-years for 32 cancer groups, 1990 to 2015 [J]. *JAMA Oncology*, 2017, 3(4):524-548.
- [2] Dong Y, Feng Q, Yang W, et al. Preoperative prediction of sentinel lymph node metastasis in breast cancer based on radiomics of T2-weighted fat-suppression and diffusion-weighted MRI [J]. *European Radiology*, 2017, 28(2):582-591.
- [3] Torre L A, Bray F, Siegel R L, et al. Global cancer statistics, 2012 [J]. *CA: A Cancer Journal for Clinicians*, 2015, 65(2):87-108.
- [4] Aleskandarany M A, Sonbul S N, Mukherjee A, et al. Molecular mechanisms underlying lymphovascular invasion in invasive breast cancer [J]. *Pathobiology*, 2015, 82(3/4):113-123.
- [5] Paduch R. The role of lymphangiogenesis and angiogenesis in tumor metastasis [J]. *Cellular Oncology*, 2016, 39(5):397-410.
- [6] Bae M S, Moon H G, Han W, et al. Early stage triple-negative breast cancer: imaging and clinical-pathologic factors associated with recurrence [J]. *Radiology*, 2016, 278(2):356-364.
- [7] Uematsu T, Kasami M, Watanabe J, et al. Is lymphovascular invasion degree one of the important factors to predict neoadjuvant chemotherapy efficacy in breast cancer? [J]. *Breast Cancer*, 2011, 18(4):309-313.
- [8] Ryu J M, Lee S K, Kim J Y, et al. Predictive factors for non-sentinel lymph node metastasis in patients with positive sentinel lymph nodes after neoadjuvant chemotherapy: nomogram for predicting nonsentinel lymph node metastasis [J]. *Clinical Breast Cancer*, 2017, 17(7):550-558.
- [9] Kumar V, Gu Y, Basu S, et al. Radiomics: the process and the challenges [J]. *Magnetic Resonance Imaging*, 2012, 30(9):1234-1248.
- [10] 刘桐桐, 李佳伟, 胡雨舟, 等. 基于影像组学预测乳腺癌雌激素受体表达情况的可行性分析 [J]. 生物医学工程杂志, 2017, 34(4):597-601.
(Liu Tong-tong, Li Jia-wei, Hu Yu-zhou, et al. Feasibility analysis of predicting expression of estrogen receptor in breast cancer based on radiomics [J]. *Journal of Biomedical Engineering*, 2017, 34(4):597-601.)
- [11] 吴佩琪, 刘再毅, 梁长虹. 基于 MRI 的影像组学特征在鉴别乳腺浸润性导管癌病理分级中的价值 [J]. 医学研究生学报, 2018, 31(9):938-942.
(Wu Pei-qi, Liu Zai-yi, Liang Chang-hong. MRI-based radiomic features in histological grading of breast invasive ductal carcinoma [J]. *Journal of Medical Postgraduates*, 2018, 31(9):938-942.)
- [12] Huang Y H, Chang Y C, Huang C S, et al. Computerized breast mass detection using multi-scale hessian-based analysis for dynamic contrast-enhanced MRI [J]. *Journal of Digital Imaging*, 2014, 27(5):649-660.
- [13] Fan M, He T, Zhang P, et al. Diffusion-weighted imaging features of breast tumours and the surrounding stroma reflect intrinsic heterogeneous characteristics of molecular subtypes in breast cancer [J]. *NMR in Biomedicine*, 2018, 31(2):e0189302.
- [14] 孙利雷, 徐勇. 基于深度学习的乳腺 X 射线影像分类方法研究 [J]. 计算机工程与应用, 2018, 54(21):13-19.
(Sun Li-lei, Xu Yong. Research on classification method of mammography based on deep learning [J]. *Computer Engineering and Applications*, 2018, 54(21):13-19.)
- [15] Boné B, Szabó B K, Perbeck L G, et al. Can contrast-enhanced MR imaging predict survival in breast cancer [J]. *Acta Radiologica*, 2015, 44(4):373-378.
- [16] Achuthan A, Rajeswari M, Ramchandram D, et al. Wavelet energy-guided level set-based active contour: a segmentation method to segment highly similar regions [J]. *Computers in Biology and Medicine*, 2010, 40(7):608-620.
- [17] Gianluca M, Matthias E, Katharina M, et al. Vessel suppressed chest computed tomography for semi-automated volumetric measurements of solid pulmonary nodules [J]. *European Journal of Radiology*, 2018, 101(5):97-102.
- [18] Kokkinos I, Maragos P. Synergy between object recognition and image segmentation using the expectation maximization algorithm [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009, 31(8):1486-1501.
- [19] Kwedlo W. A new random approach for initialization of the multiple restart EM algorithm for Gaussian model-based clustering [J]. *Pattern Analysis & Applications*, 2015, 18(4):757-770.
- [20] Dong Y, Feng Q, Yang W, et al. Preoperative prediction of sentinel lymph node metastasis in breast cancer based on radiomics of T2-weighted fat-suppression and diffusion-weighted MRI [J]. *European Radiology*, 2018, 28(2):582-591.
- [21] Rose C J, Mills S, O' Connor J P B, et al. Quantifying heterogeneity in dynamic contrast-enhanced MRI parameter maps [J]. *Medical Image Computing and Computer-Assisted Intervention*, 2010, 62(2):488-499.
- [22] Fan M, Wu G, Cheng H, et al. Radiomic analysis of DCE-MRI for prediction of response to neoadjuvant chemotherapy in breast cancer patients [J]. *European Journal of Radiology*, 2017, 94(2):140-147.
- [23] 张文华, 陈韬, 张明慧, 等. 基于放射组学的胃肠道间质瘤分类模型 [J]. 南方医科大学学报, 2018, 38(1):55-61.
(Zhang Wen-hua, Chen Tao, Zhang Ming-hui, et al. A radiomics-based model for differentiation between benign and malignant gastrointestinal stromal tumors [J]. *Journal of Southern Medical University*, 2018, 38(1):55-61.)
- [24] 李祥霞, 李彬, 田联房, 等. 基于放射影像组学和随机森林算法的肺结节良恶性分类 [J]. 华南理工大学学报, 2018, 46(8):78-86.
(Li Xiang-xia, Li Bin, Tian Lian-fang, et al. Classification of benign and malignant pulmonary nodules based on radiomics and random forests algorithm [J]. *Journal of South China University of Technology*, 2018, 46(8):78-86.)
- [25] Dougherty E, Hua J, Sima C. Performance of feature selection methods [J]. *Current Genomics*, 2009, 10(6):365-374.
- [26] Pripp A H, Stanišić M. Association between biomarkers and clinical characteristics in chronic subdural hematoma patients assessed with lasso regression [J]. *PLoS One*, 2017, 12(11):e0186838.
- [27] Breiman L. Random forest [J]. *Machine Learning*, 2001, 45(1):5-32.
- [28] 郑强, 董恩清. 一种新的基于二值水平集和形态学的局部分割方法 [J]. 电子与信息学报, 2012, 34(2):375-381.
(Zheng Qiang, Dong En-qing. A new local segmentation method based on binary level set and morphological operation [J]. *Journal of Electronics & Information Technology*, 2012, 34(2):375-381.)
- [29] 陈明, 赵云, 范能胜, 等. 医学图像分割中的主动轮廓模型研究现状 [J]. 北京生物医学工程, 2010, 29(4):426-431.
(Chen Ming, Zhao Yun, Fan Neng-sheng, et al. A review of the active contour models in medical image segmentation [J]. *Beijing Biomedical Engineering*, 2010, 29(4):426-431.)