

❖ 医学物理与工程学

Breast cancer risk prediction model based on improved local ternary pattern algorithm

YIN Kaiming, YAN Shiju*, SONG Chengli

(School of Medical Instrument and Food Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China)

[Abstract] **Objective** To explore the value of new and fused conventional texture features extracted from mammograms using improved local ternary patterns (LTP) in predicting risk of breast cancer. **Methods** Mammograms were segmented. Based on improved LTP, the new and conventional texture features were extracted from segmented mammograms of bilateral breasts. Then the features of bilateral breasts were merged. The high dimensional characteristics were reduced with principal component analysis (PCA). Finally, the new texture features were classified with k-nearest neighbor (KNN), and the fusion features were clustered with logistic alternating decision tree (LADTree) algorithm. **Results** The area under ROC curve (AUC) of new texture features for predicting breast cancer was $0.732\ 4 \pm 0.042\ 8$, and the sensitivity, specificity and prediction accuracy was 72.04% (67/93), 74.51% (76/102) and 73.33% (143/195), respectively. Furthermore, AUC of fusion features was $0.865\ 5 \pm 0.014\ 8$, the sensitivity, specificity and prediction accuracy was 84.95% (79/93), 88.23% (90/102) and 86.67% (169/195), respectively. **Conclusion** The new texture features based on improved LTP have high prediction accuracy for breast cancer, and the prediction efficacy can be improved after fusion with conventional features.

[Key words] Mammography; Texture feature; Prediction model; Local ternary patterns

DOI:10.13929/j.1003-3289.201710047

基于改进局部三元模式的乳腺癌预测模型

殷恺铭, 闫士举*, 宋成利

(上海理工大学医疗器械与食品学院, 上海 200093)

[摘要] **目的** 探讨基于改进局部三元模式(LTP)算法提取的乳腺新型纹理特征及其与常规特征融合预测乳腺癌的价值。**方法** 对钼靶图像进行乳腺分割,采用基于改进 LTP 算法提取双侧乳腺的新型纹理特征和常规特征;合并左右侧乳腺纹理特征;采用主成分分析法对提取的高维纹理特征降维;以 K 最近邻(KNN)和 LADTree 算法分别对纹理特征及融合纹理特征进行分类。**结果** 基于改进 LTP 算法提取的新型纹理特征预测乳腺癌的 ROC 曲线下面积(AUC)为 $0.732\ 4 \pm 0.042\ 8$, 敏感度为 72.04% (67/93), 特异度为 74.51% (76/102), 准确率为 73.33% (143/195);融合常规特征后 AUC 为 $0.865\ 5 \pm 0.014\ 8$, 敏感度为 84.95% (79/93), 特异度为 88.23% (90/102), 准确率为 86.67% (169/195)。**结论** 基于 LTP 算法提取的新型纹理特征预测乳腺癌的精度较高,与常规特征融合后可进一步提高预测效能。

[关键词] 乳房 X 线摄影术;纹理特征;评估模型;局部三值模式

[中图分类号] R737.9; R814.41 **[文献标识码]** A **[文章编号]** 1003-3289(2018)04-0616-05

[第一作者] 殷恺铭(1993—),男,江苏高邮人,在读硕士。研究方向:医学图像处理。E-mail: 491899030@qq.com

[通信作者] 闫士举,上海理工大学医疗器械与食品学院,200093。E-mail: yanshiju@usst.edu.cn

[收稿日期] 2017-10-16 **[修回日期]** 2018-01-19

每年全球有超过 55 万女性死于乳腺癌,超过 110 万女性罹患乳腺癌,且发病率呈不断上升趋势^[1-2]。近年来,计算机辅助诊断(computer assisted diagnosis, CAD)技术快速发展,其中乳腺癌风险评估模型在乳腺癌发病预测方面发挥着重要作用^[3-6]。基于钼靶图像量化特征的乳腺癌近期发病预测模型无需采用基因检测技术,对个体女性近期发病预测效果较好^[7]。

基于钼靶图像量化特征的 CAD 模型广泛采用均值、偏度、峰度等灰度统计特征,基于分形维数的特征^[8-10],基于灰度共生、行程矩阵的纹理特征^[11]或基于旋转均匀局部二进制(local binary pattern, LBP)模式的纹理特征^[12]。既往研究^[8-12]多使用图像的整体纹理特征,而乳腺密度不均匀,且正常组织与病变组织的局部纹理特征存在较大差异,仅通过整体纹理特征不足以有效表征患者的个体图像特性。而 LBP 特征虽具有灰度不变性和旋转不变性等显著优点,但对噪声和光照条件较为敏感,可能影响预测精度。本研究基于局部三元模式(local ternary pattern, LTP),提出一种新型提取乳腺结构纹理特征并进行分类的方法。

1 设计与方法

1.1 实验图像 收集匹兹堡大学医学中心临床数据库中 195 名女性筛查者的 390 幅双侧头尾(craniocaudal, CC)位全数字化乳腺钼靶图像,均采用 Hologic Selenia FFDM 系统(Hologic Inc., Bedford, MA, USA)采集。

1.2 方法 ①对 CC 位钼靶图像进行乳腺分割;②于双侧乳腺区提取新型纹理特征和常规特征;③基于双侧乳腺不对称特性,合并左右侧乳腺纹理特征;④以主成分分析(principal component analysis, PCA)法对高维特征进行降维,以 K 最近邻(k-nearest neighbor, KNN)分类算法对新型纹理特征进行分类,并将其与

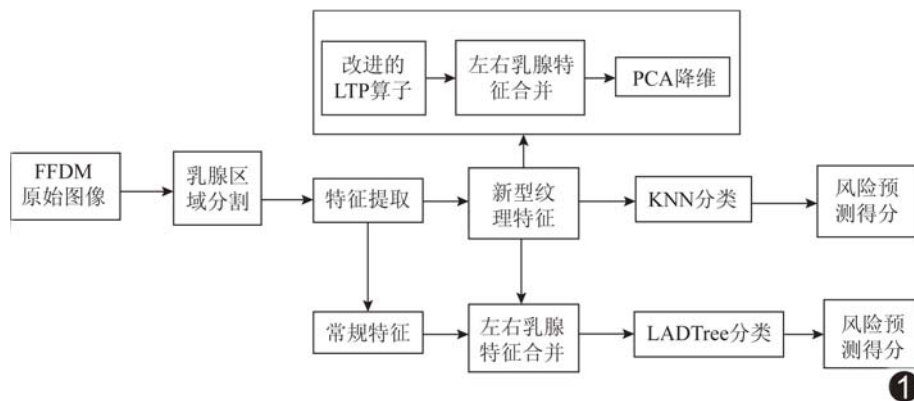


图 1 改进的 LTP 模式工作流程图

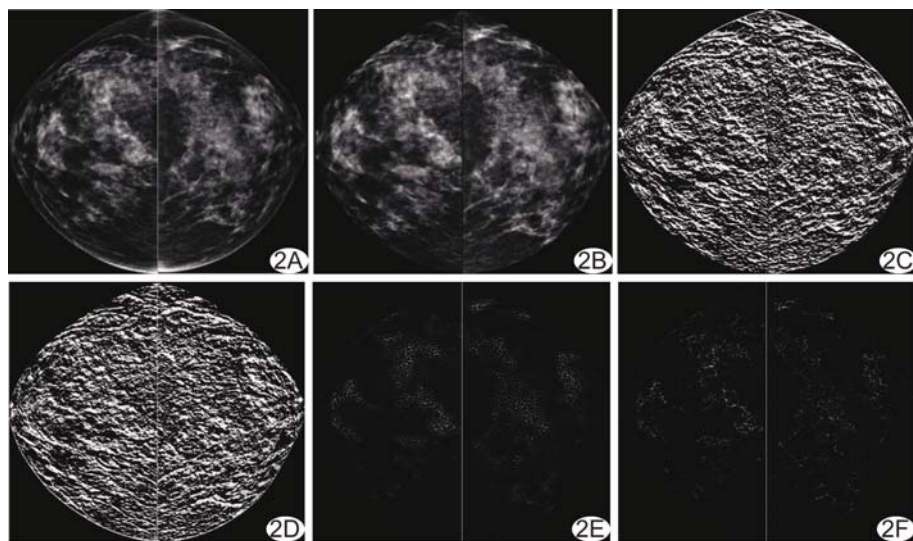


图 2 钼靶图像经分割、改进 LTP 算子变换后图像和 VAR 图 A. 原始钼靶图像; B. 分割后图像; C. $LTP^{highP,R}$ 图像; D. $LTP^{lowP,R}$ 图像; E. $VAR^{highP,R}$ 图; F. $VAR^{lowP,R}$ 图

常规纹理特征的 AUC 值进行对比以供优选;⑤将新型纹理特征与常规特征融合,采用 LADTree(logistic alternating decision tree)算法进行分类,获得乳腺癌风险预测得分(图 1)。

1.2.1 图像分割 通过钼靶图像的强度直方图,采用迭代阈值法检测乳腺与背景图像间最优曲线,移除图像背景区域;通过定向区域生长法去除钼靶图像中胸肌和皮肤区域(图 2A、2B)。

1.2.2 特征提取 首先分别从左右侧乳腺区域中提取图像常规特征,包括灰度统计特征(标准差、均值、偏度、峰度等)、基于图像分形维数特征和基于图像的直方图特征。

纹理特征是一种反映图像中同质现象的视觉特征,体现物体表面中具有缓慢变化或周期性变化的结构组织排列属性。LTP 是由 Tan 等^[13]提出的一种结构纹理

特征,该算法是对局部二值模式的改进和范化,对噪声具有较强的鲁棒性,在乳腺密度均匀区相比于 LBP 具有更强的区分能力,且可均衡乳腺经不同强度 X 线照射得到钼靶图像的像素值。LTP 算法中提出了阈值区间,将 LBP 算子的邻域中心值 m_c 作为中心值,并引入中心值区间 $[m_c - t, m_c + t]$ 为阈值区间。若邻域值在阈值区间,则将该邻域值编码为 0;若邻域值大于中心值,则将该邻域值编码为 1;若邻域值比中心区间小,则将该邻域值编码为 -1。具体流程如下:

$$G(m_i, m_c, t) = \begin{cases} 1, (m_i - m_c) \geq t \\ 0, |m_i - m_c| < t \\ -1, (m_i - m_c) < -t \end{cases} \quad (1)$$

其中 G 为算子邻域的权值,阈值 t 由实验者自定义获得,权值的编码方式与原始 LBP 算子^[14]相同,编码过程见图 3:

采用原始 3×3 邻域 LBP 编码模式时,在计算邻域与中心像素差值的过程中,由于各邻域与中心像素的距离不同,无法准确反映邻域像素点与中心像素点的位置关系。Ojala 等^[14]提出了圆域 LBP,即以圆形邻域代替正方形邻域,最终得到的新算子在半径为 R 的圆域内可选 P 个像素点,可极大地涵盖邻域像素点的信息,故圆域 LTP 算子可提高纹理特征的提取精度(公式 2):

$$LTP_{P,R} = \sum_{i=0}^{P-1} 3^i gG(m_i, m_c, t) \quad (2)$$

在原始 LTP 算子中,阈值 t 由实验者根据特定应用领域图像的特征估计获得。针对钼靶图像,由于乳腺密度区域存在不规则性,且正常组织和病变组织的灰度、纹理分布存在很大差异,单一阈值无法保证适应整幅乳腺图像。因此,本研究提出一种自适应阈值,以充分表征不同乳腺组织的局部纹理特征。

在钼靶图像的每个圆域内,通过计算中心像素与邻域像素的离散程度来确定相应阈值,其中离散度 t_m 会随着不同邻域变化而变化,表明离散度对不同的样本邻域具有自适应性。因此,将离散度作为阈值能更好地量化像素与邻域之间的关系。具体实现步骤如下:

(1) 计算邻域的平均对比度:

$$\overline{\Delta m} = (\sum_{i=0}^{P-1} \Delta m_i) / P \quad (3)$$

其中 Δm_i 为邻域各像素与中心点的差值, P 为邻域点数;

(2) 通过邻域的对比度和差值计算邻域的波动程度 V :

$$V = (\sum_{i=0}^{P-1} (\Delta m_i - \overline{\Delta m})^2) / P \quad (4)$$

(3) 通过波动程度定义离散值 t_m :

$$t_m = \sqrt{V} \quad (5)$$

LTP 特征对局部像素值的变化敏感,但不能反映灰度变化,即若邻域中图像灰度值的大小顺序相同,则得到的编码值相同,可很大程度上忽略钼靶图像中的灰度变化信息,而钼靶图像灰度变化在纹理特征中非常重要。

钼靶图像的纹理是空间结构(模式)和灰度对比度(强度)两者的正交,其中灰度值的变化不会改变模式,对比度代表灰度值的变化强度。此外,对比度具有旋转不变性,而模式受图像旋转的影响。为实现二者的优势互补,本研究提出的方法中加入了表示图像强度(局部对比度)的参数,即方差。通过方差和 LTP 算子互补,可更好地体现钼靶图像的模式和强度,从而提取出更具价值的纹理特征。由于 LTP 算子分为 $LTP^{highP,R}$ 和 $LTP^{lowP,R}$,因此其对应的局部方差强度也分为 $VAR^{highP,R}$ 和 $VAR^{lowP,R}$ (公式 6、7)。

$$VAR^{highP,R} = \frac{1}{P} \sum_{h=0}^{P-1} \left(m_h - \frac{1}{P} \sum_{h=0}^{P-1} m_h \right)^2 \quad (6)$$

$$VAR^{lowP,R} = \frac{1}{P} \sum_{l=0}^{P-1} \left(m_l - \frac{1}{P} \sum_{l=0}^{P-1} m_l \right)^2 \quad (7)$$

则钼靶图像的纹理特征通过 LTP 与 VAR 的联合分布表示为:

$$F_h = LTP^{highP,R} / VAR^{highP,R} \quad (8)$$

$$F_l = LTP^{lowP,R} / VAR^{lowP,R} \quad (9)$$

再将得到的 F_h 和 F_l 特征直方图连接,即为钼靶图像的纹理特征。由原始图像经乳腺组织分割及改进 LTP 算子变换后的图像可清晰直观地显示乳腺组织的纹理(图 2C、2D),VAR 图谱可反映乳腺钼靶图像强度的变化(图 2E、2F)。

1.2.3 左右侧乳腺纹理特征合并及降维 乳腺病变往往始于一侧,因此左右侧乳腺的不对称性对预测乳腺癌具有重要意义。本研究将左右侧乳腺特征值相减,再进行特征归一化,获得最终纹理特征^[15]。本研究对每幅钼靶图像进行新型纹理特征提取,获得 199 680 维数据量;如直接采用 KNN 分类器训练该数据,不仅分类效率极低,且分类精度将受影响,故本研究采用 PCA 方法对提取出的特征进行降维。

PCA^[16]法是将原有特征通过线性投影形成新的低维数,从而达到降维目的。针对本研究提取的乳腺纹理特征,该方法具有以下优点:①可根据数据点集中建立 1 个低维的特征子空间,反映数据点集相对于乳腺纹理特征的平均差异;②本研究提取的纹理特征维

度过于冗长,包含大量无用特征,经 PCA 法变换后,可保留乳腺纹理特征的主要信息,且具有很好的稳定性;③随着纹理特征维数降低,图像模式之间的距离相应缩小,避免了在高维空间上进行分类的复杂性;④在降维过程中,PCA 法可消除模式相关性,降低误分率。

1.2.4 特征分类 KNN 分类算法是一种有效的机器学习算法,在类别决策时仅与极少量的相邻样本相关,而非通过判断类域的方式来聚类,可高效、准确分析类域交叉或重叠较多的样本。本研究提取出的新型纹理特征实则为图像直方图,多个直方图间存在较多重叠及交叉,适用于 KNN 分类算法。

针对融合特征族,即新型纹理特征与常规特征的集合分类,本研究采用 LADTree 算法,是基于 LogitBoost 策略的类别交替分类器树。LogitBoost 算法在 Boost 算法的基础上进行了改进,其优点为将数个弱分类器通过组合,搭建为强分类器,可明显提高分类精度、避免数据的过拟合。

为对分类方法的预测性能进

行客观比较,本研究采用留一法对分类结果进行验证,即每次验证从全体样本中选取一个样本作为测试集,其余样本作为训练集训练分类器,重复此过程,直到遍历所有样本。

1.3 统计学方法 采用 weka 统计分析软件,采用 ROC 曲线分析不同纹理特征预测乳腺癌的效能,获得曲线下面积(area under the curve, AUC)和准确率、敏感度以及特异度。 $P < 0.05$ 为差异有统计学意义。

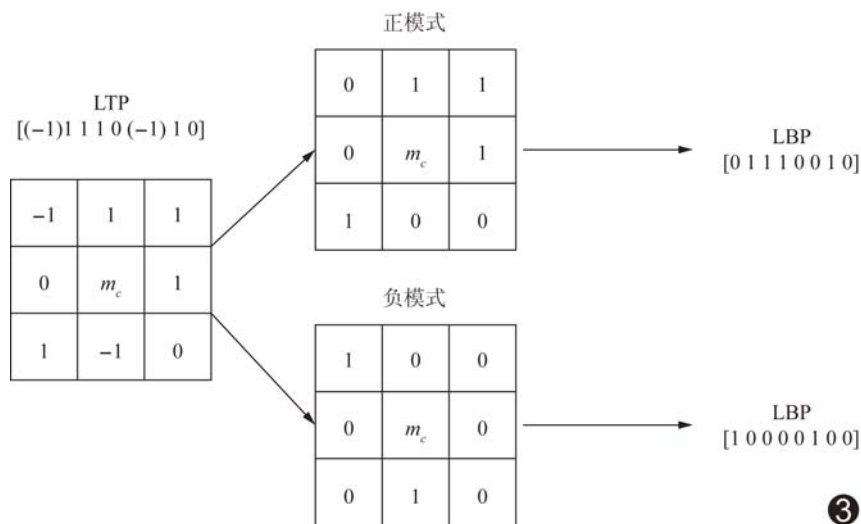


图 3 LTP 算子编码过程图

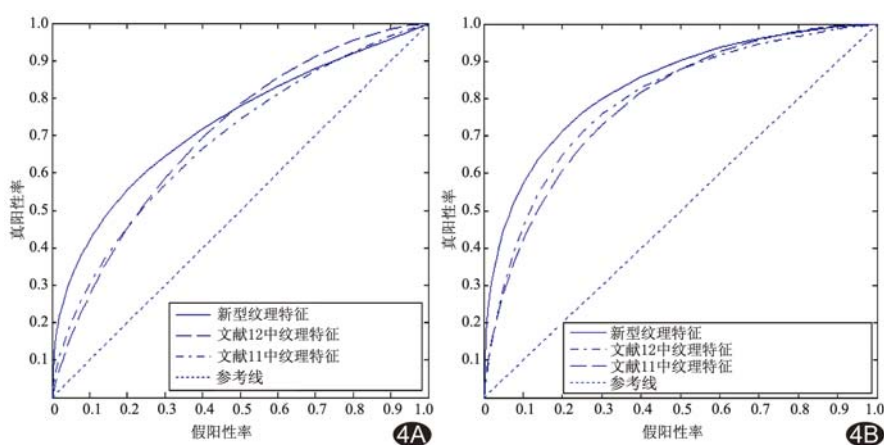


图 4 不同纹理特征(A)及融合纹理特征(B)预测乳腺癌的 ROC 曲线

2 结果

针对采用本研究方法提取出的新型纹理特征、基于灰度共生矩阵和游程矩阵提取的纹理特征以及基于旋转不变 LBP 算子提取的纹理特征,分别经 KNN 分类后进行 AUC、准确率、敏感度、特异度的诊断指标对比的结果见表 1、图 4A。通过集合方式将 3 种纹理特征分别与常规特征进行融合,经 LADTree 分类器进行分类,其效能对比见表 2、图 4B。

表 1 不同纹理特征预测乳腺癌效能对比

特征种类	AUC	准确率(%)	敏感度(%)	特异度(%)
新型纹理特征	0.732 4±0.042 8	73.33(143/195)	72.04(67/93)	74.51(76/102)
基于灰度共生矩阵和游程矩阵提取的纹理特征	0.679 4±0.041 1	65.64(128/195)	61.29(57/93)	69.61(71/102)
基于旋转不变 LBP 算子提取的纹理特征	0.705 3±0.041 2	68.72(134/195)	70.97(66/93)	66.67(68/102)

表 2 不同融合纹理特征预测乳腺癌效能对比

特征种类	AUC	准确率(%)	敏感度(%)	特异度(%)
融合新型纹理特征	0.865 5±0.014 8	86.67(169/195)	84.95(79/93)	88.23(90/102)
融合基于灰度共生矩阵和游程矩阵提取的纹理特征	0.803 4±0.021 4	80.51(157/195)	78.49(73/93)	82.35(84/102)
融合基于旋转不变 LBP 算子提取的纹理特征	0.832 3±0.014 7	82.05(160/195)	79.57(74/93)	84.31(86/102)

3 讨论

针对钼靶图像乳腺密度不均匀、纹理分布不规则的问题,本研究采用改进的 LTP 算子结合 PCA 法,提出了一种新型的纹理特征。与以往模型中的纹理特征比较,本研究提出的新型纹理特征更强调提取区域纹理特征,更贴合乳腺组织的解剖特征;同时,结合 PCA 法删除影响预测精度的干扰特征,并采用 KNN 算法对单一纹理特征进行分类及 LADTree 算法对新型特征与常规特征进行融合后分类。本研究基于改进算法提取的新型纹理特征和融合常规特征预测乳腺癌的效能均高于常规纹理特征,提示本研究方法可提取出更多有价值的纹理特征,且与常规特征融合可进一步提高预测精度。虽然本研究采用的纹理特征提取算法的运行效率较低,但相信随着计算机硬件技术的发展,尤其是图形处理器的普遍应用,可逐渐得以克服和解决。

[参考文献]

- [1] Yang Q, Li L, Zhang J, et al. A new quantitative image analysis method for improving breast cancer diagnosis using DCE-MRI examinations. *Med Phys*, 2015, 42(1):103-109.
- [2] Aghaei F, Tan M, Hollingsworth AB, et al. Computer-aided breast MR image feature analysis for prediction of tumor response to chemotherapy. *Med Phys*, 2015, 42(11):6520-6528.
- [3] Dai J, Hu Z, Jiang Y, et al. Breast cancer risk assessment with five independent genetic variants and two risk factors in Chinese women. *Breast Cancer Res*, 2012, 14(1):R17.
- [4] Varesco L, Viassolo V, Viel A, et al. Performance of BOADICEA and BRCAPRO genetic models and of empirical criteria based on cancer family history for predicting BRCA mutation carrier probabilities: A retrospective study in a sample of Italian cancer genetics clinics. *Breast*, 2013, 22(6):1130-1135.
- [5] Lee AJ, Cunningham AP, Kuchenbaecker KB, et al. BOADICEA breast cancer risk prediction model: Updates to cancer incidences, tumour pathology and web interface. *Br J Cancer*, 2014, 110(2):535-545.
- [6] Gail MH, Mai PL. Comparing breast cancer risk assessment models. *J Natl Cancer Inst*, 2010, 102(10):665-668.
- [7] Tan M, Pu J, Cheng S, et al. Assessment of a four-view mammographic image feature based fusion model to predict near-term breast cancer risk. *Ann Biomed Eng*, 2015, 43(10):2416-2428.
- [8] Chang YH, Wang XH, Hardesty LA, et al. Computerized assessment of tissue composition on digitized mammograms. *Acad Radiol*, 2002, 9(8):899-905.
- [9] Wang X, Lederman D, Tan J, et al. Computerized prediction of risk for developing breast cancer based on bilateral mammographic breast tissue asymmetry. *Med Eng Phys*, 2011, 33(8):934-942.
- [10] Wang X, Lederman D, Tan J, et al. Computerized detection of breast tissue asymmetry depicted on bilateral mammograms: A preliminary study of breast risk stratification. *Acad Radiol*, 2010, 17(10):1234-1241.
- [11] Filipczuk P, Fevens T, Krzyżak A, et al. GLCM and GLRLM based texture features for computer-aided breast cancer diagnosis. *Journal of Medical Informatics & Technologies*, 2012, 19. <http://yadda.icm.edu.pl/yadda/element/bwmetal.element.baztech-article-pwa4-0026-0012>
- [12] Reyad YA, Berbar MA, Hussain M. Comparison of statistical, LBP, and multi-resolution analysis features for breast mass classification. *Journal of Medical Systems*, 2014, 38(9):100.
- [13] Tan X, Triggs B. Enhanced local texture feature sets for face recognition under difficult lighting conditions. *IEEE Trans Image Process*, 2010, 19(6):1635-1650.
- [14] Ojala T, Pietikäinen M, Mäenpää T. Gray scale and rotation invariant texture classification with local binary patterns// *Computer Vision-ECCV 2000*. Berlin: Springer-Verlag, 2000: 404-420.
- [15] Zheng B, Tan M, Ramalingam P, et al. Association between computed tissue density asymmetry in bilateral mammograms and near-term breast cancer risk. *Breast J*, 2014, 20(3):249-257.
- [16] Yamamoto H, Yamaji H, Abe Y, et al. Dimensionality reduction for metabolome data using PCA, PLS, OPLS, and RFDA with differential penalties to latent variables. *Chemometrics & Intelligent Laboratory Systems*, 2009, 98(2):136-142.