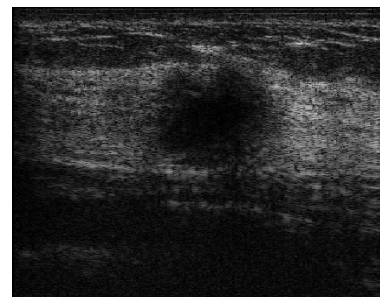




DOI: 10.12086/oe.2019.180368

基于超声 RF 信号的乳腺肿瘤分级检测方法

童莹¹, 严郁^{2,3*}¹南京工程学院通信工程学院, 江苏 南京 211167;²南京理工大学计算机科学与工程学院, 江苏 南京 210094;³南京中医药大学附属医院(江苏省中医院), 江苏 南京 210029

摘要: 为解决超声乳腺肿瘤分级检测问题, 从超声射频(RF)信号的角度提出了一种有效的乳腺肿瘤分级检测方法。首先, 采用 Shearlet 变换提取乳腺超声 RF 信号的多尺度、多方向特征; 其次, 考虑 Shearlet 特征的高维冗余性, 采用多尺度方向二值模式(MDBP)对其进行编码, 在不损失特征信息的条件下降低特征维度; 最后, 依据医生阅片经验以及不同分级乳腺肿瘤的特征差异性, 设计出适合乳腺病变分级检测的层级二叉树 SVM 分类器(CBT-SVM)。在 928 个乳腺肿瘤患者的超声 RF 信号上进行验证, 大量结果表明, 提出方法可以有效实现 3 级、4A 级~4C 级、5 级乳腺肿瘤的分级检测, 准确度、敏感度、特异度、PPV、NPV 以及 MCC 分别达到 89.29%、75.62%、94.54%、97%、98.3%和 81.01%。

关键词: 计算机辅助诊断; 超声 RF 信号; 支持向量机; Shearlet 变换

中图分类号: TP391

文献标志码: A

引用格式: 童莹, 严郁. 基于超声 RF 信号的乳腺肿瘤分级检测方法[J]. 光电工程, 2019, 46(1): 180368

The grade classification algorithm of breast tumor based on ultrasound RF signals

Tong Ying¹, Yan Yu^{2,3*}¹Department of Communication Engineering, Nanjing Institute of Technology, Nanjing, Jiangsu 211167, China;²School of Computer Science & Engineering, Nanjing University of Science & Technology, Nanjing, Jiangsu 210094, China;³Medical Equipment Department, Affiliated Hospital Nanjing University of TCM (Jiangsu Province Hospital of TCM), Nanjing, Jiangsu 210029, China

Abstract: A novel efficient method based on the ultrasound radio frequency (RF) signals is proposed to distinguish the breast tumors grades. First, we utilize the multi-scale geometric characteristic of Shearlet transformation to extract the multi-scale and multi-directional features of ultrasound RF signal, and then reduce the high-dimensional Shearlet features by multi-scale directional binary pattern which can effectively preserve the sufficient discriminated information. At last, we draw on the feature difference between different grades of breast tumors to design a cascade binary tree SVM classifier which not only overcome the problem of sample quantity disequilibrium but also conform to the subjective diagnosis rule of sonographer. Extensive experiments on 928 breast ultrasound RF signals col-

收稿日期: 2018-07-11; 收到修改稿日期: 2018-10-08

基金项目: 国家自然科学基金项目(61703201); 江苏省自然科学基金项目(BK20170765); 南京工程学院青年基金面上项目(CKJB201602)资助

作者简介: 童莹(1979-), 女, 副教授, 主要从事信号与信息处理的研究。E-mail: tongying@njit.edu.cn

通信作者: 严郁(1979-), 男, 高级工程师, 主要从事生物医学信息处理的研究。E-mail: yanyucan@126.com

lected from the hospital demonstrate the effectiveness of the new proposed method and its precision, sensitivity, specificity, PPV, NPV and MCC are 89.29%, 75.62%, 94.54%, 97%, 98.3% and 81.01%, respectively.

Keywords: CAD; ultrasound RF signal; support vector machine; Shearlet transformation

Citation: Tong Y, Yan Y. The grade classification algorithm of breast tumor based on ultrasound RF signals[J]. *Opto-Electronic Engineering*, 2019, 46(1): 180368

1 引言

据世界卫生组织统计,每年全世界约有138万妇女被诊断患有乳腺癌,这个数据量占所有癌症病例的23%。目前,超声成像技术已成为早期筛查乳腺癌的常用手段之一,它具有无创、简易、可重复等特点。在传统超声诊断中,放射科医生根据临床实践经验,分析超声图像中的乳腺肿瘤特征,并与BI-RADS (breast imaging reporting and data system)标准^[1-2]相匹配,从而实现乳腺病变的良恶性分级诊断。可见,传统人工阅片方式是一个主观判断过程,诊断结果受医生的临床经验、知识水平、疲劳程度限制,不同医师对同一超声图像,可能诊断结果完全不同,甚至同一医师在不同时间、不同状态下对同一超声图像的诊断结果也不相同。因此,将计算机辅助诊断(computer aided diagnosis, CAD)系统^[3-5]应用在超声诊断中,可有效减少医生的主观因素影响,提高疾病解释一致性,使诊断结果更客观、更准确。

近年来,国内外广大学者在超声乳腺CAD系统领域已开展大量研究^[6-9]。2012年, Moon等人^[10]分别采用一阶、二阶统计方法得到B超图像散斑特征,并将其应用在超声诊断中,用以区分乳腺良恶性病变;2013年,Alma等人^[11]利用回声反射性、不均匀性、边缘不规则性等特征,对130例活检定期患者的超声图像进行分析,得到曲下面积(area under curve, AUC)高达 0.947 ± 0.045 ;Huang等人^[12]则采用全局变分模型减少超声图像斑点噪声,结合图像纹理特征和主动轮廓模型,实现乳腺肿瘤区域的自动检测;2016年,Cheng等人^[13]又将深度学习框架应用于超声乳腺图像的自动特征探索,性能表现良好。上述方法均是基于超声B-mode图像进行数据分析和处理。传统超声B-mode图像是由超声射频(radio frequency, RF)信号中的包络线幅度信息转换得到,忽略了频率信息和相位信息,导致基于超声B-mode图像的CAD系统只能分辨直径大于超声波长的组织结构,而对小于超声波长的组织微结构特征无法评价。鉴于此,本文摒弃传

统超声B-mode图像,而采用未经处理的原始超声RF信号进行分析研究,可以充分利用RF信号中的幅度、频率和相位信息,实现乳腺组织微结构的病变特征的准确描述。

目前,基于超声RF信号的研究多集中于前列腺癌检测、淋巴结疾病鉴别、肝组织前后变化监测等方面。Masotti^[14]等利用超声射频局部定征参数分析法对前列腺疾病进行了离体组织定征;Moradi^[15-16]等提出了利用超声RF信号诊断前列腺癌的新方法;四川大学刘志东等^[17]采用时域、频域分析方法,探讨了一种新的脂肪肝分级量化评分方法。近年来,乳腺超声RF信号的分析研究也越来越受到国内外学者广泛关注。庄淑莲等^[18]采用分形维数、谱斜率、谱截距等九个特征对病灶感兴趣区进行定量分析,实现乳腺肿瘤良恶性鉴别;严郁等^[19]采用均值、方差、熵等统计方法描述超声RF信号的乳腺肿瘤特征,实现正常乳腺组织与良性肿瘤的鉴别;2015年,Nishant等^[20]又将灰度共生矩阵引入超声RF信号特征提取中,进一步提高了乳腺病变特征描述的准确度,采用支持向量机(support vector machine, SVM)分类器和随机森林(random forests, RF)分类器分别达到0.86和0.81的曲线下面积。

分析发现,当前基于乳腺超声RF信号的研究^[18-20]多集中于肿瘤的良恶性鉴别,或正常组织与良性组织鉴别,而对于乳腺肿瘤的分级检测研究仍没有涉猎。并且,现有特征提取算法不能准确描述不同分级的乳腺肿瘤的特征差异,导致传统分类器分级检测性能不佳。本文引入多尺度几何分析方法,采用Shearlet变换提取不同分级的乳腺超声RF信号的多尺度、多方向特征,并提出一种多尺度方向二值模式(multi-scale directional binary pattern, MDBP)对其进行降维,在不损失特征信息的条件下,有效降低特征维数。同时,充分利用不同分级的乳腺肿瘤的特征差异性,采用从整体到局部、层层递进的原则,设计层级二叉树SVM(cascade binary tree SVM, CBT-SVM)分类器,实现了3级、4A级~4C级、5级乳腺肿瘤的有效判别。

2 BI-RADS 标准

BI-RADS 标准是对乳腺病灶特征标准化的描述和报告方法,由美国放射学会(American College of Radiology, ACR)提出,对回声模式、钙化等进行了严格的定义,制定了标准化的分级评价标准,有效提高了乳腺肿瘤超声诊断的准确度、敏感度、特异度。BI-RADS 分级标准与对应特征如表 1 所示。

可以看出,虽然 BI-RADS 标准对回声模式、钙化等进行了严格的定义,但这些特征描述具有主观性,导致准确阅读超声图像需要丰富的临床实践经验,医生人工阅片的准确率仅为 69% ~ 75%^[20]。

3 基于 Shearlet 变换和多尺度方向二值模式的乳腺肿瘤特征提取算法

3.1 Shearlet 变换原理

剪切波(Shearlet)变换由合成小波理论衍生而来^[21],继承了曲波(Curvelet)变换和轮廓波(Contourlet)变换的优点,具有各向异性和平移不变性等特点,是一种新型的多尺度几何分析工具,多应用于图像融合^[22-24]、图像去噪^[25-26]、图像恢复^[27-28]等方面,在医学图像特征提取领域应用较少。

剪切波(Shearlet)变换通过对基函数进行缩放、剪切和平移等仿射变换,生成具有不同特征的剪切波函数。函数集为

$$\{\psi_{ast}(x) = a^{-3/4} \psi(A_a^{-1} B_s^{-1} x - t), a \in R^+, s \in R, t \in R^2\}, \quad (1)$$

其中: A_a 为各向异性膨胀矩阵,它控制了 Shearlet 变换的尺度, B_s 为剪切矩阵,它控制了 Shearlet 变换的方向, $\psi \in L^2(R^2)$ 表示函数集属于平方可积空间。Shearlet 变换时,采用非下采样金字塔分解方式实现多尺度分解,图像经过 k 级多尺度分解后,可以生成和原图像大小相同的 k 幅高频子带图像;而多方向分解过程中,在伪极化网格上对高频子带图像进行二维离散傅里叶变换,将图像转换至频率域,实现 Shearlet 多方向滤波。每一层高频子带图像经过 l 级方向分解,会产生 $2^l + 2$ ($l = 0, 1, 2, \dots$) 个和原图像大小相同的方向子带图像,具体实现步骤如图 1 所示。乳腺超声 RF 信号经多尺度、多方向 Shearlet 变换后,可以提取丰富病灶细节特征。

3.2 基于多尺度方向二值模式的特征降维方法

由前面分析可知,一个尺寸大小为 $m \times n$ 的乳腺超声 RF 信号经过 p 个尺度和 q 个方向的 Shearlet 变换后,得到 $p \times q$ 个 $m \times n$ 大小的高频系数矩阵。将这些矩阵拉成列向量,则得到一个 RF 信号所对应的高维

表 1 BI-RADS-US 分级标准及特征描述

Table 1 BI-RADS-US standard and characteristic description

分级	评价	常见判别特征
0	需要附加的影像评价	-
1	阴性	-
2	良性(发现物)	圆形或者椭圆形,边缘光整,无回声,后方回声增强
3	可能良性(发现物),建议短期间隔继续检查	椭圆或稍不规则,边界清晰,边缘光整,平行皮肤方向,后方回声增强或无变化
4A	低度疑似恶性	不规则形状,边缘不光整、分叶毛刺、模糊、成角中一
4B	中度疑似恶性	到两项模糊,边界不清楚,不确定方向,低回声,后方
4C	高度疑似恶性	回声衰减或部分衰减
5	高度提示恶性,需要采取适当措施	不规则形状,边缘不光整、分叶、毛刺、模糊、成角中两项以上,边界模糊,与皮肤不平行,强回声晕征,后方衰减或部分衰减
6	已行活检,并有恶性病理诊断	-

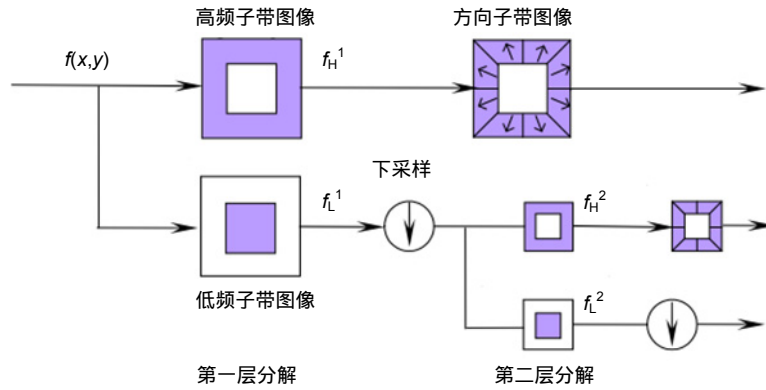


图 1 Shearlet 变换示意图

Fig. 1 The flow diagram of Shearlet transformation

Shearlet 特征。很明显,高维特征空间呈现复杂扭曲分布,不同特征子空间之间边界模糊甚至交融。因此,通常需要进行降维处理,一方面使得不同特征子空间之间具有更明显的可区分性,提高分类精度,另一方面又降低分类检测时间,提高系统实用性。

本文针对 Shearlet 特征的多尺度、多方向特性,提出一种多尺度方向二值模式(multi-scale directional binary pattern, MDBP)用于特征降维。首先,不同尺度下 Shearlet 特征的细节丰富程度不同,变换尺度越大,Shearlet 特征越稀疏。为了避免无效编码,降低特征冗余, MDBP 对 p 个尺度下产生的 Shearlet 高频系数矩阵,依次采用 $(2i+1) \times (2i+1)$ ($i=1,2,\dots,p$) 邻域大小进行二值编码,邻域分布如图 2 所示,即第 1 个尺度下的高频系数矩阵采用 3×3 邻域(蓝色区域)编码,第 2 个尺度下的高频系数矩阵采用 5×5 邻域(黄色区域)编码,第 3 个尺度下的高频系数矩阵采用 7×7 邻域(红色区域)编码,依此类推,实现 MDBP 的多尺度编码;其次,为了不损失 Shearlet 特征的方向性,又不增加模型复杂度, MDBP 仅从 0° 、 45° 、 90° 、 135° 四个关键方向对 Shearlet 高频系数矩阵进行编码,同时为了保留不同尺度下高频系数的整体分布特性, MDBP 又编码了中心值与邻域平均值大小,实现 MDBP 的多方向编码。MDBP 编码(用 $C_{MDBP}(i)$ 表示)公式如下:

$$C_{MDBP}(i) = S(g_0 - g_{im})2^4 + S(g_{i1} - g_{i8})2^3 + S(g_{i4} - g_{i5})2^2 + S(g_{i3} - g_{i6})2^1 + S(g_{i2} - g_{i7})2^0, (2)$$

$$S(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases}, (3)$$

其中:下标 i 表示编码邻域尺度,与 Shearlet 变换尺度

一致, $i=1,2,\dots,p$ 。 g_0 为中心系数值, g_{im} 为 $(2i+1) \times (2i+1)$ 邻域内的平均系数值, $g_{i1} \sim g_{i8}$ 则为该邻域内四个方向位置的高频系数值,如图 2 所示。

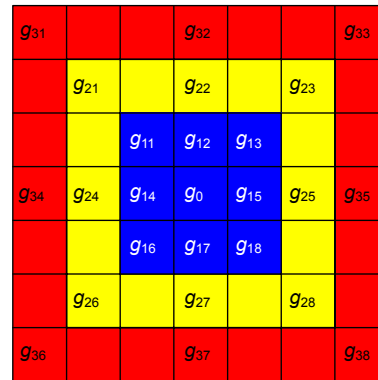


图 2 MDBP 邻域示意图

Fig. 2 The sketch map of MDBP neighborhood

3.3 算法实现步骤

本文中超声 RF 信号的乳腺肿瘤特征提取流程如图 3 所示,具体实现步骤如下:

Step 1: 输入原始乳腺超声 RF 信号;

Step 2: 依次进行 p 个尺度、 q 个方向的 Shearlet 变换,得到 $p \times q$ 个高频系数矩阵;

Step 3: 将第 1 个尺度下的高频系数矩阵划分为 $k_1 \times k_2$ 个子块,利用式(2)对每一子块进行 $C_{MDBP}(1)$ 编码计算,统计每一子块的编码直方图;将 $k_1 \times k_2$ 个统计直方图首尾级联,得到一个高频系数矩阵对应的 MDBP(1)统计直方图特征;重复上述操作,依次得到第 1 个尺度下 q 个高频系数矩阵对应的 q 个 MDBP(1)统计直方图特征;

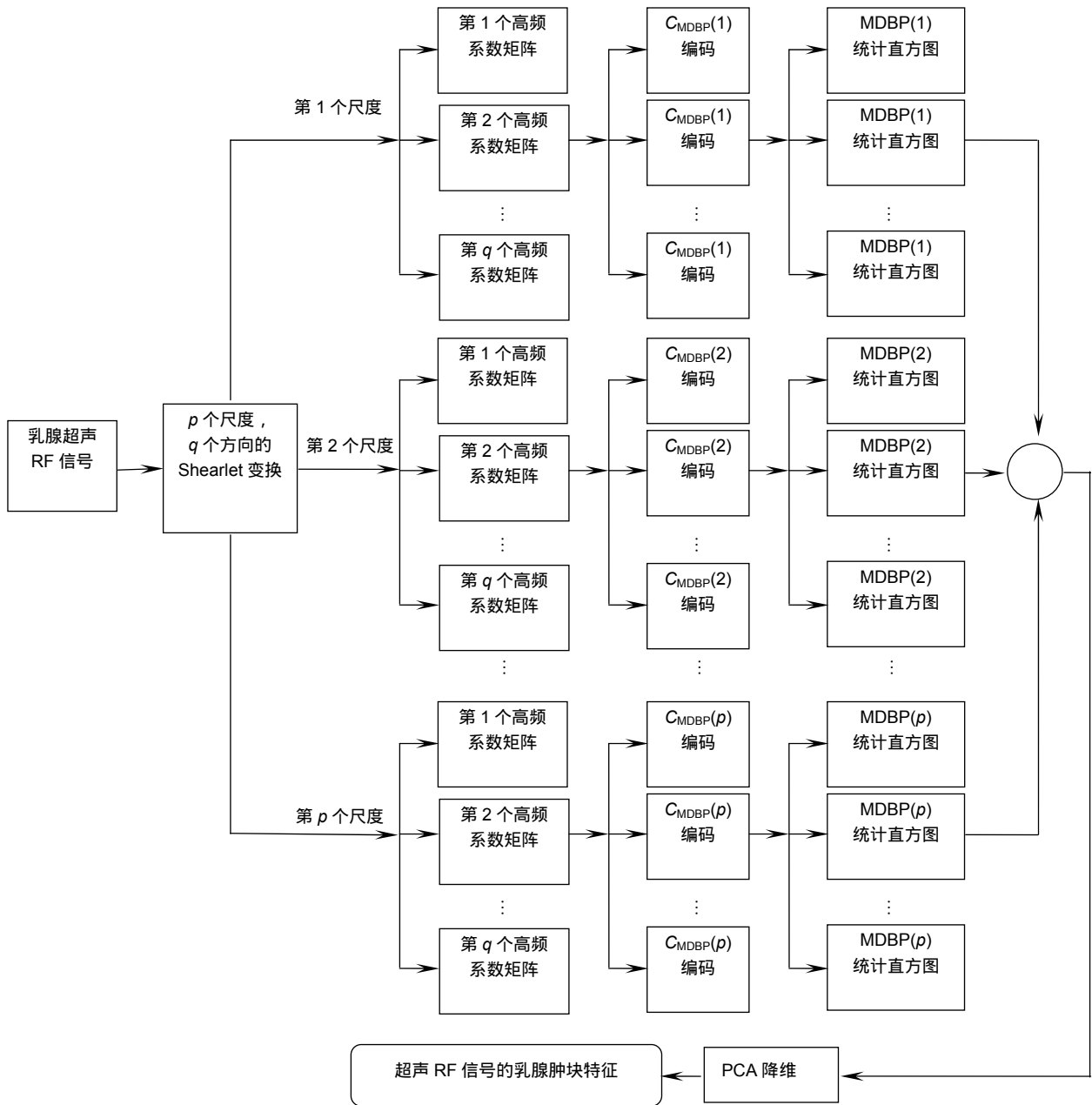


图 3 算法步骤框图

Fig. 3 The flowchat of the proposed feature extraction method

Step 4: 重复 Step 3, 对第 2 个尺度下所有高频系数矩阵进行 $C_{MDBP}(2)$ 编码计算, 编码邻域如图 2 所示, 得到第 2 个尺度下所有高频系数矩阵对应的 $C_{MDBP}(2)$ 统计直方图特征; 依此类推, 得到 $p \times q$ 个高频系数矩阵对应的 $p \times q$ 个统计直方图特征;

Step 5: 将 $p \times q$ 个统计直方图全部首尾级联, 得到一个 RF 信号对应的 MDBP 统计直方图特征;

Step 6: 对 MDBP 统计直方图特征进行主成分分析(principal component analysis, PCA)降维和归一化处

理, 得到最终的超声 RF 信号乳腺肿瘤特征。

4 层级二叉树 SVM 分类器的设计

考虑实际诊断中, 多数患者被诊断为 3 级良性, 仅少数患者为 4 级或 5 级疑似恶性, 导致不同级别的乳腺 RF 信号数据量不均衡, 采用传统分类器检测效果不理想。因此, 本文充分利用不同级别的乳腺肿瘤的特征差异性, 采用从整体到局部、层层递进的原则, 设计一个层级二叉树 SVM(cascade binary tree SVM,)

CBT-SVM)分类器,既有效克服了样本分布不均衡的问题,又符合医生阅片经验,可大大提高乳腺肿瘤的分级检测精度。需要强调的是,由于 CBT-SVM 分类器是由多个 SVM 二分类器级联而成,因此,如何设计 margin 距离较远的 SVM 二分类器且有效级联,是提高 CBT-SVM 分类性能的关键,具体设计思路如下。

首先,分析众多乳腺超声 RF 数据特征,良性肿瘤通常具有形状规则、边界清晰等特征,与恶性肿瘤特征相比,具有明显差异性。因此,我们采用类间距离较大的良性数据(3 级)和疑似恶性数据(4A 级、4B 级、4C 级、5 级)训练 SVM 二分类器,用于首先判断乳腺肿瘤的良好性。若任一待测 RF 数据被判为良性肿瘤(3 级),则检测结束;若判为疑似恶性肿瘤(4A 级、4B 级、4C 级、5 级),则还需要继续检测。

接下来,分析不同级别的疑似恶性乳腺肿瘤(4A 级、4B 级、4C 级、5 级)特征,高度可能恶性肿瘤(5 级)具有明显的形状不规则、边界模糊、分叶、毛刺等特征,相比疑似可能恶性肿瘤(4A 级、4B 级、4C 级)特征,5 级恶性肿瘤更容易辨识。因此,基于上述两类有差异的乳腺超声 RF 数据训练 SVM 二分类器,用于进一步分析判断待测数据是否属于高度可能恶性肿瘤(5 级)。

最后,完成良性肿瘤(3 级)和高度可能恶性肿瘤(5 级)的检测后,对乳腺超声 RF 数据进行疑似可能恶性

肿瘤(4A 级、4B 级、4C 级)分级判断。发现 4A 级~4C 级乳腺肿瘤的超声信号特征界定模糊,无法直接基于此构建 margin 距离较远的 SVM 二分类器。但分析活检结果可知,4A 级和 4B 级乳腺肿瘤患者,活检结果为良性的数量多于恶性的数量,而 4C 级患者,活检结果多为恶性,这也间接反应了疑似可能恶性肿瘤之间的特征差异性。因此,首先采用高度疑似恶性肿瘤数据(4C 级)和非高度疑似恶性肿瘤数据(4A 级、4B 级)训练 SVM 二分类器,又采用轻度疑似恶性肿瘤数据(4A 级)和中度疑似恶性肿瘤数据(4B 级)训练 SVM 二分类器,依次级联,实现疑似可能恶性肿瘤的有效分级。层级二叉树 SVM 分类器的设计结构如图 4 所示。

可以看出, CBT-SVM 分类器从全局角度划分样本,可以有效避免不同分级的乳腺肿瘤的数据量分布不均衡问题,同时, CBT-SVM 分类器的分级检测顺序也完全符合医生在诊断过程中先粗后细的阅片规律,即首先初步判断乳腺肿瘤的良好性,若为疑似恶性,则再进一步细致阅片,利用经验分析判断乳腺肿瘤的恶性程度。在实验部分,将与有向无环图 SVM(directed acyclic graph SVM, DAG-SVM)分类器、一对一 SVM(one against one SVM, OAO-SVM)分类器等多个分类器进行实验比较,进一步分析说明本文设计的层级二叉树 SVM 分类器在超声乳腺肿瘤分级检测中的有效性。

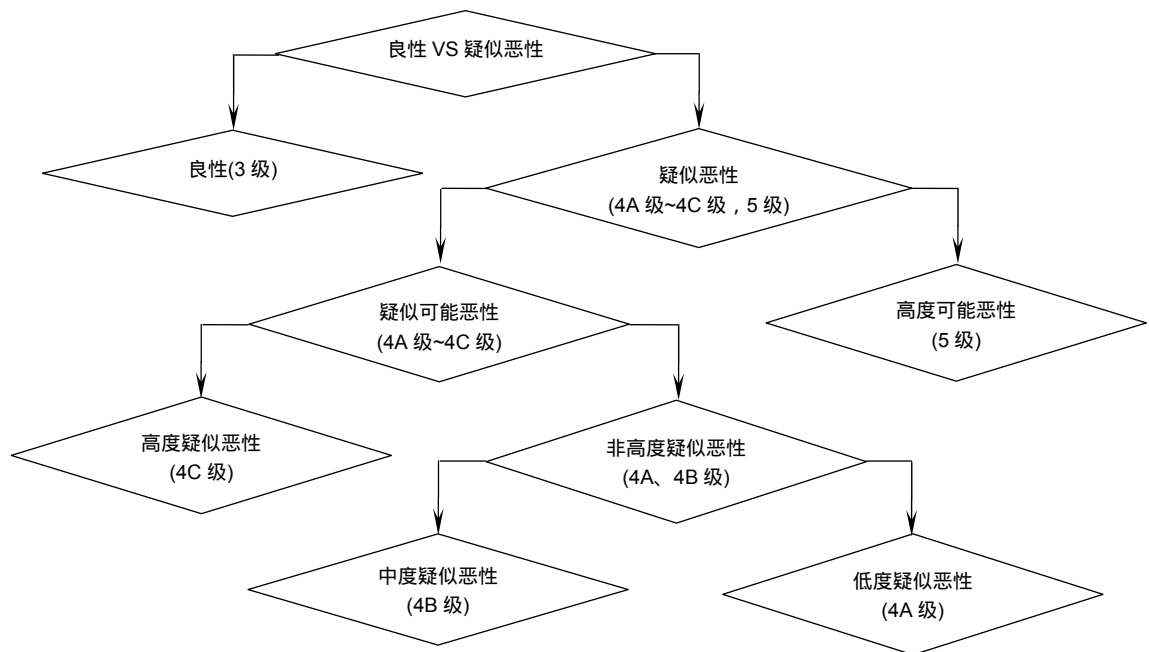


图 4 层级二叉树 SVM 分类器的结构图

Fig. 4 The structure of cascade binary tree SVM classifier

5 实验及结果分析

5.1 数据来源

在实验中, 乳腺超声 RF 数据来源于超声诊断仪 (VINNO 70, 飞依诺科技有限公司, 苏州), 超声探头共 312 条扫描线, 发射频率为 5 MHz ~ 14 MHz。实验中共采集 3 级 RF 数据 584 条, 4A 级 RF 数据 128 条, 4B 级 RF 数据 44 条, 4C 级 RF 数据 50 条, 5 级 RF 数据 122 条, 共 928 组数据, 其中部分样本如图 5 所示。所有数据都获得受试者书面同意, 且符合医院人体伦理认可。本文选取 704 组作为训练样本, 其中 3 级有 448 条, 4A 级有 96 条, 4B 级有 32 条, 4C 级有 32 条, 5 级有 96 条, 剩余的 136 组 3 级数据, 32 组 4A 级数据, 12 组 4B 级数据, 18 组 4C 级数据, 以及 26 组 5 级数据, 共 224 组数据作为测试样本, 实验仿真结果及分析如下。

5.2 Shearlet 变换的参数选择

在 Shearlet 变换中方向和尺度这两个参数的选择, 对特征描述是否准确尤为重要, 若方向和尺度过多, 则容易造成特征信息冗余, 反之, 则容易造成特征信息不足。这里, 选择尺度参数为 3、4、5, 方向参数为 4 和 8, 共六种情况进行分析研究。首先, 提取超声乳腺 RF 信号在不同尺度、不同方向参数下的 Shearlet 特征, 并对其进行 MDBP 编码, 得到 MDBP 统计直方图特征; 然后, 对 MDBP 特征进行 PCA 降维和归一化处理; 最后, 采用层级二叉树 SVM 分类器进行分级检测, 实验结果如表 2 所示。这里对 Shearlet 特征进行 MDBP 编码降维时, MDBP 分块尺寸均为 28 pixels × 26 pixels。

从表中可以看出, 随着 Shearlet 变换尺度和方向数的增加, 乳腺肿瘤分级检测准确率并没有随之呈线性增加, 当尺度参数为 4, 方向参数为 8 时, 达到最

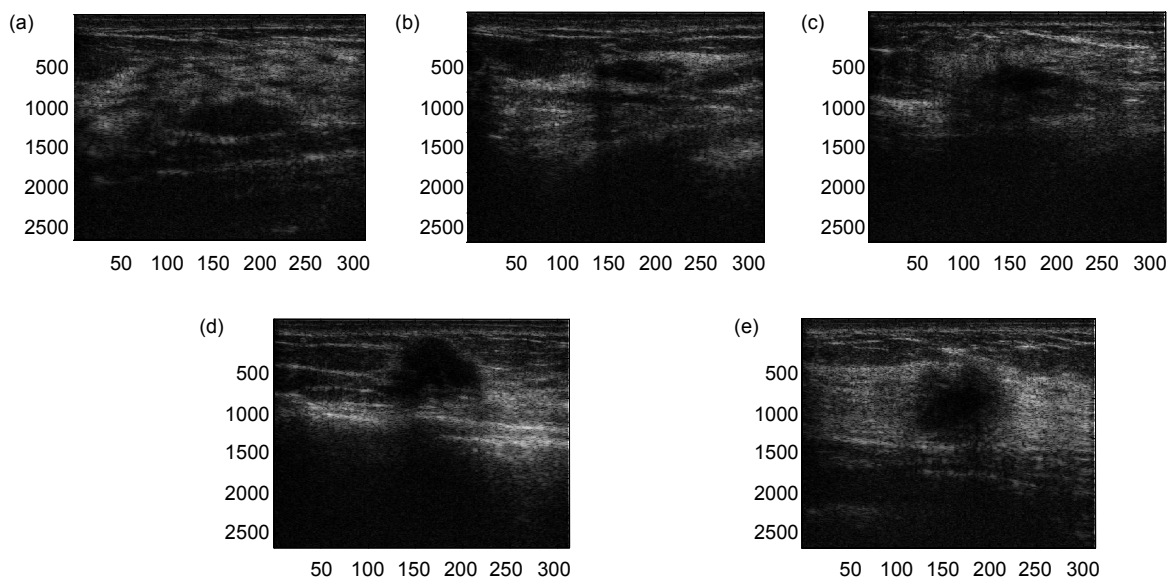


图 5 不同分级的乳腺超声 RF 数据。(a) 3 级; (b) 4A 级; (c) 4B 级; (d) 4C 级; (e) 5 级

Fig. 5 The different grades of breast tumors of ultrasound RF signal. (a) 3 grade; (b) 4A grade; (c) 4B grade; (d) 4C grade; (e) 5 grade

表 2 不同方向与尺度参数下的识别结果

Table 2 Performance comparison of different parameters of Shearlet transformation

Shearlet 参数	Shearlet 特征维数	MDBP 特征维数	识别率/(%)
3 个尺度 4 个方向	10692864	470016	79.02
3 个尺度 8 个方向	21385728	940032	81.25
4 个尺度 4 个方向	14257152	626688	79.91
4 个尺度 8 个方向	28514304	1253376	89.29
5 个尺度 4 个方向	17821440	783360	79.46
5 个尺度 8 个方向	35642880	1566720	83.93

佳检测效果, 尺度或方向数再增加, 分级检测正确率则下降。分析原因可知, Shearlet 特征维数与 RF 信号深度、采集信号探头数、尺度参数以及方向参数有关, 当尺度参数为 5, 方向参数为 8 时, Shearlet 变换的特征维数高达 35642880, 虽然采用 MDBP 算子对其进行降维, 但特征维数仍较高, 特征信息冗余, 导致乳腺肿瘤分级检测准确率下降。

同时, 进一步分析发现, 在对乳腺超声 RF 信号进行 Shearlet 特征提取时, 方向参数对病灶特征描述准确度的影响大于尺度参数。从表 2 可以看出, 尺度参数相同, 方向参数为 8 的检测结果均高于方向参数为 4 的检测结果。例如, 尺度参数为 3 时, 方向参数为 8 的检测结果比方向参数为 4 的检测结果高 2.23%, 其他尺度情况下类似。造成这种现象的原因在于, 随着方向参数增加, Shearlet 变换可以提取更多方向上的乳腺肿瘤纹理细节特征, 从而可以更准确区分不同病灶之间的特征差异, 分级检测的准确率也随之提高。反之, 当方向参数一致, 尺度参数变化时, 分级检测准确率并不随尺度参数的增加而增加, 例如, 5 个尺度 4 个方向的检测结果与 4 个尺度 4 个方向的检测结果相当。综合考虑分级检测系统的精度和速度, 本文选用 4 个尺度 8 个方向进行 Shearlet 变换, 此时提取的病灶特征最佳, 分级检测准确率最高。

5.3 MDBP 的参数选择

当 Shearlet 变换中方向与尺度参数固定时(这里取 4 尺度 8 方向), MDBP 编码中对高频系数矩阵进行分块的个数多少则会影响特征的准确度。若分块个数太多, 则特征信息冗余, 反之, 则特征信息不足。这里, 取每个子块大小为 14 pixels×13 pixels, 28 pixels×26 pixels 和 42 pixels×39 pixels 三种情况, 并采用层级二叉树 SVM 分类器进行分级检测, 实验结果如表 3 所

示。可以看出, 当子块尺寸为 14 pixels×13 pixels 时, 分块个数过多, 而当子块尺寸为 42 pixels×39 pixels 时, 分块个数又过少, 这两种情况下特征描述均不如子块尺寸为 28 pixels×26 pixels 时准确, 此时分块个数最适中, 特征描述最准确, 识别率最高。

表 3 不同分块尺寸下的识别率

Table 3 Performance comparison of different block sizes of MDBP

分块尺寸/pixels	14×13	28×26	42×39
识别率/(%)	81.25	89.29	73.66

5.4 本文特征提取方法的有效性验证

为了验证本文方法在提取乳腺肿瘤特征中的有效性, 这里选用 Curvelet 变换、Contourlet 变换、Gabor 变换和 Log_Gabor 变换分别提取乳腺超声 RF 信号特征, 并用 PCA 对其进行特征降维。由于 Curvelet 变换的中频系数体现了方向信息, 因此设置尺度参数为 3, 方向参数为 8, 选取 Detail 层作为细节特征进行后续分级检测。除 Curvelet 变换外, 其余变换均采用与 Shearlet 变换相同的参数, 即 4 个尺度、8 个方向。本文特征提取方法中参数选择见 5.2~5.3 节。为了丰富比较结果, 这里又采用庄淑莲提出的传统 RF 信号特征^[18]和 Nishant 特征^[20]进行比较。传统 RF 信号特征主要包括低频(S1)、中低(S2)、中高(S3)、高频(S4), 谱斜率 Slope、谱截距 Intercept、Higuchi 分形维数, 以及校准功率谱斜率 Slope 和谱截距 Intercept, 共九个特征。Nishant 特征^[20]则是在上述特征的基础上又增加了 164 个特征, 包括均值、标准差、偏度、峰度以及灰度共生矩阵等。这里统一采用层级二叉树 SVM 器对乳腺超声 RF 信号进行分级检测, 训练样本取 704 组, 测试样本取 224 组, 实验结果如表 4 所示。

表 4 不同特征提取算法下的识别率

Table 4 Recognition rates of different feature extraction algorithms

算法	特征维数	PCA 降维特征维数	识别率/(%)
传统特征 ^[18]	9	/	43.75
Nishant 特征 ^[20]	173	/	66.07
Curvelet 变换	7128576	743	74.11
Contourlet 变换	28514304	849	73.21
Gabor 变换	28514304	612	71.43
Log Gabor 变换	28514304	634	60.27
Shearlet 变换	28514304	854	78.13
本文方法	1253376	762	89.29

可以看出,相比传统特征和纹理特征,以 Shearlet 为代表的多尺度几何分析方法可以提取出丰富的病灶细节特征,具有较好的识别效果,同时,与 Curvelet 变换、Contourlet 变换以及 Gabor 和 Log_Gabor 两个频域变换相比较,Shearlet 变换不受支撑集大小限制,具有更好的方向性和稀疏性。在本文特征提取方法中将其与 MDBP 编码相结合,既降低了特征维度,也没有损失特征信息,更适合乳腺肿瘤特征提取。

5.5 层级二叉树 SVM 分类器的有效性验证

为了验证层级二叉树 SVM(cascade binary tree SVM, CBT-SVM)分类器的有效性,这里选用有向无环图 SVM(directed acyclic graph SVM, DAG-SVM)分类器、一对一 SVM(one against one SVM, OAO-SVM)分类器、K 近邻(K near neighbors, KNN)分类器和随机森林(random forest, RF)分类器对测试样本进行乳腺肿瘤分级检测,特征提取方法和参数选择一致,均采用 4 尺度 8 方向的 Shearlet 变换和分块尺寸为 28 pixels × 26 pixels 的 MDBP 编码获取超声 RF 信号的乳腺肿瘤特征。

由于根节点选择是设计 DAG-SVM 分类器的关键,考虑 3 级乳腺肿瘤(良性)与 5 级乳腺肿瘤(恶性)的特征差异最大,则训练得到的 SVM 二分类器的 margin 值最大,因此这里取 3 级和 5 级样本训练 SVM 二分类器,作为 DAG-SVM 分类器的根节点,剩余节

点则根据特征差异由大到小的原则依次构建,得到 DAG-SVM 分类器的结构如图 6 所示。OAO-SVM 分类器相比 DAG-SVM 分类器,仅缺少二叉树判断路径,它逐个采用 10 个 SVM 二分类器(与图 5 中 10 个 SVM 二分类器一致)进行乳腺肿瘤分级判断,然后统计 10 个 SVM 二分类器的输出标签,统计标签个数最多者即为最终分级结果。KNN 分类器和 RF 分类器均采用 Matlab 中库函数进行实验仿真。

采用上述四种不同分类器分别对乳腺超声 RF 信号进行分级检测,结果如表 5 所示。可以看出,虽然 DAG-SVM 和 OAO-SVM 从局部角度依次分析各个不同级别乳腺肿瘤的特征差异性,但相比 CBT-SVM 分类器从全局角度判别的性能,分类效果不理想。同时,在不考虑分类器训练代价的基础上,CBT-SVM 分类器的检测时间最短。综合考虑各个分类器的检测精度和速度,CBT-SVM 分类器性能最佳,具有实用价值。

5.6 分类器性能比较

为了进一步说明层级二叉树分类器(CBT-SVM)的有效性,这里使用 6 个指标评估分类器的性能,分别为准确度(accuracy,用 $P_{Accuracy}$ 表示)、敏感度(sensitivity,用 $P_{Sensitivity}$ 表示)、特异度(specificity,用 $P_{Specificity}$ 表示)、阳性预测值(positive predictive value,用 P_{PPV} 表示)、阴性预测值(negative predictive value,用 P_{NPV} 表示)、马修斯相关性系数(Matthews correlation

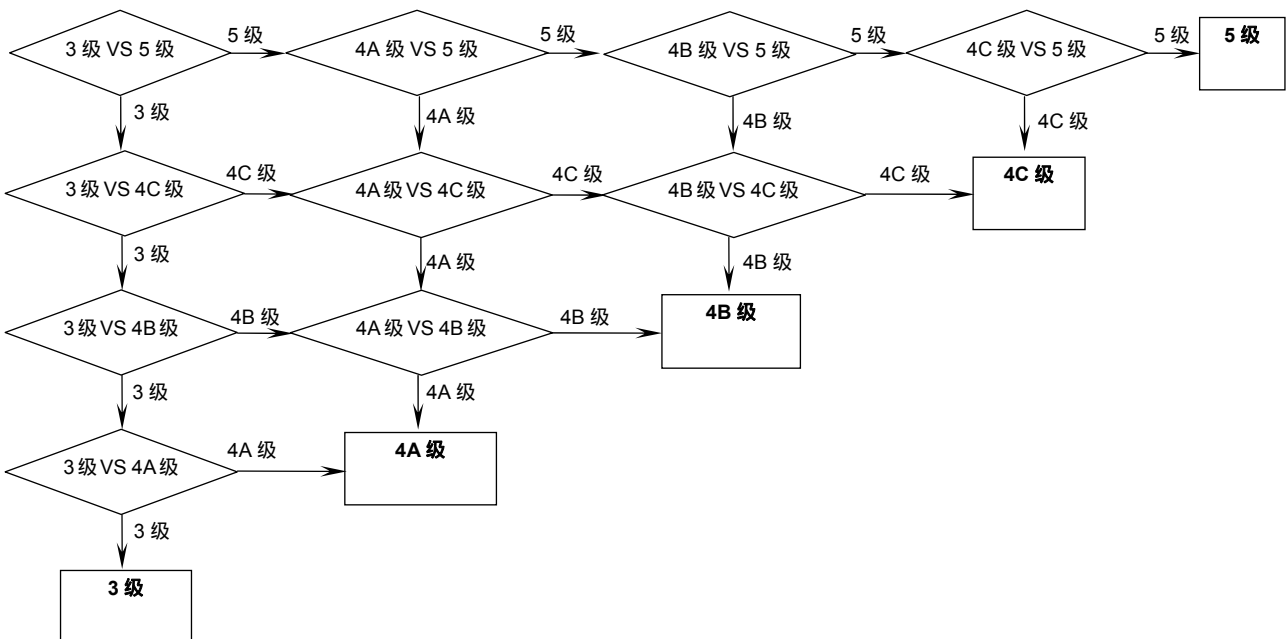


图 6 有向无环图 SVM 分类器的结构图
Fig. 6 The structure of directed acyclic graph SVM classifier

表 5 不同分类器下的识别率和时间

Table 5 Recognition rate and classification time of different classifiers

	识别率/(%)	训练时间/s	测试时间/s
DAG-SVM	58.03	58.93	1.79
OAO-SVM	58.03	58.93	3.93
KNN	79.91	1.81	0.44
RF	79.46	37.69	8.09
CBT-SVM	89.29	206.89	0.37

coefficient, 用 P_{MCC} 表示)。定义如下:

$$P_{Accuracy} = \frac{TN + TP}{TN + FN + TP + FP}, \quad (4)$$

$$P_{Sensitivity} = \frac{TP}{TP + FN}, \quad (5)$$

$$P_{Specificity} = \frac{TN}{TN + FP}, \quad (6)$$

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

$$P_{NPV} = \frac{TN}{TN + FN}, \quad (8)$$

$$P_{MCC} = \frac{TP \times TN - FP \times FN}{[(TP + FP)(TP + FN)(TN + FP)(TN + FN)]^{1/2}}. \quad (9)$$

其中: TP 为真阳性例数, TN 为真阴性例数, FP 为假阳性例数, FN 为假阴性例数。这里分别计算每个肿瘤分级的评价指标, 最终评价指标为 5 个级别的平均值。

由于有向无环图 SVM 分类器(DAG-SVM)与一对一 SVM 分类器(OAO-SVM)的检测结果一致, 所以这里仅分析有向无环图 SVM 分类器(DAG-SVM)、K 近邻分类器(KNN)、随机森林分类器(RF)和层级二叉树

SVM 分类器(CBT-SVM)四种分类器的性能, 根据式(4)~式(9), 计算各个性能指标, 统计结果如表 6 所示。可以看出, CBT-SVM 的六个性能指标均高于其他分类器, 尤其是 PPV 值高达 97%, NPV 值高达 98.3%, 这说明 CBT-SVM 分类器在检测阳性和阴性结果时具有卓越的查准率, 即被检测为阳性或阴性的患者确实患病或无病的可能性最大。同时, CBT-SVM 分类器的敏感度为 75.62%, 特异度为 94.54%, 均高于其他分类器, 这也进一步说明 CBT-SVM 分类器漏检和误检的可能性最小。

四种分类器的混淆矩阵分别如表 7~表 10 所示。根据表 7~表 10, 将 3 级表示为良性病变, 4 级和 5 级表示为疑似恶性病变, 画出受试者工作特征曲线(receiver operating characteristic curve, 简称 ROC 曲线), 并计算曲下面积(area under curve, AUC), 如图 7~图 10 所示。可以看出, 层级二叉树 SVM 分类器的 AUC 最大, 表明分离器性能最佳, 由此也进一步验证了层级二叉树分类器的有效性。

表 6 四种分类器性能比较

Table 6 Performance comparison of four classifiers

分类器	准确度/(%)	敏感度/(%)	特异性/(%)	PPV/(%)	NPV/(%)	MCC/(%)
DAG-SVM	58.03	64.71	88.97	72.56	88.03	54.28
KNN	79.91	63.84	92.69	79.53	88.03	64.31
RF	79.46	63.07	92.59	77.88	86.73	63.14
CBT-SVM	89.29	75.62	94.54	97.00	98.30	81.01

表 7 有向无环图 SVM 分类器的混淆矩阵

Table 7 The confusion matrix of DAG-SVM classifier

输入	3 级	4A 级	4B 级	4C 级	5 级
3 级	66	29	0	0	41
4A 级	0	24	0	0	8
4B 级	0	0	8	0	4
4C 级	12	0	0	6	0
5 级	0	0	0	0	26

表 8 KNN 分类器的混淆矩阵

Table 8 The confusion matrix of KNN classifier

输入	3 级	4A 级	4B 级	4C 级	5 级
3 级	130	0	0	0	6
4A 级	4	20	0	0	8
4B 级	2	0	8	0	2
4C 级	6	0	0	8	4
5 级	10	0	0	3	13

表 9 随机森林分类器的混淆矩阵

Table 9 The confusion matrix of random forest classifier

输入	3 级	4A 级	4B 级	4C 级	5 级
3 级	130	0	0	0	6
4A 级	4	20	0	0	8
4B 级	2	0	8	0	2
4C 级	6	0	0	8	4
5 级	10	0	0	4	12

表 10 层级二叉树分类器的混淆矩阵

Table 10 The confusion matrix of CBT-SVM classifier

输入	3 级	4A 级	4B 级	4C 级	5 级
3 级	136	0	0	0	0
4A 级	8	24	0	0	0
4B 级	4	0	8	0	0
4C 级	10	0	0	8	0
5 级	2	0	0	0	24

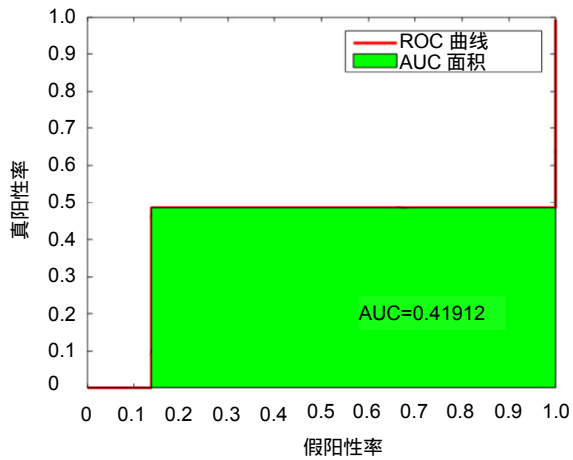


图 7 有向无环图 SVM 分类器的 ROC 曲线及 AUC 面积

Fig. 7 The ROC and AUC of DAG-SVM classifier

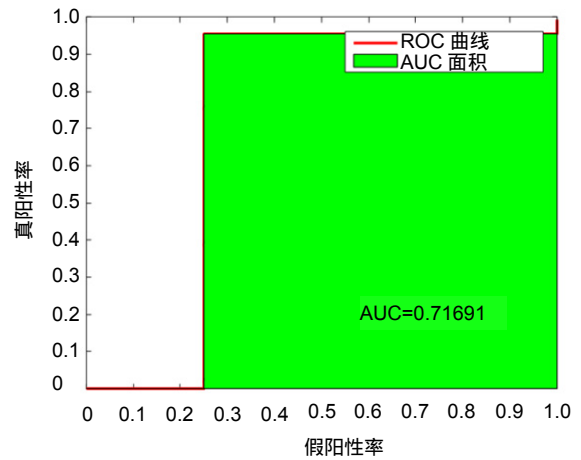


图 8 KNN 分类器的 ROC 曲线及 AUC 面积

Fig. 8 The ROC and AUC of KNN classifier

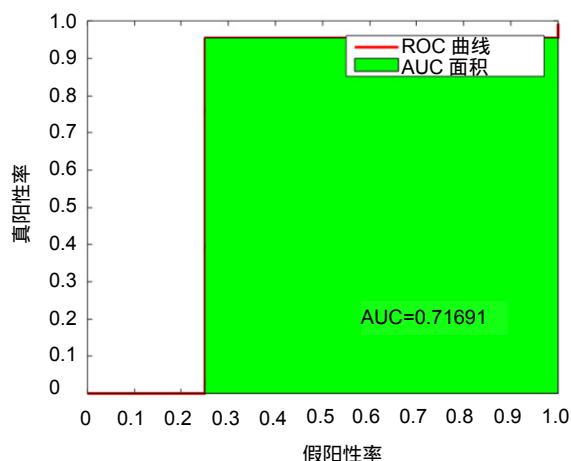


图9 随机森林分类器的 ROC 曲线及 AUC 面积
Fig. 9 The ROC and AUC of random forest classifier

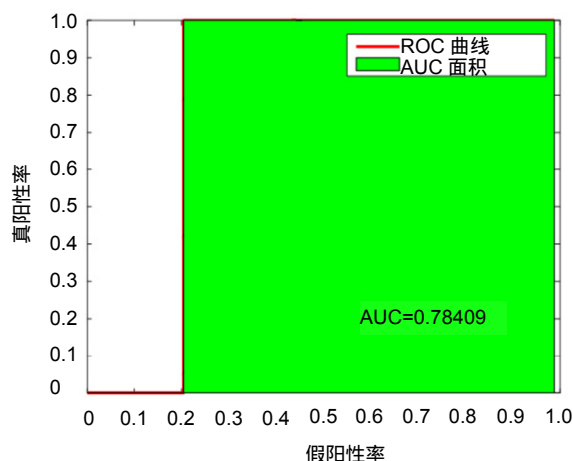


图10 层级二叉树 SVM 分类器的 ROC 曲线及 AUC 面积
Fig. 10 The ROC and AUC of CBT-SVM classifier

6 结论

本文从超声 RF 信号角度出发,在特征提取和分类器两个方面进行改进,提出一种有效的乳腺肿瘤分级检测方法。首先,引入 Shearlet 多尺度几何特性,提取不同分级的乳腺肿瘤的多尺度、多方向特征;其次,针对 Shearlet 特征的方向性、稀疏性,采用多尺度方向二值模式(multi-scale directional binary pattern, MDBP)对其进行降维,在不损失特征信息的条件下,有效降低特征维数;最后,充分利用不同分级的乳腺肿瘤的特征差异性,设计层级二叉树 SVM(cascade binary tree SVM, CBT-SVM)分类器,既避免了样本分布不均问题,又符合医生主观阅片规律。在真实超声 RF 数据集上进行实验仿真,大量实验结果表明,本文方法在超声乳腺分级检测中具有优越性,尤其是阳性查准率和阴性查准率分别达到 97%和 98.3%。

由于超声成像具有动态特性,我们接下来将充分利用超声 RF 数据的时空信息,进一步准确提取不同分级的乳腺肿瘤特征,提高乳腺肿瘤准确分级检测准确率。

参考文献

[1] Zhou S C, Zeng W, Fan Y W, et al. Discussing of using breast grades in ultrasound[J]. *Chinese Journal of Ultrasound in Medicine*, 2008, **24**(6): 19–23.
周世崇, 曾炜, 范亦武, 等. 乳腺超声分级方法应用的初步探讨[J]. *中国超声医学杂志*, 2008, **24**(6): 19–23.

[2] Hou X Y, Gao Y, Huang X L, et al. Application value of breast

image report data system for ultrasonography in mammary gland[J]. *Chinese Journal of Medical Ultrasound (Electronic Edition)*, 2011, **8**(6): 1227–1233.

侯新燕, 高宇, 黄晓玲, 等. 乳腺影像报告数据系统在乳腺超声中的应用价值[J]. *中华医学超声杂志(电子版)*, 2011, **8**(6): 1227–1233.

[3] Suzuki K. A review of computer-aided diagnosis in thoracic and colonic imaging[J]. *Quantitative Imaging in Medicine and Surgery*, 2012, **2**(3): 163–176.

[4] He J T, Chen M H, Jia W Y, et al. Segmentation of diabetic macular edema in Oct retinal images[J]. *Opto-Electronic Engineering*, 2018, **45**(7): 170605.
何锦涛, 陈明惠, 贾文字, 等. 眼底 OCT 图像中糖尿病性黄斑水肿的分割[J]. *光电工程*, 2018, **45**(7): 170605.

[5] Lv W, Zhai Q W, Chu J H, et al. Automated detection of diabetic retinopathy exudates in color fundus images[J]. *Opto-Electronic Engineering*, 2016, **43**(12): 183–192, 199.
吕卫, 翟庆伟, 褚晶辉, 等. 彩色眼底图像糖网渗出物的自动检测[J]. *光电工程*, 2016, **43**(12): 183–192, 199.

[6] Wang X W, Li L H, Liu W, et al. An interactive system for computer-aided diagnosis of breast masses[J]. *Journal of Digital Imaging*, 2012, **25**(5): 570–579.

[7] Jen C C, Yu S S. Automatic detection of abnormal mammograms in mammographic images[J]. *Expert Systems with Applications*, 2015, **42**(6): 3048–3055.

[8] Sharma S, Khanna P. Computer-aided diagnosis of malignant mammograms using Zernike moments and SVM[J]. *Journal of Digital Imaging*, 2015, **28**(1): 77–90.

[9] Tan W C, Wang J H, Cai H M, et al. The performance of computer-aided diagnosis for DCIS based on classification of clustered microcalcifications[J]. *Journal of Clinical Radiology*, 2016, **35**(9): 1352–1356.
谭婉婷, 王金霞, 蔡洪明, 等. 基于微钙化检测的计算机辅助诊断系统对于乳腺导管原位癌的诊断价值[J]. *临床放射学杂志*, 2016, **35**(9): 1352–1356.

[10] Moon W K, Lo C M, Huang C S, et al. Computer-aided diagnosis based on speckle patterns in ultrasound images[J]. *Ultrasound in Medicine & Biology*, 2012, **38**(7): 1251–1261.

- [11] Alam S K, Feleppa E J, Rondeau M, *et al.* Computer-Aided diagnosis of solid breast lesions using an ultrasonic multi-feature analysis procedure[J]. *Bangladesh Journal of Medical Physics*, 2013, **4**(1): 1–10.
- [12] Huang Q H, Yang F B, Liu L Z, *et al.* Automatic segmentation of breast lesions for interaction in ultrasonic computer-aided diagnosis[J]. *Information Sciences*, 2015, **314**: 293–310.
- [13] Cheng J Z, Ni D, Chou Y H, *et al.* Computer-Aided diagnosis with deep learning architecture: Applications to breast lesions in us images and pulmonary nodules in CT scans[J]. *Scientific Reports*, 2016, **6**: 24454.
- [14] Masotti L, Biagi E, Granchi S, *et al.* Tissue differentiation based on radiofrequency echographic signal local spectral content[C]//*Proceedings of the IEEE Symposium on Ultrasonics*, 2003, **1**: 1030–1033.
- [15] Moradi M, Mahdavi S S, Nir G, *et al.* Ultrasound RF time series for tissue typing: first in vivo clinical results[J]. *Proceedings of SPIE*, 2013, **8670**: 867011.
- [16] Moradi M, Abolmaesumi P, Mousavi P. Tissue typing using ultrasound RF time series: experiments with animal tissue samples[J]. *Medical Physics*, 2010, **37**(8): 4401–4413.
- [17] Liu Z D, Luo Y, Lin J L, *et al.* Quantizing and grading of fatty liver based on ultrasonic RF signals[J]. *Journal of Sichuan University (Engineering Science Edition)*, 2011, **43**(S1): 160–164.
刘志东, 罗燕, 林江莉, 等. 基于超声射频 RF 信号的脂肪肝分级量化方法[J]. *四川大学学报(工程科学版)*, 2011, **43**(S1): 160–164.
- [18] Zhuang S L, Zhou J H, Wang J W, *et al.* Differential diagnosis of breast lesions with RF time-series signal based on ultrasonic radio-frequency flow[J]. *Chinese Journal of Medical Ultrasound (Electronic Edition)*, 2016, **13**(5): 393–397.
庄淑莲, 周建华, 王建伟, 等. 基于超声射频流的 RF 时间序列信号在乳腺病变良恶性鉴别中的价值[J]. *中华医学超声杂志(电子版)*, 2016, **13**(5): 393–397.
- [19] Yan Y, Fang G, Cai R Q, *et al.* Ultrasound radio-frequency characteristic signal processing for breast lump recognition[J]. *China Medical Equipment*, 2016, **13**(9): 20–22.
严郁, 方舫, 蔡润秋, 等. 基于乳腺肿块的超声射频信号特征算法识别[J]. *中国医学装备*, 2016, **13**(9): 20–22.
- [20] Uniyal N, Eskandari H, Abolmaesumi P, *et al.* Ultrasound RF time series for classification of breast lesions[J]. *IEEE Transactions on Medical Imaging*, 2015, **34**(2): 652–661.
- [21] Guo K H, Labate D. Optimally sparse multidimensional representation using shearlets[J]. *SIAM Journal on Mathematical Analysis*, 2007, **39**(1): 298–318.
- [22] Gao G R, Xu L P, Feng D Z. Multi-focus image fusion based on non-subsampled shearlet transform[J]. *IET Image Processing*, 2013, **7**(6): 633–639.
- [23] Chang L H. Fusion method based on shearlet transform and sparse representation[J]. *Acta Scientiarum Naturalium Universitatis Sunyatseni*, 2017, **56**(4): 16–19.
常莉红. 基于剪切波变换和稀疏表示理论的图像融合方法[J]. *中山大学学报(自然科学版)*, 2017, **56**(4): 16–19.
- [24] Zhou F, Jia Z H, Yang J, *et al.* Medical image enhancement method based on improved Gamma correction in shearlet domain[J]. *Journal of Optoelectronics · Laser*, 2017, **28**(5): 566–572.
周飞, 贾振红, 杨杰, 等. 基于剪切波域改进 Gamma 校正的医学图像增强算法[J]. *光电子·激光*, 2017, **28**(5): 566–572.
- [25] Xu C, Chen X, Ji Q Y. Shearlet domain image denoising via sparse coding[J]. *Laser Journal*, 2017, **38**(10): 96–100.
徐畅, 陈晓, 季仟亿. 基于稀疏编码的 Shearlet 域图像去噪[J]. *激光杂志*, 2017, **38**(10): 96–100.
- [26] Easley G R, Labate D, Colonna F. Shearlet-based total variation diffusion for denoising[J]. *IEEE Transactions on Image Processing*, 2009, **18**(2): 260–268.
- [27] Vagharshakyan S, Bregovic R, Gotchev A. Light field reconstruction using Shearlet transform[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, **40**(1): 133–147.
- [28] Xu Z L, Deng C Z, Zhang Y S. Shearlet sparsity regularized image reconstruction based on nonlocal self-similarity[J]. *Journal of University of Electronic Science and Technology of China*, 2016, **45**(1): 43–47, 101.
许志良, 邓承志, 张运生. 非局域自相似约束的 Shearlet 稀疏正则化图像恢复[J]. *电子科技大学学报*, 2016, **45**(1): 43–47, 101.
- [29] Zhou S C, Shi J, Zhu J, *et al.* Shearlet-based texture feature extraction for classification of breast tumor in ultrasound image[J]. *Biomedical Signal Processing and Control*, 2013, **8**(6): 688–696.

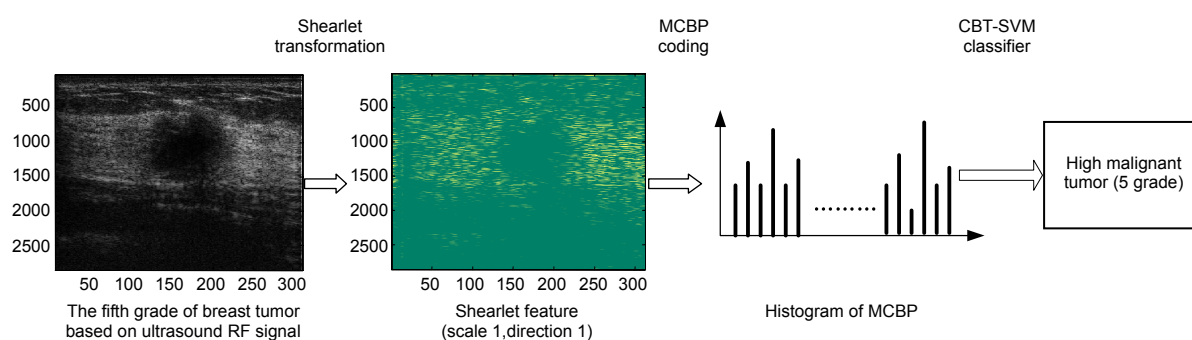
The grade classification algorithm of breast tumor based on ultrasound RF signals

Tong Ying¹, Yan Yu^{2,3*}

¹Department of Communication Engineering, Nanjing Institute of Technology, Nanjing, Jiangsu 211167, China;

²School of Computer Science & Engineering, Nanjing University of Science & Technology, Nanjing, Jiangsu 210094, China;

³Medical Equipment Department, Affiliated Hospital Nanjing University of TCM (Jiangsu Province Hospital of TCM), Nanjing, Jiangsu 210029, China



The flowchart of the proposed method

Overview: According to the statistics published by the American Cancer Society (ACS) in 2015, it is estimated that breast cancer is one of the most common types of cancer in women's patients accounting for 29% of all cancer cases. Early detection and better diagnosis methods play a significant role in reducing the number of fatalities induced by breast cancer. Current sonography has become one of the common methods for early screening breast cancer which are widely used to evaluate doubtful masses based on breast imaging-reporting and data system (BI-RADS). However, this method is limited by low contrast of B-mode images and high subjectivity of sonographers which may make the diagnosis results inaccurate and inconsistent. To address these limitations, ultrasound-based computer aided diagnosis (CAD) system is proposed to assist sonographers in breast tumor diagnosis for achieving higher accuracy and consistency. Since most of the existing CAD systems only can distinguish benign tumors and malignant tumors, and their processing data are all B-mode images which are obtained by ultrasound radio frequency signals, the existing CAD systems still need further researches and improvements. In view of this, we present a new method for distinguishing the grades of breast tumors based on the original ultrasound radio frequency signals which have richer tumor lesion information compared to B-mode images. First, we utilize the multi-scale geometric characteristic of Shearlet transformation to extract the multi-scale and multi-directional features of ultrasound RF signal. Second, multi-scale directional binary pattern (MDBP) is designed to code the texture information of high-frequency Shearlet features in different directions and different scales, which can not only reduce the dimension of Shearlet features but also preserve the sufficient discriminated information of breast tumors for the subsequent grade detection. At last, we draw on the feature difference between different grades of breast tumors to put forward a cascade binary tree SVM classifier, which not only overcome the problem of unbalance samples but also conform to the diagnosis rule of sonographer. Extensive experiments on 928 breast ultrasound RF signals collected from the hospital demonstrate the effectiveness of the proposed method and its precision, sensitivity, specificity, PPV, NPV and MCC are 89.29%, 75.62%, 94.54%, 97%, 98.3% and 81.01%, respectively. A point worth emphasizing that the higher values of PPV and NPV further show that the diagnosis results of the proposed method are close to the biopsy gold standard.

Citation: Tong Y, Yan Y. The grade classification algorithm of breast tumor based on ultrasound RF signals[J]. *Opto-Electronic Engineering*, 2019, 46(1): 180368

Supported by National Natural Science Foundation of China (61703201), NSF of Jiangsu Province (BK20170765), and NIT fund for Young Scholar (CKJB201602)

* E-mail: yanyucan@126.com