

# 基于多尺度多色域特征融合的乳腺癌图像分类

张莉<sup>1</sup>, 张成<sup>1</sup>, 郝岩<sup>2</sup>, 程蓉<sup>1</sup>, 白艳萍<sup>1</sup>

(1. 中北大学理学院, 山西太原 030051;

2. 中北大学信息与通信工程学院, 山西太原 030051)

**摘要:**乳腺癌是世界上女性发病率最高的癌症,而组织病理图像是鉴定乳腺癌的“黄金标准”。为了实现对乳腺癌组织病理图像的精确分类,提出了一种基于多尺度多色域特征融合的乳腺癌组织病理图像分类方法。此方法能实现良、恶性病理图像的有效分类。首先进行多色域特征提取,将病理图像从RGB空间转换到HSV空间,分别提取H、S、V三个色彩分量的9维颜色矩特征和24维灰度共生矩阵特征(GLCM);其次进行多尺度特征提取,利用Haar两层小波分解提取病理图像的高频分量(水平、垂直、对角),共得高频分量的48维灰度共生矩阵特征。将最终形成的81维特征向量输入到不同训练集训练所生成的7类支持向量机(SVM)中进行分类,将分类结果采取多数投票策略,获得最终识别准确率。通过BrecaHis公开数据集的实验表明,4个放大倍数图像的分类准确率分别达到约95.31%、94.34%、93.07%和91.94%。

**关键词:**乳腺癌;颜色矩;灰度共生矩阵;Haar小波;支持向量机;多数投票策略

中图分类号:TP391.4

文献标识码:A

文章编号:1673-629X(2022)04-0175-06

doi:10.3969/j.issn.1673-629X.2022.04.030

## Breast Cancer Histopathological Image Classification Based on Multi-scale and Multi-gamut Feature Fusion

ZHANG Li<sup>1</sup>, ZHANG Cheng<sup>1</sup>, HAO Yan<sup>2</sup>, CHENG Rong<sup>1</sup>, BAI Yan-ping<sup>1</sup>

(1. School of Science, North University of China, Taiyuan 030051, China;

2. School of Information and Communication Engineering, North University of China, Taiyuan 030051, China)

**Abstract:** Breast cancer is the most frequent cancer in women in the world, and histopathological images are the "gold standard" for identifying breast cancer. To achieve precise classification of breast cancer histopathological images, a method for breast cancer image classification based on multi-scale multi-color feature fusion is proposed. Such method can effectively classify benign and malignant pathological images. First, multi-color domain feature extraction was performed to transform the pathological images from RGB space to HSV space, and 9-dimensional color moment features and 24 dimensional gray level co-occurrence matrix features (GLCM) of the three color components of H, S, and V were extracted, respectively. Next, multi-scale feature extraction was performed, high-frequency components (horizontal, vertical, diagonal) of breast cancer pathological images were extracted using Haar two-layer wavelet decomposition, and 48 dimensional gray level co-occurrence matrix features with high-frequency components were obtained. The final formed 81 dimensional feature vector was input into the 7-class support vector machine (SVM) generated by training with different training sets for classification, and the classification results were taken a majority voting strategy to obtain the final identification accuracy. The classification accuracy of the four magnification images reached approximately 95.31%, 94.34%, 93.07%, 91.94%, respectively, as determined by the results of experiments with BrecaHis publicly available datasets.

**Key words:** breast cancer; color moments; gray level co-occurrence matrix; Haar wavelet; support vector machine; majority voting strategy

## 0 引言

根据世界卫生组织的数据,乳腺癌是世界上第二大常见的癌症,同时也是女性中发病率最高的癌症,严

重危害着女性同志的身心健康<sup>[1]</sup>。随着乳腺癌患者年轻化趋势的不断增强,患者的人数也不断增加,因此迫切需要借助计算机辅助诊断系统建立针对乳腺癌病理

收稿日期:2021-05-10

修回日期:2021-09-13

基金项目:国家自然科学基金项目(61774137);山西省重点研发计划项目(201903D121156);山西省自然科学基金(201801D121026);山西省基础研究计划资助项目(202103021223189)

作者简介:张莉(1997-),女,硕士研究生,从事现代优化算法、神经网络在组合优化中的应用研究;白艳萍,教授,博导,从事神经网络在函数逼近与预测中的应用、神经网络在组合优化中的应用、数据处理与数据融合、非线性动力学的数学建模与研究。

图像的分类方法,缓解大量病理图像给医生带来的负担。

目前,针对乳腺癌组织病理图像识别主要有两种方法:一种是基于手工提取特征结合机器学习的算法,另一种是基于深度学习的算法。虽然第二种算法是以卷积层为核心的网络结构,能够更有效地实现特征学习过程,但第一种算法耗费时间更少、质量更高,也能获得良好的识别效果。因此该文主要对第一种算法进行研究。Spanhol 等<sup>[2]</sup>公布了 BreaKHis 乳腺癌数据集并研究了结合传统纹理特征与 SVM 等分类器的分类效果。Gupta 等<sup>[3]</sup>运用了 Gabor 特征、复杂小波特征和对立色局部二值模式等多特征融合以及利用多数投票策略将支持向量机、最近邻分类器、决策树等分类器进行集成,实验准确率达到了 87%。Wang 等<sup>[4]</sup>通过融合 4 个形状特征和 138 个颜色特征,实现了对乳腺癌图像的分类。张鹏飞等<sup>[5]</sup>通过融合形状、灰度以及纹理特征构造了三个不同的样本空间,空间中构造 SVM 分类器,采用相对多数投票法进行集成输出结论,实现对肺癌图像的分类。Kausar 等<sup>[6]</sup>利用卷积神经网络对 Haar 小波两层分解后的乳腺癌图像进行深度特征提取,实现了对乳腺癌组织病理图像的分类。

特征提取与融合是实现图像分类的重要途径,它保留了图像的大量信息,对乳腺癌组织病理图像的分类也具有非常重要的意义。在计算机辅助诊断阶段,常用的分类器为 SVM,但单个 SVM 不能精准判别对象类别,降低了稳定性和泛化性能,故将 SVM 作为成员分类器,进行 SVM 集成以提高泛化性能。本研究基于多尺度多色域特征融合的性质,丰富了模型中的特征信息,有效融合了乳腺癌组织病理图像颜色以及纹理特征,运用集成 SVM 模型进行分类识别。

## 1 颜色空间转换

对于彩色图像的识别,颜色空间的选择是一个至关重要的环节。在 RGB 彩色空间中,所有颜色通过  $R$  (红色)、 $G$  (绿色)、 $B$  (蓝色) 三个分量组合形成。这种方式与人类视觉感知并不贴近。所以需要将 RGB 空间转换到 HSV 空间。HSV 空间是一个均匀的六角锥体模型,它是通过  $H$  (色调)、 $S$  (饱和度)、 $V$  (明度) 表示颜色的。其模型的三个坐标是独立的,能把 HSV 空间各个分量的变化表现得很清晰<sup>[7]</sup>。转换公式如下:

$$H = \frac{\arccos\left[\frac{(R - G) + R}{2\sqrt{(R - G)^2 + (R - B)(G - B)}}\right]}{2\pi - \arccos\left[\frac{(R - G) + R}{2\sqrt{(R - G)^2 + (R - B)(G - B)}}\right]} \quad (1)$$

$$S = \frac{\max(R, G, B) - \min(R, G, B)}{\max(R + G + B)} \quad (2)$$

$$V = \frac{\max(R, G, B)}{255} \quad (3)$$

式中,  $\arccos$  为反三角函数中的反余弦;  $R, G, B$  分别为红色、绿色、蓝色的像素值;  $\max(R, G, B)$  为  $R, G, B$  三个分量的最大值;  $\min(R, G, B)$  为  $R, G, B$  三个分量的最小值。

## 2 颜色矩特征提取

颜色矩是一种以数字为基础的方法,通过计算矩来描述颜色的分布。颜色的主要信息一般集中在低阶矩中,因此常采用一阶矩(均值)、二阶矩(方差)、三阶矩(偏度)来描述颜色的分布特征。一阶矩表示颜色分量的均值,即平均强度;二阶矩表示颜色的方差;三阶矩表示颜色的偏斜度,即不均匀性<sup>[8-9]</sup>。计算颜色矩的公式如下:

一阶矩:

$$M_i = \frac{1}{N} \sum_{j=1}^N P_{ij} \quad (4)$$

二阶矩:

$$V_i = \left[ \frac{1}{N} \sum_{j=1}^N (P_{ij} - M_i)^2 \right]^{\frac{1}{2}} \quad (5)$$

三阶矩:

$$S_i = \left[ \frac{1}{N} \sum_{j=1}^N (P_{ij} - M_i)^3 \right]^{\frac{1}{3}} \quad (6)$$

式中,  $i$  为颜色分量,  $i = 1$  为  $H$  分量,  $i = 2$  为  $S$  分量,  $i = 3$  为  $V$  分量;  $P_{ij}$  为颜色分量  $i$  的像素值为  $j$  的概率值;  $N$  为像素点个数。

根据上述所提到的 HSV 空间的特性,分别提取  $H$ 、 $S$ 、 $V$  颜色分量的三个颜色矩特征,共得九个颜色矩分量<sup>[10]</sup>。选取良性与恶性中具有代表性的病理图像进行颜色矩统计。统计结果见表 1。

由表 1 可以看出,良性与恶性肿瘤图像  $H$ 、 $S$ 、 $V$  三个颜色分量的颜色矩有较大差异,为良、恶性乳腺癌组织病理图像分类提供重大依据。

## 3 Haar 小波的多尺度图像分析

### 3.1 小波变换原理

小波变换是图像处理中一种比较常见的基于傅里叶变换而发展的小波分析方法,是具有多分辨率特点的时频分析方法<sup>[11]</sup>。其基本原理为:利用低通和高通滤波器对原始图像的水平方向和垂直方向分别进行卷积操作,将原始图像分解为低频图像(LL)、水平细节(HH)、垂直细节(HV)、对角细节(HD)这四个低尺度分量。

图 1 为小波分解示意图。

表 1 良性与恶性病理图像 HSV 空间颜色矩特征参数值

类别	H 分量			S 分量			V 分量		
	$M_1$	$V_1$	$S_1$	$M_2$	$V_2$	$S_2$	$M_3$	$V_3$	$S_3$
良性	腺瘤	0.805 3	0.043 9	0.144 1	0.165 0	0.012 1	0.144 6	0.780 0	0.012 8
	纤维腺瘤	0.881 2	0.006 0	0.006 8	0.401 3	0.024 6	0.042 0	0.810 3	0.012 0
	叶状肿瘤	0.878 9	0.053 4	0.116 9	0.183 8	0.018 1	0.116 4	0.799 4	0.012 8
	管状腺瘤	0.850 8	0.008 6	0.022 5	0.449 8	0.013 5	0.045 2	0.845 1	0.007 1
恶性	导管癌	0.800 8	2.006 9	0.020 0	0.449 8	3.273 1	0.041 0	0.845 1	1.919 7
	小叶癌	0.836 1	3.051 8	0.066 3	0.133 4	4.677 6	0.059 5	0.822 8	2.915 5
	粘液癌	0.841 3	4.710 8	0.128 1	0.199 8	7.059 4	0.128 6	0.814 6	4.167 7
	乳突癌	0.720 6	6.864 9	0.137 1	0.189 8	11.354 8	0.139 0	0.745 5	7.263 4

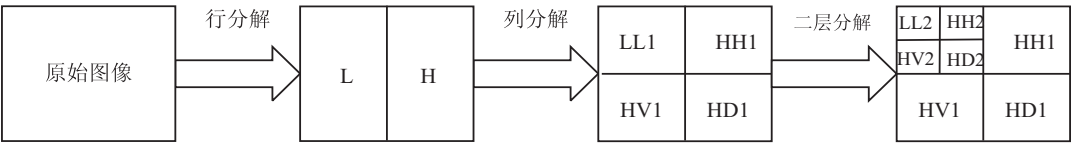


图 1 图像二层小波分解示意图

3.2 乳腺癌组织病理图像多尺度分析

由图 1 可知,随着分解层数的增加,图像尺寸随之减小,因此分解层数不宜太高,利用 Haar 小波对乳腺癌病理图像进行二层小波分解。每一次分解都可以形成三个高频分量 (HH、HV、HD) 和一个低频分量 (LL)。高频分量表示图像的边缘和噪声。低频分量

表示图像的近似,它保留了原始图像 90% 以上的能量。在进行下一层分解时,仅对分解得到的低频图像进行。

图 2 显示了不同层数下的小波分解所得乳腺癌病理图像。

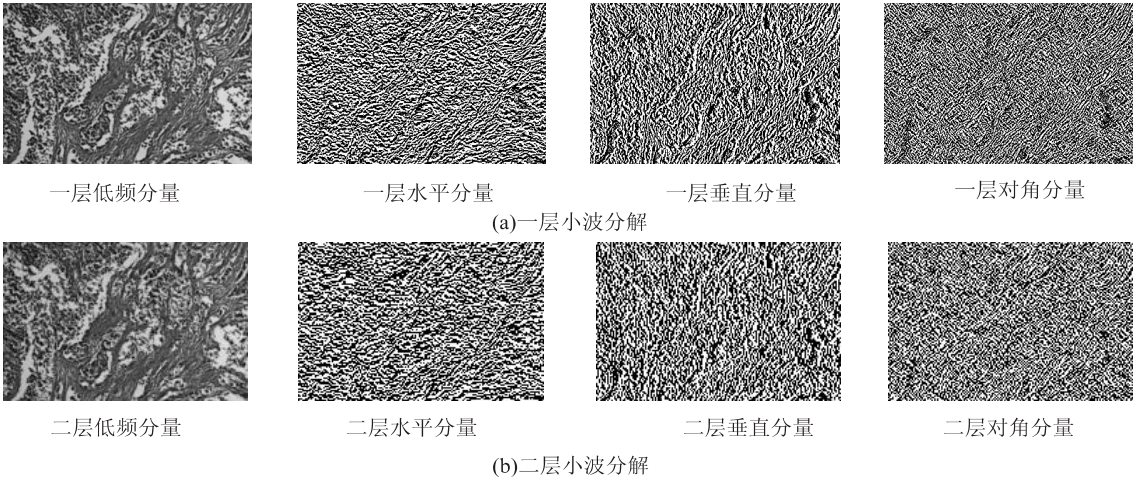


图 2 Haar 小波分解示意图

由于低频信息中包含了原始乳腺癌病理图像大部分的能量,高频信息中较好地保留了低频信息中丢失的细节信息,因此,在本实验中选用经过两次小波分解所得到的低频分量图像进行多尺度灰度共生矩阵特征提取<sup>[6]</sup>。

4 灰度共生矩阵

灰度共生矩阵是在 20 世纪 70 年代由 Haralick<sup>[12]</sup>首次提出的,自此被广泛应用于纹理特征提取。它是通过研究具有一定距离像素对之间的空间相关性来表

述图像纹理特征。由于灰度共生矩阵计算量较大,一般不直接应用,而是在此基础上采用二次统计量描述纹理信息。

给定一幅数字图像  $I(x,y)$ , 其灰度级为  $L$ , 则灰度共生矩阵  $G(x,y)$  为  $I(x,y)$  中沿方向  $\theta$ , 间隔为  $d$  的两个像素点对  $(i,j)$  的统计数目  $n$ 。 $G(x,y)$  的尺寸为  $L \times L$ ,  $n$  为灰度共生矩阵在  $G(i,j)$  的像素值。

乳腺癌组织病理图像多尺度灰度共生矩阵求取:

在早期的研究中,乳腺癌病理图像主要是在单一尺度下进行灰度共生矩阵特征提取,但在不同尺度下,



纹理层次信息呈现出不同的特征,因此单一尺度不能很好地描述乳腺癌纹理特征。于是该文将多尺度的思想与乳腺癌病理图像融合,在多尺度下提取上述 8 个灰度共生矩阵纹理特征,即两层小波分解得到的高频分量、颜色转换后的  $H$  分量、 $S$  分量、 $V$  分量。

对于乳腺癌组织病理图像灰度共生矩阵特征提取,为保留图像的最大信息,选择乳腺癌图像灰度级  $L=256$ ,步长  $d=1$ ,方向参数选取  $0^\circ, 45^\circ, 90^\circ, 135^\circ$  这四个方向特征值的平均值,从而大大减少了方向对特征参数的影响<sup>[13]</sup>。

该文使用了 8 种灰度共生矩阵特征。分别是二阶矩、对比度、相关性、逆差距、均值和、熵、和方差与差方差。纹理特征公式如下:

(1) 二阶矩:表示图像纹理的粗细程度。纹理越粗,值越大。

$$F_1 = \sum_{i=1}^L \sum_{j=1}^L P_{ij}^2 \quad (7)$$

(2) 对比度:表示图像的清晰程度以及纹理的变化程度。图像越清晰,纹理越明显,值越大。

$$F_2 = \sum_{i=1}^L \sum_{j=1}^L (i-j)^2 p_{ij} \quad (8)$$

(3) 相关性:表示纹理的非均匀性。哪个方向值大,纹理就指向该方向。

$$F_3 = \sum_{i=1}^L \sum_{j=1}^L \frac{(i-\mu_x)(j-\mu_y)}{\sigma_x \sigma_y} P_{ij} \quad (9)$$

(4) 逆差距:度量图像局部灰度均衡性。

$$F_4 = \sum_{i=1}^L \sum_{j=1}^L \frac{P_{ij}}{1 + |i-j|^2} \quad (10)$$

(5) 均值和。

$$F_5 = \sum_{K=2}^{2L} K \times P_{x+y}(K) \quad (11)$$

(6) 熵:表示纹理复杂度。纹理越复杂,值越大。

$$F_6 = - \sum_{i=1}^L \sum_{j=1}^L P_{ij} \log_2 P_{ij} \quad (12)$$

(7) 和方差:表示纹理变化的快慢及其周期性大小。值越大,纹理周期越大。

$$F_7 = \sum_{K=2}^{2L} (K - F_5)^2 \times P_{x+y}(K) \quad (13)$$

(8) 差方差:表示邻近像素灰度级差异。值越大,对比度越强烈。

$$F_8 = - \sum_{K=0}^{L-1} \left\{ \left[ K - \sum_{i=0}^{L-1} i \times P_{x-y}(i) \right]^2 P_{x-y}(K) \right\} \quad (14)$$

## 5 乳腺癌组织病理图像特征提取与分析

针对乳腺癌组织病理图像的分类,提出一种多尺度与多色域特征融合的特征提取方法。首先对病理图

像进行多色域处理,提取 HSV 空间的三个色彩分量,分别对三个色彩分量提取 9 维颜色矩特征以及 24 维灰度共生矩阵纹理特征。其次利用 Haar 小波对病理图像进行两层分解,提取高频分量的 48 维灰度共生矩阵特征。最后将多尺度特征与多色域特征串联融合。

具体操作步骤如下:

(1) 对原始乳腺癌病理图像进行两层 Haar 小波分解得到 6 个高频分量,分别记为  $[HH_{n1}, HV_{n1}, HD_{n1}, HH_{n2}, HV_{n2}, HD_{n2}]$ 。 $n$  为乳腺癌病理图像个数。

(2) 将原始彩色图像转换到 HSV 色彩空间,并提取  $H$ 、 $S$ 、 $V$  三个色彩分量,分别记为  $[H_n, S_n, V_n]$ 。

(3) 对步骤 1、2 所得的 9 类图像分别提取其灰度共生矩阵的 8 个特征参数。以第一层小波分解所得高频水平分量为例,记为:

$$[HH_{n11}, HH_{n12}, HH_{n13}, HH_{n14}, HH_{n15}, HH_{n16}, HH_{n17}, HH_{n18}]。$$

(4) 重复上述步骤,获取 9 个分量图像的 8 维灰度共生矩阵纹理特征,并将其串联融合为纹理特征矩阵  $F_1$ 。记为:

$$\begin{pmatrix} HH_{111} & \cdots & HH_{118} & HV_{111} & \cdots & HV_{118} & \cdots & V_{11} & \cdots & V_{18} \\ HH_{211} & \cdots & HH_{218} & HV_{211} & \cdots & HV_{218} & \cdots & V_{21} & \cdots & V_{28} \\ \vdots & & \vdots & & \ddots & & & \vdots & & \vdots \\ HH_{n11} & \cdots & HH_{n18} & HV_{n11} & \cdots & HV_{n18} & \cdots & V_{n1} & \cdots & V_{n8} \end{pmatrix} \quad (15)$$

得到一个  $n \times 72$  维的特征矩阵。

(5) 对  $H$ 、 $S$ 、 $V$  三个分量分别提取颜色矩特征。以  $H$  分量为例,记为  $[H_{n9}, H_{n10}, H_{n11}]$ 。

(6) 重复上述步骤,获取三个分量子图的颜色矩特征,并将其串联融合成颜色特征矩阵  $F_2$ 。

$$\begin{pmatrix} H_{19} & H_{110} & H_{111} & S_{19} & S_{110} & S_{111} & V_{19} & V_{110} & V_{111} \\ H_{29} & H_{210} & H_{211} & S_{29} & S_{210} & S_{211} & V_{29} & V_{210} & V_{211} \\ \vdots & & \ddots & & \vdots & & \vdots & & \vdots \\ H_{n9} & H_{n10} & H_{n11} & S_{n9} & S_{n10} & S_{n11} & V_{n9} & V_{n10} & V_{n11} \end{pmatrix} \quad (16)$$

得到一个  $n \times 9$  的特征矩阵。

(7) 将纹理特征矩阵与颜色特征矩阵串联融合得到一个  $n \times 81$  的特征矩阵。

## 6 乳腺癌组织病理图像分类识别

### 6.1 数据集

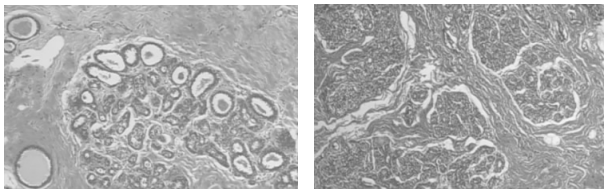
实验选用的是由 Spanhol 发布的 BreaKHis 数据集。该数据集中包括了从 82 名患者中所获取的 7 909 张乳腺癌显微活检图像,其中良性有 2 480 张,恶性有 5 429 张。包含了 4 种放大倍数 (40x, 100x, 200x, 400x),尺寸大小为  $700 \times 460$  像素。

表 2 显示了乳腺癌数据集样本分布。

图 3 显示了 40x 放大倍数下乳腺癌样本图像。

表2 乳腺癌图像样本数据

放大倍数	乳腺癌图像样本数据		
	良性 (Benign)	恶性 (Malignant)	总计
40x	625	1 370	1 995
100x	644	1 437	2 081
200x	623	1 390	2 013
400x	588	1 232	1 820
总数	2 480	5 429	7 909
病人人数	24	58	82



(a) 良性乳腺肿瘤图像

(b) 恶性乳腺肿瘤图像

图3 乳腺癌良性、恶性组织病理图像样本

## 6.2 基于支持向量机的多数投票策略

### 6.2.1 支持向量机理论

由于对乳腺癌组织病理图像进行二分类(良性、恶性),于是主要描述二分类 SVM。分类思想是将低维空间中线性不可分的数据样本映射到高维空间使其线性可分。并在高维空间中找到一个最优超平面,使得不同类别数据点之间的距离最大化<sup>[14-15]</sup>。

设  $n$  个样本集  $\{(x_i, y_i) | i = 1, 2, \dots, n\}$ ,  $x_i \in R^n$ ,  $y_i \in \{-1, +1\}$ ,  $x_i$  为训练样本点,  $y_i$  为其对应标签。使  $y = +1$  类与  $y = -1$  类数据点的间隔最大的超平面如下:

$$\omega^T x_i - b = 0, \quad i = 1, 2, \dots, n \quad (17)$$

为了保证数据点  $x_i$  所允许的偏差最小,引入松

弛变量  $\xi_i \geq 0$ , 惩罚因子  $c$ , 得到了如下目标函数:

$$\min \frac{1}{2} \|\omega\|^2 + c \sum_{i=1}^n \xi_i \quad (18)$$

为处理非线性数据,选用 RBF 核函数对 SVM 进行推广, RBF 核函数公式如下:

$$K(x, x_i) = \exp(-g \| (x - x_i)^2 \|) \quad (19)$$

其中,  $g$  为核函数半径。

$$g = 1/2\sigma^2 \quad (20)$$

根据给定的 RBF 核函数, SVM 分类器表示为:

$$F(x) = \text{sgn}(f(x)) \quad (21)$$

$$f(x) = \sum_{i=1}^l \alpha_i y_i K(x_i, x) + b \quad (22)$$

其中,  $f(x)$  为 SVM 分类超平面决策函数。

### 6.2.2 集成 SVM 的多数投票策略

以数据集中 40x 放大倍数的 1 995 张图像为例, 随机选取 70% 训练集 (Trainx0) 和 30% 测试集 (Testx)。将 Trainx0 的 1 395 张图像平均划分成三份训练集 (Trainx1、Trainx2、Trainx3), 将这三种训练集两两组合又形成三种训练集 (Trainx4 = [Trainx1; Trainx2]、Trainx5 = [Trainx1; Trainx3]、Trainx6 = [Trainx2; Trainx3])。对上述 7 种训练集训练形成 7 种 SVM 分类器, 并用 Testx 进行测试。

为了融合不同训练样本所训练分类器的输出结果, 选择了多数投票策略。多数投票策略是一种决策规则, 其中新的乳腺癌组织病理图像样本的类别是根据不同 SVM 分类器分类所获得的标签通过投票 (标签) 来决定的, 规定收到投票最多的类用作测试样本的最终标签<sup>[3,5]</sup>。

图4显示了乳腺癌组织病理图像分类的总过程。

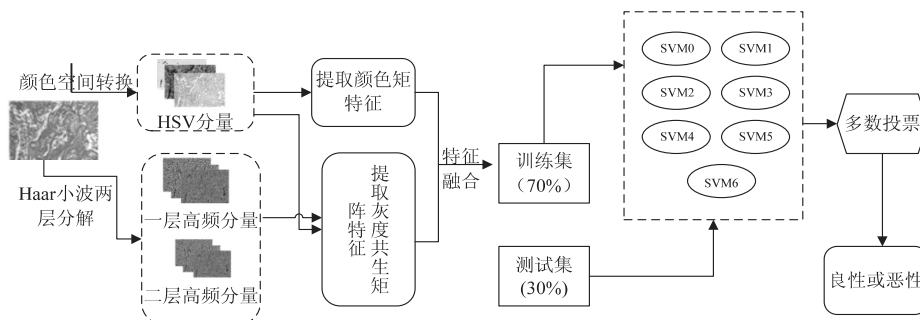


图4 图像分类的总过程

## 6.3 实验结果

为了验证多尺度与多色域特征提取与融合算法的分类效果, 对比了传统的单一尺度灰度共生矩阵与其他特征提取方法对实验结果的影响。为了更全面更客观地对分类效果进行评价, 针对不同的特征提取方法, 根据指标体系(准确率 (Acc)、敏感性 (Sen)、特异性 (Spe)) 进行评价<sup>[16]</sup>。指标体系的公式如下:

$$\text{Acc} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{TN} + \text{FP}} \quad (23)$$

$$\text{Sen} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (24)$$

$$\text{Spe} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (25)$$

其中, TP 为正确分类的良性; TN 为正确分类的恶性; FP 为错误分类的良性; FN 为错误分类的恶性。

表 3 显示了不同特征提取方法对实验结果的影响。实验结果表明,本研究方法优于其他传统的特征

提取方法,并且针对不同放大倍数下的乳腺癌组织病理图像,该方法取得了较好的分类结果。

表 3 不同特征提取方法实验结果对比 %

方法	40x			100x			200x			400x		
	Acc	Sen	Spe	Acc	Sen	Spe	Acc	Sen	Spe	Acc	Sen	Spe
颜色矩	85.94	81.92	87.39	84.57	81.36	86.13	84.12	82.36	85.79	83.98	81.19	82.67
GLCM	72.30	70.00	72.00	70.49	71.46	69.91	70.15	72.36	69.43	69.43	71.38	72.16
多尺度多色域 GLCM	91.53	92.89	91.02	89.91	83.82	92.14	88.67	83.26	91.57	89.41	87.32	82.19
颜色矩+ 多尺度多色域 GLCM	94.83	94.53	93.00	93.63	91.94	94.98	92.57	92.16	93.51	91.31	92.03	91.97

为了探讨基于 SVM 分类器的多数投票策略的分类效果,针对准确率(Acc),与 7 种 SVM 分类器的分类结果作对比(见表 4)。实验结果表明,多数投票策略集成了单个 SVM 的良好性能,提高了泛化能力。

表 4 基于 SVM 分类器的多数投票

策略的分类结果 %

方法	40x	100x	200x	400x
SVM0	94.83	93.63	92.57	91.31
SVM1	91.50	90.26	90.65	89.97
SVM2	90.00	89.73	88.73	88.46
SVM3	88.30	89.25	88.79	87.49
SVM4	93.67	93.12	92.13	92.81
SVM5	93.83	92.89	92.05	90.06
SVM6	94.83	93.42	92.69	90.46
多数投票策略	95.31	94.34	93.07	91.94

将文中方法与采用同一数据集所使用的方法进行比较,见表 5,该方法识别率更高,鲁棒性和泛化性能更好。

表 5 文中方法与其他方法识别率比较 %

方法	40x	100x	200x	400x
Spanhol et. al. <sup>[2]</sup>	81.6	79.9	85.1	82.3
Gupta V et al. <sup>[3]</sup>	87.2	88.22	88.89	85.82
P. J. Sudharshan et al. <sup>[17]</sup>	87.8	85.6	80.8	82.9
Hou Y <sup>[18]</sup>	90.89	90.99	91.00	91.00
文中方法	95.31	94.34	93.07	91.94

## 7 结束语

提出一种多尺度多色域特征融合的乳腺癌组织病理图像分类算法。分别介绍了颜色空间转换、颜色矩、Haar 小波分析、灰度共生矩阵等原理。利用集成 SVM 的多数投票策略对该算法获取的 BreaKHis 乳腺癌数据集特征进行分类识别。实验结果表明,该算法在乳腺癌图像分类识别中的准确率均优于单一尺度以及传统的 GLCM 算法,对于不同放大倍数下的乳腺癌组织

病理图像也取得了较好的识别准确率。通过基于 SVM 的多数投票策略,集成了单一 SVM 的良好性能,提升了实验准确率。与其他乳腺组织病理图像分类方法进行对比分析,证明了该分类方法优于其他传统的分类方法。在未来的研究工作中,可以研究多分类任务的乳腺癌组织病理图像分类方法。

## 参考文献:

- [1] MOURA D C, LÓPEZ M A G. An evaluation of image descriptors combined with clinical data for breast cancer diagnosis[J]. International Journal of Computer Assisted Radiology & Surgery, 2013, 8(4): 561-574.
- [2] SPANHOL F A, OLIVEIRA L S, PETITJEAN C, et al. A dataset for breast cancer histopathological image classification[J]. IEEE Transactions on Biomedical Engineering, 2016, 63(7): 1455-1462.
- [3] GUPTA V, BHAVSAR A. Breast cancer histopathological image classification: is magnification important? [C]//Proceedings of IEEE conference on computer vision and pattern recognition workshops. Honolulu, USA: IEEE, 2017: 769-776.
- [4] WANG P, HU X, LI Y, et al. Automatic cell nuclei segmentation and classification of breast cancer histopathology images[J]. Signal Processing, 2016, 122: 1-13.
- [5] 张飞飞, 周涛, 陆惠玲, 等. 基于集成 VPRS-RUGGA 支持向量机的多模态肺部肿瘤计算机辅助诊断模型[J]. 生物医学工程研究, 2019, 38(1): 48-53.
- [6] KAUSAR T, WANG M J, IDREES M, et al. HWDCCN: multi-class recognition in breast histopathology with Haar wavelet decomposed image based convolution neural network[J]. Biocybernetics and Biomedical Engineering, 2019, 39(4): 967-982.
- [7] 邓秋菊, 王宁, 钟怡. 基于 HSV 空间的多相隐写图像阈值分割方法仿真[J]. 计算机仿真, 2020, 37(9): 167-170.
- [8] MALIK F, BAHARUDIN B. Features analysis for content-based image retrieval based on color moments[J]. Research

(下转第 185 页)