第46卷 第12期 湖南大学学报(自然科学版) Vol.46,No.12
2019年12月 Journal of Hunan University(Natural Sciences) Dec. 2019

文章编号:1674-2974(2019)12-0001-08 DOI:10.16339/j.cnki.hdxbzkb.2019.12.001

宫颈细胞图像的特征选择与分类识别算法研究

董娜†,赵丽,常建芳,吴爱国
(天津大学 电气自动化与信息工程学院,天津 300072)

摘 要:为了提高宫颈细胞识别速度,以最少的特征数量获得最高的识别准确率,运用分类与回归树算法(Classification and Regression Trees,CART)进行特征的选择,并采用粒子群算法(Particle Swarm Optimization,PSO)对分类器支持向量机(Support Vector Machine,SVM)进行优化,形成了PSO-SVM分类算法对细胞进行分类.使用Herlev数据集对文中提出的算法进行验证.通过CART特征选择方法,成功地从20个特征中提取出9个更具代表性的特征,并且二分类和七分类的准确率均达到99%以上.并引入其他几种宫颈癌细胞的分类识别算法进行仿真比较,结果表明,本文算法在特征数目较少的情况下识别准确率依然具有明显优势,从而验证了该算法的有效性.所述方法有效降低了人工特征选择的难度,在减少了识别用时的情况下,依然保证了细胞的识别准确率与之前几乎无异,为宫颈癌疾病诊断提供了一套有效的方法框架.

关键词:特征提取;特征选择;CART;PSO-SVM;宫颈细胞检测
中图分类号:TP18 **文献标志码**:A

Research on Feature Selection and Classification Recognition Algorithm of Cervical Cell Image

DONG Na†,ZHAO Li,CHANG Jianfang,WU Aiguo
(School of Electrical Automation and Information Engineering,Tianjin University,Tianjin 300072,China)

Abstract:In order to improve the recognition speed of cervical cell and obtain the highest recognition accuracy with the least number of features,this paper innovatively uses the Classification and Regression Trees (CART) algorithm to select features,and then the Particle Swarm Optimization (PSO) algorithm is used to optimize the Support Vector Machine (SVM). Therefore,the PSO-SVM classification algorithm is formed to classify the cells. This paper uses the Herlev dataset to verify the validity of the proposed algorithm. Through the CART feature selection method, 9 representative features are successfully extracted from 20 features,and the accuracy of two classifications and seven classifications are above 99%. Further,this paper introduces several other classification and recognition algorithms of cervical cancer cells for simulation comparison. It can be founds that the recognition accuracy of this algorithm is obviously superior when the number of features is small,which indicates that the proposed algorithm is effective. The method effectively reduces the difficulty of artificial feature selection,and ensures that the recognition accuracy of the cells is almost the same as before when the recognition time is reduced. Thus,the proposed algorithm provides an effective method for the diagnosis of cervical cancer diseases.

Key words:feature extraction;feature selection;CART;PSO-SVM;cervical cell detection

* 收稿日期:2019-01-27
基金项目:国家自然科学基金资助项目(61773282),National Natural Science Foundation of China(61773282)
作者简介:董娜(1983—),女,天津人,天津大学副教授
† 通讯联系人,E-mail:dongna@tju.edu.cn

根据全球的病理报告显示,宫颈癌的发病率逐年增加,每年的新增患者达到 50 多万. 一些发达国家定期对妇女进行宫颈筛查,有效地降低了宫颈癌的发病率[1].目前最传统的方法是通过人工阅片检查是否患有宫颈癌,然而这种方法识别效率低,准确率不高. 即使使用图像处理的方法对宫颈细胞进行识别,也需要提取大量的细胞特征,冗余的特征不仅会造成识别速度缓慢,也容易对识别效果产生影响.

分割是细胞识别的基础,根据分割结果可以提取细胞的各类特征. 癌变细胞与正常细胞在颜色、形态等方面具有很大差异,细胞病理学专家就是通过观察细胞的颜色、形状、大小等特征对细胞进行判定.

细胞的特征主要是依据细胞病理学专家的诊断经验来提取的. 因此,选择的细胞特征,既要满足医师的诊断经验,又要以计算机理解的方式输入. 文献[2]对宫颈细胞特征提取进行了详细的分析,提取了 87 个细胞特征. 并非所有特征都是有效特征, 特征数量过多或过少,对分类精度都有影响,因此,选出最有用的特征,以最少的特征获得最高的准确率才是关键[3].

宫颈细胞图像识别的最终目的是判定该细胞是否为癌细胞,经过分类器的训练,可以准确地将细胞分为正常细胞和癌细胞. 尽管现有分类器种类较多,但是 SVM 因操作简单,分类准确率高而受到广大学者关注.

为了缓解高维细胞特征导致的欠拟合,并提高宫颈细胞诊断的效率和准确率,本文提出了一种基于 CART 特征选择的宫颈细胞分类方法. 首先提取宫颈细胞的颜色特征、形态特征和纹理特征共 20 个特征,然后进行特征选择,选定 9 个特征进行训练,最后将 SVM 与 PSO 算法结合起来,形成 PSO-SVM 分类器,从而实现宫颈细胞的正确分类.

1 宫颈细胞的特征提取

特征提取是从细胞图片中寻找有效的特征. 样本差异越明显,则有效特征的差异越明显. 有些特征可以通过肉眼观察,如颜色、大小等,有些则需要进行一定的变换,如直方图、纹理等. 一般情况下主要从以下 3 个方面提取特征:

1)颜色特征. 主要指细胞核或细胞质的颜色. 细胞在染色之后,细胞核和细胞质会变成不同的颜色,因此更便于观察. 一般情况下颜色特征的提取主要是基于 RGB 颜色[4-5].此外,HSI 颜色空间也应用相对广泛 [6-8]. RGB 颜色特征主要是提取 Red、Green、Blue 3 个颜色空间上的数值特征,如均值、方差、能量,熵等.

2)形态特征. 此特征的参数提取种类较多,Jantzen 等[9]对宫颈细胞进行仔细地观察后,提出 20 个特征,其中包括面积、周长[10-11]以及圆形度等. 此外,其他特征用来描述细胞,如核质比等[11-12].

3)纹理特征. 在提取细胞的纹理特征时,灰度共生矩阵[13]是常用的方法,该方法提取的特征主要是对比度、能量、熵等. Walker 等[14-15]利用灰度共生矩阵实现了宫颈细胞的分类. 赵晖等[16]用灰度共生矩阵求出 6 个纹理特征,再对其进行归一化,在细胞分类中取得了较好的结果. Plissiti 等[17]采用局部二进制模式特征,来分析宫颈细胞的纹理特征.

本文算法的流程图如图 1 所示.

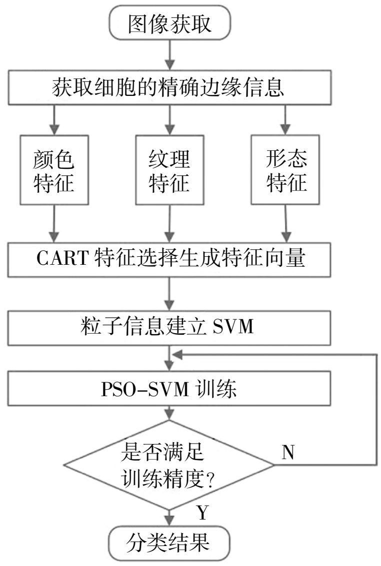


图 1 本文算法流程图

Fig.1 Algorithm flow chart of this paper

综合以上颜色特征、纹理特征和形态特征,本文共选择了 20 个特征,汇总如表 1 所示.

表 1 20 个特征分类

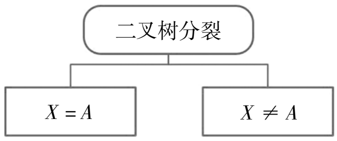
Tab.1 20 feature categories

特征	个数	名称
颜色特征	6	均值、方差、倾斜度、峰态、能量、熵
纹理特征	8	能量、熵、惯性矩、相关性、逆差矩、粗糙度、对比度、方向度
形态特征	6	周长、面积、圆形度、长宽比、矩形度、核质比

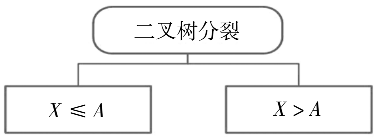
虽然表 1 所含的 20 个特征中有部分特征的名称相同,如颜色特征和纹理特征中的能量和熵,但是两者意义不同,它们分别从不同的角度来定义,因此各自起着一定作用.

2　宫颈细胞的特征选择

特征选择即选择出最具代表性的特征. Liu 等[18]对特征选择做出总结,选择一组最优的特征子集,该特征子集需保证元素个数最少,既降低特征空间维数，又保证细胞识别的准确率. 因此本文选择 CART 算法对数据进行降维处理,其本质是对特征空间进行二元划分，即进行二叉树分裂. 二叉树分裂有标量属性分裂和连续属性分裂. 标量属性是指进行分裂的条件为不等于某个值. 连续属性是指进行分裂的条件为不大于某个值 ,如 取相邻两个属性中间值. 图 2 为标量属性和连续属性进行分裂时的示意图.



(a)标量属性分裂



(b)连续属性分裂

图 2　标量属性分裂和连续属性分裂图

Fig.2　Scalarattribute split and continuous attribute split

为了保证分类效果,需要确定选择哪种特征属性. 为了衡量各个特征的重要性,CART 采用 Gini 指数观察分裂时的不纯度. Gini 指数计算公式如下:

$$\mathrm{Gini}(t)=1-\sum_{k}[p(c_k|t)]^2 \qquad (1)$$

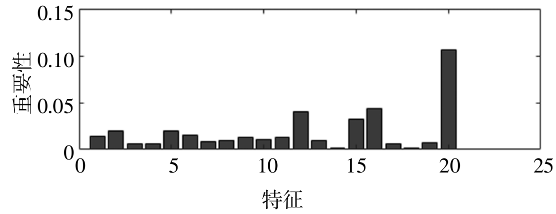
式中:t 为决策树的节点;c_k 为第 k 个样本的个数;p 为概率.

观察该计算公式,发现 Gini 指数的实质是 1 与类别 c_k 的概率平方和的差值,即反映了样本的不确定程度. Gini 指数通过计算特征对每个节点上观测值的异质性的影响,观察特征的重要性. 该值越大表示该特征越重要,因此应该选择具有较大 Gini 指数的特征. 假设父节点对应的样本集合为 D,CART 选择特征 A 分裂为两个子节点,对应集合为 D_{L} 与 D_{R},分裂后的 Gini 指数定义如下:

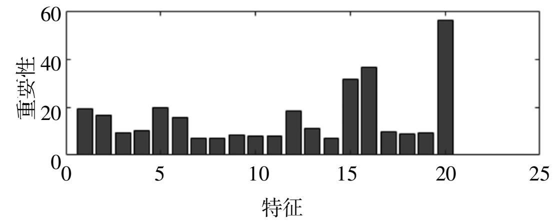
$$G(D,A)=\frac{|D_{\mathrm{L}}|}{D}\mathrm{Gini}(D_{\mathrm{L}})+\frac{|D_{\mathrm{R}}|}{D}\mathrm{Gini}(D_{\mathrm{R}}) \qquad (2)$$

式中:|·|表示样本集合的记录数量.

本文采用该算法将 20 个特征分别进行特征选择,二分类的比较结果如图 3 所示.



(a)Mean decrease in accuracy



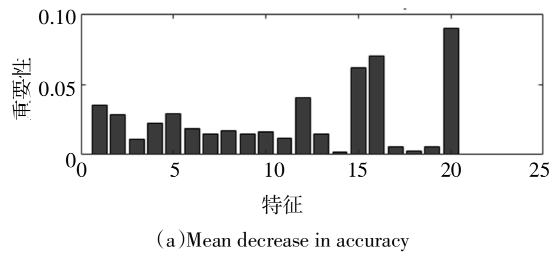
(b)Mean decrease in Gini index

图 3　CART 二分类比较结果

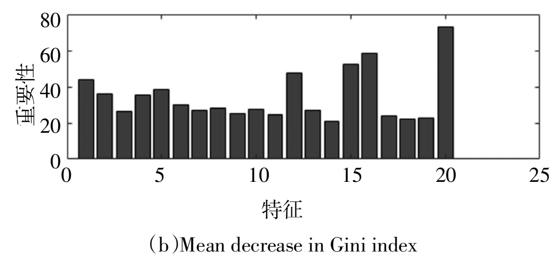
Fig.3　CART two classification comparison results

观察二分类的结果，可以发现一共有两种衡量指标:Mean decrease in accuracy 和 Mean decrease in Gini index. Mean decrease in accuracy 指去掉该特征后实验结果准确率的下降量. 假设把特征 20 去掉，分类准确率降低 11.05%,因此该值越大表示该特征越重要. Mean decrease in Gini index 是计算特征对每个节点观测值的异质性的影响，该值越大也表示该特征越重要.

本文对七分类进行特征选择，比较结果如图 4 所示.



(a)Mean decrease in accuracy



(b)Mean decrease in Gini index

图 4　CART 七分类比较结果

Fig.4　CART seven classification comparison results

观察图 3 和图 4，可以发现七分类的特征性能表现与二分类近乎一致. 对细胞而言无论是二分类还是七分类,正常细胞之间的差异和异常细胞之间的差异是很小的,二分类中正常细胞与异常细胞之间的差异就类似于七分类中正常柱状上皮细胞与轻度癌变细胞之间的差异,因此特征的表现能力相似.

为了对特征的重要性进行排序,以便观察每个特征,本文对 20 个特征的表现能力进行了排序,更为直观的准确率表示如图 5 所示.

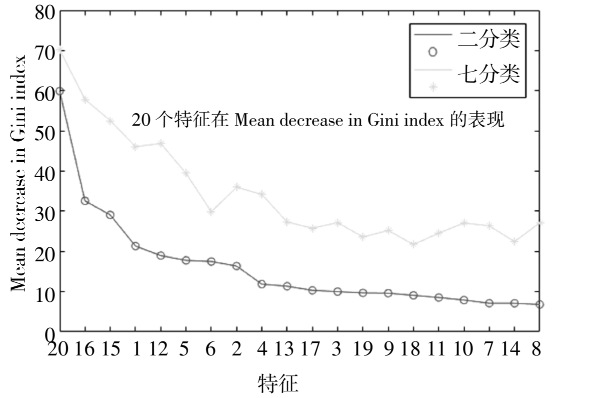


图 5 20 个特征在 Mean decrease in Gini index 的表现

Fig.5 Performance of 20 features in Mean decrease in Gini index

观察图 5 可以发现,前 6 个特征中两种分类情况的 Gini 系数性能表现一致. 七分类中每个特征的 Gini 系数比二分类大，因为随着分类类别数目的增加,类别之间细微的差异就使得某个特征发挥了作用. 另外,七分类中特征 6 的表现能力没有特征 2 和 4 表现好,但是在二分类中其表现能力较好,因此其重要性是可以确定的. 综合考虑每个特征的表现能力,本文选择前 9 个特征送入分类器进行训练.

以下为 20 个特征在 Mean decrease in accuracy 的表现,具体情况如图 6 所示.

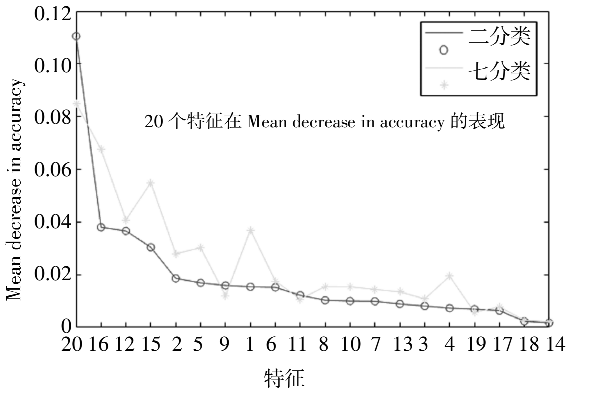


图 6 20 个特征在 Mean decrease in accuracy 的表现

Fig.6 Performance of 20 features in Mean decrease in accuracy

图 6 中,20 个特征在 Mean decrease in accuracy 和 Mean decrease in Gini index 中的表现近乎一致. 虽然个别顺序有所改变，但整体上特征性质改变不大. Gini 系数是 CART 进行特征选择的重要依据,因此本文着重分析该系数的变化. 通过观察图 6 可以发现,特征 20 所占的比重最大,效果较为明显,其次是特征 16，依次类推. 观察特征 4 及之后特征可以发现，这些特征的缩减对 Gini 系数的影响并不大，因此综合两类情况，本文选择特征 1、2、4、5、6、12、15、16、20 共 9 个特征作为选择后的特征，即均值，方差,峰态,能量,熵,粗糙度,周长,面积,核质比,进行后续的分类器训练.

3 PSO-SVM 分类器训练

对于线性不可分的样本,SVM 通过将其映射到一个高维的线性空间,使其线性可分,并且加入了核函数的使用,既减少了计算的复杂性,又减少了维数灾难[19-20]. 虽然传统的 SVM 模型效果不错,但是求解最优化问题依然效果欠佳,如参数的选取. 本文引入了 PSO 算法优化 SVM 的超参数,构建 PSO-SVM 模型,通过全局寻优,寻找更好的分类效果.

PSO 的数学描述[21]为:假设搜索空间是 M 维,粒子数为 n，其中第 i 个粒子的位置表示为 $x_i=(x_{i1},x_{i2},\cdots,x_{iM})$,$i=1,2,\cdots,n$,把 x_i 带入目标函数就可以算出适应度值 fit,根据适应度值的大小可以判断该粒子的好坏. 第 i 个粒子的飞行速度为 $v_i=(v_{i1},v_{i2},\cdots,v_{iM})$,搜索到的最优位置为 $p_i=(p_{i1},p_{i2},\cdots,p_{iM})$,所有粒子群搜到的最优位置为 $p_g=(p_{g1},p_{g2},\cdots,p_{gM})$,当两个最优位置都找到时，每个粒子可以根据更新公式来更新自己的位置和速度.

为了实现全局寻优，需要将本文的数据分为 3 部分(训练集、测试集和验证集). 具体的 PSO-SVM 算法如下：

1)在初始状态下根据经验初步确定惩罚因子 c 和径向基函数参数 σ 的取值范围.

2)PSO 初始化. 初始化算法的参数，一般情况下,搜索空间设置为 2 维,粒子数为 20~50. 随机初始化粒子参数并形成粒子群，随机生成粒子的开始速度，并确定参数 p_{best} 和 g_{best},p_{best} 设置为粒子的当前位置,g_{best} 为所有粒子中最好的粒子的当前位置.

3)加入训练集训练 SVM,用适应度函数计算每个粒子的适应度值 fit，并随时更新第 i 个粒子的最优位置 p_i 和所有粒子的最优位置 p_g,如果某个粒子

当前适应度高于 p_{best},则新的适应度值取代 p_{best},如果所有粒子当前适应度值高于 g_{best},则当前最优位置的适应度值取代 g_{best}. 对 c 和 σ 进行迭代寻优,从而获取最优参数 c 和 σ 的值.

4)观察训练集训练结果是否满足精度,是则继续向下进行;否则转向步骤 3).

5)加入验证集观察训练结果,并计算误差和适应度函数,如果算法达到结束条件,继续向下进行;否则,对粒子的速度和位置进行更新,并返回 3)继续向下进行. 算法的终止条件是达到分类精度或迭代次数.

6)至此,就可保证得到了适应度最优的粒子信息,加入测试集进行测试,输出分类结果即可.

如图 7 所示,为 PSO 训练 SVM 的流程图.

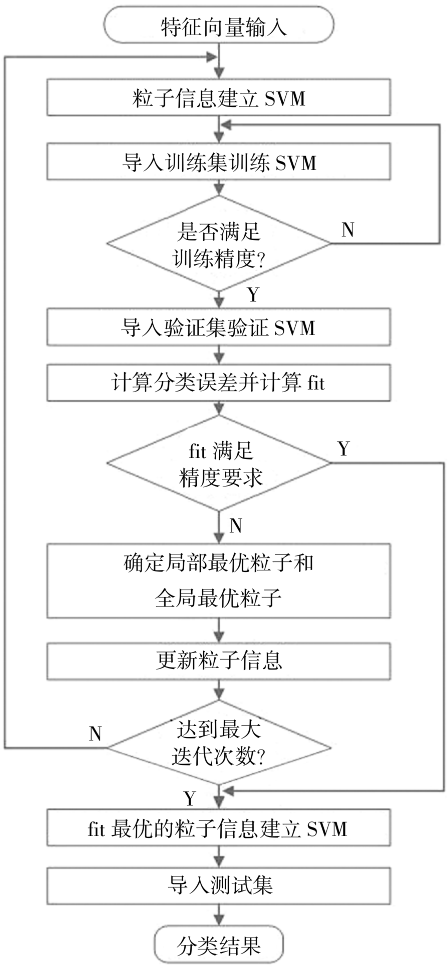


图 7　PSO-SVM 模型

Fig.7　PSO-SVM model

经过以上算法,得到使 SVM 误差最小的参数 c 和 σ 的最优值.

4　宫颈细胞的准确率比较

本文采用十折交叉验证对 917 张 Herlev 数据集进行训练和测试，即每次选择数据集中的 9 份作为训练集,剩下的 1 份作为测试集,共进行 10 次. 其中分别进行了 20 个特征和 9 个特征的分类器训练,训练效果如表 2 所示.

表 2　20 个特征和 9 个特征的训练结果比较

Tab.2　Comparison of the training result of 20 features and 9 features

	类别	准确率/%	识别正确数目/个	运行时间/s
二分类	20 个特征	99.89	916	7.15
	9 个特征	99.78	915	4.46
七分类	20 个特征	99.56	913	22.70
	9 个特征	99.35	911	13.55

在二分类中,20 个特征的识别准确率为 99.89%,9 个特征的识别准确率为 99.78%. 七分类中,20 个特征的识别准确率为 99.56%,9 个特征的识别准确率为 99.35%. 在特征数目减少的过程中，细胞识别的准确率几乎没有发生变化，识别所消耗的时间却得到了有效地降低. 二分类的运行时间减少了 3 s 左右,七分类的识别时间减少了 9 s 左右. 确定了 9 个特征后，将 9 个特征放进 PSO-SVM 进行二分类和七分类的训练，其准确率如表 3 和表 4 所示.

表 3　二分类识别准确率

Tab.3　Two classifications recognition accuracy

类别	总数/个	识别准确数目/个	准确率/%
正常细胞	242	241	99.59
异常细胞	675	674	99.85

表 3 中,1 个正常细胞被识别为异常细胞,1 个异常细胞被识别为正常细胞. 表 4 中,有 4 个正常柱状细胞被识别轻度癌变细胞,2 个轻度病变细胞被识别为正常柱状细胞. 对病人而言,实际意义上的七分类识别错误个数只有 2 个，将正常细胞识别为癌细胞不会对患者造成真正的伤害，因此识别准确率为 99.78%,与二分类几乎一致.

表 4 七分类识别准确率

Tab.4 Seven classifications recognition accuracy

类别	总数/个	识别准确数目/个	准确率/%
正常上皮鳞状细胞	74	74	100
正常中层鳞状细胞	70	70	100
正常柱状细胞	98	94	95.92
轻度病变细胞	182	180	97.8
中度病变细胞	146	146	100
重度病变细胞	197	197	100
原位癌细胞	150	150	100

为了证实所提算法的有效性，本文同时采用了主成分回归分析(PCR)、主成分分析(PCA)-PSO-SVM、核主成分分析(KPCA)-PSO-SVM、最小冗余最大相关(MRMR)[22]-PSO-SVM 和 ReliefF[23]-PSO-SVM 共 5 种方法作为对照. 并且，本文采用均方根误差(Root Mean Square Error,RMSE)、平均绝对误差（Average Absolute Error,AAE）和最大绝对误差(Maximal Absolute Error,MAE)、准确率(Accuracy,ACC)、敏感度(Sensitivity,SEN)和特异度(Specificity,SPE）作为模型性能评价指标，其中 RMSE 和 ACC 为主要的评价指标.

$$\mathrm{RMSE}=\sqrt{\frac{1}{N}\sum_{1}^{N}(\hat{y}_i-y_i)^2} \tag{3}$$

$$\mathrm{AAE}=\frac{1}{N}\sum_{i=1}^{N}|y_i-\hat{y}_i| \tag{4}$$

$$\mathrm{MAE}=\max(|y_i-\hat{y}_i|) \tag{5}$$

式中:y_i 为第 i 个样本的真实值；$\hat{y}_i$ 为第 i 个样本的估计值;N 为测试集的样本个数.

本文在将特征数降维为 9 个的基础上对各个评价指标进行统计,各个算法表现性能如表 5、表 6 所示.

表 5 二分类中不同测试方法测试误差比较

Tab.5 Comparison of test errors of different test methods in the two classifications

方法	AAE	RMSE	SPE/%	ACC/%	SEN/%	MAE
PCR	0.122 1	0.349 5	83.06	87.79	89.48	1
PCA-PSO-SVM	0.073 1	0.270 3	85.95	92.69	95.11	1
KPCA-PSO-SVM	0.022 9	0.151 3	95.04	97.71	98.67	1
MRMR-PSO-SVM	0.043 6	0.208 9	92.56	95.64	96.74	1
ReliefF-PSO-SVM	0.093 78	0.306 2	88.43	90.62	91.41	1
CART-PSO-SVM	0.002 18	0.046 7	99.59	99.78	99.85	1

表 6 七分类中不同测试方法测试误差比较

Tab.6 Comparison of test errors of different test methods in the seven classifications

方法	AAE	RMSE	SPE/%	ACC/%	SEN/%	MAE
PCR	0.524 7	0.732 6	69.01	82.01	86.67	4
PCA-PSO-SVM	0.235 2	0.624 7	83.06	88.88	90.96	4
KPCA-PSO-SVM	0.055 6	0.325 2	92.98	96.07	97.19	3
MRMR-PSO-SVM	0.112 3	0.461 1	86.78	92.69	94.81	4
ReliefF-PSO-SVM	0.354 2	0.697 3	88.89	85.71	76.86	5
CART-PSO-SVM	0.006 54	0.114 4	99.70	99.35	98.35	1

观察 6 个评价指标可以发现，在宫颈癌细胞的二分类和七分类研究中，本文提出的特征选择方法的性能指标最优，即本文的 CART 模型具有较高的准确性,且各类性能指标均优于对照实验方法. 本文算法的准确率比 KPCA 算法高出 3%左右，比 PCR 高出 30%,灵敏度和特异度也均为最高.

观察表 5 和表 6 可以发现，本文算法的 RMSE 和 AAE 均为最低. 本文的评价指标为均方根误差和平均绝对误差,误差越小,分布越均匀,识别错误的可能性越小,识别越稳定. MAE 为最大绝对误差,在二分类中,如果识别出现错误,MAE 值只能为 1,在七分类中,该值具有不确定性. 该值越大,证明该算法的不稳定性越大,在 ReliefF 算法中 MAE 值为 5,识别中出现了将第一类正常细胞识别为第六类异常细胞的情况. 而本文 MAE 值为 1，即只在第三类正常细胞和第四类异常细胞之间识别出现误差，但仍保证了较高的识别精度，因此也体现了本文识别算法的优异性.

为了更加全面地验证本文算法的有效性，本文列举了其它几种算法的特征选择数目以及分类准确率,具体的比较结果如表 7 所示.

在其它 6 种算法中,文献[26]识别准确率为 98.98%,特征数为 11 个,本文识别算法准确率比其高了 1%左右,特征数少用 2 个,效果较为明显. 从文献[9]和[24]中可以看出,特征数多识别效果不一定高. 文献[27]和[28]中虽然选择了相同的 9 种特征,识别准确率却比本文算法低了 6%和 4%,因此特征数目的选择不仅要准确反映细胞的差异，还要具有较高的准确率. 综合 7 种算法的识别准确率和特征数,本文算

法能够以最少的特征数实现最高的准确率,识别效果较好,对于人工智能识别癌细胞的推广具有极大的价值.

表 7　本文算法与其它算法的比较

Tab.7　Comparison of the algorithm and other algorithms in this paper

算法	准确率/%	特征数/个
文献[9]	93.41	20
文献[24]	96.91	24
文献[25]	95.2	11
文献[26]	98.98	11
文献[27]	93.72	9
文献[28]	95.36	9
本文	99.78	9

5 结　论

针对目前宫颈细胞识别过程中出现的特征冗余和准确率低的问题，本文建立了基于 CART 特征选择算法的宫颈细胞分类识别模型,从 20 个特征中选择了 9 个有效的特征，建立了 PSO-SVM 模型进行细胞的分类,使得细胞识别的准确率达到了 99%以上. 为了验证本文算法的有效性，引入了 RMSE、AAE 等评价指标，并与其它分类方法进行比较,结果表明,本文算法不仅提取了更加精准有效的特征,而且实现了较高的识别准确率. 该研究结果有效地提高了二分类和七分类的识别效率,同时对于降低宫颈癌的误诊率有良好的应用价值.

参考文献

[1] FERLAY J,SOERJOMATARAM I,DIKSHIT R,*et al*. Cancer incadence and mortality worldwide:sources,methods and major patterns [J]. International Journal of Cancer,2015,136(5):359—386.

[2] 洪继光,成德信,于光普. 细胞图像的特征描述[J]. 信息与控制,1983,12(2):28—33,45.
HONG J G,CHENG D X,YU G P. Characteristics of cell image [J]. Information and Control,1983,12(2):28—33,45.(In Chinese)

[3] 段洁,胡清华,张灵均,等. 基于邻域粗糙集的多标记分类特征选择算法[J].计算机研究与发展. 2015,52(1):56—65.
DUAN J,HU Q H,ZHANG L J,*et al*. Multi-label classification feature selection algorithm based on neighborhood rough set [J]. Computer Research and Development,2015,52 (1):56—65.(In Chinese)

[4] NUNOBIKI O,SATO M,TANIGUCHI E,*et al*. Color image analysis of cervical neoplasia using RGB computer color specification [J]. Analytical and Quantitative Cytology and Histology,2002,24(5):289—294.

[5] VIJAYASHREE R,RAO K. A semi-automated morphometric assessment of nuclei in pap smears using image[J]. Journal of Evolution of Medical and Dental Sciences,2015,4(53):63—70.

[6] 张勇. 一种有效的白细胞图像彩色空间序贯分割方法[J]. 西安交通大学学报,1998,32(8):52—56.
ZHANG Y. An effective segmentation method for color space of white blood cell image [J]. Journal of Xi′an Jiaotong University,1998,32(8):52—56.(In Chinese)

[7] 花蕾,叶玉坤. 基于知识的肺癌早期细胞诊断系统[J]. 计算机应用研究,2000,17(2):90—92.
HUA L,YE Y K. Knowledge-based early cell diagnosis system for lung cancer [J]. Journal of Computer Applications,2000,17(2):90—92.(In Chinese)

[8] 陆新泉,李宁,陈世福. 形态,颜色特征及神经网络在肺癌细胞识别中的应用研究 [J]. 计算机辅助设计与图形学学报,2001,13(1):87—92.
LU X Q,LI N,CHEN S F. Application of morphology,color features and neural network in recognition of lung cancer cells [J]. Journal of Computer-Aided Design & Computer Graphics,2001,13(1):87—92.(In Chinese)

[9] JANTZEN J,NORUP J,DOUNIAS G,*et al*. Pap-smear benchmark data for pattern classification [J]. Nature Inspired Smart Information Systems (NiSIS 2005),2005:1—9.

[10] HALLINAN J,JACKWAY P. Detection of malignancy associated changes in thionin stained cervical cells [J]. Digital Image Computing and Applications,1995,27(5):426—431.

[11] PLISSITI M E,NIKOU C. A review of automated techniques for cervical cell image analysis and classification [M]. Netherlands: Biomedical Imaging and Computational Modeling in Biomechanics,2013:1—18.

[12] CHEN Y F,HUANG P C,LIN K C,*et al*. Semi-automatic segmentation and classification of pap smear cells [J]. Biomedical and Health Informatics,2014,18(1):94—108.

[13] HARALICK R M,SHANMUGAM K,DINSTEIN I H. Textural features for image classification [J]. Systems,Man and Cybernetics,1973, 12(6):610—621.

[14] WALKER R F, JACKWAY P, LOVELL B,*et al*. Classification of cervical cell nuclei using morphological segmentation and textural feature extraction [C]//Conference on Intelligent Information Systems. New Zealand:IEEE, 1994:297—301.

[15] WALKER R F, JACKWAY P T, LOVELL B. Cervical cell classification via co-occurrence and Markov random field features[C]// Proceedings of Digital Image Computing:Techniques and Applications. Australia Brisbane:IEEE, 1995:294—299.

[16] 赵晖,鲍莉,梁光明,等. 基于综合灰度共生矩阵的显微细胞图像纹理研究[J].自动化技术与应用,2005,23(10):27—29.

ZHAO H, BO L, LIANG G M,*et al*. Research on microscopic cell image texture based on integrated gray level co-occurrence matrix [J]. Automation Technology and Application, 2005, 23(10):27—29. (In Chinese)

[17] PLISSITI M E,NIKOU C,CHARCHANTI A. Automated detection of cell nuclei in Pap smear images using morphological reconstruction and clustering [J]. Information Technology in Biomedicine, 2011, 15(2):233—241.

[18] LIU H,MOTODA H. Feature selection for knowledge discovery and data mining [M]. Netherlands:Kluwer Academic Publishers, Springer Science & Business Media,1998:1—15.

[19] 孙磊,陈阳,黄洋文,等. 支持向量机算法对鼻咽癌与正常鼻咽细胞株拉曼光谱分析 [J], 光谱学与光谱分析,2013,33(6):1566—1569.

SUN L, CHEN Y, HUANG Y W,*et al*. Raman spectroscopy analysis of nasopharyngeal carcinoma and normal nasopharyngeal cell line by support vector machine algorithm [J]. Spectroscopy and Spectral Analysis, 2013, 33 (6):1566—1569. (In Chinese)

[20] 张问银,金宁德,刘印锋,等. 基于支持向量机的 CD4 细胞图像识别方法[J].计算机工程与科学,2009,31(7):150—152.

ZHANG W Y, JIN N D, LIU Y F,*et al*. Image recognition method based on support vector machine for CD4 cells [J]. Computer Engineering and Science,2009,31(7):150—152. (In Chinese)

[21] 张英杰,许伟,汤龙波,等. 基于滚动时间窗的 PSO-LSSVM 的通信基站能耗建模[J]. 湖南大学学报(自然科学版), 2017, 44(2):122—128.

ZHANG Y J, XU W, TANG L B,*et al*. Modeling of communication base station energy consumption based on PSO-LSSVM based on rolling time window [J].Journal of Hunan University (Natural Sciences),2017,44(2):122—128. (In Chinese)

[22] PENG H, LONG F, DING C. Feature selection based on mutual information criteria of max-dependency,max-relevance,and min-redundancy [J]. Pattern Analysis & Machine Intelligence IEEE Transactions, 2005, 27(8):1226—1238.

[23] KONONENKO I,ŠIMEC E,ROBNIK-ŠIKONJA M. Overcoming the myopia of inductive learning algorithms with relieff [J]. Applied Intelligence,1997,7(1):39—55.

[24] EDWIN J M, ALLWIN S. Nominated texture based cervical cancer classification [J]. Computational and Mathematical Methods in Medicine,2015,2015:1—10.

[25] KANGKANA B, MANISH CB, LIPI B. Automated classification of Pap smear images to detect cervical dysplasia [J]. Computer Methods and Programs in Biomedicine,2017,138:31—47.

[26] ZHAO M,WU A G. Automatic screening of cervical cells using block image processing [J]. BioMedical Engineering OnLine, 2016,15(1):14—33.

[27] SAJEENA T A,JEREESH A S. Automated cervical cancer detection through RGVF segmentation and SVM classification [C]// International Conference on Computing & Network Communications. Kauai,Hawai:IEEE,2016:16—19.

[28] CHANKONG T,THEERA-UMPON N,AUEPHANWIRIYANKUL S. Automatic cervical cell segmentation and classification in Pap smears [J]. Computer Methods Programs in Biomedicine,2014, 113(2):539—556.