



(12) 发明专利

(10) 授权公告号 CN 113177927 B

(45) 授权公告日 2022.12.09

(21) 申请号 202110536794.3 G06V 10/44 (2022.01)
(22) 申请日 2021.05.17 G06V 10/764 (2022.01)
(65) 同一申请的已公布的文献号 G06V 10/82 (2022.01)
申请公布号 CN 113177927 A G06N 3/04 (2006.01)
G06N 3/08 (2006.01)
(43) 申请公布日 2021.07.27 审查员 沈晴
(73) 专利权人 西安交通大学
地址 710049 陕西省西安市咸宁西路28号
(72) 发明人 祖建 葛飞 王雯娟 吴晓明
(74) 专利代理机构 西安通大专利代理有限责任
公司 61200
专利代理师 高博
(51) Int.Cl.
G06T 7/00 (2017.01)
G06T 7/40 (2017.01)
G06T 7/90 (2017.01)

权利要求书3页 说明书13页 附图7页

(54) 发明名称

基于多特征和多分类器的骨髓细胞分类识别方法及系统

(57) 摘要

本发明公开了一种基于多特征和多分类器的骨髓细胞分类识别方法及系统,基于3个迁移学习分类器对骨髓细胞进行训练测试;分别提取骨髓细胞图像数据集的纹理特征LBP,形状特征HOG和颜色特征HSV并进行融合得到特征融合图像,利用Keras模型融合算法对3个迁移分类器进行融合,得到融合分类器,并利用在单个分类器中取得最好测试准确率的特征融合图像对融合分类器进行训练,得到测试准确率最好的多特征多分类器融合模型;对骨髓细胞图像中的骨髓细胞进行定位分割,得到模型应用数据集;并利用测试效果最好的多特征多分类器融合模型对该数据集进行测试,得到最终的模型应用效果。本发明结合不同特征和不同分类器的优势,提高了骨髓细胞的测试分类准确率。

1. 基于多特征和多分类器的骨髓细胞分类识别方法,其特征在于,包括以下步骤:

S1、根据骨髓细胞原始图像建立原始图像数据集;对原始图像数据集进行定位、分割和数据增强处理,得到骨髓细胞图像数据集;

S2、构建VGG16,DenseNet121和ResNet50迁移分类器,并用步骤S1中的骨髓细胞图像数据集进行训练和测试,得到三种迁移分类器的测试准确率;

S3、从步骤S1得到的骨髓细胞图像数据集中的图像提取纹理特征LBP,形状特征HOG和颜色特征HSV;利用加权图像融合算法对提取的纹理特征LBP,形状特征HOG和颜色特征HSV特征图像进行融合得到特征融合图像,构建特征融合图像数据集,利用步骤S2构建的VGG16,DenseNet121和ResNet50迁移分类器对特征融合图像数据集进行训练并测试,得到每个迁移分类器中测试准确率最高的特征融合图像;

S4、利用Keras模型融合算法对步骤S2构建的VGG16,DenseNet121和ResNet50迁移分类器进行融合得到4个融合分类器,利用4个融合分类器对步骤S3得到的测试准确率最高的特征融合图像进行训练和测试,确定测试准确率最高的多特征多分类器融合模型;利用准确率最高的多特征多分类器融合模型对细胞图像进行分类识别。

2. 根据权利要求1所述的方法,其特征不在于,步骤S1具体为:

S101、获取骨髓细胞数据集BCI和ASDML,建立骨髓细胞数据集;

S102、从ASDML数据集和BCI数据集中选择嗜酸性粒细胞,淋巴细胞,单核细胞,原始粒细胞和嗜中性粒细胞作为原始骨髓细胞数据集;

S103、骨髓细胞定位和分割;

S104、对步骤S103分割后的骨髓细胞图像进行中值滤波处理;

S105、对步骤S104滤波后的骨髓细胞图像进行去雾处理;

S106、采用图像翻转,图像旋转和图像平移操作对步骤S105去雾处理后的骨髓细胞图像进行扩增,构建骨髓细胞图像数据集。

3. 根据权利要求1所述的方法,其特征不在于,步骤S103具体为:

S1031、利用LabelImg标注软件步骤S102中50%的对骨髓细胞进行标注,得到标注文件XML;

S1032、利用YOLOv3目标检测网络对步骤S1031得到的标注文件XML进行训练,得到YOLOv3训练模型;

S1033、利用步骤S1032得到的YOLOv3训练模型对步骤S102中所有的骨髓细胞图像数据集进行目标检测和定位,得到目标定位的候选框及四个像素点如下:左上角像素点 (x_1, y_1) ,右上角像素点 (x_2, y_1) ,左下角像素点 (x_1, y_2) ,右下角像素点 (x_2, y_2) ;

S1034、对步骤S1033得到的四个像素点进行计算 $x_1 - x_2$ 和 $y_1 - y_2$ 并划分出候选框之间的区域,最终将细胞从整个图像中分割出来,得到分割骨髓细胞图像。

4. 根据权利要求1所述的方法,其特征不在于,步骤S2具体为:

S201、构建VGG16,DenseNet121,ResNet50迁移分类器,并采用冻结和训练策略,分别冻结VGG16,DenseNet121,ResNet50迁移分类器网络的前面层数,并将公开自然图像数据集ImageNet训练的VGG16,DenseNet121,ResNet50迁移分类器网络参数和权重迁移到三个迁移分类器网络中,最后训练新设置的网络层数,得到训练好的三个单分类器模型;

S202、利用步骤S201训练好的三个单分类器模型对步骤S1中骨髓细胞图像数据集中的

测试集进行测试,采用准确率,精确率,召回率和F1分数分类指标对三个单分类器模型的测试结果进行评价。

5. 根据权利要求4所述的方法,其特征在于,步骤S201建立的三个迁移分类器中,优化器为SGD,激活函数为Softmax,学习率为0.001,迭代次数为100,输入尺寸为64*64,批尺寸为128,预训练权重为ImageNet,损失函数为交叉熵分类损失函数categorical_crossentropy loss function。

6. 根据权利要求1所述的方法,其特征在于,步骤S3具体为:

S301、对步骤S1骨髓细胞图像数据集中的图像进行局部二值模式特征提取,采用LBP等价模型的特征提取方法,得到骨髓细胞的LBP特征图像;

S302、对步骤S1骨髓细胞图像数据集中的图像进行方向梯度直方图特征提取,采用cell为1*1,block为2*2的特征提取方法,得到骨髓细胞的HOG特征图像;

S303、对步骤S1骨髓细胞图像数据集中的图像进行色调-饱和度-明度(HSV)特征提取,得到骨髓细胞的HSV特征图像;

S304、对步骤S301,步骤S302和步骤S303得到的骨髓细胞的LBP特征图像,骨髓细胞的HOG特征图像和骨髓细胞的HSV特征图像进行特征融合,采用加权图像融合算法对图像进行融合,得到8类特征融合图像;

S305、利用步骤S304得到的8类特征融合图像对步骤S2中构建的三个迁移分类器进行训练,并对步骤S1构建的骨髓细胞图像数据集中的测试集进行测试,得到测试准确率结果,选择测试准确率最高的特征融合图像作为基准特征融合图像。

7. 根据权利要求6所述的方法,其特征在于,步骤S304中,融合后的特征融合图像I(x)具体为:

$$I(x) = \omega_1 I_1(x) + \omega_2 I_2(x)$$

其中, $I_1(x)$ 和 $I_2(x)$ 表示需要融合的图像, ω_1 和 ω_2 表示融合图像的权重。

8. 根据权利要求1所述的方法,其特征在于,步骤S4具体为:

S401、利用Keras模型融合算法对VGG16,DenseNet121和ResNet50迁移分类器进行融合,加载分类模型1的训练权重,并加载分类模型2的训练权重,最后将模型1和模型2的网络进行融合,并加上新构建的全连接层和输出层,完成4个分类器融合;

S402、利用步骤S3得到的基准特征融合图像对步骤S401融合后的4个融合分类器进行训练,并对步骤S1构建的骨髓细胞图像数据集中的测试集进行测试,最终选取准确率最高的多特征多分类器融合模型用于骨髓细胞进行分类识别。

9. 根据权利要求8所述的方法,其特征在于,步骤S401中,4个分类器融合后的模型m计算如下:

$$m = \sum_{i=1}^n \omega_i m_i$$

其中, m_i 表示第i个分类器模型, ω_i 表示第i个分类器模型的权重。

10. 一种基于多特征和多分类器的骨髓细胞分类识别系统,其特征在于,包括:

数据模块,根据骨髓细胞原始图像建立原始图像数据集;对原始图像数据集进行定位、分割和数据增强处理,得到骨髓细胞图像数据集;

分类器模块,构建VGG16,DenseNet121和ResNet50迁移分类器,并用数据模块中的骨髓

细胞图像数据集进行训练和测试,得到三种迁移分类器的测试准确率;

特征提取模块,从数据模块得到的骨髓细胞图像数据集中的图像提取纹理特征LBP,形状特征HOG和颜色特征HSV;利用加权图像融合算法对提取的纹理特征LBP,形状特征HOG和颜色特征HSV特征图像进行融合得到特征融合图像,构建特征融合图像数据集,利用分类器模块构建的VGG16,DenseNet121和ResNet50迁移分类器对特征融合图像数据集进行训练并测试,得到每个迁移分类器中测试准确率最高的特征融合图像;

识别模块,利用Keras模型融合算法对分类器模块构建的VGG16,DenseNet121和ResNet50迁移分类器进行融合得到4个融合分类器,利用4个融合分类器对提取模块得到的测试准确率最高的特征融合图像进行训练和测试,确定测试准确率最高的多特征多分类器融合模型;利用准确率最高的多特征多分类器融合模型对细胞图像进行分类识别。

基于多特征和多分类器的骨髓细胞分类识别方法及系统

技术领域

[0001] 本发明属于图像处理技术领域,具体涉及一种基于多特征和多分类器的骨髓细胞分类识别方法及系统。

背景技术

[0002] 计算机辅助技术相比于人工识别来说,具有许多天然的优势。第一,它的识别速度快,效率高。一般而言,具有丰富经验的检验人员很少,而骨髓细胞图像却很多,人工识别的话必然导致效率极其低下。而好的算法和机器能够一分钟扫描识别出几十甚至上百张图像,大大提高了效率。第二,识别准确率高。有些检验人员的经验并不丰富,在识别过程中很可能会出现错误,这样对于诊断疾病不利。而计算机辅助识别系统只要设计好模型算法,那么就能达到一个不错的识别准确率。第三,细胞分割出色。人工往往在观察细胞时,由于细胞之间相互粘连,有时很难将它们分开,大大降低了识别的效率。而计算机辅助识别系统能够将细胞之间分割开,这样就能够方便地对单独细胞进行识别。

[0003] 多种特征以及多种分类器融合模型的构建能够更加全面和有效地识别骨髓细胞,进而能够有效地计算出细胞比例和数量,具有很高的应用实践价值。

发明内容

[0004] 本发明所要解决的技术问题在于针对上述现有技术中的不足,提供一种基于多特征和多分类器的骨髓细胞分类识别方法及系统,利用纹理特征,形状特征和颜色特征等多种融合特征,同时结合多种分类器融合模型来提高测试分类准确率,解决骨髓细胞的分类识别问题。

[0005] 本发明采用以下技术方案:

[0006] 基于多特征和多分类器的骨髓细胞分类识别方法,包括以下步骤:

[0007] S1、根据骨髓细胞原始图像建立原始图像数据集;对原始图像数据集进行定位、分割和数据增强处理,得到骨髓细胞图像数据集;

[0008] S2、构建VGG16,DenseNet121和ResNet50迁移分类器,并用步骤S1中的骨髓细胞图像数据集进行训练和测试,得到三种迁移分类器的测试准确率;

[0009] S3、从步骤S1得到的骨髓细胞图像数据集中的图像提取纹理特征LBP,形状特征HOG和颜色特征HSV;利用加权图像融合算法对提取的纹理特征LBP,形状特征HOG和颜色特征HSV特征图像进行融合得到特征融合图像,构建特征融合图像数据集,利用步骤S2构建的VGG16,DenseNet121和ResNet50迁移分类器对特征融合图像数据集进行训练并测试,得到每个迁移分类器中测试准确率最高的特征融合图像;

[0010] S4、利用Keras模型融合算法对步骤S2构建的VGG16,DenseNet121和ResNet50迁移分类器进行融合得到4个融合分类器,利用4个融合分类器对步骤S3得到的测试准确率最高的特征融合图像进行训练和测试,确定测试准确率最高的多特征多分类器融合模型;利用准确率最高的多特征多分类器融合模型对细胞图像进行分类识别。

- [0011] 具体的,步骤S1具体为:
- [0012] S101、获取骨髓细胞数据集BCI和ASDML,建立骨髓细胞数据集;
- [0013] S102、从ASDML数据集和BCI数据集中选择嗜酸性粒细胞,淋巴细胞,单核细胞,原始粒细胞和嗜中性粒细胞作为原始骨髓细胞数据集;
- [0014] S103、骨髓细胞定位和分割;
- [0015] S104、对步骤S103分割后的骨髓细胞图像进行中值滤波处理;
- [0016] S105、对步骤S104滤波后的骨髓细胞图像进行去雾处理;
- [0017] S106、采用图像翻转,图像旋转和图像平移操作对步骤S105去雾处理后的骨髓细胞图像进行扩增,构建骨髓细胞图像数据集。
- [0018] 具体的,步骤S103具体为:
- [0019] S1031、利用LabelImg标注软件步骤S102中50%的对骨髓细胞进行标注,得到标注文件XML;
- [0020] S1032、利用YOLOv3目标检测网络对步骤S1031得到的标注文件XML进行训练,得到YOLOv3训练模型;
- [0021] S1033、利用步骤S1032得到的YOLOv3训练模型对步骤S102中所有的骨髓细胞图像数据集进行目标检测和定位,得到目标定位的候选框及四个像素点如下:左上角像素点 (x_1, y_1) ,右上角像素点 (x_2, y_1) ,左下角像素点 (x_1, y_2) ,右下角像素点 (x_2, y_2) ;
- [0022] S1034、对步骤S1033得到的四个像素点进行计算 $x_1 - x_2 * |y_1 - y_2|$ 并划分出候选框之间的区域,最终将细胞从整个图像中分割出来,得到分割骨髓细胞图像。
- [0023] 具体的,步骤S2具体为:
- [0024] S201、构建VGG16,DenseNet121,ResNet50迁移分类器,并采用冻结和训练策略,分别冻结VGG16,DenseNet121,ResNet50迁移分类器网络的前面层数,并将公开自然图像数据集ImageNet训练的VGG16,DenseNet121,ResNet50迁移分类器网络参数和权重迁移到三个迁移分类器网络中,最后训练新设置的网络层数,得到训练好的三个单分类器模型;
- [0025] S202、利用步骤S201训练好的三个单分类器模型对步骤S1中骨髓细胞图像数据集中的测试集进行测试,采用准确率,精确率,召回率和F1分数分类指标对三个单分类器模型的测试结果进行评价。
- [0026] 进一步的,步骤S201建立的三个迁移分类器中,优化器为SGD,激活函数为Softmax,学习率为0.001,迭代次数为100,输入尺寸为64*64,批尺寸为128,预训练权重为ImageNet,损失函数为交叉熵分类损失函数categorical_crossentropy loss function。
- [0027] 具体的,步骤S3具体为:
- [0028] S301、对步骤S1骨髓细胞图像数据集中的图像进行局部二值模式特征提取,采用LBP等价模型的特征提取方法,得到骨髓细胞的LBP特征图像;
- [0029] S302、对步骤S1骨髓细胞图像数据集中的图像进行方向梯度直方图特征提取,采用cell为1*1,block为2*2的特征提取方法,得到骨髓细胞的HOG特征图像;
- [0030] S303、对步骤S1骨髓细胞图像数据集中的图像进行色调-饱和度-明度(HSV)特征提取,得到骨髓细胞的HSV特征图像;
- [0031] S304、对步骤S301,步骤S302和步骤S303得到的骨髓细胞的LBP特征图像,骨髓细胞的HOG特征图像和骨髓细胞的HSV特征图像进行特征融合,采用加权图像融合算法对图像

进行融合,得到8类特征融合图像;

[0032] S305、利用步骤S304得到的8类特征融合图像对步骤S2中构建的三个迁移分类器进行训练,并对步骤S1构建的骨髓细胞图像数据集中的测试集进行测试,得到测试准确率结果,选择测试准确率最高的特征融合图像作为基准特征融合图像。

[0033] 进一步的,步骤S304中,融合后的特征融合图像 $I(x)$ 具体为:

$$[0034] \quad I(x) = \omega_1 I_1(x) + \omega_2 I_2(x)$$

[0035] 其中, $I_1(x)$ 和 $I_2(x)$ 表示需要融合的图像, ω_1 和 ω_2 表示融合图像的权重。

[0036] 具体的,步骤S4具体为:

[0037] S401、利用Keras模型融合算法对VGG16,DenseNet121和ResNet50迁移分类器进行融合,加载分类模型1的训练权重,并加载分类模型2的训练权重,最后将模型1和模型2的网络进行融合,并加上新构建的全连接层和输出层,完成4个分类器融合;

[0038] S402、利用步骤S3得到的基准特征融合图像对步骤S401融合后的4个融合分类器进行训练,并对步骤S1构建的骨髓细胞图像数据集中的测试集进行测试,最终选取准确率最高的多特征多分类器融合模型用于骨髓细胞进行分类识别。

[0039] 进一步的,步骤S401中,4个分类器融合后的模型 m 计算如下:

$$[0040] \quad m = \sum_{i=1}^n \omega_i m_i$$

[0041] 其中, m_i 表示第 i 个分类器模型, ω_i 表示第 i 个分类器模型的权重。

[0042] 本发明的另一技术方案是,一种基于多特征和多分类器的骨髓细胞分类识别系统,包括:

[0043] 数据模块,根据骨髓细胞原始图像建立原始图像数据集;对原始图像数据集进行定位、分割和数据增强处理,得到骨髓细胞图像数据集;

[0044] 分类器模块,构建VGG16,DenseNet121和ResNet50迁移分类器,并用数据模块中的骨髓细胞图像数据集进行训练和测试,得到三种迁移分类器的测试准确率;

[0045] 特征提取模块,从数据模块得到的骨髓细胞图像数据集中的图像提取纹理特征LBP,形状特征HOG和颜色特征HSV;利用加权图像融合算法对提取的纹理特征LBP,形状特征HOG和颜色特征HSV特征图像进行融合得到特征融合图像,构建特征融合图像数据集,利用分类器模块构建的VGG16,DenseNet121和ResNet50迁移分类器对特征融合图像数据集进行训练并测试,得到每个迁移分类器中测试准确率最高的特征融合图像;

[0046] 识别模块,利用Keras模型融合算法对分类器模块构建的VGG16,DenseNet121和ResNet50迁移分类器进行融合得到4个融合分类器,利用4个融合分类器对提取模块得到的测试准确率最高的特征融合图像进行训练和测试,确定测试准确率最高的多特征多分类器融合模型;利用准确率最高的多特征多分类器融合模型对细胞图像进行分类识别。

[0047] 与现有技术相比,本发明至少具有以下有益效果:

[0048] 本发明一种基于多特征和多分类器的骨髓细胞分类识别方法,使用YOLOv3目标检测网络对骨髓细胞进行定位和分割,保证了细胞分割结果的准确性和完整性;使用迁移学习对骨髓细胞进行训练测试,降低了对图片训练数据量的要求,仅仅只有传统机器学习数量的十分之一;基于多特征和多分类器对骨髓细胞进行训练测试,可以全面地结合不同特征和不同分类器的优势,提高骨髓细胞的测试分类准确率。

[0049] 进一步的,对原始数据进行定位分割以及数据增强等数据预处理工作,保证了骨髓细胞图像的一致性,清晰性和平衡性。骨髓细胞图像中不仅仅包括骨髓细胞,而且还包括红细胞,血小板等杂质,这些杂质的存在会很大程度上影响骨髓细胞分类识别的结果,同时原始图像中的骨髓细胞由于采集手段或设备的影响,其图像质量并不高,其低质量的图像容易造成分类识别结果的不准确性;再次由于原始骨髓细胞图像各类别之间数目不平衡,容易造成骨髓细胞分类的倾向性,数量多的骨髓细胞分类效果要高于数量少的骨髓细胞。为了避免以上三种原因造成的骨髓细胞分类效果的误差,本发明使用了图像预处理方法,利用YOLOv3对骨髓细胞进行定位分割,保证了其一致性;利用去雾滤波操作对骨髓细胞进行处理,保证了其清晰性;利用图像增强方法对骨髓细胞进行图像扩增,保证了其平衡性。

[0050] 进一步的,利用YOLOv3目标检测网络对骨髓细胞进行定位分割保证了分割结果的准确性和完整性,对后续处理起了重要的作用。YOLOv3网络是一种良好的目标检测网络,其最大的优势是构建多尺度预测,通过多尺度预测对骨髓细胞进行三种尺度的特征提取,分别是大尺度特征,中尺度特征和小尺度特征,通过特征金字塔对多尺度特征进行处理并构建多种预测框,同时利用Logisitic算法来对不同的预测框进行概率预测,从而得到预测概率最大的预测框作为最终定位后的候选框。由于多尺度预测,使得YOLOv3对不同大小和尺寸的骨髓细胞能够进行精确和完整的定位,再结合候选框坐标分割算法,将候选框内的骨髓细胞完整地分割出来。该分割方法相对于传统的人工分割方法,其效率和准确率大大提高。

[0051] 进一步的,利用迁移学习思想构建的迁移分类器对骨髓细胞进行分类识别,能够大大地降低训练数据量的要求,同时能够达到相应的准确率。传统的机器学习或者深度学习方法对训练样本量有一定的要求,若样本量太小,会对整个分类识别的效果产生不良的影响。在某些领域内,其公开的图像数据比较少,故使用传统的分类方法难以达到良好的效果。由于医院的特殊性和病人的隐私性,骨髓细胞图像难以大量获取,同时存在的公开骨髓细胞图像数据集也比较稀少,故使用传统的分类方法对骨髓细胞进行分类识别,很难达到一定的效果。而迁移学习作为一种比较前沿的深度学习方法,其核心思想是将相似领域的分类器结构、参数和权重迁移到当前分类任务中,这样便可以解决样本量不足的情况下,分类识别效果能够达到良好的结果。

[0052] 进一步的,利用冻结和训练的迁移学习分类方法对骨髓细胞进行分类识别,能够大大减少训练的时间和空间。迁移学习包括两种迁移策略,一种是微调的方法,一种是冻结和训练的方法。冻结和训练策略实质上是冻结分类器模型中被迁移的部分,保留迁移的参数和权重,不对该部分进行训练,同时构建出适合骨髓细胞分类识别任务的全连接层和输出层,仅仅对新建的层数进行训练,这样避免从头进行训练并保留和迁移的思想,大大减少训练的时间和空间,并达到相应的识别准确率。

[0053] 进一步的,构建多特征融合图像,并利用其对骨髓细胞进行分类识别,能够结合不同特征的优势,提高了分类识别的准确性和有效性。骨髓细胞作为一种血液性细胞,其形状、纹理和颜色各不相同,形状方面呈现月牙形,圆形,类圆形,镰刀型等形状,而纹理方面呈现有颗粒和无颗粒等纹理,而颜色方面呈现淡蓝色,深蓝色,天蓝色等颜色,而原始细胞图像并没有将这些特征良好的显现出来。LBP是一种经典的纹理特征,其等价模型的效果能够显现出图像最本质的纹理特性;而HOG特征作为重要的形状特征,其提取的特征图像能够

很好地表示物体的形状;HSV是一种重要的颜色系统,不同于RGB颜色系统,该颜色系统能够对颜色中的重要特性(色调,饱和度和明度)进行提取,而这些特性在细胞中产生重要的作用。通过对提取后的特征图像进行融合,能够结合不同特征的优势和细胞最本质的特性,提高了骨髓细胞分类识别的准确率。

[0054] 进一步的,利用加权融合算法对特征图像进行融合,能够保证特征信息的保留,为后续分类识别做好了基础。加权融合算法是一种经典的图像融合算法,其核心思想是对图像间的像素点进行加权并求和,其结果为最后融合后图像的像素点。其方法最大的优势是简单好用,不需要繁杂的步骤。而骨髓细胞特征图像融合需要对骨髓细胞间的特征进行信息间的结合,这样才能实现特征间的融合。而加权融合算法通过设置权重均为1能够最大程度上保留细胞特征的信息,为后续的处理奠定了良好的基础。

[0055] 进一步的,构建多分类器融合模型并对多种特征融合图像进行分类识别,能够结合多个分类器和多个特征的优势和特性,大大提高了骨髓细胞分类识别的准确率。上述所说的三个单分类器VGG16,ResNet50和DenseNet121具有不同的结构和参数,其对骨髓细胞分类有着不同的效果,其单一分类器效果必然无法完成全面地分类识别任务。而多种分类器的融合能够对分类器之间进行互补,同时利用多分类器融合模型对特征融合图像进行分类识别,能够同时结合特征间的优势,保证了多种分类器和特征间的互补和融合,大大提高了分类识别的准确率。

[0056] 进一步的,利用Keras多分类器融合算法对分类器进行融合,能够结合多种分类器之间的优势,保证了分类结果的全面性,提高了分类识别效果。Keras作为经典的深度学习框架,在深度学习中具有重要的作用。其连接方法(concatenate)是将分类器中的网络部分进行融合,其结合效果为线性融合。其原理是首先保留输入层的不变,接着将迁移分类器1的网络结构和参数连接进去,同时连接迁移分类器2的网络结构和参数,最后连入适合本分类任务的全连接层和输出层,最终完成了分类器融合的思想。利用Keras分类器融合算法能够将各分类器中的优势结合起来,提高了分类的准确性。

[0057] 综上所述,本发明使用YOLOv3目标检测网络对骨髓细胞进行定位和分割,保证了细胞分割结果的准确性和完整性;使用迁移学习对骨髓细胞进行训练测试,降低了对图片训练数据量的要求;基于多特征和多分类器对骨髓细胞进行训练测试,可以全面地结合不同特征和不同分类器的优势,提高骨髓细胞的测试分类准确率。

[0058] 下面通过附图和实施例,对本发明的技术方案做进一步的详细描述。

附图说明

[0059] 图1为本发明的总体流程图;

[0060] 图2为本发明中数据集的建立流程示意图;

[0061] 图3为细胞定位分割流程示意图;

[0062] 图4为基于迁移学习的骨髓细胞分类识别流程示意图;

[0063] 图5为基于多特征和多分类器融合的骨髓细胞分类识别流程示意图;

[0064] 图6为基于医院血涂片的模型应用实践流程示意图;

[0065] 图7为单分类器模型在原始图像上的实验训练图,其中,(a)为VGG16分类器,(b)为ResNet50分类器,(c)为DenseNet121分类器;

- [0066] 图8为单分类器模型在测试集上的评价指标图；
- [0067] 图9为单分类器在特征融合图像上的实验训练图，其中，(a)为VGG16分类器，(b)为ResNet50分类器，(c)为DenseNet121分类器；
- [0068] 图10为VGG16分类器在在测试集上的评价指标图；
- [0069] 图11为ResNet50分类器在在测试集上的评价指标图；
- [0070] 图12为DenseNet121分类器在在测试集上的评价指标图；
- [0071] 图13为多特征多分类器融合模型的实验训练图；
- [0072] 图14为多特征多分类融合模型在测试集上的评价指标图；
- [0073] 图15为模型在医院血涂片上的检验效果图。

具体实施方式

[0074] 下面将结合本发明实施例中的附图，对本发明实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例是本发明一部分实施例，而不是全部的实施例。基于本发明中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都属于本发明保护的范围。

[0075] 应当理解，当在本说明书和所附权利要求书中使用时，术语“包括”和“包含”指示所描述特征、整体、步骤、操作、元素和/或组件的存在，但并不排除一个或多个其它特征、整体、步骤、操作、元素、组件和/或其集合的存在或添加。

[0076] 还应当理解，在本发明说明书中所使用的术语仅仅是出于描述特定实施例的目的而并不意在限制本发明。如在本发明说明书和所附权利要求书中所使用的那样，除非上下文清楚地指明其它情况，否则单数形式的“一”、“一个”及“该”意在包括复数形式。

[0077] 还应当进一步理解，在本发明说明书和所附权利要求书中使用的术语“和/或”是指相关联列出的项中的一个或多个的任何组合以及所有可能组合，并且包括这些组合。

[0078] 在附图中示出了根据本发明公开实施例的各种结构示意图。这些图并非是按比例绘制的，其中为了清楚表达的目的，放大了某些细节，并且可能省略了某些细节。图中所示出的各种区域、层的形状及它们之间的相对大小、位置关系仅是示例性的，实际中可能由于制造公差或技术限制而有所偏差，并且本领域技术人员根据实际所需可以另外设计具有不同形状、大小、相对位置的区域/层。

[0079] 本发明提供了一种基于多特征和多分类器的骨髓细胞分类识别方法及系统，利用YOLOv3对骨髓细胞进行定位和分割，保留了细胞的完整性，使得细胞分割更加精确；使用迁移学习对骨髓细胞进行训练测试，降低了对图片训练数据量的要求，仅仅只有传统机器学习数量的十分之一；基于多特征和多分类器对骨髓细胞进行训练测试，可以全面地结合不同特征和不同分类器的优势，提高骨髓细胞的测试分类准确率。

[0080] 请参阅图1，本发明一种基于多特征和多分类器的骨髓细胞分类识别方法，包括以下步骤：

[0081] S1、获取骨髓细胞原始图像，建立原始图像数据集；对获取的图像进行定位和分割，并对图像进行数据增强，建立供训练和测试用的骨髓细胞图像数据集；

[0082] 请参阅图2，建立基于YOLOv3的骨髓细胞数据集，具体包括以下步骤：

[0083] S101、数据集获取

[0084] 本发明获取了两大公开骨髓细胞数据集BCI和ASDML用于骨髓细胞数据集的建立,其中BCI数据集共有4类细胞,训练集共有9957张,验证集共有2487张;ASDML数据集共有15类细胞,训练集共有12848张,验证集共有3675张。

[0085] S102、数据的选择

[0086] 依据两大数据集中的细胞数量和细胞图像质量,最终从中选择了5类细胞,分别是嗜酸性粒细胞,淋巴细胞,单核细胞,原始粒细胞和嗜中性粒细胞作为原始骨髓细胞数据集,其中训练集共17500张,测试集共7102张。

[0087] S103、骨髓细胞定位和分割

[0088] 为了对骨髓细胞图像中的杂质进行剔除和过滤,需要对骨髓细胞进行分割。

[0089] 请参阅图3,实施过程如下:

[0090] S1031、利用LabelImg标注软件对骨髓细胞进行标注,这里选择数据集中的一半即8750张骨髓细胞作为标注数据集,得到8750个标注文件XML;

[0091] S1032、利用XML标注文件对YOLOv3进行训练,得到YOLOv3训练模型,其中YOLOv3模型损失函数的目标函数如下:

$$\begin{aligned}
 \text{Loss} = & \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{\text{obj}} [(x_i - \hat{x}_i^j)^2 + (y_i - \hat{y}_i^j)^2] \\
 & + \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{\text{obj}} [(\sqrt{a_i^j} - \sqrt{\hat{a}_i^j})^2 + (\sqrt{b_i^j} - \sqrt{\hat{b}_i^j})^2] \\
 [0092] \quad & - \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{\text{obj}} [\hat{C}_i^j \log(C_i^j) + (1 - \hat{C}_i^j) \log(1 - \hat{C}_i^j)] \\
 & - \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{\text{noobj}} [(\hat{C}_i^j \log(C_i^j) + (1 - \hat{C}_i^j) \log(1 - \hat{C}_i^j))] \\
 & - \sum_{i=0}^{S^2} I_{ij}^{\text{obj}} \sum_{c \in \text{classes}} \hat{P}_i^j \log(P_i^j) + (1 - \hat{P}_i^j) \log(1 - \hat{P}_i^j)
 \end{aligned}$$

[0093] 其中,前一种损失采用误差平方和损失函数,后两者损失采用二分类交叉熵损失函数,S代表特征图长度,而B代表每个网络中预测的边界框个数。 x_i, y_i, a_i^j, b_i^j 分别表示预测框的中心横纵坐标以及预测框的长和宽,而 $\hat{x}_i^j, \hat{y}_i^j, \hat{a}_i^j, \hat{b}_i^j$ 表示实际标记框中心横纵坐标以及长和宽。 $I_{ij}^{\text{obj}}, I_{ij}^{\text{noobj}}$ 分别表示第i个网格中第j个候选框是否对该目标进行检测, C_i^j, \hat{C}_i^j 分别表示第j个目标的预测置信度和真实置信度,而 P_i^j, \hat{P}_i^j 表示了第i个网格中第j个目标属于某一类别的预测概率大小和真实概率大小。

[0094] S1033、利用YOLOv3训练模型对骨髓细胞数据集进行目标检测和定位,得到目标定位的候选框及其四个像素点如下:左上角像素点 (x_1, y_1) ,右上角像素点 (x_2, y_1) ,左下角像素点 (x_1, y_2) ,右下角像素点 (x_2, y_2) ;

[0095] S1034、对四个像素点进行计算 $|x_1 - x_2| * |y_1 - y_2|$ 并划分出候选框之间的区域,最终将细胞从整个图像中分割出来,得到分割骨髓细胞图像,计算公式如下:

$$[0096] \quad I = |x_1 - x_2| * |y_1 - y_2|$$

[0097] 其中, I表示分割后的图像, (x_1, y_1) 表示左上角像素点, (x_2, y_1) 表示右上角像素点, (x_1, y_2) 表示左下角像素点, (x_2, y_2) 表示右下角像素点。

[0098] S104、图像滤波操作

[0099] 对分割后的图像进行中值滤波,其计算公式如下:

[0100] $h(x, y) = \text{mid}(f(x, y) \in s)$

[0101] 其中, $h(x, y)$ 为中值滤波值, s 为周围正方形区域, $f(x, y)$ 为区域内像素点的值。

[0102] S105、图像去雾操作

[0103] 对滤波后的图像进行去雾,目的是让图像更加清晰可见,其计算公式如下所示:

[0104]
$$J(x) = \frac{I(x) - A}{\max[t(x), t_0]} + A$$

[0105] 其中, $I(x)$ 为原始图像, A 为全球大气光成分, $t(x)$ 指的是大气折射率, t_0 为阈值, $J(x)$ 为去雾后的图像。

[0106] S106、图像扩增

[0107] 采用图像翻转,图像旋转和图像平移操作最终对图像进行扩增,并最终构建了供训练测试的骨髓细胞图像数据集,该数据集一共有训练集25000张,验证集7500张,测试集792张。

[0108] S2、对步骤S1中的骨髓细胞图像数据集进行迁移学习,利用骨髓细胞图像数据集对三个迁移分类器(VGG16, ResNet50, DenseNet121)进行训练和验证,如图7所示,并利用训练好的模型对测试集进行测试得到测试准确率,如图8所示。

[0109] 请参阅图4,具体步骤如下:

[0110] S201、构建三个迁移分类器(VGG16, ResNet50, DenseNet121),并采用冻结和训练策略,即只冻结网络的前面层数,并将ImageData数据集训练的网络参数和权重迁移到三个迁移网络中,最后训练新设置的上述三个迁移分类器网络层数,得到训练和验证结果。其参数设置为:选择优化器为SGD,激活函数为Softmax,学习率为0.001,迭代次数为100,输入尺寸为64*64,批尺寸(Batch_Size)为128,预训练权重为ImageNet,损失函数为交叉熵分类损失函数。激活函数的计算公式如下:

[0111]
$$S_i = \frac{e^{v_i}}{\sum_i e^{v_i}}$$

[0112] 其中, v_i 表示判别种类中第*i*类数据出现的概率, S_i 的值表示被预测为第*i*类在所有类中的比例,该值越大则证明属于该类的可能性越大,该值越小说明属于该类的可能性越小。损失函数的计算公式如下:

[0113]
$$\text{Loss} = -\frac{1}{n} \sum_x [y \ln a + (1 - y) \ln(1 - a)]$$

[0114] 其中, y 为整个期望的输出,而 a 为神经元的实际输出。交叉熵主要是用来评估概率分布和真实分布之间的差异。即交叉熵的值越小,其概率分布和真实分布就越接近,即分类识别的效果也就越好。交叉熵损失函数主要是处理多分类问题,通过对多种类别形成一系列概率,然后将这些概率进行交叉熵操作,最终得到整个模型的损失。

[0115] S202、利用训练好的模型对测试集中的数据进行测试,采用四个分类指标对其进

行评价测试结果。其分类指标分别为准确率 (Accuracy), 精确率 (Precision), 召回率 (Recall) 和F1分数 (F1), 利用四个评价指标对模型测试结果进行评价, 其评价指标计算公式如下:

$$[0116] \quad Accuracy = \frac{TP+TN}{P+N} = \frac{TP+TN}{TP+FN+FP+TN}$$

$$[0117] \quad Precision = \frac{TP}{TP+FP}$$

$$[0118] \quad Recall = \frac{TP}{TP+FN}$$

$$[0119] \quad \frac{2}{F1} = \frac{1}{Precision} + \frac{1}{Recall}$$

$$[0120] \quad F1 = \frac{2TP}{2TP+FP+FN}$$

[0121] 其中, TP表示正样本中被正确分类的样本量; FN表示正样本中被错误分类的样本量; FP表示负样本中被错误分类的样本量; TN表示负样本中被正确分类的样本量; P表示正样本的数量; N表示负样本的数量; P+N表示正负样本的数量。

[0122] S3、对步骤S1中的骨髓细胞图像数据集进行特征提取, 分别提取纹理特征LBP, 形状特征HOG和颜色特征HSV; 利用加权图像融合算法对提取到的特征图像进行融合, 分别得到8个特征融合图像, 并分别利用步骤S2中的3个迁移分类器进行训练验证, 如图9所示; 并利用训练好的模型对测试集进行测试, 得到各测试准确率, 如图10, 图11, 图12所示。

[0123] 请参阅图5, 具体步骤如下:

[0124] S301、对步骤S106中的骨髓细胞图像数据集中的骨髓细胞图像进行局部二值模式 (LBP) 特征提取, 采用LBP等价模型的特征提取方法, 得到骨髓细胞的LBP特征图像, 其计算公式如下:

$$[0125] \quad U(LBP_{P,R}) = |s(i_{P-1} - i_c) - s(i_0 - i_c)| + \left| \sum_{j=1}^{P-1} (s(i_j - i_c) - s(i_{j-1} - i_c)) \right|$$

$$[0126] \quad s(i_j - i_c) = \begin{cases} 1, & \text{if } i_j \geq i_c \\ 0, & \text{else} \end{cases}$$

[0127] 其中, i_j 为邻域内像素点的灰度值, i_c 为中心像素点的灰度值, P为总采样点, R为圆形的半径值, c为中心点的值, $U(LBP_{P,R})$ 为等价模式的LBP特征值。

[0128] S302、对S106中骨髓细胞图像数据集中的骨髓细胞图像进行方向梯度直方图 (HOG) 特征提取, 采用cell为1*1, block为2*2的特征提取方法, 其中cell表示图像划分的细胞单元, block是由若干个细胞单元组成的, 从而得到骨髓细胞的HOG特征图像, 其计算公式如下:

$$[0129] \quad G'_x(x, y) = H(x+1, y) - H(x-1, y)$$

$$[0130] \quad G'_y(x, y) = H(x, y+1) - H(x, y-1)$$

$$[0131] \quad G'(x, y) = \sqrt{G'_x(x, y)^2 + G'_y(x, y)^2}$$

$$[0132] \quad \alpha(x, y) = \tan^{-1} \left(\frac{G'_y(x, y)}{G'_x(x, y)} \right)$$

[0133] 其中, $G'_x(x, y)$, $G'_y(x, y)$, $H(x, y)$ 分别表示输入图像中的像素点 (x, y) 处的水平方向的梯度, 垂直方向的梯度以及该点的像素值, 而 $G'(x, y)$ 和 $\alpha(x, y)$ 表示像素点 (x, y) 处的梯度幅度和梯度方向。

[0134] S303、对S106中骨髓细胞图像数据集中的骨髓细胞图像进行色调-饱和度-明度(HSV)特征提取, 得到骨髓细胞的HSV特征图像, 其计算公式如下:

$$[0135] \quad h = \begin{cases} 0, & \text{if } \max(r, g, b) = \min(r, g, b) \\ 60 \times \frac{g-b}{\max-\min} + 0, & \text{if } \max = r \text{ and } g \geq b \\ 60 \times \frac{g-b}{\max-\min} + 360, & \text{if } \max = r \text{ and } g < b \\ 60 \times \frac{g-b}{\max-\min} + 120, & \text{if } \max = g \\ 60 \times \frac{g-b}{\max-\min} + 240, & \text{if } \max = b \end{cases}$$

$$[0136] \quad v = \max(r, g, b)$$

$$[0137] \quad s = \begin{cases} 0, & \text{if } \max=0 \\ \frac{\max-\min}{\max}, & \text{otherwise} \end{cases}$$

[0138] 其中, r, g, b 分别表示原始RGB图像中的red值, green值和blue值, 而 h, s, v 分别表示转换为hsv颜色空间中的色调值, 饱和度值和明度值。

[0139] S304、对步骤S301、S302和S303中的特征图像进行特征融合, 采用加权图像融合算法对图像进行融合, 最终得到8类特征融合图像, 其计算公式如下:

$$[0140] \quad I(x) = \omega_1 I_1(x) + \omega_2 I_2(x)$$

[0141] 其中, $I(x)$ 表示融合后的图像, $I_1(x)$ 和 $I_2(x)$ 表示需要融合的图像, ω_1 和 ω_2 表示融合图像的权重, 在本发明中为了保证所有特征信息尽可能的保留, 设置的权重均为1。

[0142] S305、利用三个迁移分类器对8类特征融合图像进行训练验证, 并对测试集进行测试, 得到测试准确率结果, 并且将测试准确率最好的特征融合图像作为基准特征融合图像。

[0143] S4、利用Keras模型融合算法对步骤S2中的3个迁移分类器进行融合, 得到4个融合分类器, 并利用4个融合分类器对步骤S3中在单个分类器中取得最好测试准确率的特征融合图像进行训练验证, 如图13所示, 并利用训练好的模型对测试集进行测试, 如图14所示, 得到测试准确率最好的多特征多分类器融合模型。

[0144] S401、利用Keras中的连接方法对分类器进行融合, 首先加载分类模型1的训练权重, 然后加载分类模型2和训练权重, 最后将模型1和模型2的网络进行融合, 并加上新建的全连接层和输出层即完成分类器融合。

[0145] 最终融合的分类器模型m计算如下:

$$[0146] \quad m = \sum_{i=1}^n \omega_i m_i$$

[0147] 其中, m_i 表示第 i 个分类器模型, ω_i 表示第 i 个分类器模型的权重, 在这里, 为了尽可能保证保留分类器模型的优势, 本发明设置权重均为 1。

[0148] S402、将融合后的 4 个融合分类器对基准特征融合图像进行训练并对测试集进行测试, 最终得到测试准确率最好的多特征多分类器融合模型, 如图 13 所示。

[0149] 本发明再一个实施例中, 提供一种基于多特征和多分类器的骨髓细胞分类识别系统, 该系统能够用于实现上述基于多特征和多分类器的骨髓细胞分类识别方法, 具体的, 该基于多特征和多分类器的骨髓细胞分类识别系统包括数据模块、分类器模块、特征提取模块以及识别模块。

[0150] 其中, 数据模块, 根据骨髓细胞原始图像建立原始图像数据集; 对原始图像数据集进行定位、分割和数据增强处理, 得到骨髓细胞图像数据集;

[0151] 分类器模块, 构建 VGG16, DenseNet121 和 ResNet50 迁移分类器, 并用数据模块中的骨髓细胞图像数据集进行训练和测试, 得到三种迁移分类器的测试准确率;

[0152] 特征提取模块, 从数据模块得到的骨髓细胞图像数据集中提取纹理特征 LBP, 形状特征 HOG 和颜色特征 HSV; 利用加权图像融合算法对提取的纹理特征 LBP, 形状特征 HOG 和颜色特征 HSV 特征图像进行融合得到特征融合图像, 构建特征融合图像数据集, 利用分类器模块构建的 VGG16, DenseNet121 和 ResNet50 迁移分类器对特征融合图像数据集进行训练并测试, 得到每个迁移分类器中测试准确率最高的特征融合图像;

[0153] 识别模块, 利用 Keras 模型融合算法对分类器模块构建的 VGG16, DenseNet121 和 ResNet50 迁移分类器进行融合得到 4 个融合分类器, 利用 4 个融合分类器对提取模块得到的测试准确率最高的特征融合图像进行训练和测试, 确定测试准确率最高的多特征多分类器融合模型; 利用准确率最高的多特征多分类器融合模型对细胞图像进行分类识别。

[0154] 本发明再一个实施例中, 提供了一种终端设备, 该终端设备包括处理器以及存储器, 所述存储器用于存储计算机程序, 所述计算机程序包括程序指令, 所述处理器用于执行所述计算机存储介质存储的程序指令。处理器可能是中央处理单元 (Central Processing Unit, CPU), 还可以是其他通用处理器、数字信号处理器 (Digital Signal Processor, DSP)、专用集成电路 (Application Specific Integrated Circuit, ASIC)、现成可编程门阵列 (Field-Programmable Gate Array, FPGA) 或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件等, 其是终端的计算核心以及控制核心, 其适于实现一条或一条以上指令, 具体适于加载并执行一条或一条以上指令从而实现相应方法流程或相应功能; 本发明实施例所述的处理器可以用于基于多特征和多分类器的骨髓细胞分类识别的操作, 包括:

[0155] 根据骨髓细胞原始图像建立原始图像数据集; 对原始图像数据集进行定位、分割和数据增强处理, 得到骨髓细胞图像数据集; 构建 VGG16, DenseNet121 和 ResNet50 迁移分类器, 并用步骤 S1 中的骨髓细胞图像数据集进行训练和测试, 得到三种迁移分类器的测试准确率; 从骨髓细胞图像数据集中的图像提取纹理特征 LBP, 形状特征 HOG 和颜色特征 HSV; 利用加权图像融合算法对提取的纹理特征 LBP, 形状特征 HOG 和颜色特征 HSV 特征图像进行融合得到特征融合图像, 构建特征融合图像数据集, 利用 VGG16, DenseNet121 和 ResNet50 迁

移分类器对特征融合图像数据集进行训练并测试,得到每个迁移分类器中测试准确率最高的特征融合图像;利用Keras模型融合算法对VGG16,DenseNet121和ResNet50迁移分类器进行融合得到4个融合分类器,利用4个融合分类器对测试准确率最高的特征融合图像进行训练和测试,确定测试准确率最高的多特征多分类器融合模型;利用准确率最高的多特征多分类器融合模型对细胞图像进行分类识别。

[0156] 本发明再一个实施例中,本发明还提供了一种存储介质,具体为计算机可读存储介质(Memory),所述计算机可读存储介质是终端设备中的记忆设备,用于存放程序和数据。可以理解的是,此处的计算机可读存储介质既可以包括终端设备中的内置存储介质,当然也可以包括终端设备所支持的扩展存储介质。计算机可读存储介质提供存储空间,该存储空间存储了终端的操作系统。并且,在该存储空间中还存放了适于被处理器加载并执行的一条或一条以上的指令,这些指令可以是一个或一个以上的计算机程序(包括程序代码)。需要说明的是,此处的计算机可读存储介质可以是高速RAM存储器,也可以是非不稳定的存储器(non-volatile memory),例如至少一个磁盘存储器。

[0157] 可由处理器加载并执行计算机可读存储介质中存放的一条或一条以上指令,以实现上述实施例中有关基于多特征和多分类器的骨髓细胞分类识别的相应步骤;计算机可读存储介质中的一条或一条以上指令由处理器加载并执行如下步骤:

[0158] 根据骨髓细胞原始图像建立原始图像数据集;对原始图像数据集进行定位、分割和数据增强处理,得到骨髓细胞图像数据集;构建VGG16,DenseNet121和ResNet50迁移分类器,并用步骤S1中的骨髓细胞图像数据集进行训练和测试,得到三种迁移分类器的测试准确率;从骨髓细胞图像数据集中的图像提取纹理特征LBP,形状特征HOG和颜色特征HSV;利用加权图像融合算法对提取的纹理特征LBP,形状特征HOG和颜色特征HSV特征图像进行融合得到特征融合图像,构建特征融合图像数据集,利用VGG16,DenseNet121和ResNet50迁移分类器对特征融合图像数据集进行训练并测试,得到每个迁移分类器中测试准确率最高的特征融合图像;利用Keras模型融合算法对VGG16,DenseNet121和ResNet50迁移分类器进行融合得到4个融合分类器,利用4个融合分类器对测试准确率最高的特征融合图像进行训练和测试,确定测试准确率最高的多特征多分类器融合模型;利用准确率最高的多特征多分类器融合模型对细胞图像进行分类识别。

[0159] 为使本发明实施例的目的、技术方案和优点更加清楚,下面将结合本发明实施例中的附图,对本发明实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例是本发明一部分实施例,而不是全部的实施例。通常在此处附图中的描述和所示的本发明实施例的组件可以通过各种不同的配置来布置和设计。因此,以下对在附图中提供的本发明的实施例的详细描述并非旨在限制要求保护的本发明的范围,而是仅仅表示本发明的选定实施例。基于本发明中的实施例,本领域普通技术人员在没有作出创造性劳动前提下所获得的所有其他实施例,都属于本发明保护的范围。

[0160] 获取标注好的血涂片骨髓细胞图像,并对图像中的骨髓细胞进行定位分割,得到模型应用数据集;并利用步骤S4中测试效果最好的多特征多分类器融合模型对该数据集进行测试,得到最终的模型应用效果,如图15所示。

[0161] 请参阅图6,基于医院血涂片的模型应用实践步骤如下:

[0162] 1) 从西安交通大学一附院血液科获取医生标注好的血涂片骨髓细胞图像,一共20

张；

[0163] 2) 利用YOLOv3网络对血涂片骨髓细胞图像中的细胞进行定位分割,得到分割后的骨髓细胞；

[0164] 3) 对分割后的骨髓细胞进行中值滤波,去雾操作以及数据扩增等操作,得到了供模型应用检验的骨髓细胞图像检验数据集；

[0165] 4) 利用步骤S4中得到的多特征多分类器融合模型对检验数据集进行模型应用,并利用分类指标进行评估,最终得到模型应用效果,如图15所示。

[0166] 综上所述,本发明一种基于多特征和多分类器的骨髓细胞分类识别方法及系统,首先对骨髓细胞进行了完整并准确地分割;其次利用迁移学习的思想构建了迁移分类器,克服了样本数据量较少的缺点;然后构建了多特征和多分类器融合模型,结合了多种分类器和多种特征的特点和优势,提高了骨髓细胞分类识别的准确率;最后将模型应用到现实场景下,验证了本发明的准确性和可行性。

[0167] 本领域内的技术人员应明白,本申请的实施例可提供为方法、系统、或计算机程序产品。因此,本申请可采用完全硬件实施例、完全软件实施例、或结合软件和硬件方面的实施例的形式。而且,本申请可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质(包括但不限于磁盘存储器、CD-ROM、光学存储器等)上实施的计算机程序产品的形式。

[0168] 本申请是参照根据本申请实施例的方法、设备(系统)、和计算机程序产品的流程图和/或方框图来描述的。应理解可由计算机程序指令实现流程图和/或方框图中的每一流程和/或方框、以及流程图和/或方框图中的流程和/或方框的结合。可提供这些计算机程序指令到通用计算机、专用计算机、嵌入式处理机或其他可编程数据处理设备的处理器以产生一个机器,使得通过计算机或其他可编程数据处理设备的处理器执行的指令产生用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的装置。

[0169] 这些计算机程序指令也可存储在能引导计算机或其他可编程数据处理设备以特定方式工作的计算机可读存储器中,使得存储在该计算机可读存储器中的指令产生包括指令装置的制造品,该指令装置实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能。

[0170] 这些计算机程序指令也可装载到计算机或其他可编程数据处理设备上,使得在计算机或其他可编程设备上执行一系列操作步骤以产生计算机实现的处理,从而在计算机或其他可编程设备上执行的指令提供用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的步骤。

[0171] 以上内容仅为说明本发明的技术思想,不能以此限定本发明的保护范围,凡是按照本发明提出的技术思想,在技术方案基础上所做的任何改动,均落入本发明权利要求书的保护范围之内。

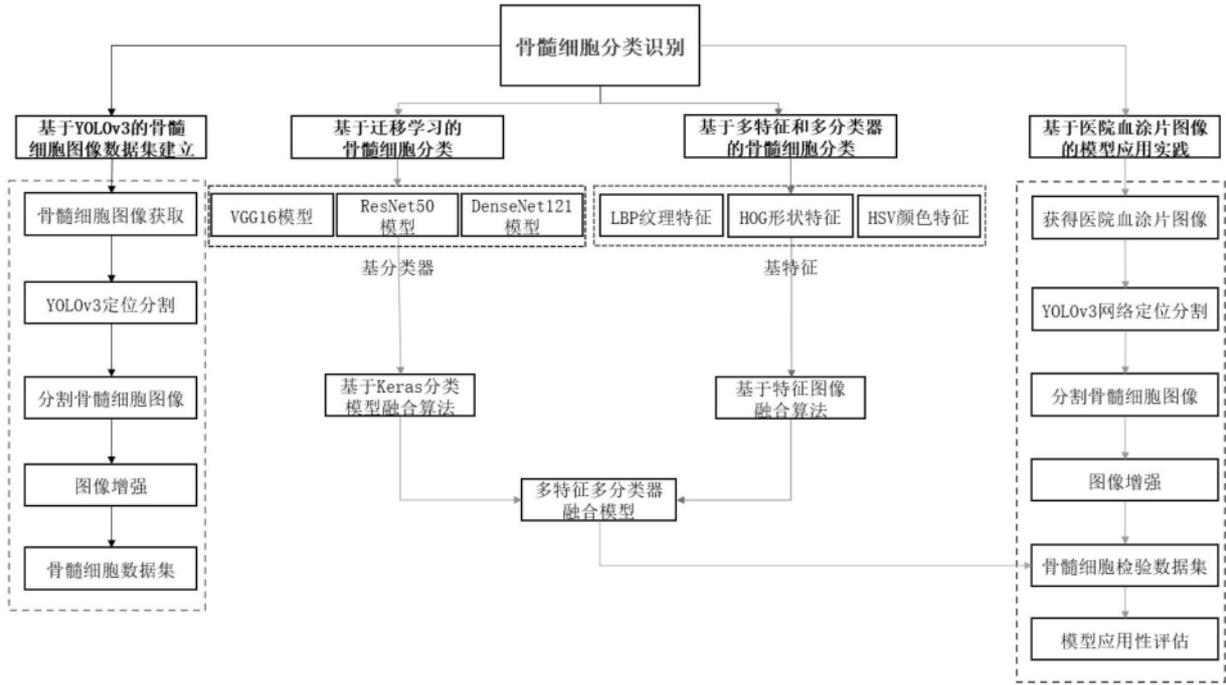


图1

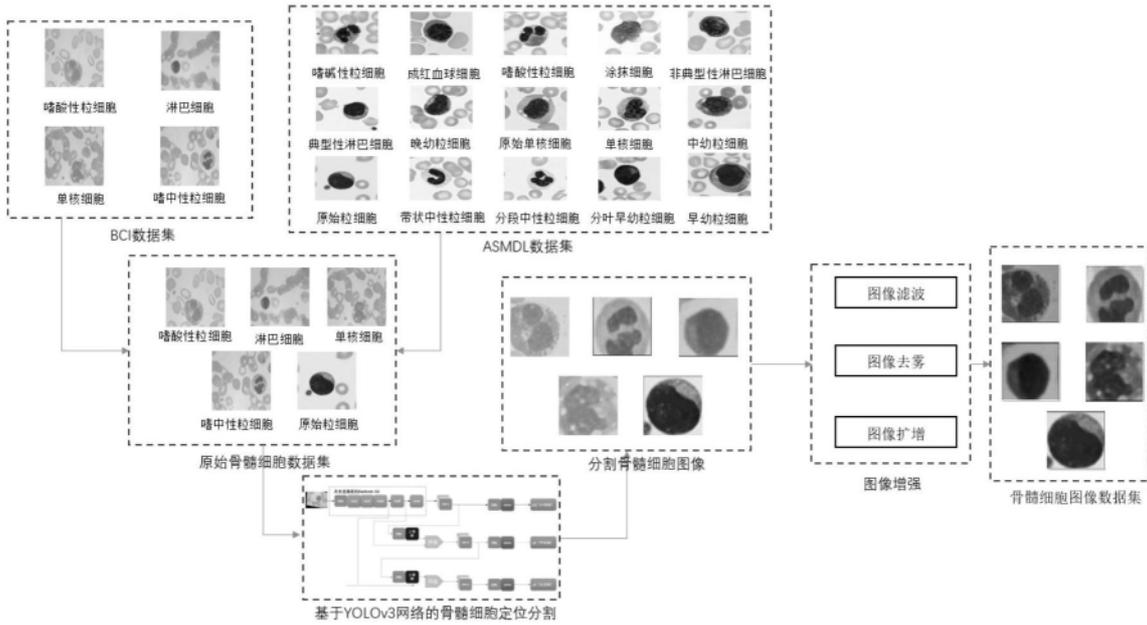


图2

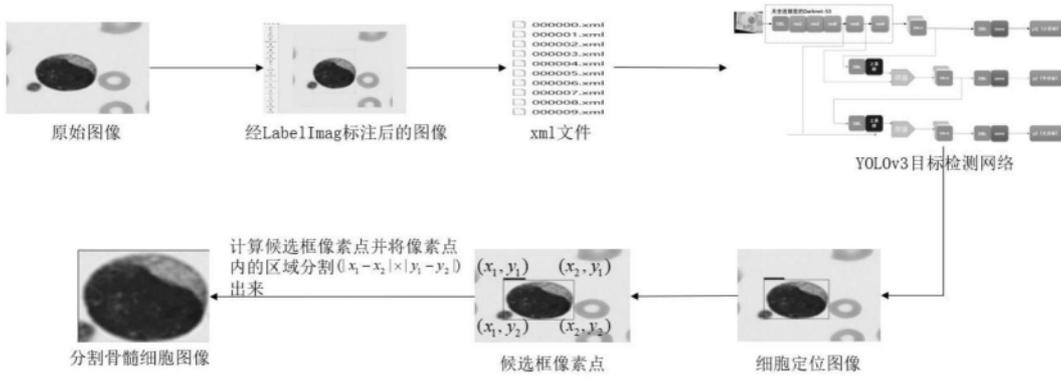


图3

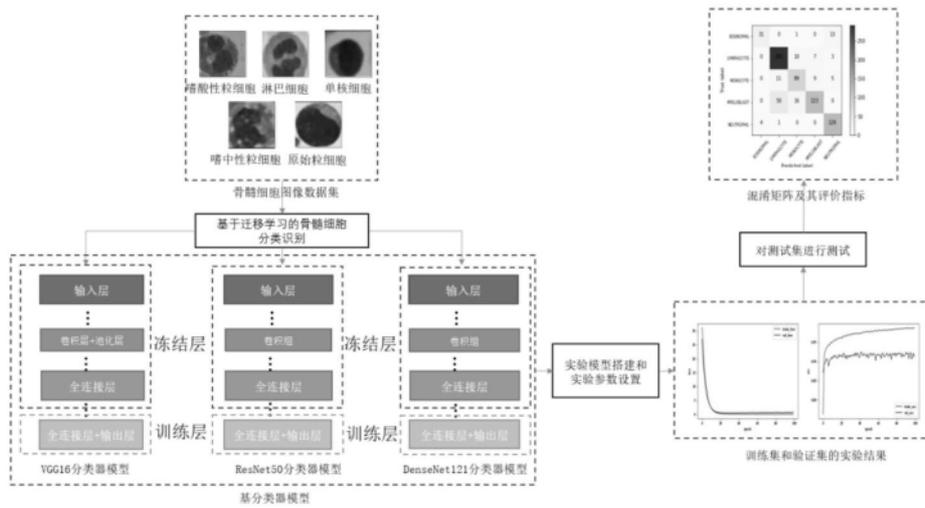


图4

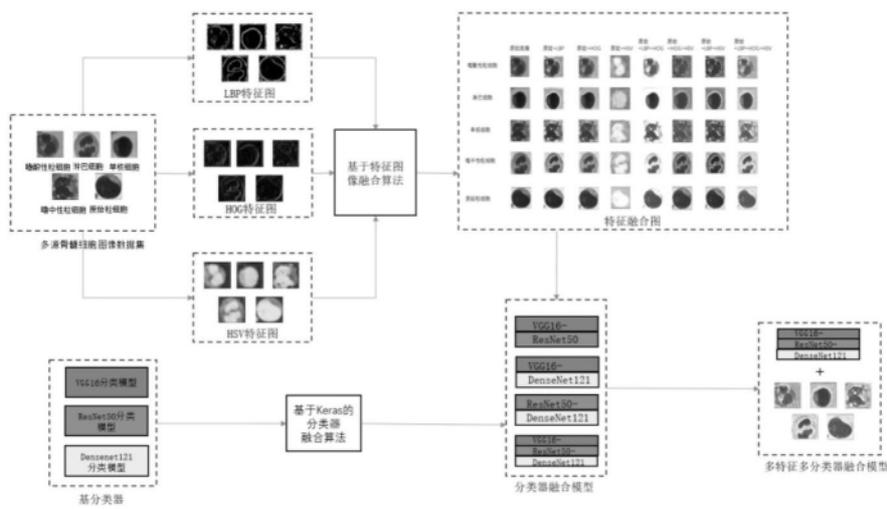


图5

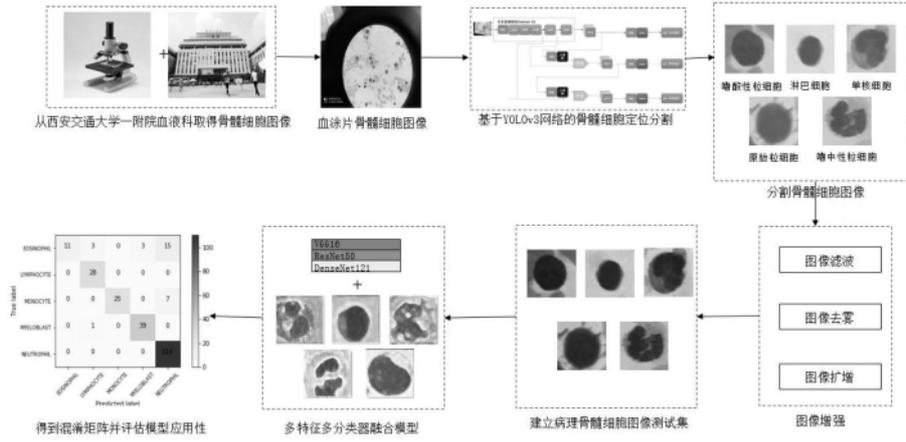


图6

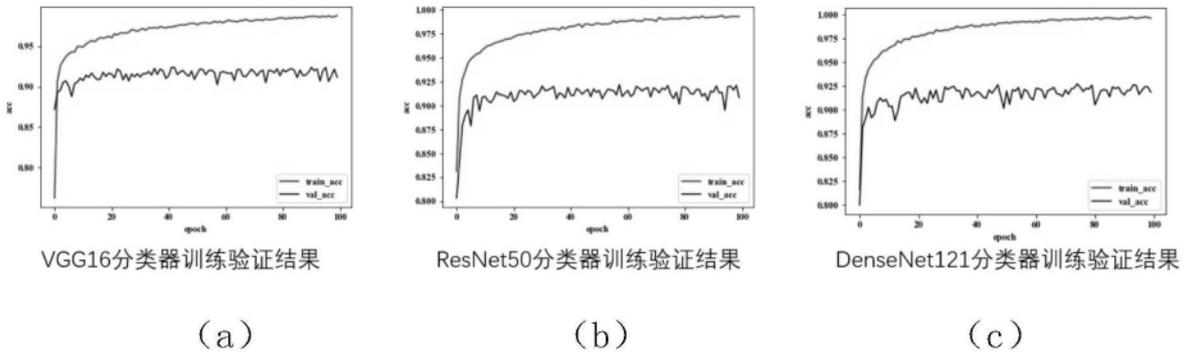


图7

| 不同分类器模型分类指标 | | | | | | | |
|---------------|------|--------|--------|--------|--------|--------|--------|
| | | 嗜酸性粒细胞 | 淋巴细胞 | 单核细胞 | 原始粒细胞 | 嗜中性粒细胞 | 平均值 |
| VGG16模型 | 准确率 | 0.6889 | 0.9357 | 0.7807 | 0.6508 | 0.9624 | 0.8359 |
| | 精确率 | 0.8857 | 0.8244 | 0.7672 | 0.8849 | 0.8591 | 0.8399 |
| | 召回率 | 0.6889 | 0.9357 | 0.7807 | 0.6508 | 0.9624 | 0.8359 |
| | F1分数 | 0.7750 | 0.8765 | 0.7739 | 0.7500 | 0.9078 | 0.8310 |
| ResNet50模型 | 准确率 | 0.9111 | 0.8585 | 0.9123 | 0.9365 | 0.9774 | 0.9078 |
| | 精确率 | 0.9111 | 0.9745 | 0.8387 | 0.8469 | 0.9286 | 0.9132 |
| | 召回率 | 0.9111 | 0.8585 | 0.9123 | 0.9365 | 0.9774 | 0.9078 |
| | F1分数 | 0.9111 | 0.9128 | 0.8739 | 0.8894 | 0.9524 | 0.9082 |
| DenseNet121模型 | 准确率 | 0.8889 | 0.9678 | 0.8772 | 0.8307 | 0.9774 | 0.9192 |
| | 精确率 | 0.9302 | 0.9012 | 0.9091 | 0.9458 | 0.9353 | 0.9203 |
| | 召回率 | 0.8889 | 0.9678 | 0.8772 | 0.8307 | 0.9774 | 0.9192 |
| | F1分数 | 0.9091 | 0.9333 | 0.8929 | 0.8845 | 0.9559 | 0.9183 |

图8

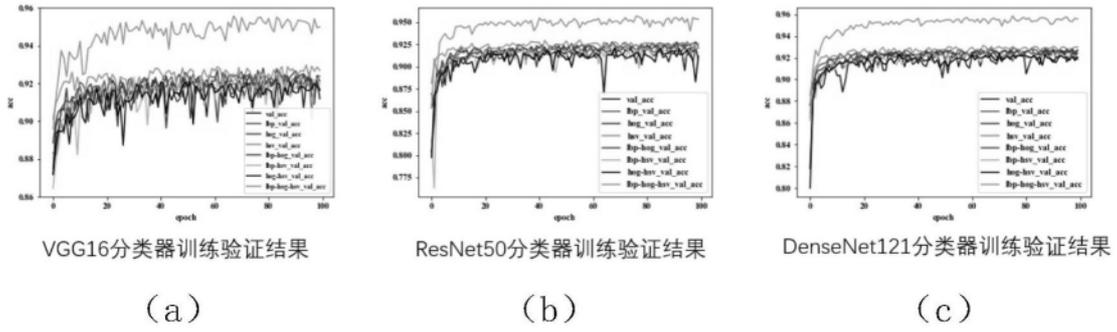


图9

| VGG16特征融合模型分类指标 | | | | | | | |
|-----------------|------|--------|--------|--------|--------|--------|--------|
| | | 嗜酸性粒细胞 | 淋巴细胞 | 单核细胞 | 原始粒细胞 | 嗜中性粒细胞 | 平均值 |
| 原始特征 | 准确率 | 0.6889 | 0.9357 | 0.7807 | 0.6508 | 0.9624 | 0.8359 |
| | 精确率 | 0.8857 | 0.8244 | 0.7672 | 0.8849 | 0.8591 | 0.8399 |
| | 召回率 | 0.6889 | 0.9357 | 0.7807 | 0.6508 | 0.9624 | 0.8359 |
| | F1分数 | 0.7750 | 0.8765 | 0.7739 | 0.7500 | 0.9078 | 0.8310 |
| LBP特征 | 准确率 | 0.9111 | 0.9293 | 0.8333 | 0.9206 | 0.9699 | 0.9192 |
| | 精确率 | 0.8913 | 0.9293 | 0.9406 | 0.8923 | 0.9281 | 0.9197 |
| | 召回率 | 0.9111 | 0.9293 | 0.8333 | 0.9206 | 0.9699 | 0.9192 |
| | F1分数 | 0.9011 | 0.9293 | 0.8837 | 0.9062 | 0.9485 | 0.9189 |
| HOG特征 | 准确率 | 0.9333 | 0.9646 | 0.8860 | 0.8042 | 0.9850 | 0.9167 |
| | 精确率 | 0.9545 | 0.8850 | 0.9018 | 0.9441 | 0.9632 | 0.9186 |
| | 召回率 | 0.9333 | 0.9646 | 0.8860 | 0.8042 | 0.9850 | 0.9167 |
| | F1分数 | 0.9438 | 0.9231 | 0.8938 | 0.8686 | 0.9740 | 0.9156 |
| HSV特征 | 准确率 | 0.9333 | 0.9100 | 0.8246 | 0.9577 | 0.9624 | 0.9192 |
| | 精确率 | 0.8750 | 0.9626 | 0.9495 | 0.8538 | 0.9209 | 0.9228 |
| | 召回率 | 0.9333 | 0.9100 | 0.8246 | 0.9577 | 0.9624 | 0.9192 |
| | F1分数 | 0.9032 | 0.9355 | 0.8826 | 0.9027 | 0.9412 | 0.9192 |
| LBP+HOG特征 | 准确率 | 0.9778 | 0.9550 | 0.8596 | 0.9418 | 0.9624 | 0.9407 |
| | 精确率 | 0.8800 | 0.9581 | 0.9423 | 0.9128 | 0.9624 | 0.9413 |
| | 召回率 | 0.9778 | 0.9550 | 0.8596 | 0.9418 | 0.9624 | 0.9407 |
| | F1分数 | 0.9263 | 0.9565 | 0.8991 | 0.9271 | 0.9624 | 0.9405 |
| HOG+HSV特征 | 准确率 | 0.9111 | 0.9389 | 0.8421 | 0.9524 | 0.9699 | 0.9318 |
| | 精确率 | 0.9111 | 0.9481 | 0.9231 | 0.9000 | 0.9556 | 0.9322 |
| | 召回率 | 0.9111 | 0.9389 | 0.8421 | 0.9524 | 0.9699 | 0.9318 |
| | F1分数 | 0.9111 | 0.9435 | 0.8807 | 0.9254 | 0.9627 | 0.9315 |
| LBP+HSV特征 | 准确率 | 0.9111 | 0.9453 | 0.9035 | 0.8677 | 0.9850 | 0.9255 |
| | 精确率 | 0.9535 | 0.9216 | 0.9035 | 0.9425 | 0.9225 | 0.9260 |
| | 召回率 | 0.9111 | 0.9453 | 0.9035 | 0.8677 | 0.9850 | 0.9255 |
| | F1分数 | 0.9318 | 0.9333 | 0.9035 | 0.9036 | 0.9527 | 0.9251 |
| LBP+HOG+HSV特征 | 准确率 | 0.9556 | 0.9453 | 0.8860 | 0.9735 | 0.9925 | 0.9520 |
| | 精确率 | 0.9556 | 0.9767 | 0.9352 | 0.9200 | 0.9565 | 0.9526 |
| | 召回率 | 0.9556 | 0.9453 | 0.8860 | 0.9735 | 0.9925 | 0.9520 |
| | F1分数 | 0.9556 | 0.9608 | 0.9099 | 0.9460 | 0.9742 | 0.9519 |

图10

| ResNet50特征融合模型分类指标 | | | | | | | |
|--------------------|------|--------|--------|--------|--------|--------|--------|
| | | 嗜酸性粒细胞 | 淋巴细胞 | 单核细胞 | 原始粒细胞 | 嗜中性粒细胞 | 平均值 |
| 原始特征 | 准确率 | 0.9111 | 0.8585 | 0.9123 | 0.9365 | 0.9774 | 0.9078 |
| | 精确率 | 0.9111 | 0.9745 | 0.8387 | 0.8469 | 0.9286 | 0.9132 |
| | 召回率 | 0.9111 | 0.8585 | 0.9123 | 0.9365 | 0.9774 | 0.9078 |
| | F1分数 | 0.9111 | 0.9128 | 0.8739 | 0.8894 | 0.9524 | 0.9082 |
| LBP特征 | 准确率 | 0.8444 | 0.9453 | 0.8772 | 0.8889 | 0.9549 | 0.9179 |
| | 精确率 | 0.9048 | 0.9274 | 0.9009 | 0.9130 | 0.9203 | 0.9177 |
| | 召回率 | 0.8444 | 0.9453 | 0.8772 | 0.8889 | 0.9549 | 0.9179 |
| | F1分数 | 0.8736 | 0.9363 | 0.8889 | 0.9008 | 0.9373 | 0.9176 |
| HOG特征 | 准确率 | 0.9556 | 0.9646 | 0.8158 | 0.8783 | 0.9699 | 0.9230 |
| | 精确率 | 0.9149 | 0.9036 | 0.9394 | 0.9222 | 0.9627 | 0.9238 |
| | 召回率 | 0.9556 | 0.9646 | 0.8158 | 0.8783 | 0.9699 | 0.9230 |
| | F1分数 | 0.9348 | 0.9331 | 0.8732 | 0.8997 | 0.9663 | 0.9222 |
| HSV特征 | 准确率 | 0.8889 | 0.9582 | 0.8684 | 0.8201 | 0.9850 | 0.9129 |
| | 精确率 | 0.9524 | 0.8817 | 0.9252 | 0.9337 | 0.9424 | 0.9146 |
| | 召回率 | 0.8889 | 0.9582 | 0.8684 | 0.8201 | 0.9850 | 0.9129 |
| | F1分数 | 0.9195 | 0.9183 | 0.8959 | 0.8732 | 0.9632 | 0.9120 |
| LBP+HOG特征 | 准确率 | 0.9556 | 0.8947 | 0.9453 | 0.9206 | 0.9774 | 0.9381 |
| | 精确率 | 0.9348 | 0.9189 | 0.9515 | 0.9255 | 0.9420 | 0.9381 |
| | 召回率 | 0.9556 | 0.8947 | 0.9453 | 0.9206 | 0.9774 | 0.9381 |
| | F1分数 | 0.9451 | 0.9067 | 0.9484 | 0.9231 | 0.9594 | 0.9380 |
| HOG+HSV特征 | 准确率 | 0.9556 | 0.9550 | 0.8509 | 0.8995 | 0.9850 | 0.9318 |
| | 精确率 | 0.9556 | 0.9310 | 0.9327 | 0.9043 | 0.9632 | 0.9317 |
| | 召回率 | 0.9556 | 0.9550 | 0.8509 | 0.8995 | 0.9850 | 0.9318 |
| | F1分数 | 0.9556 | 0.9429 | 0.8899 | 0.9019 | 0.9740 | 0.9314 |
| LBP+HSV特征 | 准确率 | 0.9111 | 0.9228 | 0.8772 | 0.9259 | 0.9925 | 0.9280 |
| | 精确率 | 0.9535 | 0.9535 | 0.8850 | 0.9067 | 0.9296 | 0.9285 |
| | 召回率 | 0.9111 | 0.9228 | 0.8772 | 0.9259 | 0.9925 | 0.9280 |
| | F1分数 | 0.9318 | 0.9379 | 0.8811 | 0.9162 | 0.9600 | 0.9279 |
| LBP+HOG+HSV特征 | 准确率 | 0.9778 | 0.9518 | 0.9211 | 0.9206 | 0.9850 | 0.9470 |
| | 精确率 | 0.9778 | 0.9457 | 0.9545 | 0.9206 | 0.9704 | 0.9470 |
| | 召回率 | 0.9778 | 0.9518 | 0.9211 | 0.9206 | 0.9850 | 0.9470 |
| | F1分数 | 0.9778 | 0.9487 | 0.9375 | 0.9206 | 0.9776 | 0.9469 |

图11

| DenseNet121特征融合模型分类指标 | | | | | | | |
|-----------------------|------|--------|--------|--------|--------|--------|--------|
| | | 嗜酸性粒细胞 | 淋巴细胞 | 单核细胞 | 原始粒细胞 | 嗜中性粒细胞 | 平均值 |
| 原始特征 | 准确率 | 0.8889 | 0.9678 | 0.8772 | 0.8307 | 0.9774 | 0.9192 |
| | 精确率 | 0.9302 | 0.9012 | 0.9091 | 0.9458 | 0.9353 | 0.9203 |
| | 召回率 | 0.8889 | 0.9678 | 0.8772 | 0.8307 | 0.9774 | 0.9192 |
| | F1分数 | 0.9091 | 0.9333 | 0.8929 | 0.8945 | 0.9559 | 0.9183 |
| LBP特征 | 准确率 | 0.9333 | 0.9486 | 0.8421 | 0.9101 | 0.9774 | 0.9280 |
| | 精确率 | 0.9545 | 0.9306 | 0.9143 | 0.9005 | 0.9630 | 0.9279 |
| | 召回率 | 0.9333 | 0.9486 | 0.8421 | 0.9101 | 0.9774 | 0.9280 |
| | F1分数 | 0.9438 | 0.9395 | 0.8767 | 0.9053 | 0.9701 | 0.9277 |
| HOG特征 | 准确率 | 0.9333 | 0.9421 | 0.8246 | 0.9365 | 0.9774 | 0.9293 |
| | 精确率 | 0.9333 | 0.9361 | 0.9307 | 0.9031 | 0.9489 | 0.9294 |
| | 召回率 | 0.9333 | 0.9421 | 0.8246 | 0.9365 | 0.9774 | 0.9293 |
| | F1分数 | 0.9333 | 0.9391 | 0.8744 | 0.9195 | 0.9630 | 0.9288 |
| HSV特征 | 准确率 | 0.9111 | 0.9035 | 0.8596 | 0.9471 | 0.9699 | 0.9192 |
| | 精确率 | 0.9762 | 0.9656 | 0.8991 | 0.8443 | 0.9348 | 0.9225 |
| | 召回率 | 0.9111 | 0.9035 | 0.8596 | 0.9471 | 0.9699 | 0.9192 |
| | F1分数 | 0.9425 | 0.9336 | 0.8789 | 0.8928 | 0.9520 | 0.9196 |
| LBP+HOG特征 | 准确率 | 0.9333 | 0.9678 | 0.8070 | 0.9048 | 0.9850 | 0.9306 |
| | 精确率 | 0.9767 | 0.9094 | 0.9485 | 0.9396 | 0.9424 | 0.9316 |
| | 召回率 | 0.9333 | 0.9678 | 0.8070 | 0.9048 | 0.9850 | 0.9306 |
| | F1分数 | 0.9545 | 0.9377 | 0.8720 | 0.9218 | 0.9632 | 0.9297 |
| HOG+HSV特征 | 准确率 | 0.9556 | 0.9293 | 0.9211 | 0.9418 | 0.9699 | 0.9394 |
| | 精确率 | 0.9556 | 0.9764 | 0.8824 | 0.9128 | 0.9416 | 0.9406 |
| | 召回率 | 0.9556 | 0.9293 | 0.9211 | 0.9418 | 0.9699 | 0.9394 |
| | F1分数 | 0.9556 | 0.9522 | 0.9013 | 0.9271 | 0.9556 | 0.9396 |
| LBP+HSV特征 | 准确率 | 0.9333 | 0.9614 | 0.8333 | 0.9365 | 0.9850 | 0.9394 |
| | 精确率 | 0.9545 | 0.9522 | 0.9500 | 0.9031 | 0.9493 | 0.9398 |
| | 召回率 | 0.9333 | 0.9614 | 0.8333 | 0.9365 | 0.9850 | 0.9394 |
| | F1分数 | 0.9438 | 0.9568 | 0.8879 | 0.9195 | 0.9668 | 0.9389 |
| LBP+HOG+HSV特征 | 准确率 | 0.9333 | 0.9582 | 0.9386 | 0.9418 | 0.9850 | 0.9545 |
| | 精确率 | 0.9545 | 0.9675 | 0.9224 | 0.9519 | 0.9562 | 0.9547 |
| | 召回率 | 0.9333 | 0.9582 | 0.9386 | 0.9418 | 0.9850 | 0.9545 |
| | F1分数 | 0.9438 | 0.9628 | 0.9304 | 0.9468 | 0.9704 | 0.9545 |

图12

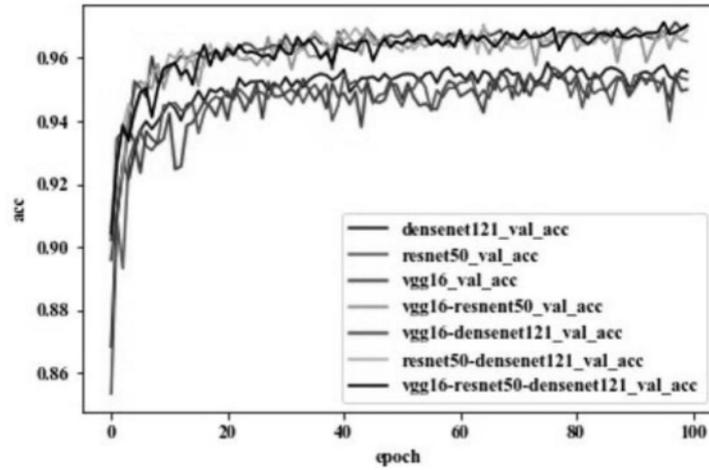


图13

| 多特征多分类器融合模型分类指标 | | | | | | | |
|------------------------------------|------|--------|--------|--------|--------|--------|--------|
| | | 嗜酸性粒细胞 | 淋巴细胞 | 单核细胞 | 原始粒细胞 | 嗜中性粒细胞 | 平均值 |
| vgg16多特征融合模型 | 准确率 | 0.9556 | 0.9453 | 0.8860 | 0.9735 | 0.9925 | 0.9520 |
| | 精确率 | 0.9556 | 0.9767 | 0.9352 | 0.9200 | 0.9565 | 0.9526 |
| | 召回率 | 0.9556 | 0.9453 | 0.8860 | 0.9735 | 0.9925 | 0.9520 |
| | F1分数 | 0.9556 | 0.9608 | 0.9099 | 0.9460 | 0.9742 | 0.9519 |
| resnet50多特征融合模型 | 准确率 | 0.9778 | 0.9518 | 0.9211 | 0.9206 | 0.9850 | 0.9470 |
| | 精确率 | 0.9778 | 0.9457 | 0.9545 | 0.9206 | 0.9704 | 0.9470 |
| | 召回率 | 0.9778 | 0.9518 | 0.9211 | 0.9206 | 0.9850 | 0.9470 |
| | F1分数 | 0.9778 | 0.9487 | 0.9375 | 0.9206 | 0.9776 | 0.9469 |
| densenet121多特征融合模型 | 准确率 | 0.9333 | 0.9582 | 0.9386 | 0.9418 | 0.9850 | 0.9545 |
| | 精确率 | 0.9545 | 0.9675 | 0.9224 | 0.9519 | 0.9562 | 0.9547 |
| | 召回率 | 0.9333 | 0.9582 | 0.9386 | 0.9418 | 0.9850 | 0.9545 |
| | F1分数 | 0.9438 | 0.9628 | 0.9304 | 0.9468 | 0.9704 | 0.9545 |
| vgg16-resnet50多分类器融合模型 | 准确率 | 0.8222 | 0.9807 | 0.9561 | 0.9365 | 0.9925 | 0.9596 |
| | 精确率 | 1.0000 | 0.9591 | 0.9160 | 0.9779 | 0.9635 | 0.9604 |
| | 召回率 | 0.8222 | 0.9807 | 0.9561 | 0.9365 | 0.9925 | 0.9596 |
| | F1分数 | 0.9024 | 0.9698 | 0.9356 | 0.9568 | 0.9778 | 0.9593 |
| vgg16-densenet121多分类器融合模型 | 准确率 | 0.7778 | 0.9871 | 0.9561 | 0.9524 | 0.9925 | 0.9634 |
| | 精确率 | 1.0000 | 0.9654 | 0.9646 | 0.9890 | 0.9167 | 0.9647 |
| | 召回率 | 0.7778 | 0.9871 | 0.9561 | 0.9524 | 0.9925 | 0.9634 |
| | F1分数 | 0.8750 | 0.9762 | 0.9604 | 0.9704 | 0.9531 | 0.9629 |
| resnet50-densenet121多分类器融合模型 | 准确率 | 0.7556 | 0.9936 | 0.9737 | 0.9365 | 0.9850 | 0.9621 |
| | 精确率 | 1.0000 | 0.9626 | 0.9652 | 0.9833 | 0.9225 | 0.9633 |
| | 召回率 | 0.7556 | 0.9936 | 0.9737 | 0.9365 | 0.9850 | 0.9621 |
| | F1分数 | 0.8608 | 0.9778 | 0.9694 | 0.9593 | 0.9527 | 0.9614 |
| vgg16-resnet50-densenet121多分类器融合模型 | 准确率 | 0.8222 | 0.9807 | 0.9825 | 0.9471 | 0.9925 | 0.9659 |
| | 精确率 | 0.9737 | 0.9807 | 0.9256 | 0.9835 | 0.9429 | 0.9667 |
| | 召回率 | 0.8222 | 0.9807 | 0.9825 | 0.9471 | 0.9925 | 0.9659 |
| | F1分数 | 0.8916 | 0.9807 | 0.9532 | 0.9650 | 0.9670 | 0.9656 |

图14

| 医院病理图像分类指标 | | | | | | |
|------------|--------|--------|--------|--------|--------|--------|
| | 嗜酸性粒细胞 | 淋巴细胞 | 单核细胞 | 原始粒细胞 | 嗜中性粒细胞 | 平均值 |
| 准确率 | 0.7812 | 0.8571 | 0.8750 | 1.0000 | 0.9455 | 0.9132 |
| 精确率 | 0.8929 | 0.7500 | 1.0000 | 0.8696 | 0.9630 | 0.9185 |
| 召回率 | 0.7812 | 0.8571 | 0.8750 | 1.0000 | 0.9455 | 0.9132 |
| F1分数 | 0.8333 | 0.8000 | 0.9333 | 0.9302 | 0.9541 | 0.9136 |

图15