

基于多证据的血液白细胞自动分类

庄杨凯,周 平

(浙江理工大学信息学院,杭州 310018)

摘 要:对白细胞分类方法提出了一种基于形状特征学习的分类算法。该算法首先从细胞核中提取基于欧氏距离变换的不变矩特征和形态学特征;然后,所提取的特征可以实现对单核细胞、淋巴细胞、嗜碱性粒细胞的分类;最后,提取灰度共生矩阵作为纹理特征,通过支持向量机实现对剩余白细胞类别的分类。实验结果表明,该方法具有很好的分类准确率及较短的处理时间。

关键词:白细胞分类;欧氏距离变换;不变矩;形态学特征;纹理特征;支持向量机

中图分类号: TP391.4 **文献标志码:** A

0 引 言

对血液中的白细胞的检查和分类可以协助医生诊断诸如艾滋病、白血病等疾病,使得其成为血液检测程序重要的步骤之一^[1]。根据白细胞细胞核的大小及形状,可将其分为嗜碱性粒细胞、嗜酸性粒细胞、淋巴细胞、单核细胞和中性粒细胞 5 大类。对白细胞的分类可以通过人工或者自动的方法实现,然而人工分类方法很耗时,需要经过专门训练的人员,而采用包含图像处理技术的自动分类方法可以提供高质量的分类型结果。

尽管对白细胞的分类首先需要将白细胞检测出来,然后进行分割,但目前已有较多的方法可以实现^[1-7],本文采用文献[1,7]中的方法实现细胞核和细胞质的分割,重点专注于特征提取。

目前很多血液白细胞分类基于纹理特征的基础上,结合形状特征和颜色特征,但是由于特征数过多,以及部分特征鲁棒性不高,导致分类准确率低,而且都是将提取的特征让分类器直接分类,误判范围大。Rezatofghi 等^[1]、Ko 等^[8]提取颜色、形状和纹理特征,但是特征数过多且在时间耗费上跟手工方法差不多,准确率不高。Huang 等^[7]也采用了灰度共生矩阵,再结合部分形状特征,但对淋巴细胞和

中性粒细胞的分类准确率不高。Sabino 等^[9]仅提取了基于灰度共生矩阵的 5 个纹理特征,部分类别的分类效果不理想。Theera-Umpon^[10]提取了包括细胞核面积、模式频谱的峰值位置,细胞核的第一和第二不变矩 4 个特征用于分类,但平均分类准确率只有 70%。Ramirez-Cortes 等^[11]采用形态学频谱模式来提取特征进行分类,但该算法对嗜酸性粒细胞的分类准确率很低,而且时间耗费较长。

笔者提出了一个基于形状特征的白细胞分类算法。首先对细胞分割后细胞提取形状特征,以及从细胞核提取基于欧氏距离变换的一阶矩特征;其次,提取的特征实现对单核细胞、淋巴细胞和嗜碱性粒细胞的分类;最后,对于剩余的细胞,提取灰度共生矩阵,通过支持向量机实现分类。

1 特征方法

1.1 距离变换

欧氏距离变换,用来计算二值图上一个物体点到最近背景点的距离,假设图像中目标物体的轮廓是个连续的闭合曲线 C , C 的内部为目标物体 O ,那么,对于目标 O 内任意一点,计算该点到轮廓 C 的最短距离,然后转化为灰度值就得到了该目标物体的距离变换图。对欧氏平面上,任意两点 (i_1, j_1) 和 $(i_2,$

收稿日期: 2012-11-15

作者简介: 庄杨凯(1988-),男,浙江奉化人,硕士研究生,主要从事图像处理方面的研究。

通信作者: 周 平,电子邮箱: zp@zstu.edu.cn

j_2) 间的距离表示为:

$$d_c = \sqrt{(i_1 - i_2)^2 + (j_1 - j_2)^2} \quad (1)$$

由于欧氏距离变换建立在二值图像基础上, 因而需要对图像进行分割, 5类白细胞如图1所示, 分

割后图像如图2所示。由于5类细胞的形状差不多, 以及后来的实验表明采用细胞核比采用整个细胞的结果要好, 因而, 仅对细胞核区域进行距离变换, 由图2得到的距离变换图如图3所示。

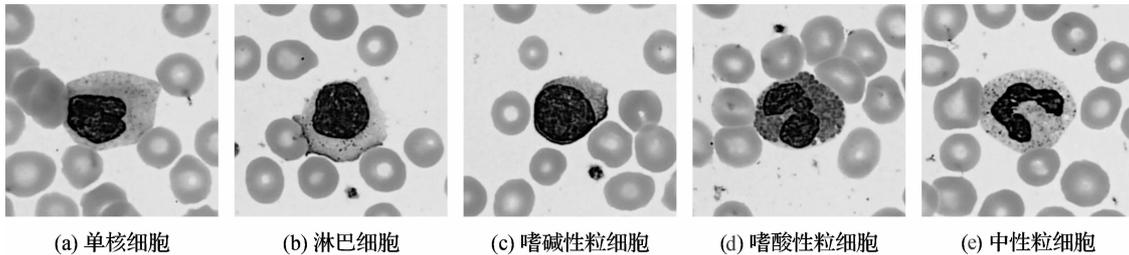


图1 5种不同类型白细胞的显微图像

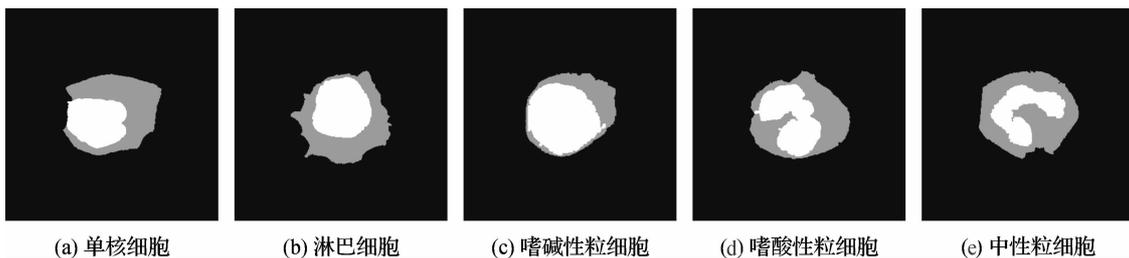


图2 根据图1得到的5类白细胞分割结果图

注: 白色区域为细胞核, 灰色为细胞质, 黑色为背景

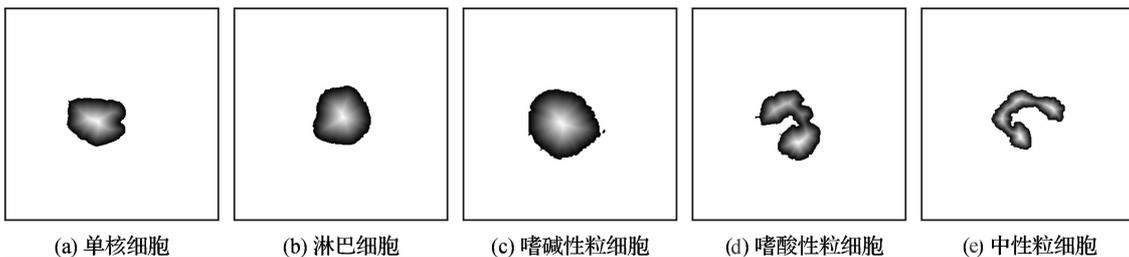


图3 由图2得到的细胞核距离变换图

1.2 特征提取

1.2.1 形态学特征

在这里, 用到的形态学特征为核质比、细胞核面积周长比。此外, 还有从细胞核距离变换图中提取的不变矩特征。

不变矩定义如下:

图像的几何矩定义为^[12]:

$$m_{p,q} = \sum_x \sum_y x^p y^q F(x,y) \quad (2)$$

式(2)中 $F(x,y)$ 为图像的二维函数, 阶数为 $(p+q)$ 。

然后, 得到中心矩定义为^[12]:

$$\mu_{p,q} = \sum_x \sum_y (x - x_c)^p (y - y_c)^q F(x,y) \quad (3)$$

式(3)中, x_c, y_c 为物体的质心坐标。

$$x_c = \frac{m_{1,0}}{m_{0,1}}, y_c = \frac{m_{0,1}}{m_{1,0}} \quad (4)$$

正态化中心矩为:

$$\eta_{p,q} = \frac{\mu_{p,q}}{\mu_{0,0}^\gamma} \quad (5)$$

式(5)中 $\gamma = \frac{p+q+2}{2}$ 。

然后得到7个不变矩的公式如下:

$$\phi_1 = (\eta_{2,0} + \eta_{0,2}) \quad (6)$$

$$\phi_2 = (\eta_{2,0} - \eta_{0,2})^2 + 4\eta_{1,1}^2 \quad (7)$$

$$\phi_3 = (\eta_{3,0} - 3\eta_{1,2})^2 + (3\eta_{2,1} - \eta_{0,3})^2 \quad (8)$$

$$\phi_4 = (\eta_{3,0} + \eta_{1,2})^2 + (\eta_{2,1} + \eta_{0,3})^2 \quad (9)$$

$$\phi_5 = (\eta_{3,0} - 3\eta_{1,2})(\eta_{3,0} + \eta_{1,2})[(\eta_{3,0} + \eta_{1,2})^2 - 3(\eta_{2,1} + \eta_{0,3})^2] + (3\eta_{2,1} - \eta_{0,3})(\eta_{2,1} + \eta_{0,3}) [3(\eta_{3,0} + \eta_{1,2})^2 - (\eta_{2,1} + \eta_{0,3})^2] \quad (10)$$

$$\phi_6 = (\eta_{2,0} - \eta_{0,2})[(\eta_{3,0} + \eta_{1,2})^2 - (\eta_{2,1} + \eta_{0,3})^2] + 4\eta_{1,1}(\eta_{3,0} + \eta_{1,2})(\eta_{2,1} + \eta_{0,3}) \quad (11)$$

$$\phi_7 = (3\eta_{2,1} - \eta_{0,3})(\eta_{3,0} + \eta_{1,2})[(\eta_{3,0} + \eta_{1,2})^2 - 3(\eta_{2,1} + \eta_{0,3})^2] + (3\eta_{2,1} - \eta_{0,3})(\eta_{2,1} + \eta_{0,3}) [3(\eta_{3,0} + \eta_{1,2})^2 - (\eta_{2,1} + \eta_{0,3})^2] \quad (12)$$

1.2.2 纹理特征

在本文中,提取的灰度共生矩阵用来实现对嗜酸性粒细胞和中性粒细胞的分类。

灰度共生矩阵建立在灰度图像基础上,与距离 d 和角度 φ 有关。假设图像的尺寸为 $M \times N$,灰度级为 N_g ,对应的矩阵为 $W = [t_{ij}]_{N_g \times N_g}$,其中 t_{ij} 是矩阵第 i 行第 j 列元素,由此可得到矩阵的定义如下^[13-14]:

$$t_{ij} = \sum_{m=1}^M \sum_{n=1}^N \delta_{nm} \quad (13)$$

式(13)中,

$$\delta_{nm} = \begin{cases} 1 & f(m,n) = i \text{ 且 } f(m,n+1) = j \\ 0, & \text{其他} \end{cases} \quad (14)$$

$f(m,n)$ 是灰度图像第 i 行第 j 列对应的像素值。

然后对应的概率密度函数通过正态化矩阵得到,定义如下:

$$p_{ij} = \frac{t_{ij}}{\sum_m \sum_n t_{nm}} \quad (15)$$

可以从共生矩阵中提取到对比度、同质性、熵等 14 个纹理特征。为使特征具有旋转不变性,取 4 个不同的角度 φ ,得到 4 个不同的矩阵,然后计算特征的平均值来实现旋转不变性,在本文中,提取对比度、熵、能量、同质性和相关度一共 5 个纹理特征用于分类。

2 实验结果

实验中所采用的血液白细胞样本图像取自瑞典 CellaVision AB 公司的血液分析仪中的数据库,共有 103 个包含 5 类白细胞的图像用于训练获取所需数据。

首先,由于细胞包含细胞核和细胞质两部分,对细胞核和细胞的一阶矩效果图进行了比较,如图 4 和图 5 所示。

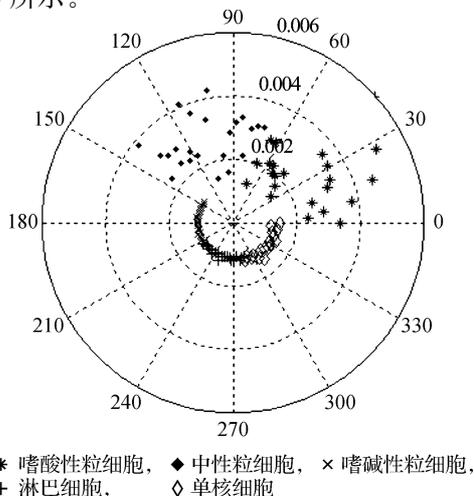


图 4 5 类白细胞细胞核距离变换图的不变矩一阶矩

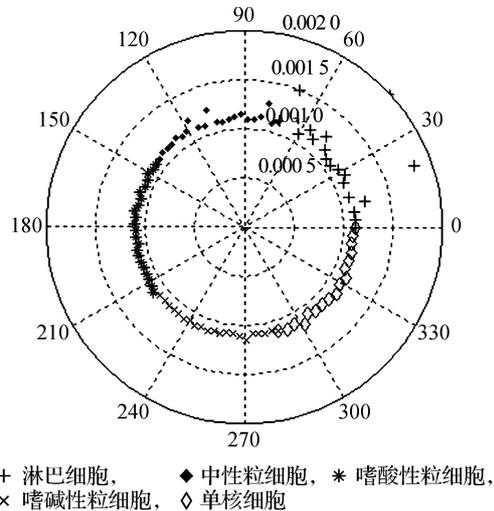


图 5 5 类白细胞整个细胞距离变换图的不变矩一阶矩

比较后可以发现,采用细胞核区域的一阶矩分类效果比用整个细胞的要好,这是由于 5 类白细胞形状大小差不多,但是细胞核差别大。根据图 4,通过贝叶斯分类器得到的分类阈值为 0.001 2 和 0.001 6,由此可将白细胞大致分为 3 类,低于 0.001 2 的,为嗜碱性粒细胞和淋巴细胞部分,同时包含个别单核细胞;在 0.001 2~0.001 6 处,为单核细胞部分,含少部分淋巴细胞、嗜酸性细胞和中性粒细胞;而大于 0.001 6 的,为嗜酸性粒细胞和中性粒细胞部分。

然后,结合 5 类细胞核质比,如图 6 所示,通过贝叶斯分类器,得到的核质比分类阈值为 1.75,大于阈值的为嗜碱性粒细胞,再结合上一步,淋巴细胞也分出来。然后由细胞核面积周长比,如图 7 所示,通过贝叶斯分类器得到的阈值为 20,将 5 类细胞分为两大类:第一大类为单核细胞,淋巴细胞及嗜碱性粒细胞,第二类为嗜酸性粒细胞和中性粒细胞。除去上两步已经分类出来的嗜碱性粒细胞和淋巴细胞,得到了单核细胞。由此,根据 3 个特征,将单核细胞、淋巴细胞和嗜碱性粒细胞分离出来了。

最后,剩余的嗜酸性粒细胞和中性粒细胞通过灰度共生矩阵和支持向量机分类实现。

根据上面通过分类器得到的 3 个形状特征的 4 个阈值,从瑞典 CellaVision AB 公司的血液分析仪中的数据库中另外选取 106 个细胞进行测试,测试结果如表 1 所示。

在同样样本下,各种分类方法与本文提出的方法的结果对各类细胞分类准确率的比较如表 2 所示,所采用的白细胞图像均取自瑞典 CellaVision AB 公司的血液分析仪中的数据库。

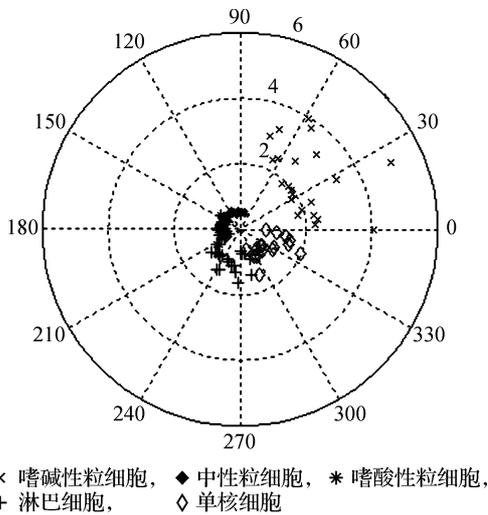


图6 细胞核和细胞质面积比

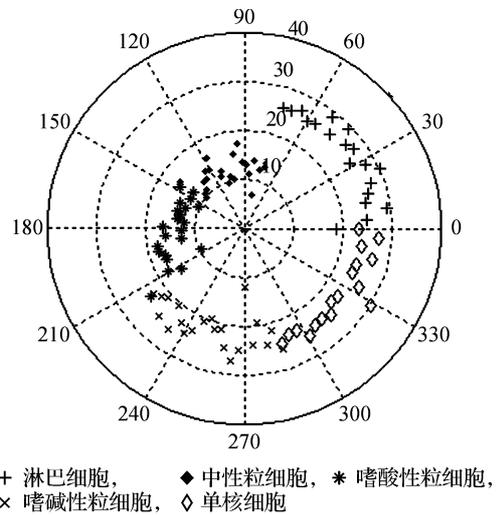


图7 细胞核面积周长比

表1 5类白细胞分类结果

	单核细胞	淋巴细胞	嗜碱性粒细胞	嗜酸性粒细胞	中性粒细胞	分类准确率/%
单核细胞	16	1	0	0	0	94.1
淋巴细胞	1	19	0	0	0	95.0
嗜碱性粒细胞	0	0	22	0	0	100
嗜酸性粒细胞	1	0	0	21	1	91.3
中性粒细胞	0	0	0	0	24	100
总分类						96.2

表2 各种分类方法的结果比较

方法	分类准确率/%				
	单核细胞	淋巴细胞	嗜碱性粒细胞	嗜酸性粒细胞	中性粒细胞
本文提出的	94.1	95.0	100	91.3	100
Rezatofighi ^[1] 等提出的	64.7	78.9	100	81.8	87.0
Ko ^[8] 等提出的	70.7	78.95	95.45	77.3	91.3
Ramirez-Cortes ^[11] 等提出的	94.1	89.4	90.9	18.2	91.3

对应的棒状图如图8所示。

从图8可以看到,本文提出的方法有较高的分类准确率,而其他的方法对部分细胞的分类效果不理想。在PC上进行实验,1 G内存,CPU为1.86 GHz,平均处理100个细胞所需时间为5 min,而相对于采用人工方法则需要15 min,耗时很少。

本文采用的特征具有很高的鲁棒性:对于基于距离变换的一阶矩,距离变换消除了染色剂对白细胞颜色的影响,一阶矩本身具有缩放和旋转不变性,因而基于距离变换的一阶矩对白细胞图像的颜色、缩放及旋转都有不变性;核质比和面积周长比特征都与白细胞的颜色、缩放及旋转无关,因而也具有颜色、缩放和旋转不变性;对于灰度共生矩阵特征,是经常使用的纹理特征,在白细胞分类中也使用,而且为保证具有旋转不变性,取四个角度的特征值的平均值作为最终分类的特征。所以,综上所述,所提出

的特征及分类算法适用于白细胞的分类。

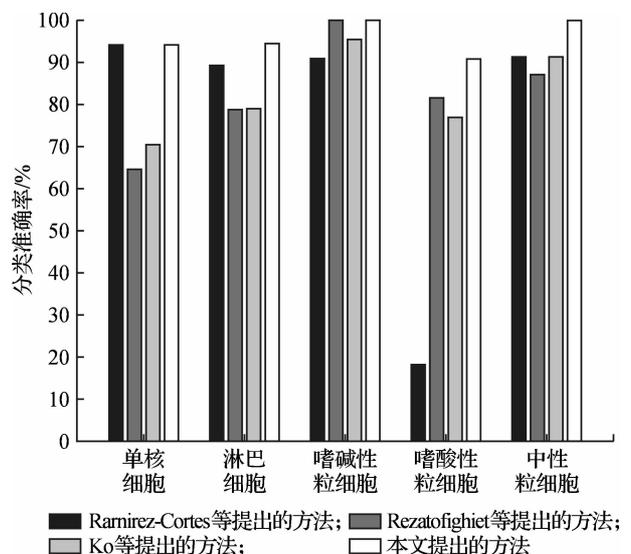


图8 本文提出的方法与其他方法实验结果的比较

虽然本文对正常白细胞分类准确率较高,但是,如果输入图像为异常白细胞图像,则会造成误判,如用于诊断急性白细胞的幼稚白细胞(如图 9 所示),以图 9 中中间的幼稚细胞为例,提取的一阶矩值为 0.001 106,核质比为 1.123 1,面积周长比为 25.83,将其判为淋巴细胞。

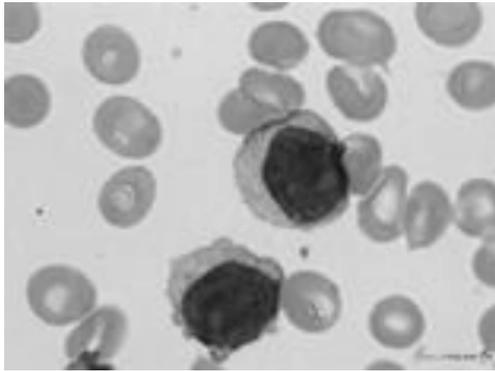


图 9 幼稚细胞

3 结 语

提出了一种新的白细胞分类方法,并对此进行了实验,该方法基于距离变换、不变矩、形状特征和灰度共生矩阵。平均分类准确率可以达到 96.2%,从特征提取到得到分类结果的处理时间短。但是对异常白细胞的分类效果不好,因而,提高区分异常白细胞的分类将是新的研究方向。

参考文献:

- [1] Rezaatofighi S H, Soltanian-Zadeh H. Automatic recognition of five types of white blood cells in peripheral blood[J]. *Computerized Medical Imaging and Graphics*, 2011, 35: 333-343.
- [2] Guo N, Zeng L, Wu Q. A method based on multispectral imaging technique for white blood cell segmentation[J]. *Computers in Biology and Medicine*, 2006, 37: 70-76.
- [3] Ghosh M, Das D, Chakraborty C, et al. Automated leukocyte recognition using fuzzy divergence[J]. *Micron*, 2010, 41: 840-846.
- [4] Ko B C, Gim J W, Nam J Y. Automatic white blood cell segmentation using stepwise merging rules and gradient vector flow snake[J]. *Micron*, 2011, 42: 695-705.
- [5] Li K, Lu Z, Liu W, et al. Cytoplasm and nucleus segmentation in cervical smear images using Radiating GVF Snake[J]. *Pattern Recognition*, 2012, 45(4): 1255-1264.
- [6] Pan C, Park D S, Yoon S, et al. Leukocyte image segmentation using simulated visual attention[J]. *Expert Systems with Applications*, 2012, 39(8): 7479-7494.
- [7] Huang D C, Huang K D, Chan Y K. A computer assisted method for leukocyte nucleus segmentation and recognition in blood smear images[J]. *The Journal of Systems and Software*, 2012, 85: 2104-2118.
- [8] Ko B C, Gim J W, Nam J Y. Cell image classification based on ensemble features and random forest[J]. *Electronics Letters*, 2011, 47(11): 638-639.
- [9] Sabino D M U, Costa L F, Rizzatti E G, et al. A texture approach to leukocyte recognition[J]. *Real-Time Imaging*, 2004, 10: 205-216.
- [10] Theera-Umpon N. White blood cell segmentation and classification in microscopic bone marrow images[J]. *Fuzzy Systems and Knowledge Discovery, Lecture Notes in Computer Science*, 2005, 3614: 787-796.
- [11] Ramirez-Cortes J M, Gomez-Gil P, Alarcon-Aquino V, et al. Neural networks and svm-based classification of leukocytes using the morphological pattern spectrum[J]. *Soft Computing for Recognition Based on Biometrics*, 2010, 312: 19-35.
- [12] Jin L, Tianxu Z. Fast algorithm for generation of moment invariants[J]. *Pattern Recognition*, 2004, 37: 1745-1756.
- [13] Honeycutt C E, Plotnick R. Image analysis techniques and gray-level co-occurrence matrices (GLCM) for calculating bioturbation indices and characterizing biogenic sedimentary structures[J]. *Computers & Geosciences*, 2008, 34: 1461-1472.
- [14] Nie F, Gao C, Guo Y, et al. Two-dimensional minimum local cross-entropy thresholding based on co-occurrence matrix[J]. *Computers and Electrical Engineering*, 2011, 37: 757-767.

(下转第 382 页)

Exploration on Induction of Breast Cancer Bcap37 Apoptosis by Oridonin

YU Yue-mei, YAN Qiao-ling, ZHANG Kang-jian, YUE Xue-tian, ZHOU Xiu-mei

(Xinyuan Institute of Medicine and Biotechnology, School of Life Sciences,
Zhejiang Sci-Tech University, Hangzhou 310018, China)

Abstract: This paper discusses the role of oridonin in the proliferation and apoptosis of breast cancer cell Bcap-37 and the mechanism causing apoptosis; treats Bcap-37 cell with Ori of different concentrations; detects cell growth inhibition rate with MTT method; detects apoptosis by using flow cytometry and the mechanism of apoptosis with Western blot. The result is that 48~96 $\mu\text{mol/L}$ Ori can significantly inhibit the proliferation of Bcap-37 cell ($P < 0.05$). After 24 h of operation of 48 $\mu\text{mol/L}$ Ori, the proliferation inhibition rate reaches 40%, while 96 $\mu\text{mol/L}$ Ori only needs 12 h treatment and the proliferation inhibition rate reaches 50%. After 24-hour treatment of 96 $\mu\text{mol/L}$ Ori, the apoptosis rate of Bcap-37 cell is 21.6%. Western blot detects up-regulation of Fas, release of cytochrome c, activation of Caspase 9 and shear of its downstream signaling protein Caspase3 and its substrates PARP and ICAD. It further detects the down-regulation of Survivin and up-regulation of GADD34. Based on these experimental results, it is inferred that the induction of apoptosis of breast cancer cell Bcap37 by oridonin as a potential anti-breast cancer preparation might be related to Survivin and GADD34.

Key words: oridonin; apoptosis; breast cancer; Caspase; survivin; GADD34

(责任编辑: 许惠儿)

(上接第 371 页)

Automatic Classification of Leukocytes in Blood Based on Multi-evidence

ZHUANG Yang-kai, ZHOU Ping

(School of Information Science and Technology, Zhejiang Sci-Tech University, Hangzhou 310018, China)

Abstract: This paper puts forward a classification algorithm based on shape feature study for leukocytes classification method. This algorithm first extracts invariant moment feature and morphological feature based on Euclidean distance transform from nucleus; then, features extracted can realize the classification of mononuclear cell, lymphocyte and basophilic granulocyte; gray-level co-occurrence matrix is extracted as texture feature and the classification of residual leukocyte categories is realized through support vector machine. The experimental result shows that this method has a very good classification accuracy rate and short processing time.

Key words: leukocytes classification; Euclidean distance transform; invariant moment; morphological features; texture features; support vector machine

(责任编辑: 陈和榜)