

基于三维形态参数的凋亡乳腺癌细胞机器学习识别研究^{*}

胡宏亮¹, 刘清¹, 谢雅芳¹, 杨大洲¹, 刘轶¹, 王可¹, 杨紫君¹, 王文进^{1,2}, 文于华^{1,2}

(1.湖南理工学院物理与电子科学学院, 湖南 岳阳 414006; 2.湖南理工学院先进光学研究所, 湖南 岳阳 414006)

摘要:通过对诱导凋亡乳腺癌细胞的激光共聚焦显微图像进行三维重建, 获得细胞核、线粒体和整个细胞的平均体积、比表面积、等体积球半径和体积百分数。再利用获得的上述三维形态参数, 分别采用层次聚类方法和 BP 神经网络模型对细胞进行分类, 发现在区分正常细胞和凋亡细胞两大类时分别可实现 65% 和 83.3% 的准确率, 并且两者对正常细胞的区分度较高。

关键词:机器学习; 细胞凋亡; BP 神经网络模型; 高斯混合模型

中图分类号: TP181

文献标志码: A

DOI: 10.15913/j.cnki.kjycx.2020.20.049

1 引言

凋亡是细胞程序性死亡的过程, 相关机制对人类多种疾病, 包括癌症、自身免疫和退行性疾病等有重要影响, 其研究对生命科学和临床应用具有广泛的意义^[1-2]。在癌症治疗的情况下, 由于肿瘤细胞对治疗方案的敏感性和微环境之间的密切关系, 细胞凋亡深刻地影响患者对治疗和复发风险的反应。因此, 对癌细胞凋亡过程的评估对于患者个体化治疗变得越来越重要。癌细胞凋亡过程中的形态变化, 为表征凋亡提供了外型特征标记。通过激光共聚焦显微技术, 可以获得高分辨率的细胞结构层切图像。对其三维重建后, 可计算获得细胞的三维形态学参数, 并统计分析得出凋亡细胞和正常细胞之间的差异^[3]。

机器学习是一门人工智能的科学, 研究人工智能如何在经验学习中改善具体算法的性能。无监督机器学习方法层次聚类^[4]和监督机器学习方法 BP (back propagation) 神经网络模型^[5], 被广泛用于各种分类识别研究中。

本文以诱导凋亡过程中的人体乳腺癌细胞 MCF-7 为研究对象, 通过激光共聚焦显微图像三维重建技术获取细胞三维形态学参数, 然后进行上述两种机器学习分类研究, 以期对正常细胞和凋亡细胞进行准确区分。

2 研究方案

研究流程如图 1 所示, 研究中首先进行细胞的培养与诱导凋亡实验。MCF-7 细胞保存在 37 °C 浓度为 5% 的 CO₂ 加湿培养箱中, 采用含质量分数为 10% 胎牛血清 (FBS) 的 DMEM 培养基进行培养, 并用 14-羟基柔红霉素 (Doxorubicin) 诱导凋亡。然后, 分别用 Syto-61、M-7510 和 Annexin V 三种荧光染料对 MCF-7 细胞进行染色, 在激光共聚焦显微镜实验中分别标定细胞核、线粒体和凋亡特征, 对应的共聚焦显微切片图像分别呈现红、绿和蓝三种颜色, 根据蓝通道的荧光强度, 区分正常细胞和处于凋亡状态的细胞。在此基础上, 对获取的切片图像进行三维重建, 获取细胞核、线粒体和细胞膜三维结构, 并计算整个细胞的体积 (c_vol)、比表面积 (SVR_c) 和等体积球半径 (Er_c), 线粒体的体积 (m_vol)、比表面积 (SVR_m)、等体积球半径 (Er_m) 和体积百分数 (rt_m), 细胞核的体积 (n_vol)、比表面积 (SVR_n)、等体积球半径 (Er_n) 和体积百分数 (rt_n) 共 11 个三维形态参数。最后, 分别采用层次聚类和 BP 神经网络模型对上述参数进行机器学习分类研究, 获取不同模型下正常细胞和凋亡细胞识别准确的数目, 并计算准确率。

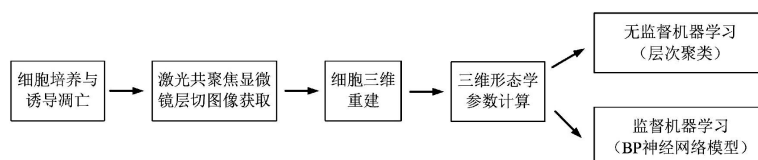


图 1 研究流程图

在机器学习研究中, 分别选取 60 个正常细胞和 60 个凋亡细胞, 随机分成 2 组, 每组含有 30 个正常细胞和 30 个凋

亡细胞。将其中一组作为 BP 神经网络模型的训练样本, 另一组作为测试样本。同时, 测试样本组用于层次聚类, 方便

^{*} [基金项目] 2019 年湖南省大学生创新训练项目 (编号: S201910543028) 资助; 湖南省教育厅优秀青年项目 (编号: 18B348) 资助

与 BP 神经网络模型分类的结果对比。

在进行机器学习研究前，对两组数据每个参数分别进行了归一化处理。进行层次聚类法和 BP 神经网络模型研究时，分别调用 MATLAB 软件中的 clusterdata 函数与 newff 函数将测试样本数据分 2 类时。其中，调用 clusterdata 函数时，linkage 参数选用 ward，其他为默认设置。调用 newff 函数时，输入层和输出层分别含有 11 和 2 个神经元，中间包含 4 个隐层，分别含有 20、8、8、8 个神经元，4 个隐层和输出层的激活函数分别为 tansig、logsig、logsig、tansig 和 tansig，多次进行训练，并输出测试样本的分类结果。

3 结果与讨论

实验得到的 60 个正常细胞和 60 个凋亡细胞的三维形态学参数，计算结果如表 1 所示。通过对比发现，正常细胞和凋亡细胞的部分三维形态学参数有明显差异。通过机器学习方法，去对比每个细胞的上述 11 个参数，可能将正常细胞

和凋亡细胞进行准确区分。MCF-7 细胞机器学习分类结果如图 2 所示。

进行层次聚类分析时，正常细胞和凋亡细胞组识别正确的数目分别为 26 和 13 个，准确率为 65%。通过 BP 神经网络模型多次训练，正常细胞和凋亡细胞组识别正确的数目分别为 29 和 21 个，准确率为 83.3%。可以发现，正常细胞组的区分度较高，其细胞三维形态特征稳定。而凋亡细胞组，可能存在早期凋亡与晚期凋亡的差异，前者更接近与正常细胞的三维形态结构，因此识别的准确率低一些。

4 结论

本文通过机器学习方法研究了诱导凋亡过程中的人体乳腺癌细胞基于三维形态参数的分类识别，分别采用层次聚类方法和 BP 神经网络模型。发现两者的准确率分别为 65% 和 83.3%，且对正常细胞组的识别率较高。采用 BP 神经网络模型进一步优化网络结构，可能获得更高的识别准确率。

表 1 MCF-7 细胞三维形态参数计算结果

参数名称	符号	单位	均值 ± 标准偏差	
			正常细胞	凋亡细胞
细胞体积	c_vol	μm^3	3 955±1 698	3 249±1 279
细胞比表面积	SVR_c	μm^{-1}	0.461±0.062	0.532±0.130
细胞等体积球半径	Er_c	μm	9.640±1.271	9.014±1.282
细胞核体积	n_vol	μm^3	1 249±566	1 172±800
细胞核比表面积	SVR_n	μm^{-1}	0.702±0.087	0.781±0.166
细胞核等体积球半径	rt_n	μm	6.550±0.922	6.333±1.123
细胞核体积百分数	Er_n	-	0.318±0.060	0.364±0.133
线粒体体积	m_vol	μm^3	310±209	143±107
线粒体比表面积	SVR_m	μm^{-1}	3.206±0.599	4.341±1.470
线粒体等体积球半径	Er_m	μm	0.973±0.207	0.750±0.198
线粒体体积百分数	rt_m	-	0.079±0.036	0.044±0.030

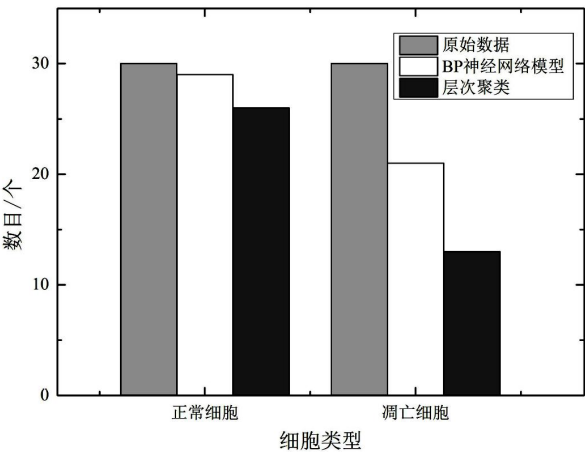


图 2 MCF-7 细胞机器学习分类结果

参考文献：

[1] BROWN J M, ATTARDI L D.The role of apoptosis in cancer development and treatment response [J] . Nat.Rev.Cancer, 2005, 5 (3) : 231-237.
[2] KERR J F, WYLLIE A H, CURRIE A R. Apoptosis: a basic biological phenomenon with wide-ranging implications in tissue kinetics [J] .Br.J.Cancer, 1972,

26 (4) : 239-257.
[3] WEN Y H, CHEN Z, LU J F, et al.Quantitative analysis and comparison of 3D morphology between viable and apoptotic MCF-7 breast cancer cells and characterization of nuclear fragmentation [J].PLoS ONE, 2017, 12 (9) : e0184726.
[4] YING Z, KARYPIS G, FAYYAD U.Hierarchical clustering algorithms for document datasets [J] . Data Mining & Knowledge Discovery, 2005, 10 (2) : 141-168.
[5] HECHT-NIELSEN R.Theory of the backpropagation neural network [C] // Neural Networks, IJCNN. International Joint Conference on, 1989.

作者简介：胡宏亮（2000—），男，微电子科学与工程专业，本科在读。
通讯作者：文于华（1981—），男，博士，讲师，研究方向为细胞光学与人工智能。

[编辑：严丽琴]