

探索 Morphogo 在多发性骨髓瘤 MRD 检测中的应用

刘思恒, 李 佳, 彭贤贵, 张 诚

(陆军军医大学附属新桥医院血液病医学中心, 重庆 400037)

【摘要】目的:探索人工智能骨髓细胞识别系统 Morphogo 应用于多发性骨髓瘤微小残留病(minimal residual disease, MRD)检测的临床价值及面临的问题。**方法:**收集已经由流式细胞术(multiparameter flow cytometry, MFC)检查后明确微小残留病结果的病例 65 例, 调取其留存的骨髓瑞氏染色涂片, 通过基于人工智能(artificial intelligence, AI)平台的分析系统 Morphogo 对所有骨髓涂片进行全自动扫描及细胞分类。AI 及细胞形态学多发性骨髓瘤 MRD 阳性阈值设为浆细胞比例大于 4.4%。按 AI 自动识别细胞数量将病例分为 I 500 组、I 1000 组、I 2000 组, 每组病例的人工智能微小残留病(AI-MRD)、细胞形态学(morphology)微小残留病(M-MRD)和流式细胞术微小残留病(MFC-MRD)结果两两行 Kappa 一致性检验, 并计算各组敏感度、特异度、准确度。分别以 MFC-MRD 和 M-MRD 结果为金标准绘制 AI-MRD 的受试者工作特征(receiver operating characteristic, ROC)曲线并计算其曲线下面积(area under the curve, AUC)。**结果:**分组后 AI-MRD vs. MFC-MRD 和 AI-MRD vs. M-MRD 的 Kappa 值、敏感度、特异度、准确度、AUC 均随识别细胞数量的增加而增高, 其中 I 2000 组 AI-MRD vs. MFC-MRD 的 Kappa 一致性检验结果为 Kappa=0.500 ($P=0.013$), 敏感度为 71%, 特异度为 80%, 准确度为 75%; AI-MRD vs. M-MRD 的 Kappa 一致性检验结果为 Kappa=0.667 ($P=0.001$), 敏感度为 100%, 特异度为 75%, 准确度为 83%。以 MFC-MRD 结果为标准, I 2000 组 AI-MRD 的 ROC AUC=0.800 ($P=0.002$, 95%CI=0.588~0.934), M-MRD 的 ROC AUC=0.779 ($P=0.005$, 95%CI=0.564~0.921)。**结论:**人工智能骨髓细胞识别系统 Morphogo 检测多发性骨髓瘤 MRD 具有细胞识别准确度高、速度快、成本低等特点, 后续开发中应加入细胞组化染色、细胞免疫等技术提高人工智能多发性骨髓瘤微小残留病诊断的准确率。

【关键词】人工智能; 骨髓细胞; 多发性骨髓瘤; 微小残留病; Morphogo

【中图分类号】R551.3

【文献标志码】A

【收稿日期】2022-01-11

Application of Morphogo in the detection of minimal residual disease in multiple myeloma

Liu Siheng, Li Jia, Peng Xiangui, Zhang Cheng

(Hematology Medical Center, Xinqiao Hospital, Army Medical University)

【Abstract】Objective: To explore the application prospects and problems of artificial intelligence bone marrow cell recognition system, Morphogo, in the detection of minimal residual disease (MRD) of multiple myeloma. **Methods:** A total of 65 cases of MRD of multiple myeloma confirmed by multiparameter flow cytometry (MFC) were collected, and their bone marrow Wright's staining smears were obtained. All bone marrow smears were automatically scanned and classified by Morphogo based on artificial intelligence platform. The positive threshold of MRD of AI and cytomorphology in multiple myeloma was set as the proportion of plasma cells, which was greater than 4.4%. The cases were divided into I 500 group, I 1000 group and I 2000 group according to the number of AI automatically recognized cells. The results of AI-MRD, morphological-MRD (M-MRD) and MFC-MRD in each group were tested by Kappa consistency test, and the sensitivity, specificity and accuracy of each group were calculated. Taking MFC-MRD and M-MRD results as the gold standard, the receiver operating characteristic (ROC) curve of AI-MRD was drawn and its area under the curve (AUC) value was calculated. **Results:** After grouping, with the increase of the number of recognized cells, the Kappa value, sensitivity, specificity, accuracy and AUC of AI-MRD vs. MFC-MRD and AI-MRD vs. M-MRD increased. The Kappa consistency test of AI-MRD vs. MFC-MRD in I 2000 group showed that the Kappa value was 0.500 ($P=0.013$), sensitivity was 71%, specificity was 80%, and accuracy was 75%. The Kappa consistency test results of AI-MRD vs. M-MRD showed that the Kappa value was 0.667 ($P=0.001$), sensitivity was 100%, specificity was 75%, and accuracy was 83%. When MFC-MRD results were taken as the diagnostic criteria, the AUC of AI-MRD in I 2000 group was 0.800 ($P=0.002$, 95%CI=0.588-0.934), and the AUC of M-MRD was 0.779 ($P=0.005$, 95%CI=0.564-0.921). **Conclusion:** The detection of MRD of multiple myeloma by Morphogo has the characteristics of high accuracy, high speed and low cost. In the follow-up development, it should be considered to de-

作者介绍: 刘思恒, Email: 26656379@qq.com,

研究方向: 血液肿瘤诊断。

通信作者: 张 诚, Email: chzhang2014@163.com。

基金项目: 重庆市技术创新与应用发展专项重点资助项目 (编号: cstc2019jscx-gksbX0016)。

优先出版: <https://kns.cnki.net/kcms/detail/50.1046.R.20220831.1155.038.html>
(2022-08-31)

velop technologies such as cell histochemical staining and cellular immunity, so as to improve the diagnostic accuracy of MRD of artificial intelligence multiple myeloma.

[Key words] artificial intelligence; bone marrow cell; multiple myeloma; minimal residual disease; Morpho

随着造血干细胞移植和新药的应用,多发性骨髓瘤的无进展生存率和总生存率均有较大提高,但其仍被认为是不可完全治愈的一类疾病。很多患者在获得完全缓解后出现疾病复发,其根源在于患者体内存在微小残留病(minimal residual disease, MRD)。研究表明,MRD 阳性是影响生存的独立不良预后因素^[1],因此 MRD 检测是判断患者预后及指导临床治疗的重要指标,目前多参数流式(multiparameter flow cytometry, MFC)、聚合酶链反应(polymerase chain reaction, PCR)、正电子发射断层成像(positron emission tomography, PET)/计算机断层扫描术(computer tomography, CT)等多种技术都已应用于多发性骨髓瘤 MRD 检测^[2]。但上述技术均存在费用高昂、用时较长等特点。细胞形态学是一项成熟的检验技术,通过显微镜观察可以明确识别浆细胞,但传统细胞形态学通过人工分类,存在分类细胞数量较少、浆细胞识别率依赖检验人员专业素质等缺陷,这使得传统细胞形态学无法运用于多发性骨髓瘤 MRD 检测。人工智能(artificial intelligence, AI)是目前蓬勃发展的一项计算机新技术,其分支包括卷积神经网络(convolutional neural network, CNN)、模糊专家系统(fuzzy expert system, FES)、进化计算和混合智能系统等技术,都已经不同的临床环境中使用^[3],特别是 CNN 技术已经应用于放射学^[4]和病理学^[5]的临床诊断。目前已有多家机构研究基于 CNN 的人工智能应用于血液病细胞形态学诊断。人工智能卷积神经网络(artificial intelligence convolutional neural network, AI-CNN)与

细胞形态学结合而来的人工智能细胞识别技术能否应用于临床多发性骨髓瘤 MRD 诊断呢? 本文进行了如下探索。

1 材料和方法

1.1 病例选择与患者一般资料

收集陆军军医大学附属新桥医院血液病医学中心 2021 年 1 月至 8 月多发性骨髓瘤治疗后复查病例 65 例,所有病例均经流式细胞术检查明确 MRD 结果,结果符合国际骨髓瘤工作组关于多发性骨髓瘤微小残留疾病评估的共识标准^[2]。患者中男性 34 例,女性 31 例,年龄 39~80 岁,中位年龄 56 岁。

1.2 AI 平台简介

本研究使用的人工智能平台分析系统 Morpho 基于 CNN 算法,其硬件条件主要包括:①摄像系统:日本奥林巴斯 BX63 显微镜,配备 4×、10×、40×、100×物镜;索尼 Exmor CMOS 传感器,最大像素 4 000×3 000;全自动显微镜传动系统。②Morpho 专用电脑:硬件条件为 CPU i9 10900K、显卡 3×RTX 2080、内存 DDR4 64G。软件条件为 Morpho 最新 1.0.0.27 版骨髓细胞识别软件。该平台根据所需分析细胞数量的多少,处理每张骨髓涂片需 10~30 min。Morpho 系统目前已在中心学习、训练超过 10 万个细胞,总灵敏度和特异度的平均值分别为 69.4% 和 97.2%,细胞分类准确率为 85.7%^[6]。其将骨髓瑞氏染色涂片扫描后,可生成数字图像,并自行判定、标记细胞种类,如图 1 所示。

1.3 研究方法

AI 细胞识别目前尚处于初步发展阶段,国内外尚无关于 AI 应用于多发性骨髓瘤 MRD 检测方面的研究,因此本文参考 Girodon F 等^[7]的研究,将浆细胞正常值上限(4.4%)规定为人工智能微小残留病(AI-MRD)及骨髓细胞形态学微小

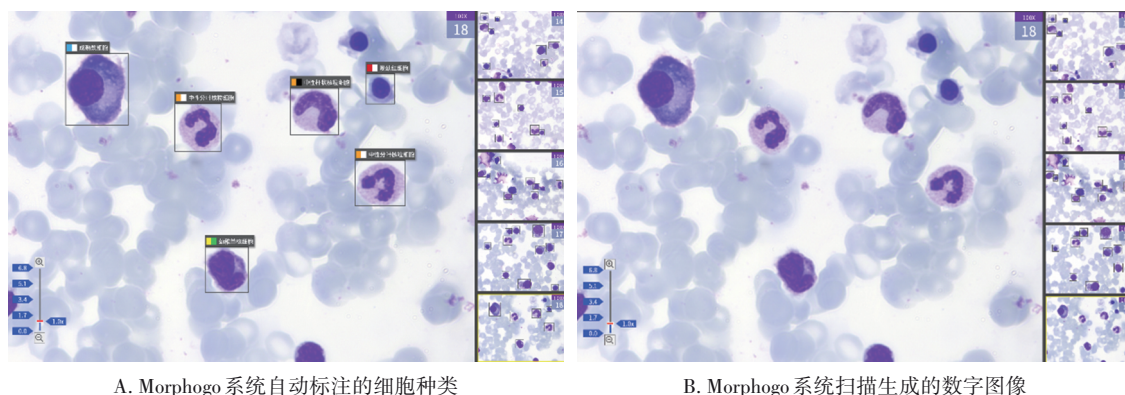


图 1 Morpho 系统扫描生成的数字图像及标注结果(瑞氏染色,1 000 ×)

残留病(M-MRD)阳性,小于等于4.4%者为阴性。

调取所有病例留存的骨髓瑞氏染色涂片,放入Morphogo摄像系统后显微镜传动系统自动运行,全片以4×物镜扫描并自动标记骨髓片识别部位,然后摄像系统切换至100×物镜拍摄识别部位内所有细胞图像,拍摄后图像传输至专用电脑进行细胞自动分类。分类完成后用Morphogo自动分类所得浆细胞数除以自动分类细胞总数,即得到AI-MRD浆细胞比例,将该比例精确至千分位。M-MRD浆细胞比例来源于病历查询。

所有病例按AI自动识别细胞数量分为识别500个细胞组(identify 500 cells,I 500组)、识别1 000个细胞组(identify 1 000 cells,I 1000组)、识别2 000个细胞组(identify 2 000 cells,I 2000组)3个亚组。其中I 500组18例,I 1000组23例,I 2000组24例。对每个亚组中AI-MRD vs. M-MRD和AI-MRD vs. MFC-MRD的结果行Kappa一致性检验,并计算其敏感度、特异度、准确度。以MFC-MRD结果为金标准绘制AI-MRD和M-MRD浆细胞比例的ROC曲线并计算AUC值。

1.4 统计学处理

采用SPSS 19.0软件进行每个亚组中AI-MRD vs. MFC-MRD和AI-MRD vs. M-MRD结果的Kappa一致性检验,并计算其敏感度、特异度、准确度。采用MedCalc 19.7统计软件绘制每个亚组AI-MRD vs. MFC-MRD和M-MRD vs. MFC-MRD的受试者工作特征(receiver operating characteristic,

ROC)曲线并计算曲线下面积(area under the curve, AUC)。检验水准 $\alpha=0.05$ 。

2 结果

2.1 三亚组相互间kappa一致性检验结果及各组敏感度、特异度、准确度

将AI-MRD及M-MRD阳性阈值规定为4.4%后,I 500组、I 1000组、I 2000组AI-MRD vs. M-MRD和AI-MRD vs. MFC-MRD的Kappa一致性检验结果及各组敏感度、特异度、准确度见表1、表2。

2.2 以MFC-MRD结果作为金标准时各组AI-MRD与M-MRD ROC曲线及AUC

以MFC-MRD结果作为诊断多发性骨髓瘤微小残留病的金标准,I 500组、I 1000组、I 2000组AI-MRD分类比例的受试者工作特征(receiver operating characteristic, ROC)曲线下面积(area under the curve, AUC)分别为0.604($P=0.479$, 95%CI=0.351~0.822)、0.739($P=0.033$, 95%CI 0.515~0.897)、0.800($P=0.002$, 95%CI=0.588~0.934);I 500组、I 1000组、I 2000组细胞形态学浆细胞(M-MRD)分类比例的ROC AUC分别为0.766($P=0.039$, 95%CI=0.511~0.929)、0.735($P=0.028$, 95%CI=0.511~0.895)、0.779($P=0.005$, 95%CI=0.564~0.921)。各组AI-MRD与M-MRD ROC曲线如图2所示。

表1 3组AI-MRD vs. M-MRD Kappa一致性检验结果及敏感度、特异度、准确度

指标		I 500组			I 1000组			I 2000组		
		AI-MRD			AI-MRD			AI-MRD		
		阳性	阴性	合计	阳性	阴性	合计	阳性	阴性	合计
M-MRD	阳性	2	0	2	5	1	6	8	0	8
	阴性	6	10	16	7	10	17	4	12	16
合计		8	10	18	12	11	23	12	12	24
Kappa值		0.270			0.319			0.667		
P值		0.094			0.076			0.001		
灵敏度/%		100			83			100		
特异度/%		63			59			75		
准确度/%		67			65			83		

表2 3组AI-MRD vs. MFC-MRD Kappa一致性检验结果及敏感度、特异度、准确度

指标		I 500组			I 1000组			I 2000组		
		AI-MRD			AI-MRD			AI-MRD		
		阳性	阴性	合计	阳性	阴性	合计	阳性	阴性	合计
MFC-MRD	阳性	3	4	7	8	4	12	10	4	14
	阴性	5	6	11	4	7	11	2	8	10
合计		8	10	18	12	11	23	12	12	24
Kappa值		-0.025			0.303			0.500		
P值		0.914			0.146			0.013		
灵敏度/%		43			67			71		
特异度/%		55			64			80		
准确度/%		50			65			75		

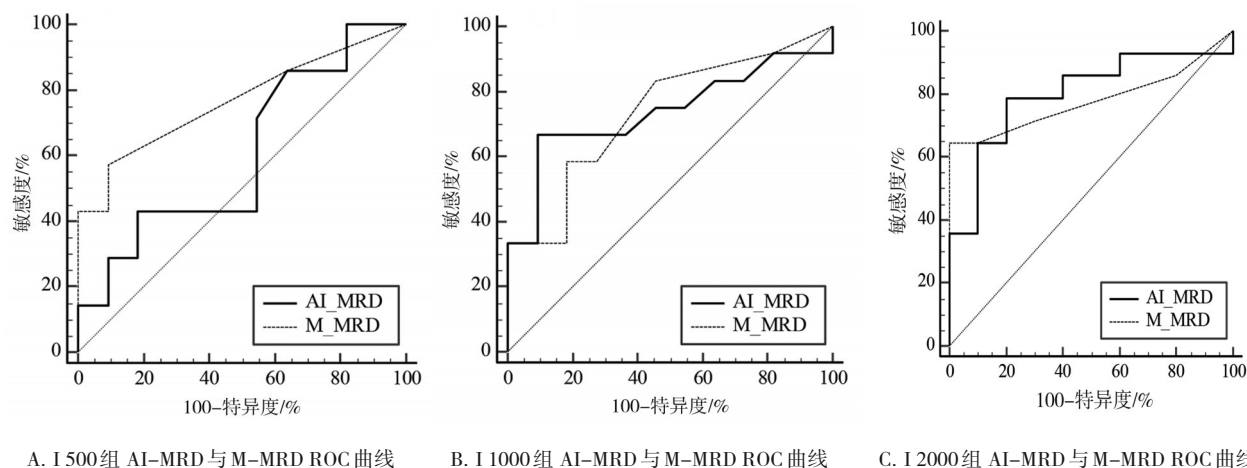


图2 各组AI-MRD与M-MRD ROC曲线图

3 讨论

多发性骨髓瘤(multiple myeloma, MM)是一种浆细胞恶性增殖性疾病,其特征为骨髓或外周组织中单克隆浆细胞异常增殖,绝大部分病例分泌单克隆免疫球蛋白或轻链,造成骨痛、贫血、肾功损害等一系列机体损伤,其发病率随人口平均寿命增加而逐年增高。随着免疫调节药物(immunomodulatory drug, IMiD)、蛋白酶体抑制剂(proteasome inhibitor, PI)、单克隆抗体(monoclonal antibody, mAb)等新药^[8]和自体造血干细胞移植的使用,越来越多的患者能够达到完全缓解,同时人们也认识到,患者达到完全缓解后MRD检测结果是决定预后的关键因素^[1]。

MM-MRD的检测目前在国内外已取得一定共识^[12,9],通常使用流式细胞术或基因测序来鉴定骨髓中残留的肿瘤细胞,并使用PET/CT等影像技术检测骨髓外是否存在残留疾病。骨髓细胞形态学是目前评估肿瘤细胞负荷最广泛的方法,由于传统骨髓细胞形态学检查使用人工分类计数200~500个细胞的方法,其灵敏度较低,不适用于MRD检测,但随着AI图像识别技术的兴起,细胞形态学检查关于数量级的问题有望得到一定程度解决。

AI在图像识别中最常用的技术是CNN算法。CNN是深度学习的一种,其优点在于可以直接在神经网络中输入图像从而避免传统算法中复杂的特征提取和数据重建过程,并且对二维图像的平移、比例缩放和其他形式的变形具有一定的不变性^[10]。

Morphogo人工智能系统采用27层CNN,已经在训练后达到较高的细胞识别准确度,在识别某些形态特征较为典型的细胞如急性早幼粒细胞白血病中的异型早幼粒细胞及急性淋巴细胞白血病中的原始/幼稚淋巴细胞时,Morphogo与人工分类细胞形态学的符合率可达100%^[11]。

本研究中,Morphogo与细胞形态学对浆细胞识别的一致性也较好,AI-MRD vs. M-MRD中Kappa一致性检验的Kappa值随识别细胞数量的增加持续增高,I 2000组Kappa=0.667($P=0.001$),表明在AI细胞识别量级达2000个时AI-MRD与M-MRD的结果具备高度一致性^[12]。进一步数据分析中,AI与细胞形态学关于MRD结果的灵敏度、准确度也随AI识别细胞数量的增加而明显增高,且特异度始终维持在90%~100%水平。这表明AI在判定患者阴性结果时可以与人工细胞形态学检查结果达到几乎完全一致,而在判定阳性结果时随识别细胞数量的增加一致性也随之增高。

在关于MM-MRD检测方面,人工智能Morphogo也有令人较为感兴趣的表现。Morphogo与MFC关于MM-MRD检测的一致性随AI识别细胞数量增高而增加,AI-MRD vs. MFC-MRD的Kappa一致性检验在I 2000组Kappa值升高到0.500($P=0.013$),敏感度、特异度、准确度均达70%以上,提示AI-MRD与MFC-MRD检查结果具备中等度一致性。此外,当以MFC-MRD结果作为判定MM-MRD是否阳性的金标准时,M-MRD的ROC AUC值维持在0.7水平,最高为0.779($P=0.005$),而AI-MRD的AUC值随识别细胞数量增加逐渐增高,在I 2000组达到

0.800 ($P=0.002$), AUC_{M-MRD} 与 AUC_{AI-MRD} 差 0.021, 差异无统计学意义 ($P=0.776$), 这表明当识别数量达到 2 000 级别时人工智能细胞识别与人工骨髓细胞形态学在 MRD 检查的效能几乎一致, 也证明 AI 在多发性骨髓瘤 MRD 诊断上具备一定效能。

AI 的研究起步较晚, 虽然常用于图像分类任务的 CNN 直至 2012 年才因 AlexNet 受到广泛关注^[13], 但由于其在图片识别方向的优势及硬件的发展, CNN 已成为当前图像理解领域的研究热点。近年来, CNN 逐渐被应用于医学, 特别是影像学领域^[14]。骨髓细胞形态学是根据细胞图像进行分类从而诊断疾病的一项技术, CNN 能够较好地应用于这一领域。本研究表明, 基于 CNN 技术的人工智能 MorphoGo 在识别 2 000 个以上细胞时获得的结果即可与传统人工计数的细胞形态学检查达到高度一致, 但在进行 MM-MRD 检测时, MorphoGo 的准确性仍与 MFC 有一定差距。

总之, 应用 AI 进行 MM-MRD 检测, 具有细胞识别准确度高、速度快、成本低等特点, 但仅依靠形态学的细胞识别, AI 与流式细胞术结果的一致性仍不够高, 但 AI 在细胞识别方面仍具有高成长性。后续开发中, 应考虑将细胞组化染色、细胞免疫、分子生物学及全玻片成像技术^[15]等技术纳入其中, 从而持续提高 AI 在 MM-MRD 诊断中的准确率, 为临床提供高效、准确、廉价的检验技术。

参 考 文 献

- [1] 杨珮钰, 刘蒙蒙, 樊红琼, 等. 105 例初诊多发性骨髓瘤患者微小残留病的动态监测及其预后价值[J]. 中华血液学杂志, 2019, 40(7): 584-588.
Yang PY, Liu MM, Fan HQ, et al. The prognostic significance of dynamic monitoring of minimal residual disease (MRD) status in patients with newly-diagnosed multiple myeloma[J]. Chin J Hematol, 2019, 40(7): 584-588.
- [2] Kumar S, Paiva B, Anderson KC, et al. International Myeloma Working Group consensus criteria for response and minimal residual disease assessment in multiple myeloma[J]. Lancet Oncol, 2016, 17(8): e328-e346.
- [3] Ramesh AN, Kambhampati C, Monson JRT, et al. Artificial intelligence in medicine[J]. Ann R Coll Surg Engl, 2004, 86(5): 334-338.
- [4] Thrall JH, Li X, Li QZ, et al. Artificial intelligence and machine learning in radiology: opportunities, challenges, pitfalls, and criteria for success[J]. J Am Coll Radiol JACR, 2018, 15(3 Pt B): 504-508.
- [5] Salto-Tellez M, Maxwell P, Hamilton P. Artificial intelligence—the third revolution in pathology[J]. Histopathology, 2019, 74(3): 372-376.
- [6] Fu XY, Fu M, Li Q, et al. MorphoGo: an automatic bone marrow cell classification system on digital images analyzed by artificial intelligence[J]. Acta Cytol, 2020, 64(6): 588-596.
- [7] Girodon F, Favre B, Carli PM, et al. Minor dysplastic changes are frequently observed in the bone marrow aspirate in elderly patients without haematological disease[J]. Clin Lab Haematol, 2001, 23(5): 297-300.
- [8] Mateos MV, Cavo M, Blade J, et al. Overall survival with daratumumab, bortezomib, melphalan, and prednisone in newly diagnosed multiple myeloma (ALCYONE): a randomised, open-label, phase 3 trial[J]. Lancet, 2020, 395(10218): 132-141.
- [9] 中国医师协会血液科医师分会, 中华医学会血液学分会, 中国医师协会多发性骨髓瘤专业委员会. 中国多发性骨髓瘤诊治指南 (2020 年修订)[J]. 中华内科杂志, 2020, 59(5): 341-346.
Chinese Hematology Association, Chinese Society of Hematology, Chinese Myeloma Committee—Chinese Hematology Association. The guidelines for the diagnosis and management of multiple myeloma in China (2020 revision)[J]. Chin J Intern Med, 2020, 59(5): 341-346.
- [10] 常亮, 邓小明, 周明全, 等. 图像理解中的卷积神经网络[J]. 自动化学报, 2016, 42(9): 1300-1312.
Chang L, Deng XM, Zhou MQ, et al. Convolutional neural networks in image understanding[J]. Acta Autom Sin, 2016, 42(9): 1300-1312.
- [11] 杜忠华, 张福明, 吴波, 等. 人工智能技术 MorphoGo 系统在骨髓细胞形态学检查的临床应用[J]. 中国实验诊断学, 2021, 25(7): 958-961.
Du ZH, Zhang FM, Wu B, et al. Clinical application of artificial intelligence technology MorphoGo system in morphological examination of bone marrow cells[J]. Chin J Lab Diagn, 2021, 25(7): 958-961.
- [12] 唐万, 胡俊, 张晖, 等. Kappa 系数: 一种衡量评估者间一致性的常用方法[J]. 上海精神医学, 2015, 27(1): 62-67.
Tang W, Hu J, Zhang H, et al. Kappa coefficient: a popular measure of rater agreement[J]. Shanghai Arch Psychiatry, 2015, 27(1): 62-67.
- [13] Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks[J]. Commun ACM, 2017, 60(6): 84-90.
- [14] Lin DJ, Johnson PM, Knoll F, et al. Artificial intelligence for MR image reconstruction: an overview for clinicians[J]. J Magn Reson Imaging, 2021, 53(4): 1015-1028.
- [15] Pantanowitz L, Farahani N, Parwani A. Whole slide imaging in pathology: advantages, limitations, and emerging perspectives[J]. Pathol Lab Med Int, 2015(7): 23-33.

(责任编辑: 唐秋姗)