

• 智慧医疗 • doi:10.3969/j.issn.1671-8348.2022.24.028

人工智能辅助多参数流式细胞术检测淋巴细胞亚群免疫表型的应用研究*

贾晓冬^{1,2,3}, 汲珊珊³, 刘蕊^{2△}, 楚玉兰³

(1. 天津医科大学研究生院 300070; 2. 天津医科大学人民医院临床学院 300121;
3. 天津金域医学检验实验室有限公司 300392)

[摘要] **目的** 通过建立人工智能(AI)模型,评价 AI 辅助多参数流式细胞术(MFC)检测人外周血中淋巴细胞亚群(LS)免疫表型方法的可行性。**方法** 运用流式细胞术对 2020 年 6—7 月天津金域医学检验实验室收集的 1 263 例患者外周血样本 T 淋巴细胞(CD3⁺、CD4⁺、CD8⁺)、B 淋巴细胞(CD3⁻CD19⁺)及 NK 细胞(CD3⁻CD16⁺CD56⁺)进行检测,基于高斯混合模型的多维数据聚类算法,应用聚类算法和核密度估计方法对淋巴细胞免疫表型自动判定和统计分析,评价两种检测方法的一致性。**结果** 经与人工分析结果比较,1 263 例人工分析数据中有 1 199 例通过 AI 分析一致性检测的通过比例为 94.93%。诊断结果差异项经高年资医师复核后与 AI 诊断结果一致。AI 单个样本平均分析时间为(1.36±0.25)s,较传统人工分析提高约 50 倍以上。且两种方法具有良好的一致性。AI 分析方法质控的重复系数为 2.833 1%,95%置信区间:2.726 8%~2.948 1%,均在临床可接受临界值(±5%)范围内,表明 AI 分析可快速检测出人外周血中淋巴细胞各亚群百分比,且重复性好。**结论** 初步建立的 AI 辅助 MFC 检测 LS 免疫表型模型重复性好,分析速度快,准确率高,已基本满足了临床使用需求,可用于基于 6 色抗体组合方案辅助 MFC 检测人外周血 LS 百分比的检测。

[关键词] 人工智能;流式细胞术;淋巴细胞亚群

[中图分类号] R331 **[文献标识码]** A **[文章编号]** 1671-8348(2022)24-4294-06

Application of multiparameter flow cytometry in detecting immunophenotyping of lymphocyte subsets by artificial intelligence assistance*

JIA Xiaodong^{1,2,3}, JI Shanshan³, LIU Rui^{2△}, CHU Yulan³

(1. Graduate School of Tianjin Medical University, Tianjin 300070, China; 2. Tianjin Union Medical Center, Tianjin Medical University, Tianjin 300121, China; 3. Tianjin Kingmed Medical Laboratory Co., Ltd., Tianjin 300392, China)

[Abstract] **Objective** To evaluate the feasibility of AI-assisted multi-parameter flow cytometry (MFC) for detecting the immunophenotype of lymphocyte subsets (LS) in human peripheral blood by establishing the artificial intelligence (AI) model. **Methods** MFC was used to detect T lymphocytes (CD3⁺, CD4⁺, CD8⁺), B lymphocytes (CD3⁻CD19⁺) and NK cells (CD3⁻CD16⁺CD56⁺) in 1 263 peripheral blood samples collected by Tianjin Kingmed Medical Laboratory. The multi-dimensional data clustering algorithm based on Gaussian mixture model was used to automatically determine and statistically analyze the lymphocyte immunophenotype by clustering algorithm and kernel density estimation method. The consistency between the two methods was evaluated. **Results** In the comparison with the manual analytical results, among 1 263 manual analysis data, 1 199 cases passed the consistency detection of AI analysis, with a passing rate of 94.93%. The difference items in the diagnosis results were consistent with the AI diagnosis after reexamination by senior physicians. The average analysis time of AI single sample was (1.36±0.25) s, which was more than 50 times faster than that of the traditional manual analysis. The two methods showed a good consistency. The quality control repetition coefficient (CR) of AI analysis method was 2.833 1%, 95% confidence interval was 2.726 8% to 2.948 1%, which all were within the clinically acceptable critical value (±5%), indicating the the AI analysis

* 基金项目:中央引导地方科技发展专项区域创新体系建设(百城百园)项目(20ZYQCSY00010)。 作者简介:贾晓冬(1986—),主管技师,本科,主要从事医学检验诊断技术的研究。 △ 通信作者, E-mail:lr3595@163.com。

could rapidly detect the LS percentage of peripheral blood LS and peripheral blood lymphocytes subsets percentage with good repeatability. **Conclusion** The preliminarily established AI-assisted MFC model for detecting LS has good repeatability with rapid analysis speed and high accuracy, which basically meets the clinical application demand and could be used to detect LS percentage of human peripheral blood based on 6-color antibody combination scheme assisted MFC.

[Key words] artificial intelligence; flow cytometry; lymphocyte subsets

运用流式细胞术检测淋巴细胞亚群(LS)观察机体细胞免疫水平,了解在不同疾病状态下患者的细胞免疫功能状态,对恶性肿瘤、自身免疫性疾病、免疫缺陷病、血液系统疾病的诊治、预后判断均具有重要意义^[1-5],已在临床得到广泛应用。但传统流式细胞术人工分析主要依靠分析者手动圈门和荧光表达强度的判读,检测灵敏度和准确性取决于分析者经验^[6-7]。近年来,随着人工智能(AI)技术在医疗领域的探索与发展,AI辅助多参数流式细胞术(MFC)可通过聚类、降维、自动分群等分析方法自动判断并进行统计分析,实现对待测细胞群的分类和快速判定,提高了检测的灵敏度和准确性^[8]。本研究应用人外周血 LS 流式检测数据建立了 AI 模型,并评价了 AI 辅助 MFC 检测 LS 免疫表型与人工分析结果的一致性,现报道如下。

1 资料与方法

1.1 一般资料

选取 2020 年 6—7 月天津金域医学检验实验室收集的 1 263 例患者外周血样本作为研究对象,其中男 588 例,女 675 例;年龄 0~81 岁,中位年龄 55 岁。

1.2 方法

1.2.1 实验步骤

采用美国 BD 公司 FACS canto 型流式细胞仪,检测用单克隆抗体购自同生时代公司,溶血素购自美国 BD 公司。按文献^[9]步骤操作进行样本制备,荧光素标记单克隆抗体组合:CD3-FITC/CD16⁺CD56-PE/CD45-PerCP-cy5.5/CD4-PC-7/CD19-APC/CD8-APC-cy7。

1.2.2 数据分析

待测样本混匀后检测并获取,使用 Kaluza Analysis 软件分析数据,传统人工分析应用 Kaluza Analysis 软件进行分析并结合抗原表达情况分析淋巴细胞各亚群百分比。AI 分析应用 AI 模型分析,计算淋巴细胞各亚群百分比,并生成可视化结果。检测前校准流式细胞仪并调整电压、补偿等参数,淋巴细胞总数小于 5 000 个的 FCS 数据不纳入本研究。

1.2.3 实验流程

AI 辅助 MFC 检测 LS 免疫表型的流程图见

图 1。

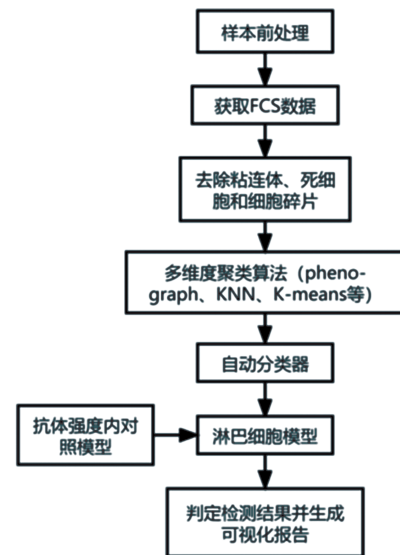


图 1 AI 辅助 MFC 检测 LS 免疫表型流程

1.2.4 AI 模型建立

1.2.4.1 无效细胞清除

(1)去除粘连体:根据非粘连体的线性分布特点建立线性回归模型: $y = ax + b$,得到线性分布主轴的位置参数 a, b ,以此参数为基准并参照人工分析特点去除粘连体。(2)去除死细胞和细胞碎片:根据细胞碎片前向散射(FSC)、侧向散射(SSC)极小,同时 CD45 阴性的特征去除细胞碎片。采用核密度估计模型计算 FSC、SSC、CD45 核密度概率密度函数。见公式 1。

$$f(f_1) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{f_1 - f_{1i}}{h}\right) \# \quad (\text{公式 1})$$

其中, f 为核密度概率密度函数, $f_{11}, f_{12} \dots f_{1n}$ 为 n 个流式细胞的指定荧光标记的荧光强度值; h 为带宽,默认设为 10; K 为核函数,采用高斯核函数(公式 2)

$$K(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) \# \quad (\text{公式 2})$$

得到细胞在 FSC、SSC 以及 CD45 上的密度分布曲线,并参照人工分析经验寻找极值来确定碎片在 FSC、SSC 以及 CD45 上的分布界限,从而去除细胞碎片。

1.2.4.2 AI 模型数据处理

采用高斯混合模型(GMM)为基础的聚类分析^[10]对数据进行高斯分布拟合。见公式 3。

$$N(x|\mu\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} = \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} \# \quad (\text{公式 3})$$

带入特定变量 x , 求得在该分布下对应的概率 N 。其中 μ 代表总体均值, σ 代表该分布的标准差, σ 越大分群数越少; 反之, σ 越小分群数越多。不符合高斯分布的数据采用改进的欧氏距离进行分析。见公式 4。

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2} = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \# \quad (\text{公式 4})$$

其中 $d(x, y)$ 代表目的细胞坐标位置 (x, y) 与固定点的距离。按最小距离原则将所有细胞点归入与其距离最小的细胞群内, 从而将有效细胞分为若干个细胞群。

1.2.4.3 抗体强度内对照模型

根据 LS 特征将粒细胞作为淋巴细胞特异性抗体的阴性内对照, 通过粒细胞平均荧光强度确认淋巴细胞各亚群的阴性表达范围。

1.2.4.4 淋巴细胞模型

根据淋巴细胞整体在 SSC/CD45 上的分布特征, 以 GMM 聚类结果为基础, 初步筛选淋巴细胞。参照粒细胞阴性内对照, 综合人工分析的诊断标准, 利用核密度分析方法对淋巴细胞荧光表达强度分布进行分析并分群^[11]。

1.2.5 AI 分析数据

经标本制作获取 FCS 数据, 将 FCS 数据输入 AI 模型并解析成与原始数据相匹配且 AI 模型可读数据。AI 模型采用多维度聚类算法 (pheno-graph、KNN、K-means 等) 将细胞分为多个集群, 再利用自动分类器在多维空间中对细胞集群进行初步分类和定义。建立抗体强度内对照模型, 参照粒细胞阴性内对照, 综合人工分析的诊断标准, 利用核密度分析方法对淋巴细胞荧光表达强度分布进行分析, 确定淋巴细胞各亚群类型并输出各类细胞群百分比。

1.2.6 AI 分析重复性试验

用 AI 模型对 1 263 份样本进行 3 次日间重复性分析, 比较各细胞群的细胞数和荧光强度差值的变异系数(CV)。

1.2.7 AI 分析与人工分析检测速度比较

比较 AI 模型分析单个样本并生成的可视化结果的平均时间与传统人工分析方法的平均时间。

1.2.8 AI 模型性能评估

以人工分析结果为参照, 对 LS 免疫表型 FCS 数

据进行 AI 分析, AI 模型分析与人工分析诊断结果按下述标准判定是否一致, 并计算检测通过比例。判断标准: 绝对误差允许 $\pm 3\%$ 的允差、细胞占父类比例大于 5% 、相对误差小于 15% 为合格; 细胞占父类比例小于或等于 5% 、相对误差小于 40% 为合格。符合上述标准即判定为一致, 否则判定为不一致。由高年资医师对结果不一致者进行人工复核, 综合判断结果差异的原因。

1.2.9 AI 分析与人工分析一致性比较

计算 1 263 例样本两种分析方法各项淋巴细胞计数百分比平均差值, 采用 SPSS17 统计软件通过配对 t 检验对 LS 细胞百分比进行统计学分析, 并计算 P 值, 以评价两种方法结果的一致性, 以 $P < 0.05$ 为差异有统计学意义。计算两种方法质控平均值, 以判定两种分析方法差异性的原因。运用 MedCalc19. 2. 1 统计软件对两种方法进行比对, 通过 Bland-Altman 图分析 95% 置信区间 ($95\% CI$)。计算 AI 分析方法质控的重复系数 (CR) 评价该方法的可重复性。

2 结 果

2.1 AI 分析结果的输出和可视化

AI 分析可快速检测出外周血中 LS 的数量和百分比, 可通过降维自动输出二维图, 清晰展现多维空间细胞群分布及抗原表达水平。见图 2。

2.2 AI 分析重复性

细胞分群类别、各群细胞的细胞数、百分比及每群细胞表达抗原的平均荧光强度均相同, 各项差值的 CV 均为 0。

2.3 AI 分析检测效率

AI 分析单个样本平均时间为 (1.36 ± 0.25) s, 而人工分析单个样本的平均时间约为 60 s, 分析疑难病例需 7~8 min。AI 分析较人工分析速度提高 50 倍以上。

2.4 AI 模型性能

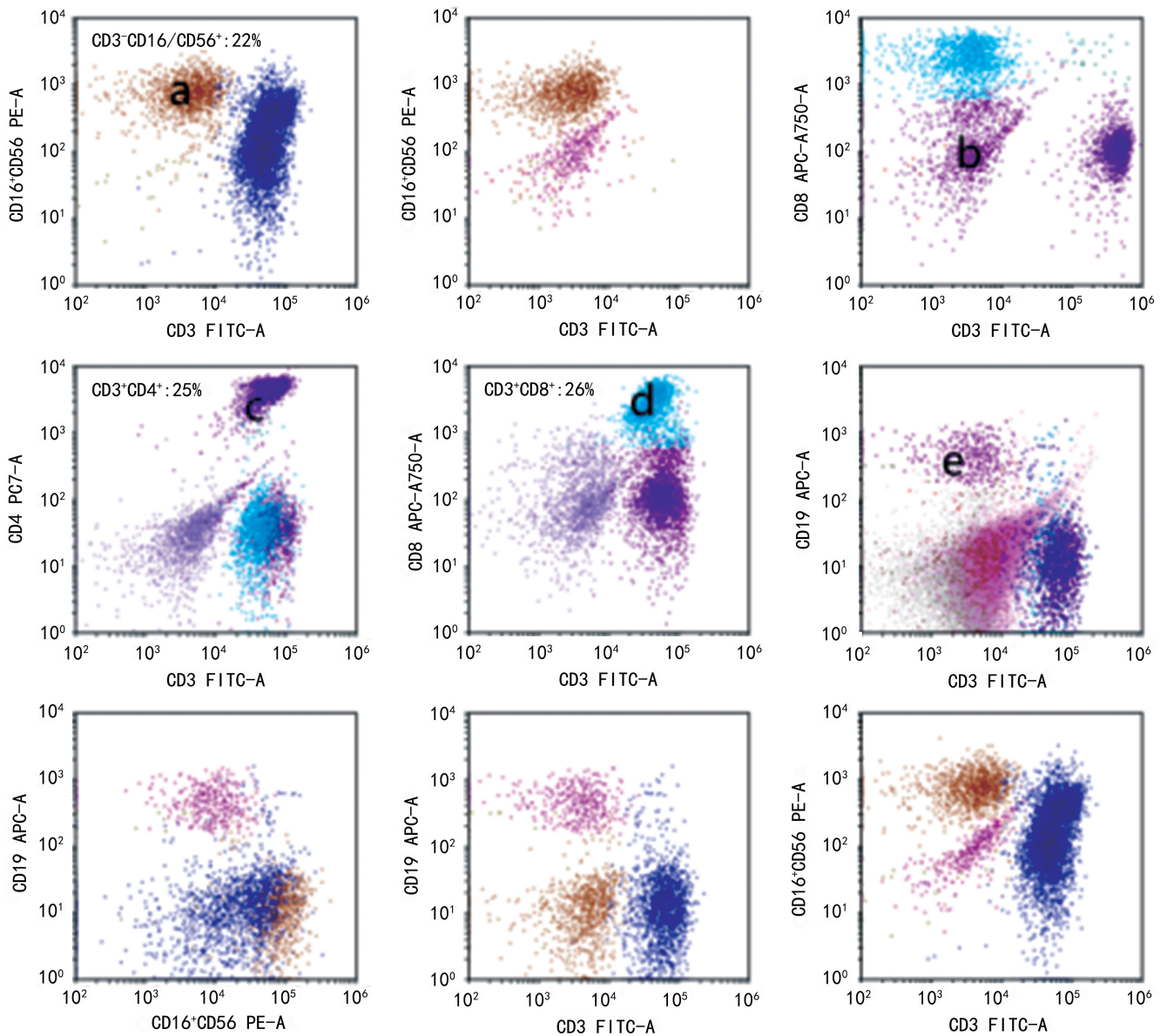
1 263 例样本中 1 199 例结果与人工结果一致, 检测通过比例为 94.93% 。64 例样本的 80 个差异项与人工分析结果有差异。差异项病例数分别为: $CD3^+ 5$ 例、 $CD3^+ CD4^+ 9$ 例、 $CD3^+ CD8^+ 23$ 例、 $CD3^- CD19^+ 11$ 例、 $CD3^- CD16^+ CD56^+ 32$ 例。

2.5 AI 分析与人工分析结果的一致性

两种方法 $CD3^+$ 、 $CD3^+ CD8^+$ 、 $CD3^+ CD4^+ / CD3^+ CD8$ 比值、 $CD3^- CD19^+ 4$ 项平均差值比较, 差异均有统计学意义 ($P < 0.05$)。经高年资医师复核, AI 分析也会因异常 T 淋巴细胞的存在导致结果分析的差异。其余 $CD3^+ CD4^+$ 和 $CD3^- CD16^+ / CD56^+ 2$ 项平均差值比较, 差异均无统计学意义 ($P > 0.05$), 两

项指标检测具有良好的一致性,可相互替代。见表 1。人工分析质控的平均值为 1.414, AI 分析质控的平均值为 0.384,提示 AI 分析比人工分析质量控制(QC)

更趋向于 0。AI 分析方法质控的 CR 为 2.833 1%, 95%CI:2.726 8~2.948 1%,均小于临床可接受的临界值范围±5%,表明 AI 分析方法可重复性好。



a: NK 细胞; b: TCR $\gamma\delta$ +T 淋巴细胞; c: 辅助性 T 淋巴细胞; d: 细胞毒性 T 淋巴细胞; e: B 淋巴细胞。

图 2 AI 分析检测患者外周血中 LS 的可视化二维图

表 1 淋巴细胞各亚群细胞百分比比较

细胞类型	差值	95%CI	t	P
CD3 ⁺	0.773	0.660~0.885	13.459	0.001
CD3 ⁺ CD4 ⁺	-0.039	-0.125~0.046	-0.901	0.368
CD3 ⁺ CD8 ⁺	-0.652	-0.749~-0.554	-13.129	0.001
CD3 ⁺ CD4 ⁺ /CD8 ⁺ CD8 ⁺	0.039	0.031~0.046	9.760	0.001
CD8 ⁻ CD19 ⁺	0.309	0.216~0.403	6.463	0.001
CD8 ⁻ CD16/CD56 ⁺	-0.052	-0.124~0.019	-1.430	0.153

3 讨论

本研究初步建立了 AI 辅助 MFC 检测人外周血 LS 含量及百分比的诊断方法,基于多维空间聚类细

胞群进行分析,可快速检测人外周血中 LS 百分比并降维后生成可视化结果。AI 分析效率较人工分析提升约 50 倍以上,极大地减少了临床诊断工作量、压缩了检测周转时间,初步满足了临床诊断的需求,并可将多维分析结果以二维散点图形式输出,提高了可视化效果,并对异常群示警允许人工订正,避免了偶然因素引起的错、误诊。日间重复性试验结果显示, CV 为 0, AI 分析不受人员疲劳程度、精神状态、人为失误等主观因素的影响。

研究发现,当总 LS 占比之和结果超过 5%~10%时考虑样本中可能包含大量异常 T 淋巴细胞亚

群,如 $\text{TCR}\gamma\delta^+$ T 淋巴细胞、双阴性 $\text{CD4}^- \text{CD8}^-$ 或双阳性 $\text{CD4}^+ \text{CD8}^+$ T 淋巴细胞。如外周血样本中存在 $\gamma\delta$ T 淋巴细胞时为满足“($\text{CD4}^+ + \text{CD8}^+$)% = ($\text{CD3}^+ \pm 5$)%”的检测标准^[12],而 AI 分析会出现圈门不准确。在人工分析 CD3/CD4 细胞分群时由于单核细胞与淋巴细胞在二维散点图上界限不清,人工分析将部分单核细胞纳入淋巴细胞计数范围内,导致人工分析较 AI 分析数据百分比降低,而 AI 会多方位空间划分,将 $\text{CD3}^- \text{CD4}^+$ 细胞群剔除整体淋巴细胞群。见图 3。对大部分单一 LS 分析表现出较好的分析速

度和准确率,但对个别复杂病例仍没有达到与人工分析完全相同的诊断效果,仍需高年资医师结合临床其他检测指标进行综合诊断,AI 分析模型的建立是基于医师的分析方法而建立的,在机器学习算法方面虽然可通过数据进行训练,但仍受制于单一检测手段分析结果而缺乏思维灵活性,进而无法完全替代高年资医师检测结果。但该 AI 模型仍可作为一种临床决策支持系统和辅助诊断工具用于临床,可为低年资医师树立信心。

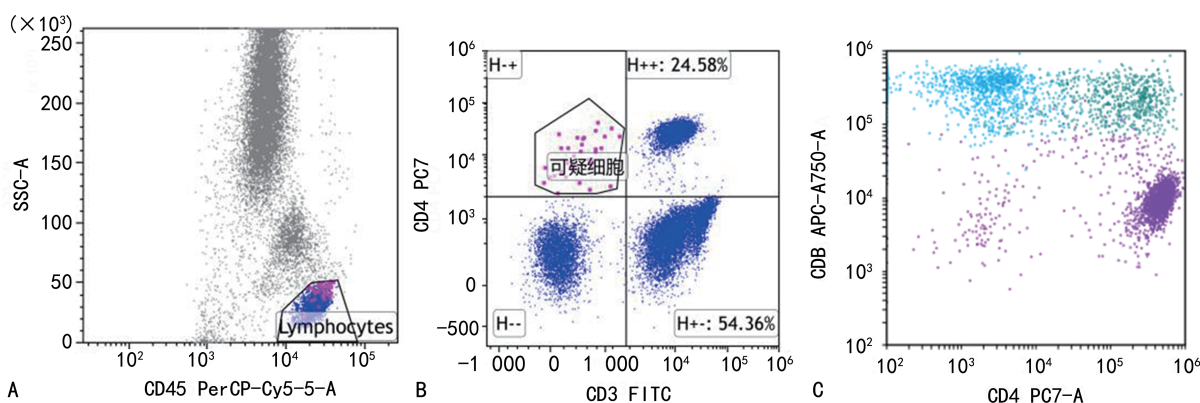


图 3 AI 分析与人工分析差异比较(CD3/CD4)
A:CD45 圈出 LS;B:人工分析的 $\text{CD3}^+ \text{CD4}^+$ 二维散点图,人工分析圈出的淋巴细胞中可能含有部分单核细胞(粉色);C:AI 分析的 $\text{CD8}^+ \text{CD4}^+$ 二维散点图。

图 3 AI 分析与人工分析差异比较(CD3/CD4)

AI 借助强大的数据分析能力和计算能力用于医学各领域,尤其是多维图像解读及大数据分析^[13-14]。研究者在疾病诊断领域挖掘 AI 辅助分析更快、更准、更便捷的方法,以满足临床疾病诊断的需要^[15]。但 AI 辅助 MFC 分析方法的相关研究较少见,不同领域和疾病诊断类型对算法的要求不尽相同,且大多研究均基于小样本量的探索性研究^[16-17],在临床实际应用中尚存在不足^[18]。有研究对比了 7 784 例淋巴瘤淋巴瘤患者和 8 493 例健康对照者的 FCS 数据,利用神经网络算法和 10 倍交叉验证建立了 AI 诊断模型,诊断准确率高达 97%^[19]。KO 等^[20]分析了 1 742 例急性髓系白血病(AML)患者和 5 333 例骨髓增生异常综合征(MDS)患者的 FCS 数据,通过对机器学习算法训练,使 AML 和 MDS 患者诊断率提高至 84.9%~92.4%,诊断准确率达 84.6%~89.7%,平均数据处理时间为 7 s,极大地提高了检测效率。周丽娜等^[17]以多维空间密度分布的非监督学习分群聚类算法为基础,应用决策树和随机森林等监督学习算法辨别细胞分类,并以二维图、降维 t-分布领域嵌入算法和热图进行可视化呈现,提高了 AI 辅助诊断微小残留病的灵敏度和准确度。本研究基于聚类算法和核密度估计方法在 AI 辅助流式细胞术检测 LS 方

面进行了初步探索及应用研究,AI 模型使用的机器学习算法与其他算法比较,其优势:(1)计算伸缩性。该算法使用多个高斯分布的组合刻画数据分布,计算伸缩性好。(2)参数依赖性。可调整参数为数据分布的均值和标准差。(3)普适性能力。描述能力和泛化能力均优于其他算法。(4)抗噪声能力。优于 K 均值聚类算法。(5)结果解释性。模型和结果均具有解释性。该算法能更好地捕捉流式细胞数据的分布特点进行细胞聚类和分群。

尽管近年来 AI 技术在医学领域取得了一定的成功,AI 机器学习方法在诊疗过程中的局限性和解决方案也被广泛地讨论^[21-23]。但探索过程中仍面临 AI 辅助诊断不流畅、检验结果重复性不够、检测机构间水平差异等诸多问题,从而阻碍了 AI 模型的推广应用。同时 AI 算法导致的医疗责任归属及其“黑匣子”“可解释性”等诸多问题也使 AI 辅助诊断在临床应用中受限而破坏了其应用价值。因此,仍需在临床更多领域、更多算法进行验证,并建立基于 AI 模型的诊疗指南和行业标准,以确保检测流程标准化和结果的准确性和灵敏度。本研究通过 1 263 例患者淋巴细胞百分比数据进行分析建立了 AI 辅助诊断模型,未来希望能应用 LS 绝对计数的 AI 分析辅助临床提供更可

靠、快速的结果,并有更多的机会对算法进行优化与训练,使诊断更精确,希望通过机器学习算法的优化和人工分析思路的高度融合不断提升 AI 模型识别复杂细胞群的能力。未来随着人类对 AI 的认知逐渐加深, AI 模型识别流式细胞数据能力将会随着不断深入的研究和探索成为医师的左膀右臂,大幅提升诊断效率。

参考文献

- [1] PICARD C, AL-HERZ W, BOUSFIHA A, et al. Primary immunodeficiency diseases: an update on the classification from the international union of immunological societies expert committee for primary immunodeficiency 2015[J]. *J Clin Immunol*, 2015, 35(8): 696-726.
- [2] VAN TILBURG C M, VAN GENT R, BIERINGS M B, et al. Immune reconstitution in children following chemotherapy for haematological malignancies: a long-term follow-up[J]. *Br J Haematol*, 2011, 152(2): 201-210.
- [3] FERNÁNDEZ-RUIZ M, LÓPEZ-MEDRANO F, ALLENDE L M, et al. Kinetics of peripheral blood lymphocyte subpopulations predicts the occurrence of opportunistic infection after kidney transplantation[J]. *Transpl Int*, 2014, 27(7): 674-685.
- [4] VANCSEA A, SZABO Z, SZAMOSI S, et al. Long-term effects of rituximab on B cell counts and autoantibody production in rheumatoid arthritis: use of high-sensitivity flow cytometry for more sensitive assessment of B cell depletion[J]. *J Rheumatol*, 2013, 40(5): 565.
- [5] HE D, YU Y L, YAN W B, et al. Individualized rituximab treatment for relapsing neuromyelitis optica: a pediatric case report[J]. *PEDIATRIC NEUROLOGY*, 2014, 51(2): 255-258.
- [6] 翁香琴, 王莉莉. 多参数流式细胞术检测急性白血病及浆细胞肿瘤微小残留病中国专家共识(2017年版)[J]. *中华血液学杂志*, 2017, 38(12): 1001-1011.
- [7] 王建中. 流式细胞术诊断急性白血病微小残留病: 当前主要的技术问题和对策[J]. *诊断学理论与实践*, 2017, 16(1): 7-11.
- [8] LI Y, MAHJOUBFAR A, CHEN C L, et al. Deep cytometry: deep learning with real-time inference in cell sorting and flow cytometry[J]. *Sci Rep*, 2019, 9(1): 11088.
- [9] 周剑峰, 楚玉兰, 钟云峰, 等. 微小残留病阳性急性 B 淋巴细胞白血病患者 106 例免疫表型分析[J]. *白血病·淋巴瘤*, 2019, 28(1): 43-46.
- [10] HYRKAS J, CLAYTON S, RIBALET F, et al. Scalable clustering algorithms for continuous environmental flow cytometry[J]. *Bioinformatics*, 2016, 32(3): 417-423.
- [11] 王先文, 王懿男, 暴洪涛, 等. 一种快速自动分析流式数据方法研究[J]. *军事医学*, 2015, 39(10): 736-741.
- [12] 中国儿童免疫与健康联盟免疫评估工作组, 中国医师协会儿科医师分会风湿免疫专委会, 中国医师协会儿科医师分会儿童过敏专委会, 等. 流式细胞术分析外周血淋巴细胞亚群在儿科的临床应用专家共识(2019版)[J]. *中华儿科杂志*, 2019, 57(6): 424-428.
- [13] MOEN E, BANNON D, KUDO T, et al. Deep learning for cellular image analysis[J]. *Nat Methods*, 2019, 16(12): 1233-1246.
- [14] 于观贞, 陈颖, 朱明华. 人工智能从数字病理切入精准医疗[J]. *临床与实验病理学杂志*, 2021, 37(4): 381-383.
- [15] HU X, YI W I, JIANG L, et al. Classification of metaphase chromosomes using deep convolutional neural network[J]. *J Comput Biol*, 2019, 26(5): 473-484.
- [16] HANSMANN L, HAN A, PENTER L, et al. Clonal expansion and interrelatedness of distinct B-Lineage compartments in multiple myeloma bone marrow[J]. *Cancer Immunol Res*, 2017, 5(9): 744-754.
- [17] 周丽娜, 贺环宇, 苗林子, 等. 人工智能辅助多参数流式细胞术诊断儿童急性 B 淋巴细胞白血病微小残留病[J]. *中华检验医学杂志*, 2020, 43(12): 1196-1204.
- [18] SALAMA M E, OTTESON G E, CAMP J J, et al. Artificial intelligence enhances diagnostic flow cytometry workflow in the detection of minimal residual disease of chronic lymphocytic leukemia[J]. *Cancers (Basel)*, 2022, 14(10): 2537.

(下转第 4303 页)

平台能够通过数据清洗,结构化、归一化处理,词汇切分,语义处理等,将原来非标准的、不完整的数据进行规范化,提升了医院医疗数据质量,完善了医院医疗质量管理。

2.4 催生智能化应用落地

通过大数据平台处理和汇聚的医疗数据并非只是为了科研、教学或管理等场景使用,更多地是为后续医院在智能化方面的建设打下基础。首先需要把这些散落的数据整合成为标准的患者诊疗模型,完成诊疗模型构建和数据处理,再根据这些整合数据,通过人工智能学习,构建智能辅诊系统,预测出患者下一步的健康变化,自动推荐诊断和治疗方案,实现更大的医疗数据生态,催生更多医学领域的智能化应用落地^[10]。

综上所述,我国许多大型医院经过几十年信息化发展,积累了大量的临床诊疗数据,这些数据以前在管理和使用上还不够规范和完善,随着大数据技术的不断发展成熟,医院构建大数据平台能够将数据进行有效汇聚和管理,提升医院诊疗数据质量,为医院临床科研的支撑、诊疗模式的转型和医疗领域的智能化应用带来机遇和动力。由于医院在大数据和人工智能领域的建设起步较晚,相关技术人员较少,经验不足,在大数据建设和管理上还会遇到很多新的难点和痛点。只有通过新技术新业务的不断学习,掌握更为有效的数据采集、清洗、加工工具和方法,在实践中不断摸索和发展,才能建设好医疗科研大数据平台,为医院创新发展打下坚实的基础^[11]。

参考文献

- [1] 郑西川,陈霆,傅一旻,等. 医疗机构大数据分析功能及应用策略研究[J]. 中国数字医学,2018,13(3):13-15.
- [2] 倪晓华. 利用 GATE 的 XML 配置文件实现病历短语抽取的机器学习方法[J]. 中国医疗设备,2017,32(7):124-125.
- [3] 赵广智. Hadoop 与关系型数据库在电信行业中的应用研究[J]. 广东通信技术,2017,37(9):40-46.
- [4] 刘晓亮,王坤,马军,等. 大数据时代的卫生信息化建设思考[J]. 中国卫生信息管理杂志,2014,11(1):43-46.
- [5] 米春香. 大数据技术助力医院精细化管理[J]. 中国医疗设备,2019,34(7):93-95.
- [6] 王才有. 大数据时代的医院数据平台建设[J]. 中国医院,2016,20(1):15-17.
- [7] 刘金晶,曹文洁. 大数据环境下的数据质量管理策略[J]. 软件导刊,2017,16(3):176-178.
- [8] 王笑笑,赵飞,梁志金,等. 基于数据挖掘技术的护理学研究现状[J]. 解放军护理杂志,2019,36(8):59-62.
- [9] 王韶锋,赵善斌,杨静,等. 医院数据治理与数据质量提升研究[J]. 现代医院,2021,21(11):1761-1763.
- [10] 汪鹏,吴昊,罗阳,等. 医疗大数据应用需求分析与平台建设构想[J]. 中国医院管理,2015,35(6):40-42.
- [11] 王红迁,汪鹏,王飞,等. 保障医疗大数据安全及其实践[J]. 医学信息学杂志,2017,38(12):43-47.
- [19] HILLEIN A, ZHAO M, SCHABATH R, et al. An artificial intelligence (AI) approach for automated flow cytometric diagnosis of B-Cell lymphoma[J]. Blood, 2018, 132:2856.
- [20] KO B S, WANG Y F, LI J L, et al. Clinically validated machine learning algorithm for detecting residual diseases with multicolor flow cytometry analysis in acute myeloid leukemia and myelodysplastic syndrome [J]. E Bio Med, 2018, 37(1):91-100.
- [21] HOLZINGER A, LANGS G, DENK H, et al. Causability and explainability of artificial intelligence in medicine[J]. Wiley Interdiscip Rev Data Min Knowl Discov, 2019, 9(4):e1312.
- [22] AMANN J, BLASIMME A, VAYENA E, et al. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective[J]. BMC Med Inform Decis Mak, 2020, 20(1):310.
- [23] CUTILLO C M, SHARMA K R, FOSCHINI L, et al. Machine intelligence in healthcare—perspectives on trustworthiness, explainability, usability, and transparency[J]. NPJ Digit Med, 2020, 3(1):47.

(收稿日期:2022-02-23 修回日期:2022-06-28)

(收稿日期:2022-04-21 修回日期:2022-09-11)

(上接第 4299 页)