

人工智能在血液疾病诊疗中的应用研究进展

张礼潜¹, 安倬玉¹, 崔丽娟², 李文倩³, 张晓辉^{1*}

1. 北京大学人民医院, 北京大学血液病研究所, 国家血液系统疾病临床医学研究中心, 血液肿瘤细胞和基因治疗北京市重点实验室, 北京 100044

2. 宁夏医科大学总医院血液科, 银川 750003

3. 青海省人民医院血液风湿科, 西宁 810007

* 联系人, E-mail: zhangxh@bjmu.edu.cn

2024-11-05 收稿, 2025-02-09 修回, 2025-02-10 接受, 2025-03-19 网络版发表

国家重点研发计划(2023YFC2507803)、国家自然科学基金(82230004, 82430006)、首都卫生发展科研专项(2022-1-4082)资助

摘要 血液疾病指原发于造血系统或主要累及血液和造血器官的疾病, 主要包括良性血液疾病和恶性血液疾病两种类型, 不仅对患者的生活质量和生命安全造成负面影响, 也给家庭和社会带来了沉重的负担. 随着计算机与机器学习等相关技术的快速发展, 人工智能已被广泛应用于医学领域和临床研究. 在血液疾病诊疗方面, 基于随机森林、决策树、支持向量机和线性回归等机器学习算法构建的人工智能模型展现出了卓越的工作效能, 在合理利用既有数据、图像识别和组学分析等任务中取得了优于传统方法的表现. 本文综述了人工智能应用于血液疾病预测、诊断、预后评估与治疗指导领域的研究进展, 总结了人工智能技术在该领域的突出成果与局限性, 以期为推动机器学习技术进一步应用于血液疾病诊疗提供参考.

关键词 血液疾病, 人工智能, 机器学习, 临床诊疗

血液疾病是指原发于造血系统或影响造血系统伴发血液异常改变的疾病. 常见的血液疾病包括贫血、免疫性血小板减少症等良性血液疾病和包括白血病(leukemia)、淋巴瘤与多发性骨髓瘤(multiple myeloma, MM)等在内的恶性血液疾病^[1]. 不同血液疾病的患病率与预后情况不尽相同^[2]. 流行病学调查显示, 仅在2022年, 我国新增血液系统肿瘤患者逾21万人, 直接因血液系统肿瘤死亡的人数也超过8000人^[3]. 近年来, 非霍奇金淋巴瘤和MM等恶性疾病的发病率更是持续升高, 给患者造成了沉重的疾病负担^[4]. 提高血液疾病的诊治水平是当前临床医学领域的重要命题之一.

人工智能(artificial intelligence, AI)是计算机科学的一门分支学科, 被定义为研究、开发用于模拟、延

伸和扩展人的智能行为的理论、方法及技术等的一门综合性科学. 在过去十余年间, 人工智能技术得益于计算能力的增长和算法的迭代进步而蓬勃发展, 其与医学领域之间的联系也随之日趋紧密^[5]. 绝大多数应用于医学研究领域的人工智能技术均属于机器学习(machine learning, ML)范畴, 这使得二者的概念在医学领域常常发生混淆^[6,7]. 机器学习被定义为计算机通过算法自主获取源数据中的核心信息并寻找最有效的途径来实现既定目标的过程^[8], 包括监督学习、半监督学习、无监督学习等类型^[9]. 在当前的血液疾病研究领域, 人工智能与机器学习技术已被广泛应用于多种疾病的预测、诊断、预后评估之中^[6], 在数据处理、图像处理、遗传学数据分析和策略优化等任务中均取得了卓越表现.

引用格式: 张礼潜, 安倬玉, 崔丽娟, 等. 人工智能在血液疾病诊疗中的应用研究进展. 科学通报, 2025, 70: 1819–1828

Zhang L, An Z, Cui L, et al. Research progress of artificial intelligence in the clinical diagnosis and treatment of hematological diseases (in Chinese). Chin Sci Bull, 2025, 70: 1819–1828, doi: 10.1360/TB-2024-1190

1 血液疾病临床诊疗视角下的人工智能

人工智能的概念最早由John McCarthy等提出，并由Alan Turing进一步完善^[7]。人工智能技术可以被理解为一项经由模拟人类认知功能而使机器能够获得与人类相似的感知力、洞察力，并做出相应判断和决策的技术^[10]。作为人工智能的一个子领域，机器学习被定义为一种能够使机器实现从既有数据中自动学习并执行预测或决策任务的算法，而非直接编程以完成目标^[7,11]。机器学习技术在处理复杂或非线性数据等方面具有突出优势，这为其被应用于具有复杂数据集的临床医学科研领域提供了可能。鉴于机器学习技术的广泛应用，在有关血液疾病诊疗的临床语境下，在血液疾病诊疗中，人工智能与机器学习的概念常被交替使用^[6,12]。需要注意的是，尽管机器学习是应用于临床领域的主要人工智能类型，大语言模型和大数据分析等人工智能分支领域在血液疾病诊疗中的巨大应用潜力也不容忽视^[13]。大语言模型以对自然语言的概率分布进行建模为目标，研究者们对其在智能问诊、辅助诊断、治疗决策等临床诊疗领域中的应用前景进行了广泛的探索。尽管当前鲜有研究将大语言模型应用于血液疾病领域，但Kumari等人^[14]的工作也揭示了大语言模型在处理复杂血液疾病案例相关信息中的良好前景。大数据分析技术立足于对大规模数据集的分析与处理，已被充分应用于数据整合、数据标准化与数据分析等临床研究场景。在血液疾病诊疗领域，大数据分析技术为研究者们实现对多中心、多组学数据的有效整合与分析提供了有效工具^[15]。基于大数据分析技术，Medeiros等人^[16]通过整合全国范围内的信息对老年AML

患者的预后情况进行了分析，Warnat-Herresthal等人^[17]则依托于多组学数据构建了准确的急性粒细胞白血病(acute myeloid leukemia, AML)预测模型。需要注意的是，机器学习、大语言模型与大数据分析等人工智能技术在应用语境下并非孤立存在，往往被联合使用以完成既定目标。例如，大语言模型的建模过程可能涉及机器学习的监督学习或无监督学习方法，而经大数据分析技术整合后的数据也可以经由机器学习方法进一步完成临床预测模型的构建。

在血液疾病诊疗领域，对机器学习的分类主要以学习方法的不同为导向。常见的机器学习类型包括监督学习、半监督学习、无监督学习和强化学习等^[9,12]，不同的机器学习类型依托于不同类型的算法以实现，表1对常见的机器学习类型与算法进行了总结。深度学习是另一种受到广泛关注的机器学习类型，其与传统机器学习在结构模式上存在不同，具体表现为对数据的特征与结构模式的学习建立在多层神经网络结构的基础上。深度学习在学习数据中的深层次特征和非线性关系方面有着突出优势，在高维数据处理、图像识别与分析等任务中取得了良好表现^[5]。

应用于血液疾病诊疗场景中的机器学习工作流程包括数据收集和预处理、算法选择与建模、模型训练、模型评估与优化以及临床应用几项内容^[7,18]。数据收集是机器学习技术应用于优化血液疾病诊疗方法的前提；研究者可能选用的算法模型与所研究的血液疾病问题共同决定了一项研究中对数据预处理的要求，既包括以恰当的方法处理原始数据中的缺失或异常数据，也包含特征选择——即识别并选择对解决目标问

表 1 机器学习分类、代表算法及应用场景

Table 1 Classification, representative algorithms, and application scenarios of machine learning

分类	主要功能	代表算法	应用场景
监督学习	训练模型根据输入数据特征预测输出标签，常用于解决分类与回归问题	随机森林	可用于解决分类问题和回归问题、处理大规模数据集等
		决策树	可用于解决分类问题和回归问题等
		支持向量机	可用于解决二分类问题、小样本问题和回归问题等
半监督学习	使用包含已标记和未标记数据的数据集进行模型训练，常用于文本分类、图像识别等领域	自训练	可用于图像识别、异常检测等
		半监督支持向量机	可用于解决分类问题和回归问题、图像识别和异常检测等
非监督学习	训练模型探索数据的内在结构和模式，常用于解决聚类、降维、异常检测问题	K-均值聚类	可用于图像处理、生信分析等
		谱聚类	可用于图像处理、文本挖掘、生信分析等
强化学习	在与环境的交互中进行模型训练，常用于机器人控制等交互领域	Deep Q-Network	可用于图像处理等
		Deep Deterministic Policy Gradient	可用于策略优化等

题贡献较大的变量;对算法的选择与以此为基础的模型构建直接决定了机器学习技术实现既定目标的路径;在完成建模后,数据集往往被分为训练集和测试集,由训练集得到的模型随后在测试集中接受验证和评估,并根据评估结果被加以校准;此外,经由上述过程构建出的人工智能模型还需要经过对模型的诠释、评估不确定性以及开放访问途径等步骤,从而最终应用于血液疾病诊疗工作^[7,12]。

基于上述工作流程,应用于血液疾病诊疗领域的人工智能和机器学习技术在对血液疾病的预测、诊断、风险分层和预后评估等方向都取得了出色表现。在不同的具体任务类型中,人工智能也展现出了优于传统方法的效能:依托于深度学习和大数据分析等技术,人工智能方法能够更加合理和充分地利用现有的血液疾病相关数据,为研究者和临床医生提供更多的指导信息;人工智能所具有的快速、准确识别图像信息能力和良好的特征提取能力等特点为满足血液疾病临床研究中病理图像处理的需求开辟了新路径;在组学数据分析方面,人工智能卓越的数据整合能力也为高效处理高维多组学数据、提高聚类精度和降低计算成本提供了可能^[19]。图1对人工智能技术应用于血液疾病诊疗的基本范式进行了概括。

2 人工智能在血液疾病预测中的应用

预测特定人群中血液疾病发生的可能性可以为血液疾病尤其是恶性血液疾病的早期诊断与提供支持,从而帮助临床医生及早干预并改善患者的预后结局。在血液疾病预测领域,人工智能已被成功应用于评估部分疾病与并发症的发生风险。在白血病复发预测的研究中,Pan等人^[20]应用随机森林(random forest, RF)、决策树(decision tree, DT)、支持向量机(support vector machine, SVM)和线性回归(linear regression)4种分类算法构建了急性淋巴细胞白血病(acute lymphocytic leukemia, ALL)的复发预测模型,而Hauser等人^[21]则应用其他机器学习算法探索了基于全血细胞计数对慢性粒细胞白血病(chronic myeloid leukemia, CML)进行预测的可能性,二者所得到的模型分别表现出了高达0.92和0.90的优异曲线下面积(area under the curve, AUC)。一些研究者对造血干细胞移植(hematopoietic stem cell transplantation, HSCT)后的白血病复发预测模型进行了深入研究,Fuse等人^[22]应用交替决策树(alternating decision tree, ADTree)算法建立起了移植后一年内的

急性白血病复发预测模型,该模型被认为具有良好的通用性与可解释性;Zhang等人^[23]则将自动图像分析检测技术应用于预测AML患者的HSCT后复发,该团队利用深度学习方法对39例AML患者的骨髓细胞涂片进行了图像处理,进而使用特征选择算法基于骨髓细胞的形状和纹理特征形成了预测模型的顶部特征,最终应用RF等算法构建了能够有效预测AML患者移植后复发的模型。

关于人工智能技术应用于其他血液肿瘤预测的研究也取得了一些进展。Goswami等人^[24]采用由谱聚类和快速节俭树算法组成的堆叠机器模型对MM患者自体HSCT后36个月的复发进行了预测,并提出了一个可应用于指导临床诊疗实践的分期方案。Radhachandran等人^[25]利用机器学习技术建立了优化的极限梯度提升(extreme gradient boosting, XGB)模型以实现诊断前预测骨髓异常增生综合征(myelodysplastic syndrome),该模型高达0.87的AUC显著优于传统算法。也有研究者关注到了HSCT术后并发症预测领域的相对空白。Arai等人^[26]应用ADTree算法在HSCT治疗后患者人群中建构了基于机器学习的急性移植抗宿主病(acute graft versus-host disease, aGVHD)预测模型,并在验证队列中确认了所得到模型的可靠性;有赖于机器学习方法,能够有效预测HSCT后静脉血栓栓塞和aGVHD发生及患者生存率的临床预测模型得以成为现实^[27,28]。Fan等人^[29]完成的用于预测HSCT后应用抗胸腺细胞球蛋白预防GVHD人群EB病毒(Epstein-Barr virus)再感染的综合机器学习模型也是这一研究方向上的成功案例。Musial等人^[30]则将机器学习算法与肾小管损伤标志物相结合,提出了新的儿童HSCT后急性肾损伤预测模型,拓展了移植后急性肾损伤的预测路径。

人工智能技术同样被成功应用于对良性血液疾病的预测。Jardim等人^[31]探索了人工智能用于预测A型血友病患者发生凝血因子抑制物生成这一并发症的可能性,尽管所得模型仍有待进一步验证,但也体现了研究者所采用的基于网络的机器学习算法的潜力。Saputra等人^[32]应用人工智能技术以解决不同贫血类型难以快速区分的难题,构建起了基于极限学习机(extreme learning machine)算法的贫血预测与分类模型;Garduno-Rapp等人^[33]使用深度学习技术方法构建了多种神经网络模型,经对比得到了能够对缺铁性贫血进行早期预测的门控循环单元模型;Schipper等人^[34]则利用血常规参数分别构建了AUC高达0.88和0.84的XGB与

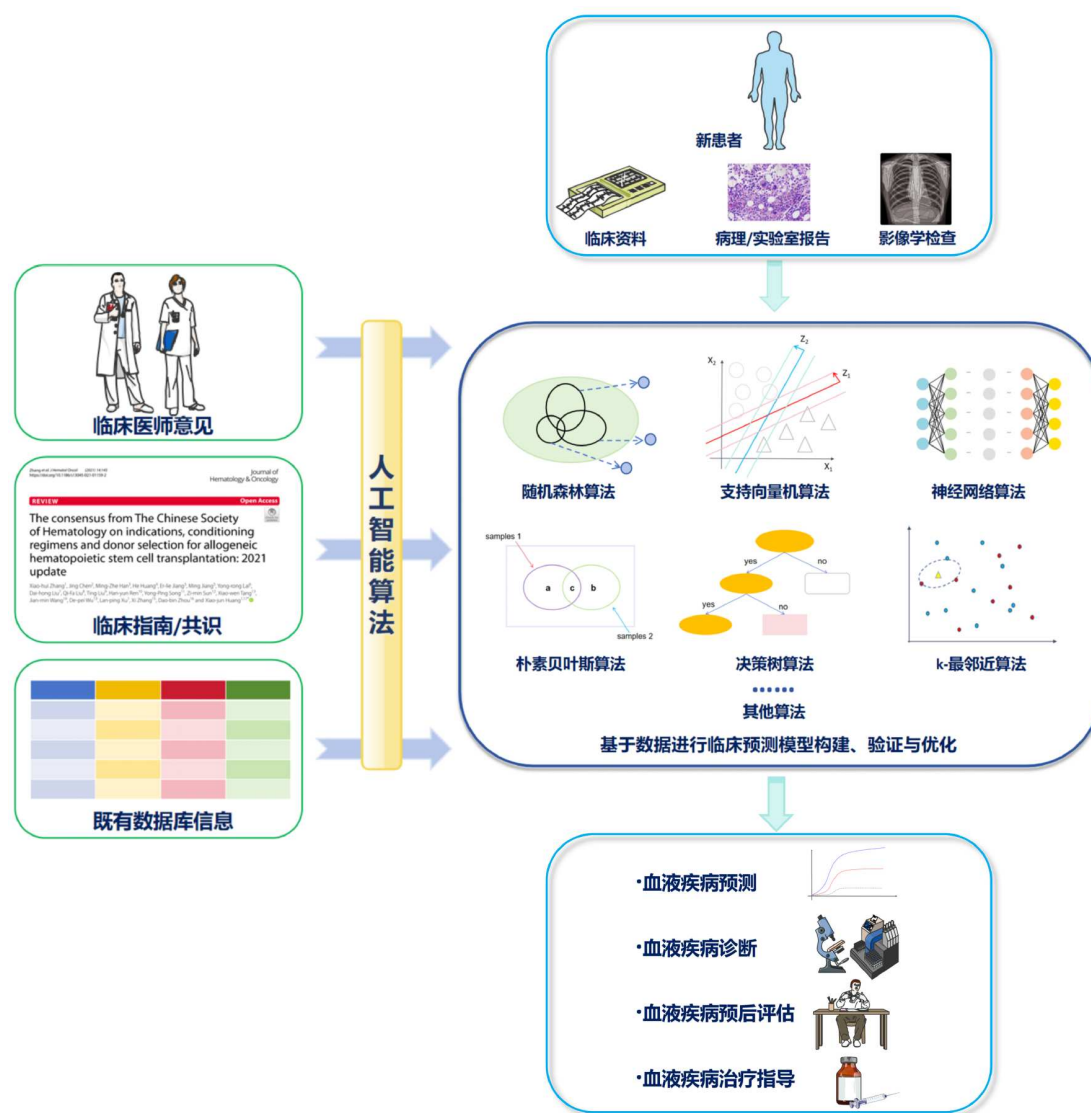


图 1 人工智能赋能血液疾病诊疗

Figure 1 Diagnostic and therapeutic models for hematologic diseases empowered by artificial intelligence

逻辑回归模型，提供了有效预测和区分地中海贫血与缺铁性贫血的方法。

3 人工智能在血液疾病诊断中的应用

对血液疾病的准确诊断是疾病诊疗过程中的重要环节，是实现血液疾病早期治疗的重要前提，不仅有利于改善患者预后，还可以减轻家庭与社会的疾病负担。人工智能赋能的新诊断方法已在各项临床试验中取得了良好反馈，在白血病领域更是取得了突出的进展。不同研究者基于不同的临床检测项目数据利用人工智能

技术构建了白血病诊断工具。Alcazer等人^[35]依托于多中心数据库，将包括血常规、血生化和凝血指标在内的共计19项常见临床实验室检测指标纳入模型中，建立起了能够对ALL和AML等急性白血病亚型进行准确诊断的急性白血病人工智能预测模型。Haider等人^[36]回顾性纳入了1577例血液肿瘤患者的血常规数据，采用径向基函数网络(radial basis function network)架构对数据集中的不同血液恶性肿瘤进行分类诊断建模，得到了AUC高达0.905的、能够有效区分和诊断包括慢性淋巴细胞白血病(chronic lymphocytic leukemia,

CLL)在内的多种白血病亚型及其他恶性血液疾病的机器学习诊断模型。El Alaoui等人^[37]基于全血细胞计数比较了RF、XGB和DT算法所得到的模型性能,发现基于DT算法构建的模型对诊断ALL有高达91.4%的准确率。

得益于人工智能在图像识别与信息提取方面的突出优势,人工智能技术被广泛应用于基于血涂片或其他病理图像的白血病诊断领域。Mohammed及其团队^[38]提出了一种能够根据血涂片图像将白细胞分类为CLL或正常细胞的系统,该系统源自SVM算法、k-最邻近(k-nearest neighbors, KNN)算法和DT算法组成的融合模型。该团队还将该系统的诊断效能与流式细胞术进行了对比并确认了二者的一致性。Steinbuss等人^[39]也基于淋巴结组织病理学图像数据开展了应用深度学习技术构建CLL等疾病诊断模型的研究。在病理图像之外,人工智能技术同样被成功应用于依托流式细胞术或遗传学特征开展的白血病诊断之中。既往综述对结合了聚类技术和机器学习的计算机驱动流式细胞术分析方法进行了总结,指出了其有助于提高恶性血液疾病的诊断准确率^[40]。借助人工智能技术,研究者们对患者的遗传学特征与白血病诊断之间的有机关联进行了发掘,其中Warnat-Herresthal等人^[17]的研究具有突出代表性。该团队使用来自105个不同研究的12029例样本的基因组学和转录组学数据开展大规模研究,应用数据驱动的高维统计方法构建了基于机器学习与深度学习技术的AML诊断模型,该模型在准确性高的同时兼具良好的可扩展性和低边际成本。

人工智能技术在其他血液疾病诊断中的应用也取得了长足的进展。Gutierrez-Rodriguez等人^[41]使用25项在患者初诊时所记录的临床或实验室变量开发了获得或遗传性骨髓衰竭(bone marrow failure)的诊断模型,从新视角出发强调了初始评估对于骨髓衰竭诊断的重要意义。Ramzan等人^[42]将兼具注意力模块和空间注意力的综合机器学习模型创新性地应用于贫血诊断中,所构建的AlexNet多重空间注意力模型具有高达99.58%的诊断准确率。RF、SVM和KNN等机器学习算法均被应用于MM诊断领域,并展现出了出色的诊断效能。在HSCT后并发症领域,Sharifi等人^[43]采用无监督方法对HSCT后肺部并发症的鉴别诊断进行了研究;Shao等人^[44]则在检测microRNA的基础上应用RF算法建立起了能够高效诊断HSCT后感染性发热的人工智能模型,该模型表现出了超过90%的诊断准确率。

4 人工智能在血液疾病预后评估与治疗指导中的应用

血液疾病尤其是血液恶性疾病的预后存在很大异质性,及时且准确的危险分层与预后评估可以为临床干预提供有效指导,帮助医生平衡治疗获益与可能的副反应或风险,降低不良预后结局的发生可能。Chen等人^[45]引入无监督聚类步骤以提供CLL人群的离散危险分层,为更好地解释应用于血液疾病预后评估中的机器学习模型提供了可行的方法。Agius等人^[46]采用恰当的预处理方式消减了数据集中缺失变量的影响,并通过整合28种机器学习算法开发出了慢性淋巴细胞白血病(chronic lymphocytic leukemia, CLL)治疗-感染模型(CLL-TIM),该模型在内部测试队列与独立的外部队列得到了验证,并展现出优于现有CLL预后评估金标准的良好评估效能。Qin等人^[47]着眼于AML发病机制中涉及的异常程序性细胞死亡机制,开发了将遗传学数据与AML患者的临床预后及药物治疗反应相关联的机器学习模型。在儿童白血病方面,来自国内外的八项临床研究也揭示了RF和LASSO(least absolute shrinkage and selection operator)等机器学习算法应用在儿童急性白血病预后结局评估领域的良好前景^[48]。

以机器学习方法为代表的人工智能同样被广泛应用于评估白血病以外的血液疾病的预后情况。Farswan等人^[49]结合常见的实验室检查参数构建了优于传统分期的MM改良分期系统,其他研究团队也分别基于不同的遗传学数据使用机器学习模型实现了对MM患者的总生存期的准确预测和评估。借助神经网络分析,研究者们也构建起了患者的遗传信息和淋巴瘤等血液系统肿瘤的预后之间的联系^[50]。Zhang及其团队^[51]将机器学习技术应用于原发性免疫性血小板减少症(primary immune thrombocytopenia, ITP)患者的危重出血事件分析中,并将基于多中心回顾性数据开发出的RF算法模型在国内的39个中心进行前瞻性验证,提出了能够快速且准确识别ITP患者出血风险概况以提供临床决策指导的人工智能方法。近期的一项研究也揭示了应用机器学习方法通过特定的凝血因子蛋白序列预测女性A型血友病患者预后的可能性^[52]。

人工智能技术在血液疾病预后评估中的应用也被拓展到多种治疗措施或方案的治疗反应预测领域,智能模型对于疗效的准确评估在临床决策中意义重大。HSCT的预后评估是学者们关注的重点,包括RF、随机

生存森林(random survival forest)和SVM等算法均被成功应用于HSCT后患者的预后分析之中。依托人工智能方法, Tislevoll等人^[53]使用多维细胞计数技术对AML患者的生存结局进行了有效评估, 以Lee等人^[54]为代表的多个团队则分别对血液肿瘤患者的基因表达情况与药物治疗反应间存在的联系进行了探索。Xu等人^[55]也在ITP人群中利用机器学习算法建立了AUC高达0.964的疗效预测评分系统, 可用于指导ITP患者的个体化治疗。Rodríguez-Belenguer等人^[56]于近期发表的工作聚焦于血液肿瘤患者接种新型冠状病毒疫苗后可能的低应答或无应答情况, 应用机器学习方法得到了准确且解释性良好的预后模型, 其相对新颖的研究方向具有借鉴意义。

5 血液疾病诊疗中人工智能的局限性

尽管在应用层面已经取得了许多成果, 但血液疾病诊疗领域中的人工智能仍存在着一定的局限性。首先, 人工智能系统的准确性高度依赖于原始数据集的质量, 这对原始数据集的完整性和可靠性提出了要求, 而临床研究所得到的复杂或存在缺失的数据集可能对基于此进行的人工智能应用研究造成不利影响。尽管如此, 借助在特征识别等领域相较传统方法有明显优势的机器学习方法, 人工智能技术不仅可能克服复杂的原始数据集限制, 还可能展现出优于既往统计方法的模型效能。过小的样本量也可能损害机器学习模型的性能^[13], 为机器学习与大数据分析技术在罕见病诊治或其他基于小规模数据库的研究中的应用造成了一定制约。这对研究者在技术层面提出了更高的要求, 能否根据待解决的临床问题选取恰当的人工智能工具与算法类型, 可能直接决定了人工智能能否成功应用于受到数据限制的血液疾病治疗场景中。需要指出的是, 尽管部分研究采用了多中心临床数据^[35,52], 但大多数有关人工智能的血液疾病研究都是在来自单中心的数据基础上完成的, 这也削弱了所得到模型的外部适用性。在预测模型拟合以外, 以大语言模型为代表的交互性人工智能技术还面临着输出幻觉、泛化等问题, 这些问题有赖于完善模型建构方法或优化评估指标等手

段加以解决^[14]。

其次, 来自伦理方面的挑战也是人工智能技术在应用于血液疾病的临床诊疗时必须面对的问题。包括大语言模型和大数据分析在内的人工智能技术对医患关系的潜在影响、患者隐私泄露的风险、社会偏见在决策过程中被放大的可能性以及人工智能被用于牟取不当利益的风险等问题均需要经过审慎的伦理考量加以解决^[57,58]。

此外, 对人工智能模型的合理解释也是决定其临床应用价值的重要因素。人机交互是人工智能技术走向应用的一项重要内容, 而许多模型对医生而言仍是黑箱状态, 这对人工智能系统应用于临床实践中产生了阻碍, 博弈论等方法乃至大语言模型技术的应用有望解决这一问题^[59,60]。人工智能工具所带来的人机交互需求也给临床医生提出了进一步的要求, 在掌握有关人工智能系统的基本知识基础上, 结合现有的临床指南、共识与自身经验依然是作出临床决策的关键。在对人工智能模型的评估方面, 传统的统计学评价指标如特异度和灵敏度可能存在局限性, 仍需要引入如精确度(真阳性数除以真阳性和假阳性数之和)等更加有效的评估指标以准确判定人工智能工具的效能^[7]。

6 讨论与展望

人工智能技术的蓬勃发展推动了其在医学领域的应用日趋成熟, 并在血液疾病这一专科领域体现出推动传统诊疗范式发生转变的巨大潜能。本文基于有关血液疾病诊疗的应用视角对人工智能进行了阐释, 并回顾了人工智能在预测、诊断、危险分层和预后评估等血液疾病诊疗领域所取得的应用性进展, 强调了以机器学习为代表的人工智能技术在图像识别和高维数据处理等方面表现出的卓越性能, 也对大语言模型、大数据分析技术等人工智能技术的应用现状进行了总结。尽管仍面临着原始数据的制约、伦理争论与人机交互障碍等挑战, 但人工智能已经展现出了良好的发展前景, 有望进一步应用于临床实践之中, 在实现科技赋能优化现有临床诊疗范式的同时改善患者预后结局。

参考文献

- 1 Zhu Y, Huang Y, Tan Y, et al. Single-cell RNA sequencing in hematological diseases. *Proteomics*, 2020, 20: 1900228
- 2 Fitzmaurice C, Abate D, Abbasi N, et al. Global, regional, and national cancer incidence, mortality, years of life lost, years lived with disability, and

- disability-adjusted life-years for 29 cancer groups, 1990 to 2017. *JAMA Oncol*, 2019, 5: 1749–1768
- 3 Xia C, Dong X, Li H, et al. Cancer statistics in China and United States, 2022: profiles, trends, and determinants. *Chin Med J*, 2022, 135: 584–590
 - 4 Huang L, He J. Trend analysis of hematological tumors in adolescents and young adults from 1990 to 2019 and predictive trends from 2020 to 2044: a Global Burden of Disease study. *Cancer Med*, 2024, 13: e70224
 - 5 Miotto R, Wang F, Wang S, et al. Deep learning for healthcare: review, opportunities and challenges. *Brief Bioinf*, 2018, 19: 1236–1246
 - 6 Radakovich N, Nagy M, Nazha A. Artificial intelligence in hematology: current challenges and opportunities. *Curr Hematol Malig Rep*, 2020, 15: 203–210
 - 7 Radakovich N, Nagy M, Nazha A. Machine learning in haematological malignancies. *Lancet Haematol*, 2020, 7: e541–e550
 - 8 Deo R C. Machine learning in medicine. *Circulation*, 2015, 132: 1920–1930
 - 9 Yang J F, Qiao P R, Li Y, et al. A review of Machine-learning classification and algorithms (in Chinese). *Stat Decis*, 2019, 35: 36–40 [杨剑锋, 乔佩蕊, 李永梅, 等. 机器学习分类问题及算法研究综述. *统计与决策*, 2019, 35: 36–40]
 - 10 An Z Y, Zhang X H. Embracing the age of artificial intelligence: paradigm shifts, opportunities, and challenges in the treatment of acute graft-versus-host disease after allogeneic hematopoietic stem cell transplantation. *Sci China Life Sci*, 2024, 67: 1302–1304
 - 11 Beam A L, Kohane I S. Big data and machine learning in health care. *JAMA*, 2018, 319: 1317–1318
 - 12 Shouval R, Fein J A, Savani B, et al. Machine learning and artificial intelligence in haematology. *Br J Haematol*, 2021, 192: 239–250
 - 13 Yuan M, Bao P, Yuan J, et al. Large language models illuminate a progressive pathway to artificial intelligent healthcare assistant. *Med Plus*, 2024, 1: 100030
 - 14 Kumari A, Kumari A, Singh A, et al. Large language models in hematology case solving: a comparative study of ChatGPT-3.5, google bard, and microsoft bing. *Cureus*, 2023, doi: 10.7759/cureus.43861
 - 15 Westin G F, Dias A L, Go R S. Exploring big data in hematological malignancies: challenges and opportunities. *Curr Hematol Malig Rep*, 2016, 11: 271–279
 - 16 Medeiros B C, Satram-Hoang S, Hurst D, et al. Big data analysis of treatment patterns and outcomes among elderly acute myeloid leukemia patients in the United States. *Ann Hematol*, 2015, 94: 1127–1138
 - 17 Warnat-Herresthal S, Perrakis K, Taschler B, et al. Scalable prediction of acute myeloid leukemia using high-dimensional machine learning and blood transcriptomics. *iScience*, 2019, 23: 100780
 - 18 Kawakami R, Nakada Y, Hashimoto Y, et al. Prevalence and prognostic significance of pulmonary function test abnormalities in hospitalized patients with acute decompensated heart failure with preserved and reduced ejection fraction. *Circ J*, 2021, 85: 1426–1434
 - 19 Liu Q. Translating “AI for omics” into precision therapy. *Med Plus*, 2024, 1: 100042
 - 20 Pan L, Liu G, Lin F, et al. Machine learning applications for prediction of relapse in childhood acute lymphoblastic leukemia. *Sci Rep*, 2017, 7: 7402
 - 21 Hauser R G, Esserman D, Beste L A, et al. A machine learning model to successfully predict future diagnosis of chronic myelogenous leukemia with retrospective electronic health records data. *Am J Clin Pathol*, 2021, 156: 1142–1148
 - 22 Fuse K, Uemura S, Tamura S, et al. Patient-based prediction algorithm of relapse after allo-HSCT for acute Leukemia and its usefulness in the decision-making process using a machine learning approach. *Cancer Med*, 2019, 8: 5058–5067
 - 23 Zhang Z, Arabyarmohammadi S, Leo P, et al. Automatic myeloblast segmentation in acute myeloid leukemia images based on adversarial feature learning. *Comput Methods Programs Biomed*, 2024, 243: 107852
 - 24 Goswami C, Poonia S, Kumar L, et al. Staging system to predict the risk of relapse in multiple myeloma patients undergoing autologous stem cell transplantation. *Front Oncol*, 2019, 9: 633
 - 25 Radhachandran A, Garikipati A, Iqbal Z, et al. A machine learning approach to predicting risk of myelodysplastic syndrome. *Leukemia Res*, 2021, 109: 106639
 - 26 Arai Y, Kondo T, Fuse K, et al. Using a machine learning algorithm to predict acute graft-versus-host disease following allogeneic transplantation. *Blood Adv*, 2019, 3: 3626–3634
 - 27 Deng R X, Zhu X L, Zhang A B, et al. Machine learning algorithm as a prognostic tool for venous thromboembolism in allogeneic transplant patients. *Transplant Cell Ther*, 2023, 29: 57.e1–57.e10
 - 28 Rowley S D, Gunning T S, Pelliccia M, et al. Using targeted transcriptome and machine learning of pre- and post-transplant bone marrow samples to predict acute graft-versus-host disease and overall survival after allogeneic stem cell transplantation. *Cancers*, 2024, 16: 1357
 - 29 Fan S, Hong H Y, Dong X Y, et al. Machine learning algorithm as a prognostic tool for Epstein-Barr virus reactivation after haploidentical hematopoietic stem cell transplantation. *Blood Sci*, 2022, 5: 51–59
 - 30 Musiał K, Stojanowski J, Miśkiewicz-Bujna J, et al. KIM-1, IL-18, and NGAL, in the machine learning prediction of kidney injury among children undergoing hematopoietic stem cell transplantation—a pilot study. *Int J Mol Sci*, 2023, 24: 15791
 - 31 Jardim L L, Schieber T A, Santana M P, et al. Prediction of inhibitor development in previously untreated and minimally treated children with

- severe and moderately severe hemophilia A using a machine-learning network. *J Thrombosis Haemostasis*, 2024, 22: 2426–2437
- 32 Saputra D C E, Sunat K, Ratnaningsih T. A new artificial intelligence approach using extreme learning machine as the potentially effective model to predict and analyze the diagnosis of anemia. *Healthcare*, 2023, 11: 697
- 33 Garduno-Rapp N E, Ng Y S, Weon J L, et al. Early identification of patients at risk for iron-deficiency anemia using deep learning techniques. *Am J Clin Pathol*, 2024, 162: 243–251
- 34 Schipper A, Rutten M, van Gammeren A, et al. Machine learning-based prediction of hemoglobinopathies using complete blood count data. *Clin Chem*, 2024, 70: 1064–1075
- 35 Alcazer V, Le Meur G, Roccon M, et al. Evaluation of a machine-learning model based on laboratory parameters for the prediction of acute leukaemia subtypes: a multicentre model development and validation study in France. *Lancet Digital Health*, 2024, 6: e323–e333
- 36 Haider R Z, Ujjan I U, Khan N A, et al. Beyond the in-practice cbc: the research CBC parameters-driven machine learning predictive modeling for early differentiation among leukemias. *Diagnostics*, 2022, 12: 138
- 37 El Alaoui Y, Padmanabhan R, Elomri A, et al. An artificial intelligence-based diagnostic system for acute lymphoblastic leukemia detection. *Stud Health Technol Inform*, 2023, 305: 265–268
- 38 Mohammed E A, Mohamed M M A, Naugler C, et al. Toward leveraging big value from data: chronic lymphocytic leukemia cell classification. *Netw Model Anal Health Inform Bioinforma*, 2017, 6: 6
- 39 Steinbuss G, Kriegsmann M, Zgorzelski C, et al. Deep learning for the classification of non-hodgkin lymphoma on histopathological images. *Cancers*, 2021, 13: 2419
- 40 Duetz C, Bachas C, Westers T M, et al. Computational analysis of flow cytometry data in hematological malignancies: Future clinical practice? *Curr Opin Oncol*, 2020, 32: 162–169
- 41 Gutierrez-Rodriguez F, Munger E, Ma X, et al. Differential diagnosis of bone marrow failure syndromes guided by machine learning. *Blood*, 2022, doi: 10.1182/blood.2022017518
- 42 Ramzan M, Sheng J, Saeed M U, et al. Revolutionizing anemia detection: integrative machine learning models and advanced attention mechanisms. *Vis Comput Ind Biomed Art*, 2024, 7: 18
- 43 Sharifi H, Lai Y K, Guo H, et al. Machine learning algorithms to differentiate among pulmonary complications after hematopoietic cell transplant. *Chest*, 2020, 158: 1090–1103
- 44 Shao W, Wang Y, Liu L, et al. Combining serum microRNAs and machine learning algorithms for diagnosing infectious fever after HSCT. *Ann Hematol*, 2024, 103: 2089–2102
- 45 Chen D, Goyal G, Go R S, et al. Improved interpretability of machine learning model using unsupervised clustering: predicting time to first treatment in chronic lymphocytic leukemia. *JCO Clin Cancer Inf*, 2019, 3: 1–11
- 46 Agius R, Brieghel C, Andersen M A, et al. Machine learning can identify newly diagnosed patients with CLL at high risk of infection. *Nat Commun*, 2020, 11: 363
- 47 Qin Y, Pu X, Hu D, et al. Machine learning-based biomarker screening for acute myeloid leukemia prognosis and therapy from diverse cell-death patterns. *Sci Rep*, 2024, 14: 17874
- 48 Gurumurthy G, Gurumurthy J, Gurumurthy S. Machine learning in paediatric haematological malignancies: A systematic review of prognosis, toxicity and treatment response models. *Pediatr Res*, 2024, doi: 10.1038/s41390-024-03494-9
- 49 Farswan A, Gupta A, Gupta R, et al. AI-supported modified risk staging for multiple myeloma cancer useful in real-world scenario. *Transl Oncol*, 2021, 14: 101157
- 50 Yuan J, Zhang Y, Wang X. Application of machine learning in the management of lymphoma: current practice and future prospects. *Dig Health*, 2024, 10: 20552076241247963
- 51 An Z Y, Wu Y J, Hou Y, et al. A life-threatening bleeding prediction model for immune thrombocytopenia based on personalized machine learning: a nationwide prospective cohort study. *Sci Bull*, 2023, 68: 2106–2114
- 52 Rawal A, Kidchob C, Ou J, et al. Application of machine learning approaches for predicting hemophilia A severity. *J Thrombosis Haemostasis*, 2024, 22: 1909–1918
- 53 Tislevoll B S, Hellesøy M, Fagerholt O H E, et al. Early response evaluation by single cell signaling profiling in acute myeloid leukemia. *Nat Commun*, 2023, 14: 115
- 54 Lee S I, Celik S, Logsdon B A, et al. A machine learning approach to integrate big data for precision medicine in acute myeloid leukemia. *Nat Commun*, 2018, 9: 42
- 55 Xu M, Liu J, Huang L, et al. A novel scoring model for predicting efficacy and guiding individualised treatment in immune thrombocytopaenia. *Br J Haematol*, 2024, 205: 1108–1120
- 56 Rodríguez-Belenguer P, Piñana J L, Sánchez-Montañés M, et al. A machine learning approach to identify groups of patients with hematological malignant disorders. *Comput Methods Programs Biomed*, 2024, 246: 108011

-
- 57 Challen R, Denny J, Pitt M, et al. Artificial intelligence, bias and clinical safety. [BMJ Qual Saf](#), 2019, 28: 231–237
 - 58 Walter W, Pohlkamp C, Meggendorfer M, et al. Artificial intelligence in hematological diagnostics: Game changer or gadget? [Blood Rev](#), 2023, 58: 101019
 - 59 Chan K Y Y, Zhang C, Wong Y T S, et al. R4 RGS proteins suppress engraftment of human hematopoietic stem/progenitor cells by modulating SDF-1/CXCR4 signaling. [Blood Adv](#), 2021, 5: 4380–4392
 - 60 Yang G, Ye Q, Xia J. Unbox the black-box for the medical explainable AI via multi-modal and multi-centre data fusion: a mini-review, two showcases and beyond. [Inf Fusion](#), 2022, 77: 29–52

Summary for “人工智能在血液疾病诊疗中的应用研究进展”

Research progress of artificial intelligence in the clinical diagnosis and treatment of hematological diseases

Liqian Zhang¹, Zhuoyu An¹, Lijuan Cui², Wenqian Li³ & Xiaohui Zhang^{1*}

¹ Peking University People's Hospital, Peking University Institute of Hematology, National Clinical Research Center for Hematologic Disease, Beijing Key Laboratory of Cell and Gene Therapy for Hematologic Malignancies, Beijing 100044, China

² Department of Hematology, General Hospital of Ningxia Medical University, Yinchuan 750003, China

³ Department of Hematology and Rheumatology, Qinghai Provincial People's Hospital, Xining 810007, China

* Corresponding author, E-mail: zhangxh@bjmu.edu.cn

Hematological diseases, characterized by bleeding, fever, and abnormal blood components, are disorders originating from or affecting the blood and the hematopoietic system. Common examples of hematological diseases include benign diseases such as anemia and malignant diseases such as leukemia, lymphoma, and multiple myeloma. The prognosis of these disorders varies and poses a significant disease burden. Improving the diagnosis and treatment of hematological diseases is a key focus in clinical medicine. Artificial intelligence (AI) is a branch of computer science that aims to simulate, study, and extend theories, methods, technologies, and applications of human intelligence. With the rapid development of relevant technologies, AI is increasingly utilized in clinical medicine. In the clinical research field, AI technology primarily involves machine learning (ML), where computers independently extract core information from data through algorithms to achieve established goals. ML encompasses supervised learning, semi-supervised learning, unsupervised learning, and reinforcement learning, all of which are implemented through various algorithms, including random forests, decision trees, support vector machines, neural networks, and Progress. The process of ML in clinical practice includes data collection, preprocessing, algorithm selection, model building, model training, evaluation, optimization, and clinical application. In current research, AI and ML have been widely utilized for prediction, diagnosis, prognosis evaluation, and treatment guidance across a variety of blood diseases, demonstrating exceptional effectiveness in tasks, such as image recognition, omics data mining, and others. For instance, some relapse prediction models for leukemia have been successfully established through different ML algorithms. Researchers have efficiently predicted other benign and malignant hematologic disorders by using AI methods. In the realm of blood disease diagnosis, accurate diagnoses of leukemia have been achieved through the use of AI algorithms with various clinical data, including laboratory test results, blood smears, and genetic information. Additionally, ML prediction models for bone marrow failure, anemia, and other blood disorders have demonstrated excellent performance. Clinical studies focusing on various leukemia subtypes have also shown the value of ML models based on random forests, decision trees, and additional algorithms in guiding treatment decisions. Furthermore, Farswan et al., Zhang et al., and other teams have applied AI to accurately assess the prognosis of other hematological disorders such as multiple myeloma and primary immune thrombocytopenia. AI tools have also demonstrated impressive performance in predicting treatment responses in patients with various blood diseases. Despite challenges such as limitations in original data, ethical debates, and barriers to human-computer interaction, the aforementioned research results have sufficiently demonstrated AI's potential for development. It is anticipated that AI could be applied in clinical practice, leading to significant improvements in patient prognosis while optimizing the current clinical diagnosis and treatment paradigm for blood diseases.

blood diseases, artificial intelligence, machine learning, clinical diagnosis and treatment

doi: [10.1360/TB-2024-1190](https://doi.org/10.1360/TB-2024-1190)