附件4

血液病流式细胞学人工智能分析软件

性能评价审评要点

本要点旨在指导注册申请人对血液病流式细胞学人工智能分析软件注册申报资料中非临床评价部分的准备及撰写，同时也为技术审评部门提供参考。  
 本要点是对血液病流式细胞学人工智能分析软件申报资料的一般要求，申请人需依据产品的具体特性确定其中内容是否适用。若不适用，需具体阐述理由及相应的科学依据，并依据产品的具体特性对注册申报资料的内容进行充实和细化。  
 本要点是供注册申请人和技术审评人员使用的指导性文件，但不包括审评审批所涉及的行政事项，亦不作为法规强制执行，需在遵循相关法规的前提下使用本指导原则。如果有能够满足相关法规要求的其他方法，也可以采用，但是需要提供详细的研究资料和验证资料。  
 本要点是在现行法规和标准体系以及当前认知水平下制定的，随着法规和标准的不断完善，以及科学技术的不断发展，本要点的相关内容也将进行适时的调整。

一、适用范围

血液病流式细胞学分析，与细胞形态学、细胞遗传学、分子遗传学等一起，用于血液病的辅助诊断、鉴别诊断、预后判断、治疗监测、免疫监测、造血干细胞移植相关检测等。样本类型可为血液或骨髓、淋巴结、脑脊液、及其他适用样本类型。

血液病流式细胞学人工智能分析软件，指通过人工智能算法帮助医生分析多参数流式细胞学技术/流式细胞仪产生的结构化数据，如数据文件或图形文件等进行血液病辅助体外诊断的软件，通常用于定性或定量分析，可以是软件组件或独立软件。通过对结构化数据的人工智能算法分析，用途包含：（一）基础分析：通过单一的人工智能算法辅助流式细胞学数据的可视化，从而辅助流式诊断人员进行自动设门和分析。（二）高阶分析：通过多个人工智能算法组合使用（如集成学习算法和卷积神经网络算法组合使用），实现自动设门辅助血液疾病诊断（良/恶性、白血病亚型分类、淋巴瘤分型等）、化疗后的微小残留病监测等。

产品在医疗机构、医学实验室使用，其结果供执业医师参考。软件作为血液病流式细胞学辅助诊断工具，其提示的诊断结果不能作为临床诊断决策的唯一依据。

产品管理类别：

独立软件分类编码：21-04-02（计算机辅助诊断/分析软件）

软件组件分类编码：22-01-08（流式细胞分析仪器）

应注意本要点主要适用于辅助决策类的人工智能医疗器械，按照第三类管理，其他类别的产品可根据其适用性参照执行。关于人工智能医用软件分类的判定，可参照《人工智能医用软件产品分类界定指导原则》。

产品名称：建议依据医用软件通用名称要求，按“特征词1（如有）+特征词2（如有）+特征词3（如有）+核心词”结构编制。具体到该类产品，建议体现处理对象和临床用途，如血液病（具体病种）+流式细胞学辅助分析/检测软件，可参考《医疗器械分类目录》、《医用软件通用名称命名指导原则》中的品名举例或已批准的同类产品进行命名。如为软件组件，则产品名称体现医疗器械名称，如“多参数流式细胞仪”，但在适用范围体现人工智能算法类型，在产品说明书体现使用限制等相关内容。

本要点基于《人工智能医疗器械注册审查指导原则》的基本框架要求，其中关于人工智能医疗器械的的术语定义、相关要求参照上述指导原则。

二、产品简介

该类软件的分析对象基于流式细胞仪产生的标准化的结构化数据文件或图形文件等。

产品的工作原理：产品为独立软件或软件组件，采用人工智能算法对流式细胞仪数据进行处理（如：数据特征提取、数据通过函数图形化以及自动设门、决策分析等），实现对流式细胞仪数据、对细胞分类及定量结果进行分析，达到辅助诊断的功能。

主要结构及组成：结构组成明确交付内容和功能模块，其中交付内容包括软件安装程序、授权文件、外部软件环境安装程序等软件程序文件，功能模块包括客户端、服务器端（若适用）等，若适用注明选装、模块版本。

服务器端：负责接收并处理请求，对请求进行数据处理，调用人工智能算法（如PCA、t-SNE、SVM等）对数据进行处理并输出分析结果。把用户请求的数据结果返回给浏览器。

客户端：将服务器端分析结果呈现到用户界面。用户 可在用户界面端进行自动设门、细胞分群、辅助分类、辅助诊断、报告管理、检索查询、用户管理、查看日志等功能操作。

预期使用环境：申报产品主要使用环境为医疗机构、医学实验室。

三、基本要求

在满足相关法规、规章、指导原则、标准的前提下，还应重点关注下列内容：

（一）综述资料

详细描述该产品的算法名称，选择该算法的依据和权威文献，数据处理的在流式细胞分析流程中的节点及辅助决策功能及功能实现方法，描述配套使用流式细胞仪器设备（厂家、型号、注册证号及主要参数设置：包括激光数、通道设置等）及仪器质量控制参数（仪器校准、电压补偿等）、配套使用的流式细胞检测试剂（名称、厂家、注册证号等），描述数据产生的样本类型。

人工智能医疗器械从成熟度角度可分为成熟和全新两种类型，其中成熟是指安全有效性已在医疗实践中得到充分证实的情形，全新是指未上市或安全有效性尚未在医疗实践中得到充分证实的情形。人工智能医疗器械的算法、功能、用途若有一项为全新则属于全新类型，反之属于成熟类型。

在资料中应当说明采用算法的成熟程度，采用全新算法的，应当在综述资料中详细说明。

阐述软件研发背景，包括对辅助决策功能的介绍，目前该血液病临床诊断参考标准、诊疗指南和分型依据。申报产品数据处理标准和分型依据应与临床标准一致。

（二）非临床资料

1.产品技术要求及检验报告

性能指标应包括“所分析的数据类型”、“分析速度”、“临床功能”等，如数据处理、数据分析、诊断提示等。

其他要求请参照《医疗器械软件注册审查指导原则（2022年修订版）》。

2.软件研究资料

按照《医疗器械软件注册审查指导原则（2022年修订版）》、《医疗器械网络安全注册审查指导原则（2022年修订版）》、《人工智能医疗器械注册审查指导原则》的要求，提交研究资料。明确发布版本号。提交的研究资料主要包含软件研究资料、算法研究资料、网络安全研究资料。研究资料内容应涵盖申报产品适用范围中描述的全部内容。

2.1风险管理

软件安全性级别越高，其生存周期质控要求越严格，注册申报资料越详尽，同时由于全新类型的潜在未知风险多于成熟类型，故需结合成熟度予以综合考虑。

人工智能医疗器械的软件安全性级别可基于产品的预期用途、使用场景、核心功能进行综合判定，其中预期用途主要考虑用途类型、重要程度、紧迫程度等因素，使用场景主要考虑使用场合、疾病特征、适用人群、目标用户等因素，核心功能主要考虑功能类型、核心算法、输入输出、接口等因素。

参考YY/T 0316、GB/T 42062和YY/T 0664等标准进行风险管理。风险管理活动应当基于软件的预期用途、使用场景、核心功能予以实施，并贯穿软件全生命周期过程。申请人重点考虑如下风险因素：此类产品算法风险包括算法选择风险、算法训练及验证的风险等，主要包括算法选择依据不充分，算法训练使用的数据质量不高，数据代表性不足或数据量不够，由于样本前处理、流式细胞仪光路信号校准和质控，流式细胞分析过程或流式荧光试剂标记发生错误等原因产生，造成过拟合和欠拟合导致的泛化能力不足，数据扩增（如适用）和数据偏移等造成的假阴性和假阳性错误等；使用中的风险主要包括未经培训的人员使用，以及样本采集、处理、仪器校准等错误，造成的假阴性和假阳性错误。其中假阴性即漏诊，可能导致后续诊疗活动延误，特别是要考虑快速进展疾病的诊疗活动延误风险，而假阳性即误诊，可能导致后续不必要的诊疗活动。

以算法特性为核心重点关注其泛化能力，以模型/数据为基础重点关注其质控情况，同时从风险管理角度兼顾算力不足与失效的影响。

注册申请人应结合人工智能医疗器械的预期用途、使用场景、核心功能开展风险管理活动，采取风险控制措施将风险降至可接受水平，并贯穿于人工智能医疗器械全生命周期过程。

此外，进口人工智能医疗器械还需考虑中外差异风险，如人种、流行病学特征、临床诊疗规范等差异。

2.2需求规范

提供软件需求规范文档，明确软件的功能、性能、接口、用户界面、运行环境、网络安全、数据采集等需求。

数据采集需考虑数据来源的合规性、充分性和多样性，数据分布的科学性和合理性，数据质控的充分性、有效性和准确性。数据应来源于分布于不同地域的机构且不少于3家，确保数据应当具备高度的多样性。机构应使用软件说明书中的分析流程采集数据，如明确的样本来源、仪器设备、样本处理（前处理、单细胞悬液的制作、荧光抗体试剂的选择等）。数据分布的科学性和合理性，应考虑包括但不限于疾病构成（如分型、分级、分期）、人群分布（如高危人群、患者，性别、年龄）等情况。

算法性能需结合产品预期用途，综合考虑分析速度、敏感性、特异性、重复性与再现性、泛化性等性能指标的适用性及其要求。

同时还需考虑因梯度消失（如适用）、梯度爆炸（如适用）、过拟合和欠拟合等影响算法性能的因素。

使用限制需考虑产品禁用、慎用等场景，准确表述产品使用场景，提供必要警示提示信息。

2.3软件性能研究

提供产品性能研究资料以及产品技术要求的研究和编制说明，给出软件相关的功能性、安全性指标的确定依据。注册申请人需在软件研究资料中提交GB/T 25000.51自测报告，亦可提交自检报告或检验报告代替自测报告。产品技术要求中各指标依据产品自身特点确定。

2.4算法研究资料

根据《人工智能医疗器械注册审查指导原则》提交算法研究资料。逐项提交每个人工智能算法或算法组合的算法研究报告。

2.4.1算法基本信息

明确算法的名称、类型、结构、输入输出、流程图、算法框架、运行环境等基本信息以及算法选用依据。

其中，算法类型从学习策略、学习方法（基于模型的算法和基于数据的算法）、可解释性（白盒算法和黑盒算法）等角度明确算法特性。算法结构部分应明确算法的层数、参数规模等超参数信息。流程图应包含输入示例、学习方法示意图输出示例。算法框架应明确所用人工智能算法框架的基本信息，包括名称、类型（自研算法框架、现成算法框架）、型号规格、完整版本、制造商等信息。若基于云计算平台，应明确云计算的名称、服务模式、部署模式、配置以及云服务商的名称、住所、服务资质。运行环境应明确算法正常运行所需的典型运行环境，包括硬件配置、外部软件环境、网络条件；若使用人工智能芯片应明确其名称、型号规格、制造商、性能指标等信息。算法选用依据详述算法或算法组合选用的理由和基本原则。

2.4.2算法风险管理

结合预期用途、使用场景、和核心功能、算法成熟度、网络安全等因素，此类产品的软件安全性级别为严重级别。提供算法风险管理资料，若无单独文档可提供软件风险管理资料，并注明算法风险管理所在位置。

2.4.3算法需求规范

提供算法需求规范文档，若无单独文档可提供软件需求规范，并注明算法需求所在位置。

2.4.4数据收集

2.4.4.1提供数据来源合规性声明，列明数据来源机构的名称、所在地域、数据收集量、伦理批件（或科研合作协议）编号等信息。

2.4.4.2数据采集

建议参考2.2需求规范的要求，提供数据采集操作规范文档，包括数据采集方案和数据采集标准操作规程。

数据采集主要由临床机构实施，应明确流式细胞分析流程的全部信息（包括全流程质控标准、panel设计、仪器和试剂的标准化SOP等）、明确样本及样本处理要求、样本的来源与分布、样本质量要求（样本保存时限、质量判断标准）、采集设备要求（应包含流式细胞仪型号、通道数量、注册证号、设备质控及补偿调节方式、数据的基本处理方式等）、流式抗体试剂要求（厂家、注册证号等）、数据质量要求、采集过程、数据脱敏、数据转移等要求。采集过程应对样本数据进行编号并加密，方案中应包含编号规则。

关于数据集采集方面建议明确偏倚的控制方法，如：  
 （1）为了保证研究对象的代表性，可从目标人群中随机抽样、多中心，保证样本量要足够大。  
 （2）为了保证研究对象选择的质量，研究设计中应有明确、具体的诊断标准、纳入标准和排除标准。  
 （3）检测标本尽量保持试验条件的前后一致性，注意试剂质量、仪器性能、样本保存和操作规定。  
 建议通过限制、配比、标准化、随机化、分层分析和多元分析等进行控制。

2.4.4.3数据整理

明确数据清洗/预处理程序，对数据处理中应用的软件进行简述，并以附件的形式提交数据处理中各软件的软件研究资料。

数据整理基于原始数据库考虑数据清洗、数据预处理的质控要求。数据清洗需明确清洗的规则、方法、结果，数据预处理需明确处理的方法、结果。数据整理所用软件工具均需明确名称、型号规格、完整版本、制造商、运行环境，并进行软件确认。

2.4.4.4数据标注

明确标注人员和仲裁人员的资质要求和培训内容，标注人员和仲裁人员应为相关专业资质的人员，数据应经过2人或以上进行标注，标注方式建议给出依据。简述标记系统信息，给出数据/图像界面截图，详细介绍标注过程、标注对象和标注标准（注意与临床诊断标准的一致性）等。明确标注过程质控、标注质量评估、数据安全保证等要求，明确室内质控、室间质评等要求。应当注意标注数据的质量评估，可抽选一定比例数据由有资质的非标注人员结合临床进行综合诊断评估。

提供原始数据库、基础数据库、标注数据库、扩增数据库（如有），关于疾病构成的数据分布情况，包括适用人群、数据来源机构、采集设备、样本类型等因素。

若数据来自公开数据库，提供公开数据库的基本信息（如名称、创建者、数据总量等）和使用情况（如数据使用量、数据质量评估、数据分布等）。

2.4.4.5数据集构建

明确各数据集划分的方法及依据。训练集应当保证样本分布具有均衡性，根据预期用途，训练样本应涵盖不同血液病样本类型、不同分群类型等。调优集应保证样本分布符合临床实际情况，如阳性比例，不同类型比例、不同临床分型比例等均应符合临床实际情况。训练集、调优集、测试集的样本应两两无交集并通过查重予以验证。

如适用，进行数据扩增验证时，应当明确扩增的方式、方法、倍数，并考虑数据扩增对软件的影响及风险。列表比较扩增数据库与标注数据库的差异，论证扩增数据库样本量的充分性以及分布的合理性。

2.4.5算法训练

依据适用人群、数据来源机构、采集设备、样本类型等因素，提供训练集、调优集（若有）关于疾病构成的数据分布情况。

算法训练基于训练集、调优集进行训练和调优，应明确算法训练所用的评估指标、训练方式、训练目标、调优方法，提供ROC曲线或混淆矩阵等证据（如：迭代次数-训练CountIOU曲线和迭代册数-召回率曲线）证明训练目标满足医疗要求，提供训练数据量-评估指标曲线（如迭代次数-Loss曲线）等证据以证实算法训练的充分性和有效性。

应当提供人工智能学习算法常用的评估函数来评估算法训练的质量。

2.4.6算法性能评估

基于测试集对算法设计进行评估，确认软件算法性能的效率、敏感性、特异性，性能应满足算法设计要求。

算法验证：算法性能评估应包括软件对样本满意度评价的能力测试、泛化能力的测试、压力测试（指采用罕见或特殊的真实数据样本开展的算法性能测试）、对抗测试、重复性与再现性测试、诊断敏感性与特异性测试、分析效率测试、算法性能影响因素分析、性能评估结果比较分析、偏差报告等研究。应以该软件和临床综合诊断的一致性进行比较.评价细胞分群的准确性。

通过样本量估算确定测试集中阴、阳性样本和压力样本的样本量，提供测试集关于疾病构成（包括年龄和血液病类型、白细胞分化抗原种类）、数据来源的数据分布情况。明确对抗样本的选择原则；明确性能评估的可接受标准和标准制定依据。提交测试报告和结果图示例。

若使用第三方数据库开展算法性能评估，提供第三方数据库的基本信息（如名称、创建者、数据总量等）和使用情况（如测试数据样本量、评估指标、评估结果等）。还应提交第三方数据库数据样本来源和本分析软件在产品技术要求、说明书中要求的流式细胞分析流程的一致性评价（包含样本要求、处理方式、流式细胞荧光抗体试剂的选择、流式细胞仪、荧光及电压补偿方式、应用的函数等）,应提交符合要求的流式细胞仪数据。

算法的确认：此类产品均需按照GCP的要求开展临床试验。临床试验的机构应具备该软件要求的流式细胞分析流程所需的人员、试剂及仪器设备，临床试验应以该软件和临床综合诊断的一致性进行比较，确保其产生的流式细胞仪数据与该软件说明书和技术要求的一致性。

2.4.7算法可追溯性分析

提供算法可追溯性分析报告，即追溯算法需求、算法设计、源代码（明确软件单元名称即可）、算法测试、算法风险管理的关系表。若无单独文档可提供软件可追溯性分析报告，需注明算法可追溯性分析所在位置。

（三）产品说明书和标签样稿

明确该产品适用范围，明确流式细胞学人工智能分析的全流程质控标准、适用的流式细胞分析设备、设备参数设置、适用的试剂及试剂盘设计、样本处理方法和本软件适用的标准化SOP等。明确软件报告内容。对产品带来的假阳/假阴性风险进行提示。

根据算法性能综合评价结果，对产品的适用范围、使用场景、核心功能进行必要限制，并在说明书中明确产品使用限制和必要警示提示信息。明确数据采集设备和数据采集过程相关要求。

若产品采用人工智能黑盒算法，根据算法影响因素分析报告，在说明书明确产品使用期限、使用限制和必要的警示提示信息。

明确人工智能算法的算法性能评估总结（测试集基本信息、评估指标与结果）、临床评价总结（临床数据基本信息、评价指标与结果）、决策指标定义（或提供决策指标定义所依据的临床指南、专家共识等参考文献）等信息。若采用基于数据的人工智能算法，说明书还应补充算法训练总结信息（训练集基本信息、训练指标与结果）列明算法训练总结和算法性能评估总结以及临床评价总结。

对于软件安全性级别为严重级别的产品，需提供用户培训材料。