参赛学生姓名：李玥呈

中学：华东师范大学第二附属中学闵行紫竹分校

省份：上海

国家/地区：中国

指导老师姓名：严一滨、黄潇潇

指导老师单位：华东师范大学第二附属中学闵行紫竹分校

论文题目：跨域血细胞异常检测与淋巴细胞误报率下限研究

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

跨域血细胞异常检测与淋巴细胞误报率下限研究

作者：李玥呈（华东师范大学第二附属中学闵行紫竹分校）

指导老师：严一滨、黄潇潇

摘要

儿童白血病早期检出是改善预后的关键，血涂片镜检作为筛查首要环节，在基层医疗的室间比对合格率却不足一半；即使结合现有的深度学习方案，也因跨域泛化能力弱、数据集依赖强、算力成本高等原因难以在受限条件下直接适用。

本研究用监督对比学习在四个公开数据集上训练“正常聚簇、异常推离”得到特征空间，权重随后冻结；目标医院无需异常标注，仅以 150 张正常图像（每类 30 张）在普通 CPU 上约 8 秒快速完成特征库构建，按马氏距离评分、超阈报警。为克服不同随机种子导致的结果漂移问题，本研究设计了多级聚合交付方案：先在种子内取 10 个过程性模型截面排名归一化分数中位数，再在五个种子间取中位数，随后以基于建库样本自身离散度的 Bank 方差标准化校准，最后由独立轻量大模型二级复核。

在三个未参与训练的跨国多中心数据集上，通过上述方案，在保证检出约 94% 真异常的前提下，最终误报率在 KU-Optofil 与 AML-Cytomorphology 两数据集上分别为 0.180 与 0.193，在 Chula-WBC-15 全谱系评估口径上为 0.173，整套推理流程在普通硬件上单片仅需分钟级。相比之下，38 万张图像预训练的通用血液学大模型在零适配条件下误报率为 0.48；本方案将其定位为轻量级的二级复核过滤器。

在机制层面，本研究揭示了形态学筛查的生物学瓶颈：12 组多维度干预实验、完全独立的特征空间复现与反向敲除应力测试一致证实：淋巴系误报约 0.30 的下界源于反应性与异型淋巴细胞形态连续谱的内在生物形态学特征，而非模型算法缺陷。因此，异常报警呈现细胞系分化规律：粒细胞报警可信度最高，淋巴细胞最易混淆。据此，本研究提出“粒系报警优先复核、淋巴系报警置后但也不放过、终检由上级医院兜底”的分级复核部署方案。

关键词：血细胞异常检测；淋巴细胞形态连续谱；监督对比学习；分布外检测；跨医疗机构泛化

2026 S.T. Yau High School Science Award
仅用于2026丘成桐中学科学奖公示

目录

一、引言 4
二、基于正常细胞特征空间构建与异常评分的检测方法 7
（一）数据构成与负例池 7
（二）特征空间学习 8
（三）特征库构建 9
（四）轻量评分与报警机制 10
三、跨国多中心数据集检出性能实验与分析 10
（一）评估指标与临床工作量基准 11
（二）测试环境一：KU-Optofil（开发域） 11
（三）测试环境二：AML-Cytomorphology（外部验证域） 13
（四）测试环境三：Chula-WBC-15（跨国盲测终验域） 13
（五）与代表性技术路线的对照 13
四、训练随机性抑制与多级聚合交付方案 14
（一）随机种子扰动与单次训练数字的随机成分 14
（二）多级聚合交付方案 15
（三）第三陌生域独立终验与方案汇总 16
五、淋巴系误报经验下界分析与分级复核部署 17
（一）误报按细胞系分化规律 17
（二）12 组干预实验与独立特征空间验证 17
（三）AML 数据集的反常现象与应力测试 18
（四）分级复核部署建议 18
六、讨论 19
（一）误报按细胞系分化的细胞病理学机理 19
（二）与现有技术路线的对比与范式转变 20
（三）研究局限与前瞻展望 20
七、结论 21
参考文献 22

2026 S.T. Yau High School Science Award
仅用于2026年度丘成桐中学科学奖公示

一、引言

白血病是儿童最常见的恶性肿瘤，约占儿童肿瘤的三分之一 [1]；我国每年新发白血病约 8.19 万例（2022 年数据）[1]。儿童急性淋巴细胞白血病（ALL）的五年生存率已达 90%[2]，但前提是及时检出。

血涂片镜检是白血病筛查的首要环节，但基层医疗机构在这一环节上特别薄弱。全国 104 万个基层医疗机构承担着超过一半的诊疗量 [3]，却普遍不具备足够的血涂片形态学检查能力：合肥地区 2016 年的一项调查显示，基层医疗机构血常规室间比对合格率仅 41.9%，而省级医院为 94.1%[4]。《柳叶刀》发表的研究数据更为直接：欠发达地区儿童及青少年癌症五年生存率为 72.6%，发达地区为 84.9%[5]。

深度学习能否缓解上述问题？本研究首先考察现有三条主要技术路径的适用性（其余近年代表性工作见表 1）。第一条路线是**封闭集分类**：训练模型将正常白细胞分为五类，公开数据集上的分类准确率已达很高水平（PBC 数据集的构建者报告其闭集分类模型为 96.2%[6]）。但该方法无法回答基层筛查的核心问题：“样本中是否存在训练中未出现的异常细胞”。闭集分类模型在遇到未见过的异常细胞时，会将其强制归入某一正常类别。第二条路线是**有监督异常检测**：利用目标医院自身的异常细胞进行训练，性能十分优异（Matek 等人 2019 年的工作，AUROC 0.992[7]），但其前提是目标医院提供数万张带标注的单细胞图像，基层机构恰恰难以满足。第三条路线是近年兴起的**生成式建模**：2025 年发表于《自然 • 机器智能》的 CytoDiffusion[8] 利用扩散模型直接学习正常血细胞的完整分布，在异常检出与跨医院鲁棒性上均报告了领先的指标数据；但由于需对全部形态分布建模，其训练与推理均需较大规模 GPU，同样超出基层的算力预算。

上述三条主要路线与近年相关路径的代表工作及其指标口径汇总于表 1。

表 1　血细胞异常检测主要技术路径与代表工作

技术路径	代表工作	关键指标（设定口径）	与基层需求的差距
封闭集分类	Acevedo 等 2019[6]	闭集分类准确率 96.2%（同域）	无法检出训练中未出现的异常
有监督异常检测	Matek 等 2019[7]	AUROC 0.992（单中心域内，数万张异常标注）	依赖目标域大量异常标注
生成式建模	CytoDiffusion 2025[8]	AUC 0.990（同域留出类，全量正常数据重训扩散模型）	训练与推理算力超出基层预算
血液学基础模型	DinoBloom 2024[9]	原文仅报告细胞分类与亚型指标，无异常检出指标	零适配直接用于异常检出能力有限（本研究实测 FPR95 0.48，见第三章）
免训练置信度评分	MSP[10]、Energy[11]	本研究复现：跨域 AUROC 约 0.88、FPR95 0.52–0.58	高检出率下误报失控，不宜作报警器

2026 S. Yau High School Science Award
仅用于2026年成都市学会公开

技术路径	代表工作	关键指标（设定口径）	与基层需求的差距
单类/重建（仅正常训练）	Margret 与 Krishnan 2026[12]	AUC 0.963、工作点误报 3.2%（同域血涂片）	同域工作点口径，无跨医院证据
少样本单类检测	FewSOME 2023[13]	约 30 张正常样本即可训练（通用领域，非血细胞）	无血细胞跨域证据

表 1 注：各工作的指标口径与测试设定互不相同，表中数字不可横向比较。

从基层医疗机构的实际情况出发，本研究将实际应用约束归纳为四方面前提：

1. **极低正常标注要求**：识别正常五类细胞属于基层检验员的基础技能，而识别异常形态属于稀缺能力；基层往往无力提供异常标注，系统必须仅依赖百张级正常细胞即可完成适配（若基层已掌握海量异常标注，则不再属于本系统的适用范畴）；

2. **零重训与微算力**：训练完成后，模型主体权重冻结，无须新的训练。新机构接入时仅需基于本地正常样本做极轻量的统计量更新，避免算力要求高的模型训练；

3. **跨医疗机构的泛化鲁棒性**：不同医院在显微镜光学系统、推片染色试剂及采集批次上普遍存在显著域偏移，系统必须在面对陌生成像条件时保持稳健；

4. **数字切片前提与评价指标的设定**：假设社区卫生服务中心或中心乡镇卫生院配备成像或数字扫描设备，提供分割后的单细胞切图；同时，依据国际血液学复检专家组共识 [14]，白细胞初筛对恶性前体的假阴性率（漏诊率）须控制在 5% 以内，即真异常检出率必须达到 95% 以上，**对应评价指标为保证 95% 检出率下的误报率（FPR95）**。

在此基础上，本研究提炼出**“跨医疗机构目标域、零异常标注、仅百张级正常图像建库、模型权重冻结、以 FPR95 为安全指标”**这一专门面向基层受限条件的血细胞异常检测设定。在现有公开文献中，尚无在同等严格受限设定下开展跨中心评估的研究报告。本研究即在这一设定下开展研究工作。

本研究的核心围绕两个相互支撑的目标展开：

1. **技术方案构建**：在目标域零异常标注、零重训的刚性约束下，探索跨医院异常形态初筛的性能边界，设计出一套能够在普通硬件上部署的可稳定表现的检测系统；

2. **病理机制探索**：系统性探索正常淋巴细胞与反应性/异型淋巴细胞之间的

2026 STI 全国青少年科学实践大赛 仅用于2026年成都市青少年科学实践大赛公示

形态连续谱，阐明其在特征空间中是否构成形态学检测方法难以逾越的误报下界，并由此推导具有临床可行性的分级复核工作流。

实验在三个互不相同的跨国多中心数据集（土耳其 KU-Optofil、德国 AML-Cytomorphology、泰国 Chula-WBC-15）上全面展开。本系统的核心定位是基层**初筛辅助**（报警→人工复核），而非终末诊断终端。研究成果相应包含两部分：一套可落地的轻量级检测系统架构（图 1），以及对淋巴细胞形态连续谱制约机制与经验下界的实证解析（第五章、第六章）。

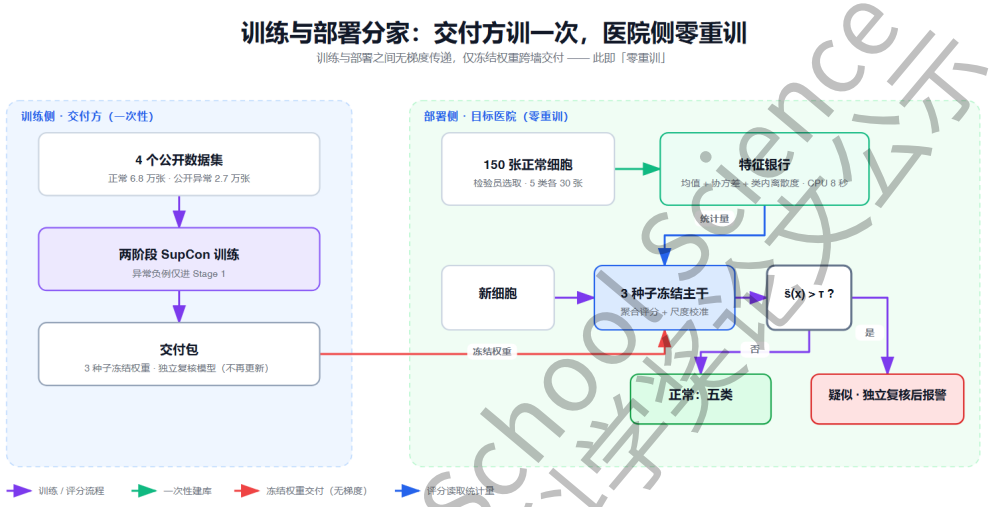


图 1　系统总体架构与两阶段跨域部署工作流。左（模型研发侧）：基于四个公开多源数据集构建通用特征空间，输出后冻结权重；右（基层部署侧）：目标医院检验员仅需选取少量正常细胞在普通 CPU 上快速生成本地特征库，待检样本经多级聚合评分与轻量二级复核后触发分流。

本研究的原创贡献可概括为四点：

1. **基层受限问题的标准化重构：**提炼出四条临床硬约束，设计了“仅学习正常类簇、以统计距离报警”的极轻量跨域异常检测路线（第二章、第三章）；

2. **训练随机性漂移的抑制方案：**针对深层网络种子扰动问题，构建了融合时间中位数、跨种子中位数、Bank 方差标准化与轻量大模型复核的多级聚合交付方案（第四章）；

3. **淋巴系误报经验下界的实证发现：**通过 12 组多维度干预、独立特征空间及反向应力测试，证实约 0.30 的淋巴误报下界源于形态连续谱的生物学本质（第五章）；

4. **分类精度与异常判别能力的解耦实证：**跨国多中心实证表明闭集分类准确率无法作为异常筛查的代理指标，揭示了度量学习在异常报警中的本质优势（第三章）。

2026 Siyuan High School Science Award
仅用于2026年成科学术论文公示

二、基于正常细胞特征空间构建与异常评分的检测方法

为实现适合基层医疗机构的极轻量跨域异常检测，本研究的设计出发点是：不对千变万化的异常细胞做显式分类，而专注于精确建模正常白细胞的低维形态流形。训练侧利用公开的多源数据与异常负例，促使正常五类细胞形成紧凑类簇并拉开类间间隔；目标域新细胞如果与特征库中任一正常类别均不相似，即判定为离群可疑并触发报警。

本章首先阐述特征空间构建与单模型基础评分层；针对随机种子扰动与多系离散度差异的多级聚合交付层将在第四章展开。

（一）数据构成与负例池

本研究的数据体系分为模型研发训练侧与目标域评估侧两部分。训练侧采用四个公开数据集的正常白细胞切图共 67,642 张，按 8:1:1 严格划分训练/验证/测试，包括：PBC 10,283 张 [15]、Raabin 14,514 张 [16]、multi-focus 23,882 张 [17] 以及 MLL23 18,963 张 [18]；公开负例池由 27,392 张异常切图构成（PBC 的幼红细胞与未成熟粒细胞 4,446 张 [15]、multi-focus 的三类异常切图 6,426 张 [17]、MLL23 异常部分的六类 16,520 张 [18]）。在第一阶段训练中，按“负例占注入后样本总量 10%”的比例随机采样注入（异常暴露训练思想源自 [19]）。负例在此处不赋予特定疾病标签，仅提供“偏离正常白细胞形态、必须在度量空间中远离”的几何排斥信号。

四个源域数据集的选取遵循严格的临床与工程标准：

1. **细胞类目口径对齐**：四个训练集的正常白细胞均完整映射至外周血标准五分类（中性粒细胞、淋巴细胞、单核细胞、嗜酸性粒细胞、嗜碱性粒细胞），使合并训练能够覆盖极其罕见的嗜碱性粒细胞的跨设备、跨染色形态变异；

2. **成像物理条件高度异质**：PBC 采自标准化数字形态分析仪 [15]，Raabin 使用手机摄像头连接显微镜目镜拍摄 [16]，multi-focus 覆盖了同一细胞从清晰到离焦的 10 个显微焦面 [17]，MLL23 则出自大型全自动数字玻片扫描系统 [18]。这种跨越光学系统、手工与自动推片染色、对焦稳定性的物理差异，提前使特征空间适应了基层最常见的制片与显微波动；

3. **异常负例的可靠可得性**：PBC、multi-focus 和 MLL23 提供了经过血液学专家病理复核的前体细胞切图，确保了负例监督信号的病理纯度。

2026 STEM High School Science Award
仅用于2026全国青少年科学论文公示

上述数据集的构成、来源及在本研究中的作用详见表 2。

表 2　模型研发源域数据集与跨域评估目标域数据集

数据集	采集地区	样本规模（正常＋异常）	类别构成	在本研究中的作用
PBC[15]	西班牙	10,283 ＋ 4,446	5 类正常＋ 2 类异常	研发训练集＋负例暴露池
Raabin[16]	伊朗	14,514	5 类正常	研发训练集
multi-focus[17]	韩国	23,882 ＋ 6,426	5 类正常＋ 3 类异常	研发训练集＋负例暴露池
MLL23[18]	德国	18,963 ＋ 16,520	5 类正常＋ 6 类异常	研发训练集＋负例暴露池
KU-Optofil[20]	土耳其	26,008 ＋ 2,097	5 类正常＋ 3 类异常	目标域建库、内控与测试（开发域）
AML-Cytomorphology[7]	德国	14,822 ＋ 3,346	5 类正常＋ 2 类异常	目标域建库、内控与测试（外部验证域）
Chula-WBC-15[21]	泰国	4,569 ＋ 406	5 类正常＋ 7 类异常	目标域建库、内控与测试（跨国盲测域）

三个测试侧的数据集的正常细胞各仅取 300 张：150 张用于构建本地特征库，150 张用作分布内正常对照，五类各 30 张。这一统一设定并非因为三家医院的正常数据规模相近，而是为了严谨模拟“目标医院仅掌握 150 张已标注正常图像”这一硬约束条件。目标域中未参与建库的其余正常图像用于测试误报率，异常细胞则全部保留作为专家病理标注真值集：KU 包含三系主指标 2,097 张，AML 包含骨髓系与幼红 3,346 张，Chula 全谱系包含 406 张、核心集 219 张。

训练侧的 MLL23 与测试侧的 AML-Cytomorphology 虽同出自德国慕尼黑白血病实验室，但二者分别发表于 2025 年与 2019 年，采集批次与时间跨度显著不同；实测仍表现出明显的域偏移（五分类准确率由同域 96% 以上降至 75.5%），跨批次外部验证的独立性由此得到保障。

（二）特征空间学习

本研究采用两阶段训练框架以实现特征分离：

第一阶段采用监督对比学习（SupCon）[22]，优化目标旨在拉近同类正常细胞在超球面上的表征距离，同时将公开负例池中的 11 类异常细胞作为推离先验。在训练中不对这些异常赋予具体的类型标签，仅将其作为同等排斥的离群点，从而将正常五类周围原本狭窄的类间缝隙与边界进一步扩大和外推，形成“正常聚簇致密、异常外围推离”的低密度几何拓扑。

第二阶段冻结全部骨干网络权重，仅训练一个极轻量的五分类线性头（覆盖五个正常分类）。此阶段存在一个关乎系统成败的关键细节：**异常负例仅参与第**

2026 Shanghai Youth High School Science Award
仅用于2026年上海高中学科竞赛论文公示

一阶段的度量对比，绝对禁止进入第二阶段的分类端。消融实验证实，一旦负例参与分类训练，交叉熵损失将强制在正常类与异常类之间划定硬性决策边界，从而极大地压缩类簇周围的缓冲带；异常样本极易被挤压并附着于分类界线两侧，导致基于距离的离群报警全面失效。

主干网络选用最新轻量架构 ConvNeXtV2-Atto[23]：参数量仅 3.7M，权重体积约 15MB。该模型是 ConvNeXtV2 系列中的最小尺寸规格；对比实验表明，稍大一档的 Femto（5.2M）在相同小样本跨域条件下反而出现过拟合，性能显著劣于 Atto。同时，其极低的参数规模确保了在基层无独立 GPU 的普通 PC 上即可实现计算要求。

经过实验探索后，训练超参数设定：单张消费级移动端 GPU（RTX 4070 Laptop，8GB 显存）即可完成全部训练。第一阶段使用 AdamW 优化器（学习率 3×10^{-4}）、批大小 50、温度系数 $\tau = 0.2$ 训练 100 个 epoch，输入分辨率为 300×300，采用颜色抖动与 TrivialAugment 增强，训练侧不做染色化学归一化；每 10 个 epoch 保存一次权重快照。第二阶段冻结骨干，以相同优化器训练分类头 50 个 epoch。交付产物包含五个独立随机种子（42、100、2026、24、1024）的冻结模型与一个独立预训练的通用复核模型。

（三）特征库构建

当冻结系统部署至新的目标机构时，无需任何基于梯度的重新训练。基层检验员仅需从本院常规涂片中选取 150 张经过复核的正常白细胞切片（5 类各 30 张），系统即刻在本地计算各类的特征均值向量 μ_c 与协方差矩阵。特征库同时记录各类 30 张样本相对于自身类中心的距离方差 σ_c，作为该细胞系特有的固有形态离散度指标（其在第四章用于跨系尺度校准）。

由于每类样本仅 30 张，在 320 维的特征空间中存在严重的样本不足（$N_c < D$），导致单一类别的样本协方差矩阵在数学上呈秩亏状态。因此，系统采用五类共享的混合协方差矩阵 Σ，即五类共用同一把刻画“正常生理形态波动”的度量标尺，并加入轻量岭正则化确保数值稳定性：

$$\Sigma = \frac{1}{N} \sum_{c=1}^{5} \sum_{i=1}^{N_c} (f(x_{c,i}) - \mu_c)(f(x_{c,i}) - \mu_c)^T + \lambda I$$

整个建库过程为纯数值统计计算，经测试，普通桌面 CPU 上仅需约 8 秒即

2026 S·T Youth High School Science Award
仅用于2026年成都中学科学奖论文公示

可为一家全新医院完成全套参数构建。

（四）轻量评分与报警机制

对于目标医院待筛查的新细胞 x，系统将其送入冻结主干提取特征 $f(x)$，并计算其到特征库中五个正常类簇的**马氏距离** [24]（图 2）：

$$d_c(x) = \sqrt{(f(x) - \mu_c)^T \Sigma^{-1} (f(x) - \mu_c)}$$

$$s(x) = \min_{c \in \{1,\dots,5\}} d_c(x)$$

马氏距离的临床意义在于“基于正常细胞自身形态变化椭球面的加权标准化距离”：如果特征偏移发生在该类别固有的形态拉伸方向（例如中性粒细胞分叶核的正常形变），则距离衰减较小；反之，若在正常细胞高度紧凑的特征方向发生微小畸变，距离将迅速放大。最终异常得分为该细胞到最近正常类簇的距离 $s(x)$，超过阈值即触发报警并导入人工复核队列。

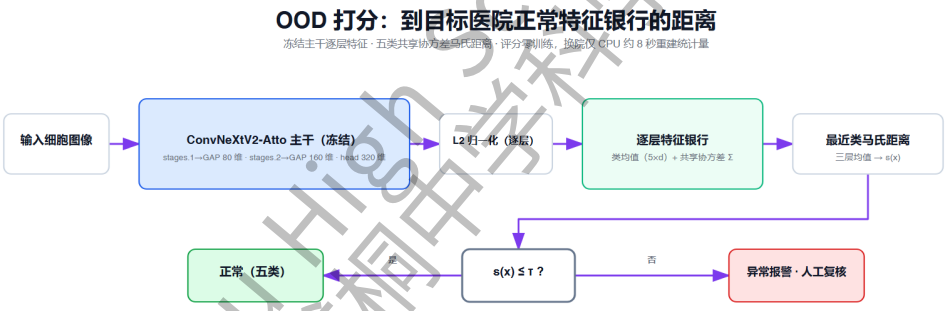


图 2　轻量特征提取、统计评分与报警工作流。新细胞切图经冻结骨干网络映射至特征空间，计算其到五个正常类中心的共享协方差马氏距离；最小距离超阈即触发报警。

经实验，**在本研究设定的共享协方差度量下，每类 30 张的标注配额已逼近边际效益饱和点**。实测增加每类标注至 60、100 张甚至按十倍规模（1,500 张）替换，目标域检测指标均不再显著提升。这表明基层检验科无需承担繁重的标注负担，150 张高质量正常样本就足以为目标机构快速锁定本院设备与试剂条件下的**几何定位中心与报警基准阈值**。

三、跨国多中心数据集检出性能实验与分析

为检验系统在异质医疗环境下的泛化表现，本研究将算法置于三个地理跨度、人群谱系、染色配方及数字化设备相互独立的外部数据集上展开测试。

2026 S.T. Yau High School Science Award
仅用于2026丘成桐中学科学奖公示

在医学形态学语境下，本研究定义的“跨域”，不只是计算机领域的数值分布迁移，还包括临床显微镜检的三方面异质性：

1. **染色试剂与化学工艺的差异**：例如土耳其 KU 使用的特定配比 May-Grünwald-Giemsa 染色、德国 AML 实验室使用的标准化全自动染色机试剂、而泰国 Chula 医院使用的手工常规推片与不同 pH 缓冲液；

2. **显微物镜与光电传感器差异**：包括科研级扫描仪、标准化数字形态分析仪与常规实验室显微镜的光学成像畸变；

3. **人群与疾病亚型构成的异质性**：涉及欧洲成年人急性髓系白血病、地中海区域病理样本以及东南亚地区样本。

（一）评估指标与临床工作量基准

根据国际血液学复检专家组共识（ICSH Recommendations）[14]，外周血细胞自动化初筛的首要临床红线是**严格控制假阴性率（漏诊率 ≤ 5%）**。对于白血病前体细胞等恶性病变，漏检将直接导致患儿错过黄金治疗期。

因此，本研究放弃了易受类别不平衡粉饰的常规准确率或单一 AUROC，将核心评价指标统一锁定为 **FPR95**：即**在保证真异常样本检出率严格达到 95%（假阴性率不超过 5%）的前提下，正常细胞被误报送检的比例**。这一 95% 工作点定义于一级聚合评分层；最终交付形态经二级大模型复核后的级联检出率及其权衡详见第四章。FPR95 越低，基层检验人员的无效复核负荷越轻。误报控制在工程上设定为低于 20%（即 FPR95 ≤ 0.20），其临床工作量含义在于：在 95% 兜底恶性前体的前提下，将 80% 以上的正常细胞安全放行，使基层有限的人力能集中于少数真正可疑的视野。

（二）测试环境一：KU-Optofil（开发域）

土耳其 KU-Optofil 数据集 [20] 采用不同显微系统与染色流程。测试所用专家病理标注异常集涵盖三系代表性病理形态：原始细胞（816 张）、有核红细胞（幼红细胞，243 张）与反应性淋巴细胞（1,038 张），共计 2,097 张。

在 10 次随机重抽样切分下，**单模型 FPR95 均值为 0.22±0.03，固定种子（42）代表性切分为 0.19，AUROC 达 0.94±0.01**（图 3）；同时跨域五分类准确率保持在 0.973。

2026 S. T. Yau High School Science Award 仅用于2026正成尚学科学论文公示

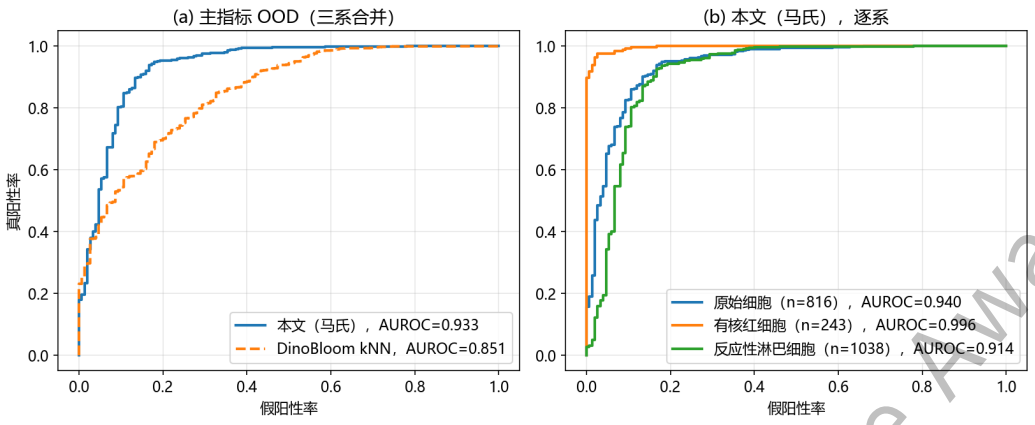


图 3　KU-Optofil 测试集 ROC 曲线分析（代表性切分）。左图 (a) 三系异常合并评估：本研究马氏距离路线 AUROC 达 0.934，显著超越零适配 DinoBloom 大模型（0.851）；右图 (b) 分系评估：原始细胞 0.940、幼红细胞 0.996、反应性淋巴细胞 0.914。横轴为正常误报率，纵轴为异常检出率；FPR95 对应纵轴 0.95 处的横轴坐标。

进一步分析特征空间分布（图 4），t-SNE 降维显示出不同病理形态的可分性差异：中性、嗜酸、嗜碱粒细胞形成了非常密实的类簇，幼红细胞被彻底排斥在正常空间远端；而原始细胞与反应性淋巴细胞呈现出部分嵌入正常淋巴簇边界的复杂拓扑，这种生理性形态重叠为第五章的误报下限分析埋下了伏笔。

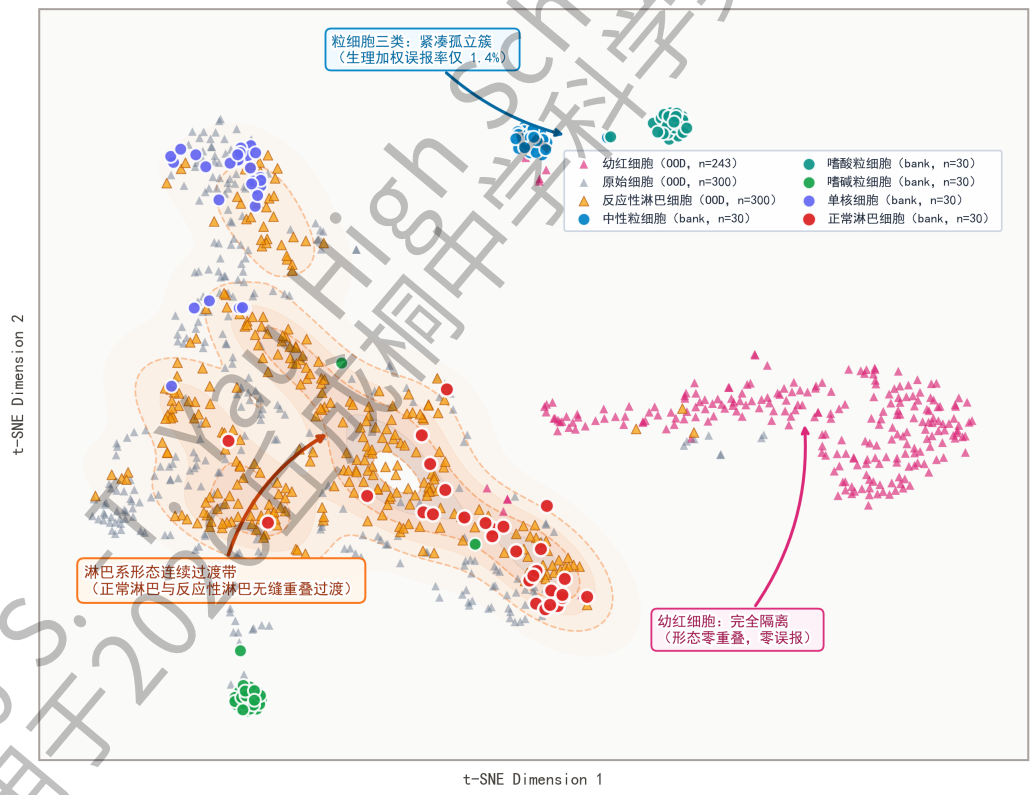


图 4　特征空间结构投影（t-SNE，KU 代表性切分）。正常粒细胞各类形成紧致簇群，幼红细胞（右侧）被完全隔离，而反应性淋巴细胞与原始细胞部分交织于正常淋巴区域边缘。

2026 S. Yat-Sen High School Science Award
仅用于2026中山大学附中科学奖公示

（三）测试环境二：AML-Cytomorphology（外部验证域）

德国慕尼黑白血病实验室 AML 数据集 [7] 的全部样本未参与算法与超参数决策。测试用专家病理标注集涵盖骨髓系原始细胞与幼红细胞共 3,346 张。

测试结果为：**FPR95 为 0.30±0.04，AUROC 为 0.93±0.01**。在该数据上，由于制片色差显著，正常白细胞分类正确率一度下滑至 0.755，**但基于距离的异常报警依然维持了高判别力**。这表明本方法对于显微色彩的剧烈漂移具有很强的抵抗力。

（四）测试环境三：Chula-WBC-15（跨国盲测终验域）

为验证系统能否跨越至陌生的亚洲人群与常规手工推片环境，本研究引入泰国常规门诊涂片数据集 Chula-WBC-15[21]。该数据集在系统超参数、主干结构及第四章聚合方案定稿锁定后方启动评测，属严格的盲测终验。

评估基准分为两套口径：

1. **核心异常评估集（主口径，219 张）**：以早期未成熟前体为主（髓系原始细胞 156、异型淋巴细胞 48、单核原始细胞 13、淋巴母细胞 2）；

2. **全谱系异常评估集（全量口径，406 张）**：在核心集基础上，纳入粒系发育早期形态（中幼粒 73、晚幼粒 83、早幼粒 31 等分化相对明确的前体）。

在单模型基准下，**全谱系口径 FPR95 为 0.293（AUROC 0.918），核心口径为 0.433**。在手工制片背景下，正常误报呈现出清晰的细胞系分化结构：粒系误报 0.16、单核系 0.40、淋巴系 0.60，与欧洲两家医院的结构完全同构。

（五）与代表性技术路线的对照

为确立本研究方案的性能坐标，表 1 中具有代表性的三类技术路线在相同硬件、同等建库约束下进行了严格横向测试：

1. **血液学通用大模型 DinoBloom（ViT-S，38 万张血液图像自监督预训练）[9]**：在无目标域微调（零适配）条件下，KU 域 FPR95 为 **0.48±0.06**，误报率达本研究单模型的 2.2 倍（0.217 对 0.48）。大模型虽然捕捉到了丰富的形态细节，但由于缺乏针对基层目标域正常流形的定标机制，直接用于边界报警时误报率较高。本研究因此在第四章将其用作二级复核过滤器；

2. **免训练置信度评分方法（MSP/Energy）[10][11]**：跨域条件下 FPR95 恶

2026 S.T.Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

化至 0.52–0.58；虽然其 AUROC 仍有 0.88 左右，但当检出要求被推至 95% 假阴性安全线时，假阳性急剧失控，无法作为严肃的临床报警器；

3. **常规交叉熵分类器（相同骨干、相同训练数据）**：尽管分类准确率看似良好，但跨域 FPR95 高达 **0.475**（KU）与 **0.833**（Chula，单核细胞误报率高达 100%）。这从反面给出了关键实证：**闭集分类精度绝不等价于异常筛查能力**。分类训练为了割裂已知类别而过度压缩类间低密度区域，破坏了度量报警所需的流形边界。

四、训练随机性抑制与多级聚合交付方案

（一）随机种子扰动与单次训练数字的随机成分

在深度学习模型研发中，随机种子对方差的影响在医学图像尾部检测中尤为剧烈 [25]。本研究在保持算法代码与数据完全不变的前提下，先后以五个随机种子复训（其中两个为验证结论稳定性的事后扩样种子，训练配置经逐键审计与前三者完全一致），发现末态模型（第 100 epoch）在 KU 上的 FPR95 竟在 0.19 至 0.44 之间剧烈摆动（图 5），摆动幅度未随种子数增加而收窄。

五个种子的末态均值期望为 0.335±0.094（三个种子时为 0.331，扩样后几乎不动）；五个种子的 KU 峰值轮次分散于第 100、40、30、50、80 轮，无一重合。若科研人员仅做一次实验并幸运地命中 0.19，并以此作为系统定型依据，所宣称的性能在临床部署中将面临巨大的复现风险。

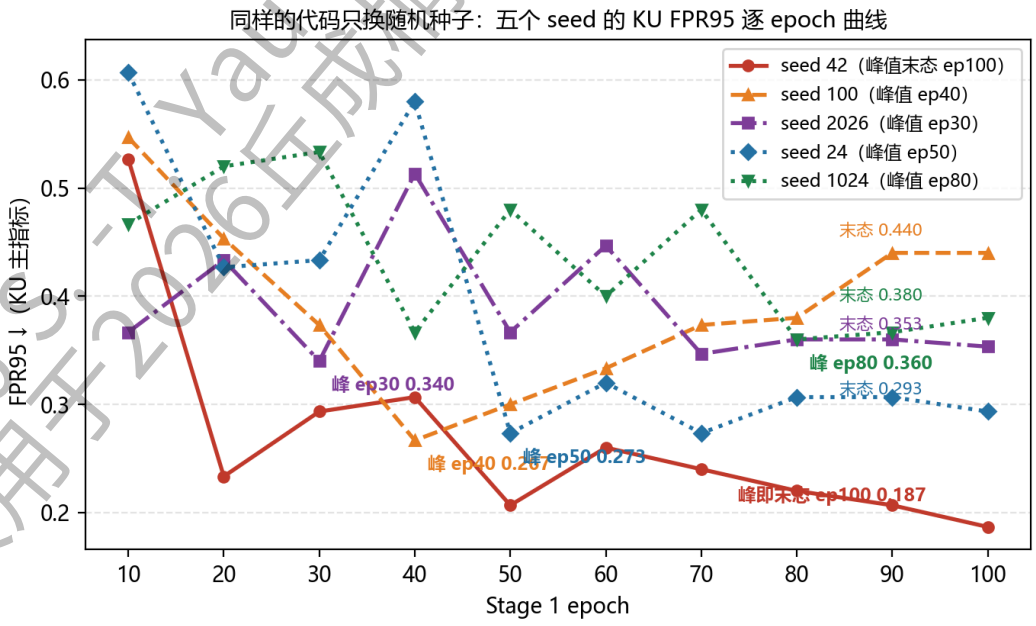


图 5　五个训练随机种子下的跨域 FPR95 演进轨迹。五个种子的 KU 峰值轮次分散于第 100、40、30、50、80 轮，无一重合；仅 seed 42 的峰值恰为末态，其余四个种子在峰值后均出现不同程度的退化。

2026 年 Yau Hing School Science Award 仅用于 2026 年成都中学科学奖论文公示

（二）多级聚合交付方案

为消除单模型偶然性，本研究设计了一套标准化的**多级聚合交付方案**，其核心思想是通过结构化中位数操作抑制方差，由四层机制级联构成（图 6）：

1. **时间维度中位数聚合（Epoch-Level）**：在单一训练种子内，保存每 10 个 epoch 的截面权重。对于输入图像，先将其各截面得分基于目标域正常集做经验分位次归一化，消除绝对量纲差异，随后取时间轨迹上的得分中位数；

2. **跨种子集成中位数聚合（Seed-Level）**：对五个种子（42、100、2026、24、1024）的时间中位数得分再次取中位数，消除特定局部极小值的偶发漂移；成员由三个增至五个后，中位数机制对病态成员的逐系容忍上限由一个提升至两个；

3. **Bank 特征方差标准化（Scale-Calibration）**：针对不同正常细胞系的固有离散度差异进行校准。单核细胞由于吞噬与变形生理特性，胞体大、核形多样，类内散度天然高于分叶核粒细胞。若使用统一距离阈值，单核细胞极易因方差大而被错推至异常端。本研究利用目标域 150 张正常样本到各自类中心的标准差 σ_c，将马氏距离重整化为无量纲倍数 $\tilde{s}(x) = \min_c d_c(x)/\sigma_c$。这一纯统计改进使 KU 与 AML 的误报率分别压低至 **0.220** 与 **0.240**，三系真异常检出率无损失；

4. **独立通用大模型轻量复核（Secondary Sifting）**：以完全独立架构预训练的 DinoBloom 作为二级安全阀门。仅当一级聚合模块发出报警时，才触发其进行前向判定（按“放行 99%”设阈），二次过滤假阳性。

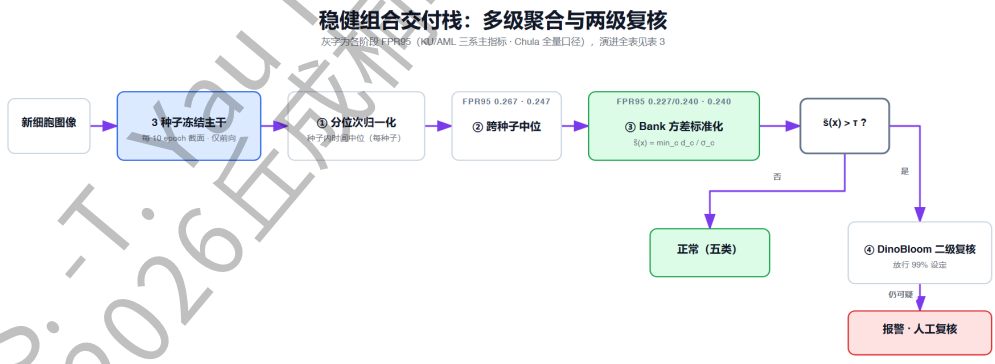


图 6　多级聚合稳健交付方案整体计算流。时间中位数滤波（①）与种子间中位数滤波（②）消除统计扰动；Bank 方差标准化（③）消除谱系尺度偏差；独立 DinoBloom 大模型（④）实施末端人工安全复核。

经该交付方案处理后，KU-Optofil 与 AML-Cytomorphology 两大数据集上的最终误报率**分别收敛至 0.180 与 0.193**，在无需人工调优的前提下，稳定通过 20% 工程设计线。两级模块级联后的真异常全局保守检出率为 $0.95 \times 0.99 \approx \mathbf{94}\%$（实测恶性原始细胞与幼红细胞检出率均在 94.0%~99.6% 之间）。二级复核的放行阈

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖公示

值为可调参数：若临床部署要求严格执行 95% 检出线，可将放行比例由 99% 收紧至 99.5% 以上，级联检出率随之回升至约 94.5% 并趋近该线，代价是误报率轻微上浮；当前 0.99 的取值是误报与检出权衡后的实测工作点。

（三）第三陌生域独立终验与方案汇总

将这套固化交付方案直接推向泰国 Chula-WBC-15 进行跨国盲测终验：

1. 经时间中位数与跨种子中位数聚合，全谱系误报率先由单模型的 0.293 降至 0.260；

2. 叠加 Bank 方差标准化后降至 0.233（AUROC 0.948）；

3. 最终经独立大模型二级复核，**全谱系评估集误报率定格于 0.173，真异常检出率 93.8%**，再次跨过 20% 的工程目标线；核心集（以前体为主）误报率为 0.280。

表 3 系统汇总了方案演进中各模块的技术贡献。

表 3　技术方案演进与跨国多中心消融总览（核心指标：FPR95，越低越好）

方案阶梯	KU-Optofil（开发域）	AML-Cytomorphology（外部验证域）	Chula-WBC-15（盲测终验域）	工程与临床定位
基准 1：常规交叉熵分类器	0.475	—	0.833	证实分类精度不等于异常检出
基准 2：DinoBloom（零适配）	0.48±0.06	—	0.567	缺乏目标域定标时误报偏高
阶段 1：本研究单模型基准	0.22±0.03（代表性 0.19）	0.30±0.04	0.293（核心集 0.433）	存在随机种子漂移风险
阶段 2：＋时间与多模型中位聚合	0.253	0.273	0.260	锁定稳态，消除单次偶然性
阶段 3：＋ Bank 特征方差标准化	0.220	0.240	0.233	校准单核等大离散生理偏差
阶段 4：＋独立通用大模型复核	**0.180**（真阳约 94%）	**0.193**（真阳约 94%）	**0.173**（真阳 93.8%）	最终稳健交付形态，跨 20% 线

表 3 注：①测试口径：KU/AML 列对应三系主要异常指标，Chula 列对应全谱系异常评估集；②统计属性：基准行与阶段 1 为 10 次随机重抽样均值，阶段 2 起为确定性多级交付稳态；③对照口径：两组基准均为零微调单模型，本研究单模型对照点为 KU 0.217、Chula 0.293。

硬件实测开销：在普通多核 CPU 上实测，单模型基础数学前向耗时约 51 ms；大模型复核耗时单张约 0.14 s 且仅对触发一级警报的约 19% 样本触发。在包含切片文件读取 I/O、50 组模型截面参数调度与大模型复核的**真实全流程端到端实测中，单个细胞平均耗时约 1.2 秒；对于一张包含 100 余个白细胞视野的标准推片，整体自动筛查与分级报告生成耗时约 2 分 10 秒至 2 分 30 秒（两分钟出头）**，完全处于基层门诊半小时取报告的充裕时间窗口内。更重要的是，**目标医院的特征库构建是一次性的（One-time Setup）**：首次接入时需对 150 张正常图像提取

2026 S.T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

特征（单模型约 8 秒，完整交付栈约 2 分钟），之后所有患者直接复用该特征库进行纯前向打分，不再产生任何重复建库开销。

五、淋巴系误报经验下界分析与分级复核部署

（一）误报按细胞系分化规律

实验数据呈现出一个跨国、跨设备高度一致的生物学规律：**系统的异常误报并非均匀散落于各类别，而是严格按白细胞谱系呈现阶梯分化。**

在代表性切分下：

1. 粒细胞系（中性粒、嗜酸性、嗜碱性）误报率极低，合计仅为 **0.08±0.03**；
2. 单核细胞系误报率居中，约为 **0.35±0.11**；
3. 淋巴细胞系误报率最高，达 **0.51±0.11**。

约 80% 的系统虚警集中在淋巴与单核两系。尤其在淋巴系中，正常成熟淋巴细胞在形态上与轻度刺激下的反应性淋巴细胞存在高度重合，目测对比两者几无二致（图 7）。

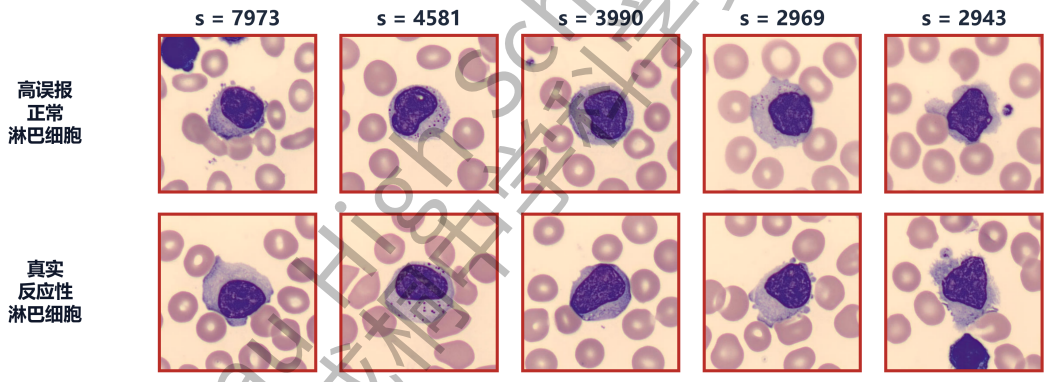


图 7　高误报正常淋巴细胞与真实反应性淋巴细胞显微形态对照。上排：触发系统高置信报警的假阳性正常淋巴细胞；下排：真实病理确认的反应性淋巴细胞。两者在核质比与胞质染色质特征上呈现出典型的重叠交织。

（二）12 组干预实验与独立特征空间验证

为探究能否通过算法调优攻破淋巴系误报，本研究设计了多达 **12 组多维度干预实验**：更换深层主干架构、调整局部协方差结构、KNN 特征融合、调整四族不同的分数归一化算法（经验分位数、Gamma 尾部建模、Copula 极值分布等）、将标注配额扩展至十倍、实施 K-T 领域微调、将清洗阶段滤除的大颗粒淋巴细胞（LGL）重新召回训练、以及引入完全不相干的独立自监督表征 DinoBloom 重测。

12 组干预无一例外地撞上了约 0.30 的淋巴误报经验下限。即使是在整体能

2026 S.T. High School Science Award 仅用于2026年比赛作品展示

力更弱的 DinoBloom 上，其淋巴系误报依然高达 0.87，稳居各谱系之首。两个相互独立的模型在同一谱系上展现出相同的瓶颈，佐证这一现象**并非局部算法欠拟合，而是源于白细胞病理形态学分布的固有几何属性**。

从形态发生学来看，正常淋巴细胞受免疫刺激转化为反应性、异型淋巴细胞，再到克隆性恶性前体，在光学显微下本就处于连续渐变的形态光谱之中。根据分布外检测的可学习性下界理论 [26][27]，当分布内支持集与异常分布存在物理不可分重叠时，任何基于形态的分类器均存在无法穿透的理论误报率下界。

（三）AML 数据集的反常现象与应力测试

在 AML 数据集的测试中，曾出现两个看似反常的读数：

1. **AML 粒系名义误报率异常升至 18%**：经亚群解剖，中性粒细胞误报率在 AML 仅为 0.0%~3.3%，真正拉高数据的是嗜碱性粒细胞（误报 40%）。其根源在于：测试为了统计平衡，采用各系 1:1:1:1:1 抽样，将真实外周血中占比仅 0.5%~1% 的罕见嗜碱细胞放大了数十倍，加之该库仅存 79 张嗜碱切图导致方差失真。若按真实生理白细胞比例折算加权，**AML 粒系真实综合误报率仅为 1.4%**；

2. **AML 淋巴误报率一度看似突破 0.30（跌至 16.7%~23.3%）**：解剖测试集构成发现，该数据集异常样本中 97.7% 为原始粒细胞，未收录反应性淋巴细胞，导致报警边界未被推入淋巴形态分布腹地。本研究利用该库中仅存的 11 张异型淋巴细胞实施**反向敲除应力测试**：一旦要求系统对这 11 张异型淋巴的召回率强制达到 95%，报警阈值被迫收紧，正常淋巴误报率立刻反弹至 **43.3%**，再次验证了 0.30 下界的存在。

（四）分级复核部署建议

既然约 0.30 的淋巴误报下界是由细胞生物学本质所决定，在算法层面强行追求将淋巴虚警归零便属于非理性投入。系统的工程重心应从“算法层消灭误报”转向“**管理层分流误报**”，本研究据此提出兼顾安全性与效率的**分级复核方案**：

1. **一级高危预警（粒系报警）**：可信度极高（虚警率仅 1%~8%）。由于儿童白血病以急性型居多且前体粒细胞极具病理特征，粒系一旦报警，大概率存在真异常，建议本地检验员立即暂停样本放行并优先复核；

2. **二级观察预警（淋巴系报警）**：可信度相对较低（约 30%~50% 虚警率）。

2026 Shanghai Science Award 仅用于2026年成都中学科学奖论文公示

但“二级”不意味着“抛弃”，所有淋巴报警样本必须完好保留在队列中；鉴于基层自身阅片能力有限，建议通过数字化远程医疗通道，打包交由区域医疗中心或上级医联体医院进行异地病理复核。

这一设计在医学逻辑上同国际公认的 CellaVision 等商用高通量数字形态分析仪同构 [28]，将传统漫无目的的“全片大海捞针式复核”改造为“重点突出的结构化分级复核”，降低了基层的医疗风险与人力负担。

六、讨论

（一）误报按细胞系分化的细胞病理学机理

本研究在三个跨国多中心数据集中稳定复现了“粒系极低、单核居中、淋巴居首”的误报阶梯分布。这一现象源于白细胞各细胞系在细胞病理学与免疫功能上的形态特征：

1. **粒细胞系（Granulocytes）**：中性粒、嗜酸性与嗜碱性粒细胞在其发育成熟终末期具有高度特异性的形态指纹。分叶核结构、特异性胞质嗜色颗粒为神经网络提供了强烈的高维判别特征，在特征流形中形成独立且超紧凑的致密类簇，因而对域偏移抵抗力最强，误报率最低；

2. **单核细胞系（Monocytes）**：作为单核-吞噬细胞系统的关键成员，细胞具有极强的阿米巴样运动与变形能力，胞体大且核形（肾形、马蹄形、不规则折叠）天然多变，造成其类内形态散度偏大。这解释了为何单核系在单模型下对随机种子高度敏感（五个种子中有两个出现该系误报陡增，同一现象两度复现）；本研究提出的 Bank 方差标准化恰恰在数学上还原了单核细胞的生理方差，从而以无损检出的方式消除了这一系统性偏差；

3. **淋巴细胞系（Lymphocytes）**：淋巴细胞是机体适应性免疫的核心。在外周血中，静息期成熟淋巴细胞在抗原刺激下会经历连续的原幼化转变，形成大小不一、核染色质疏松、胞质嗜碱性增强的反应性/异型淋巴细胞。这一过程是一个连续的生物演化光谱，在光学显微下并没有绝对清晰的生化硬边界。

临床血液学文献对此提供了互证：在针对骨髓涂片 22 类细胞的深度学习判读研究中，异型淋巴细胞与正常淋巴细胞的机器混淆错误率高达 53.33%，且即便是多位顶级血液病理专家之间的盲读一致性也仅约 60%[29]。理论层面上，这在细胞显微学中对应了分布外检测的可学习性下限公理 [26][27]。本研究不仅给

2026 S.T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

出了 0.30 这一经验下界的实证测定，更首次建立了“**形态连续谱宽度 → 空间不可分重叠 → 筛查误报下限**”的系统性理论映射。

（二）与现有技术路线的对比与范式转变

本研究成果为当前血细胞智能辅助诊断领域提供了几点对照与思考：

1. **范式转变：打破“分类精度等同于筛查能力”的传统误区**：以往大量文献追求在公开测试集上刷高闭集五分类精度。本研究通过交叉熵基准在跨国盲测中误报率高达 0.833 的事实，暴露了传统分类模型的脆弱性：分类边界会挤压正常空间，而以对比度量学习驱动的“聚簇-推离”范式才契合异常筛查的物理特点；

2. **算力与效能权衡：重估基础大模型在基层受限环境的应用定位**：拥有 38 万张高质量图像训练的 DinoBloom 大模型代表了当前形态学表征的最前沿。然而本研究实证指出，通用大模型若不针对具体机构做微调，由于缺乏局部正常分布的精确定标，其跨域异常报警误报率高达 0.48。相对而言，本研究仅 3.7M 参数的轻量小模型，凭借目标域 150 张正常样本的几何定标，实现了大幅超越大模型的筛查准确度。这证实了在基层医疗资源受限的场景下，**专精的轻量模型结合小样本定标，效果优于庞大通用的无适配大模型**。

（三）研究局限与前瞻展望

本研究仍有以下局限性：

1. **离线图像与实际验证的差距**：全部实验基于已裁剪分割的公开离线单细胞切片，缺乏真实检验科推片前处理、复杂全玻片背景干扰（如血小板聚集、细胞重叠黏连、染色沉淀）的前瞻性连续队列检验；

2. **报警阈值的临床在线标定缺口**：论文中的 FPR95 工作点阈值依赖测试集中异常样本的反推确定（离线学术基准），而在真实基层部署中，医院在零异常前提下仅能依据 150 张正常样本的尾部分布近似估计极值阈值，二者之间尚存在工程缺口；

3. **感染性核左移与恶性前体的形态鉴别**：重症感染导致的早幼粒、中幼粒细胞左移，容易与白血病前体细胞产生重叠报警；未来本系统可探索与全自动血细胞分析仪物理通道数据的多模态联合诊断；

4. **病理亚型的精细拓展**：受限于公开数据集标注，急性淋巴细胞白血病

2026 STEYouth High School Science Award
仅用于2026国际青少年科学论文大会

（ALL）特异性的淋巴母细胞尚未能作为独立类别进行孤立评估，进一步的儿童白血病专病验证是下一阶段的核心推进方向。

七、结论

面向基层医疗机构“零异常标注、缺乏重训算力、成像环境异质”的严苛现实约束，本研究设计并验证了一条跨域血细胞异常初筛技术路径。通过在源域利用监督对比学习构建“正常五类紧密聚簇、异常负例充分推离”的流形空间，目标医院仅需 150 张正常白细胞切片，在普通 CPU 上约 8 秒（单模型）至 2 分钟（完整交付栈）即可完成一次性特征库构建，实现了检测端全流程免训练与两分钟级全自动推片筛查。

针对深度学习随机种子的结果漂移问题，本研究交付了包含时间中位数、跨种子中位数、Bank 方差标准化与独立大模型复核的多级聚合交付方案，在土耳其 KU-Optofil、德国 AML-Cytomorphology 两个外部数据集上将 FPR95 分别稳定压至 0.180 与 0.193（真异常检出率约 94%），并在泰国 Chula-WBC-15 跨国盲测终验中达到 0.173，均跨过 20% 的工程设计目标线。

在病理机制层面，本研究揭示了闭集分类精度与异常判别能力的解耦现象，并凭借 12 组多维度干预、独立特征空间与反向应力测试，证实了淋巴系约 0.30 误报下界源于形态连续谱的生物学客观本质。据此提出的“粒系优先复核、淋巴系置后分流、上级医联体兜底”的分级复核部署策略，为高可靠、低成本的儿童白血病早期形态学辅助筛查在基层落地，提供了理论与工程支撑。

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

参考文献

[1] B. Han, R. Zheng, H. Zeng, et al., “Cancer incidence and mortality in China, 2022,” Journal of the National Cancer Center, vol. 4, no. 1, pp. 47–53, 2024.

[2] 国家儿童医学中心（上海），上海交通大学医学院附属上海儿童医学中心. 中国儿童血液病 2020 白皮书 [R]. 上海, 2021.

[3] 国家卫生健康委员会. 2024 年我国卫生健康事业发展统计公报 [R]. 北京, 2025.

[4] 韩菊铭, 李辉, 胡瑞, 等. 合肥地区基层医疗机构检验科血常规室间比对结果分析 [J]. 国际检验医学杂志, 2016, 37(18): 2571–2572, 2575.

[5] X. Ni, Z. Li, Y. Liu, et al., “Socioeconomic inequalities in cancer survival and access to health care among children and adolescents in China, 2018–24: a prospective nationwide cohort study,” The Lancet, vol. 407, no. 10525, pp. 237–255, 2026.

[6] A. Acevedo, S. Alférez, A. Merino, L. Puigví, and J. Rodellar, “Recognition of peripheral blood cells images using convolutional neural networks,” Computer Methods and Programs in Biomedicine, vol. 180, 105020, 2019.

[7] C. Matek et al., “Human-level recognition of blast cells in acute myeloid leukaemia with convolutional neural networks,” Nature Machine Intelligence, vol. 1, pp. 538–544, 2019.

[8] S. Deltadahl et al., “Deep generative classification of blood cell morphology,” Nature Machine Intelligence, vol. 7, no. 11, pp. 1791–1803, 2025.

[9] V. Koch et al., “DinoBloom: A foundation model for generalizable cell embeddings in hematology,” in Proc. MICCAI, 2024; arXiv:2404.05022.

[10] D. Hendrycks and K. Gimpel, “A baseline for detecting misclassified and out-of-distribution examples in neural networks,” in ICLR, 2017.

[11] W. Liu et al., “Energy-based out-of-distribution detection,” in NeurIPS, 2020.

[12] I. N. Margret and R. Krishnan, “Anomaly-aware peripheral blood smear analysis via hybrid detection and one-class learning,” Computer Modeling in Engineering & Sciences, vol. 147, no. 3, 2026, doi: 10.32604/cmes.2026.079104.

[13] N. Belton, M. T. Hagos, A. Lawlor, and K. M. Curran, “FewSOME: One-class few shot anomaly detection with Siamese networks,” in Proc. CVPR Workshops,

2026 S. T. Yau High School Science Award 仅用于2026丘成桐中学科学奖公示

pp. 2977–2986, 2023.

[14] P. W. Barnes, S. L. McFadden, S. J. Machin, and E. Simson, “The international consensus group for hematology review: suggested criteria for action following automated CBC and WBC differential analysis,” Laboratory Hematology, vol. 11, no. 2, pp. 83–90, 2005.

[15] A. Acevedo et al., “A dataset of microscopic peripheral blood cell images for development of automatic recognition systems,” Data in Brief, vol. 30, 105474, 2020.

[16] Z. M. Kouzehkanan et al., “A large dataset of white blood cells containing cell locations and types, along with segmented nuclei and cytoplasm,” Scientific Reports, vol. 12, 1123, 2022.

[17] S. Park et al., “A large multi-focus dataset for white blood cell classification,” Scientific Data, vol. 11, 1106, 2024.

[18] S. S. Shetab Boushehri et al., “A large expert-annotated single-cell peripheral blood dataset for haematological disease diagnostics,” Scientific Data, vol. 12, 1773, 2025.

[19] D. Hendrycks, M. Mazeika, and T. Dietterich, “Deep anomaly detection with outlier exposure,” in ICLR, 2019.

[20] A. E. Yarıkan et al., “A large-scale peripheral blood cell dataset for automated hematological analysis,” Scientific Data, vol. 13, no. 1, 417, 2026.

[21] L. Clifton, X. Tian, D. Palasuwan, P. Watanaboonyongcharoen, P. Rojnuckarin, and N. Anantrasirichai, “Mamba-based ensemble learning for white blood cell classification,” in Proc. IEEE ISBI, 2026.（Chula-WBC-15 数据集来源论文，doi:10.1109/ISBI61048.2026.11515403）

[22] P. Khosla et al., “Supervised contrastive learning,” in NeurIPS, 2020.

[23] S. Woo et al., “ConvNeXt V2: Co-designing and scaling ConvNets with masked autoencoders,” in CVPR, 2023.

[24] K. Lee et al., “A simple unified framework for detecting out-of-distribution samples and adversarial attacks,” in NeurIPS, 2018.

[25] L. Picard, “torch.manual_seed(3407) is all you need: The incredible sonic power of seed 42,” arXiv preprint arXiv:2109.08203, 2021.

2026 S.T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

[26] Z. Fang et al.,“Is out-of-distribution detection learnable?”in NeurIPS, 2022.

[27] Z. Fang, Y. Li, F. Liu, B. Han, and J. Lu, “On the learnability of out-of-distribution detection,”Journal of Machine Learning Research, vol. 25, pp. 1−83, 2024.

[28] E. Cornet, J.-P. Perol, and X. Troussard, “Performance evaluation and relevance of the CellaVision DM96 system in routine analysis and in patients with malignant hematological diseases,”International Journal of Laboratory Hematology, vol. 30, no. 6, pp. 536−542, 2008.

[29] S. Zhou et al.,“VFM-SSL-BMADCC-Framework: Vision foundation model and self-supervised learning based automated framework for differential cell counts on whole-slide bone marrow aspirate smears,”Frontiers in Medicine, vol. 12, 1624683, 2025.

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示