International Journal of Laboratory Hematology

WILEY

ISLH International Journal of Laboratory Hematology

REVIEW

AI In Leukemia Diagnostics: Complementing the Pathologist's Role

Sayandeep K. Das[1] | Kusal K. Das[2]

[1]Department of Pathology, BLDE (Deemed To Be University), Shri B. M. Patil Medical College Hospital and Research Centre, Vijayapura, Karnataka, India | [2]Laboratory of Vascular Physiology and Medicine, Department of Physiology, BLDE (Deemed To Be University), Shri B. M. Patil Medical College Hospital and Research Centre, Vijayapura, Karnataka, India

Correspondence: Kusal K. Das (kusaldas@bldedu.ac.in)

Received: 10 January 2026 | **Revised:** 16 June 2026 | **Accepted:** 25 June 2026

Keywords: artificial intelligence | digital pathology | flow cytometry | hematopathology | human–AI collaboration | leukemia | medical education

ABSTRACT

Artificial intelligence (AI) is reshaping every stage of leukemia diagnostics, from digital morphology and multiparameter flow cytometry to next-generation sequencing, multi-omics analysis, and emerging computational frontiers such as quantum-inspired feature selection. This review outlines how contemporary AI tools can automate labor-intensive quantitation, flag diagnostically salient patterns, and standardize interpretation, while the pathologist or hematologist retains authority over validation, context-specific integration, and clinical decision-making. We present an illustrative "human-in-the-loop" workflow that embeds AI modules within current laboratory information systems, emphasizing points where expert oversight mitigates algorithmic bias and resolves discordant findings. We further map the validator–integrator role across morphology, flow cytometry, and genomic/multi-omic interpretation and provide practical training competencies and use cases for AI-assisted hematopathology. Beyond technical deployment, the article addresses the educational transformation required for sustainable adoption. Drawing on international competency frameworks, including the Digital Health Competencies in Medical Education Framework and recently proposed AI-specific Entrustable Professional Activities, we map core skills that future hematopathologists must master: data-science literacy, critical appraisal of AI outputs, and ethical governance. We highlight evaluated training models such as the Pathology Informatics Essentials for Residents curriculum, Stanford Artificial Intelligence in Machine and Imaging workshops, and College of American Pathologists bootcamps and propose integration strategies adaptable across resource settings. By pairing rigorous validation with targeted education, AI can elevate rather than eclipse the diagnostic role of the leukemia specialist, enabling more timely, reproducible, and personalized patient care.

1 | Introduction

The diagnostic workup and final classification of leukemia require coordinated integration of morphologic, laboratory, immunophenotypic, cytogenetic, and genomic data. Establishing an accurate classification necessitates meticulous morphological assessment of peripheral blood films, bone marrow aspirate smears, trephine biopsy sections, and, in selected cases, tissue or body fluid specimens, together with comprehensive immunophenotyping via multiparameter flow cytometry, detailed cytogenetic analysis, and molecular findings [1, 2]. The accuracy and timeliness of this complex diagnostic process are paramount, directly influencing critical therapeutic decisions and shaping patient prognosis in hematological malignancies [1]. Contemporary diagnostic frameworks, including the International Consensus Classification (ICC) and World Health Organization (WHO) classification systems, increasingly require integration of morphology with immunophenotypic, cytogenetic, and genomic findings; therefore, any artificial intelligence (AI)-assisted diagnostic tool must be evaluated

according to its ability to support, rather than replace, this integrated classification process [3].

AI technologies, particularly machine learning and deep learning algorithms, demonstrate increasing utility across various medical domains [4]. Within hematopathology, AI presents considerable potential for analyzing intricate, high-dimensional datasets generated during the workup of leukemia. AI applications range from automated analysis of digital microscopic images using advanced convolutional neural networks (CNNs) and interpretation of complex flow cytometry data via sophisticated clustering and classification algorithms to identifying predictive patterns within genomic and transcriptomic profiles [5, 6].

The advent of AI systems capable of performing these sophisticated analytical tasks prompts a critical question regarding the future role of the human expert: Will AI systems eventually replace pathologists and hematologists in the diagnostic pathway, or will they rather serve as powerful augmenting tools? This study posits that the optimal and most likely future will not be shaped through competition, but through collaboration and development of a synergistic relationship between clinical expertise and AI capabilities. By examining the integration of AI into the complex diagnostic workflow of acute leukemia, this perspective review uses hematopathology as a detailed exemplar for a transformation of human–AI synergy occurring across medicine. The principles governing the evolving role of the specialist, the need for new training paradigms, and the challenges in clinical implementation are highly relevant to other data-intensive fields facing similar AI-assisted transformation. Accordingly, we put forward a conceptual model of the pathologist as a "validator–integrator" and propose an actionable competency framework designed to prepare trainees for this new paradigm, using the leukemia workflow as a detailed exemplar.

2 | Methods

To inform this review, we conducted a targeted narrative review of the literature, an approach chosen for its flexibility in synthesizing a broad and rapidly evolving field [7]. Our goal was to identify and synthesize key publications that address the application of AI in leukemia diagnostics and its subsequent impact on the pathology profession and medical training. To provide structure and rigor, our process was guided by the principles outlined in the Scale for the Assessment of Narrative Review Articles [8].

A literature search was performed using PubMed and Google Scholar for articles published between January 2013 and March 2025. The search terms included, but were not limited to, "artificial intelligence," "machine learning," "leukemia," "hematopathology," "digital pathology," "flow cytometry," and "medical education." We included original research articles demonstrating key technological capabilities, high-impact review articles, foundational clinical guidelines (e.g., WHO, ELN), and established educational frameworks.

After removal of duplicate bibliographic entries, 90 unique sources were retained in the final reference set. These comprised 36 core articles selected during the initial narrative search for their direct relevance to AI applications in leukemia diagnostics, the evolving validator–integrator role of the pathologist/hematologist, and educational reform for AI-assisted practice. Additional targeted sources were incorporated during revision to strengthen the discussion of clinical readiness, external validation, technical barriers, regulatory and governance considerations, AI reporting frameworks, quantum-inspired computational approaches, and competency-based training. The findings were synthesized narratively around five major themes: AI applications in morphology, AI-assisted flow cytometry, genomic and multi-omic interpretation, the evolving role of the pathologist/hematologist, and educational preparation for AI-assisted hematopathology.

A detailed description of the search strategy, eligibility criteria (Table S1), literature selection flow diagram (Figure S1), summary of the 36 core narrative-synthesis sources (Table S2), and category-wise summary of additional targeted sources incorporated during revision (Table S3) is provided in the File S1.

3 | AI's Role in Leukemia Diagnostics: Capabilities, Evidence Maturity, and Translational Barriers

Morphological review remains a cornerstone of leukemia diagnosis, with deep learning making significant strides in this area through digital pathology. Whole-slide imaging (WSI) scanners convert glass slides into large digital image files suitable for computational analysis. CNNs, inspired by the human visual cortex, dominate this field. Models such as ResNet, Visual Geometry Group (VGG), Inception, or custom variants are trained on extensive datasets of annotated digital smears [9]. Annotations by pathologists, including bounding boxes around blasts and cell type labels, are essential for supervised learning, despite challenges in annotation consistency and effort. Techniques like data augmentation, which include rotation, scaling, and color jittering, are used to expand the dataset size and enhance model robustness.

Significant challenges, however, can impact the real-world performance of these models. Besides annotation consistency, a primary hurdle is addressing pre-analytical variability, such as inconsistencies in slide preparation and staining across different laboratories, requiring robust stain normalization techniques [10]. This issue contributes to the broader problem of dataset shift, where a model's performance degrades when applied to data from a new institution using different WSI scanners or tissue fixation protocols. Overcoming this problem requires extensive external validation on diverse, multi-institutional datasets to ensure model robustness and safety in clinical practice, as recommended by professional guidelines [11].

Current applications have shown potential for automating differential counts with high precision; for example, an explainable AI model for white blood cell classification reported a remarkable accuracy of 99.70% [12]. Deep learning models have also been developed to differentiate myeloblasts from lymphoblasts in peripheral blood images, a critical step in distinguishing between acute myeloid leukemia (AML) and acute lymphoblastic leukemia. A notable study by Matek et al.

discusses a CNN that identified AML blasts with an F1-score of 0.897, which was on par with the performance of human experts [5]. This potential of AI in identifying differential diagnoses is now being tested in large-scale clinical settings, bridging the gap between research and practice. For example, the prospective, multi-center clinical study supporting the FDA clearance of the Scopio X100 platform evaluated its performance on thousands of peripheral blood smears across multiple sites, demonstrating high correlation with manual microscopy for differential counts and cell morphology review [13]. However, the clinical readiness of morphology-based AI varies substantially by task. Algorithms trained to classify single-cell images or detect blasts in curated datasets should be interpreted as proof-of-concept unless they have been externally validated across staining protocols, scanners, laboratories, patient populations, and rare leukemia subtypes. In contrast, digital morphology platforms used for peripheral blood smear review or differential counting may be closer to workflow deployment when evaluated prospectively and used within a restricted intended use. Even in these settings, such tools should not be presented as standalone leukemia diagnostic systems; rather, they provide triage, quantification, or region-of-interest support for expert review [14].

Other models have shown similar capabilities in identifying diagnostically relevant features such as Auer rods or subtle dysplastic changes. Furthermore, AI holds promise for screening applications by highlighting regions of interest or flagging rare suspicious events or candidate abnormal cells for expert review [15, 16]. A 2025 systematic review and meta-analysis reported that AI-assisted systems achieved a pooled sensitivity of 99% for leukemia detection [17]. While these research platforms demonstrate significant potential, their performance in large-scale, multi-center external validation studies is still under investigation, and fully integrated, clinically validated software is still emerging. In this domain, the validator–integrator role requires the pathologist or hematologist to verify AI-generated blast counts, review flagged regions, distinguish true blasts from artifacts or reactive cells, assess dysplasia and Auer rods where relevant, and reconcile automated outputs with peripheral blood, bone marrow aspirate, trephine biopsy, and clinical findings. The human expert, therefore, validates both the cellular identity and the diagnostic significance of the AI-highlighted features [18].

3.1 | Immunophenotyping: Unsupervised and Supervised AI for Flow Cytometry Data

Multiparameterflow cytometry produces complex, high-dimensional data essential for leukemia classification and minimal residual disease (MRD) assessment. The traditional analysis, which relies on subjective manual gating, becomes increasingly challenging and time-consuming as panel complexity increases [14]. As a solution to this, AI offers both unsupervised clustering algorithms, such as FlowSOM and PhenoGraph, and supervised classification methods such as Random forests, k-nearest neighbor classifiers, and support vector machines [19, 20]. Unsupervised techniques are valuable for exploratory analysis, identifying cell populations based on the data's inherent structure without requiring prior labels. Supervised methods, in contrast, are trained on expert-labeled data to automate the classification of known cell populations or to detect predefined aberrant phenotypes with high consistency [6, 21]. Methodologically, flow cytometry AI models are highly sensitive to pre-analytical and analytical variability. Differences in antibody panels, fluorochrome combinations, compensation strategies, gating conventions, instrument calibration, sample processing time, and event acquisition thresholds can alter the distribution of high-dimensional cytometry data. As a result, an algorithm trained on one institutional panel may not generalize to another laboratory without harmonization and local validation. Ground-truth labels may also vary because expert gating itself is partly subjective, especially for rare populations, regenerating marrow, therapy-associated changes, and low-level MRD-like events [22].

Cutting-edge deep learning models are now extending beyond diagnosis to predicting underlying cytogenetic and molecular abnormalities directly from raw flow cytometry data [23, 24]. For data exploration and visualization, dimensionality reduction techniques such as t-distributed Stochastic Neighbor Embedding and Uniform Manifold Approximation and Projection are frequently used alongside these analytical tools.

The primary benefits demonstrated in research studies include improved efficiency and standardization. For instance, a real-time machine learning system for AML detection reduced the median analysis time from upwards of 30 min for complex manual reviews to just 23.1 s per case [24]. This automation also enhances standardization by minimizing the inter-operator variability inherent in manual gating, a crucial factor for achieving the high sensitivity and reproducibility required for MRD monitoring [25]. Therefore, clinical readiness in flow cytometry should be defined by intended use. Automated clustering and visualization tools may be useful for exploratory analysis and standardization, whereas supervised AML detection or MRD-support systems require independent validation against expert reference standards, assessment of false-positive and false-negative patterns, and verification under local laboratory conditions before clinical reporting [26]. For flow cytometry, the validator–integrator role includes reviewing AI-suggested clusters or gates, confirming aberrant immunophenotypes, adjudicating discordant populations, and ensuring that AI outputs are consistent with morphology, cytochemistry, where used, cytogenetics, molecular data, treatment status, and MRD reporting standards. The hematopathologist remains responsible for deciding whether an AI-flagged population is diagnostically meaningful, artifactual, regenerative, or clinically indeterminate [11].

3.2 | Genomic and Molecular Landscapes: Machine Learning for Next-Generation Sequencing Interpretation and Multi-Omics

The integration of next-generation sequencing (NGS) data has transformed leukemia classification [1, 2], yet interpreting the vast genomic data it generates remains challenging. Machine learning models, such as random forests, support vector machines, and gradient boosting, are increasingly employed to analyze mutation profiles, copy number variations, and gene

expression data [27]. These models aid in refining leukemia classifications, identifying prognostic signatures by integrating multiple genetic features, and predicting treatment responses. In some retrospective studies, these machine learning models have shown improved prognostic performance; for example, a model integrating long noncoding RNA expression data with clinical variables improved outcome prediction in pediatric AML, achieving a C-index of 0.67 compared with 0.63 for standard clinical and molecular risk stratification alone [28]. However, these models face challenges with generalizability; a model trained on genomic data from one patient cohort or generated by a specific sequencing platform may not perform reliably on data from different populations or technologies without careful re-validation to mitigate the risk of dataset shift [29]. AI-assisted multi-omics integration combines genomics with transcriptomics, proteomics, clinical data, and other laboratory data, including AI-derived morphology or flow features [30, 31]. Although computationally demanding, these approaches offer a system-level understanding, potentially identifying complex interactions that drive leukemia pathogenesis or resistance and enabling more precise, personalized medicine strategies [6, 32]. The translational gap is particularly important for genomic and multi-omic AI because model performance may be affected by sequencing platform, panel design, depth of coverage, variant-calling pipeline, copy-number detection strategy, fusion-calling sensitivity, RNA quality, missing data, and population structure [33, 34]. Many models are derived from retrospective cohorts and may overfit to institution-specific testing strategies or treatment protocols. Under-representation of ancestry groups [35], pediatric or elderly patients, therapy-related leukemia, rare genetic subtypes, and relapsed/refractory disease can also produce biased risk estimates [36]. Researchers are applying numerous machine learning algorithms to multi-omics datasets to define novel molecular subtypes based on specific biological pathways, such as neutrophil extracellular traps, and to construct robust prognostic models that can predict outcomes and response to immunotherapy [37]. At present, many genomic and multi-omic AI models are best regarded as decision-support or hypothesis-generating tools unless they have been prospectively evaluated, externally validated, calibrated for the target population, and shown to improve clinically meaningful outcomes beyond established classification and risk-stratification systems [38]. Other works focus on screening for prognostic and therapeutic biomarkers from diverse patterns of programmed cell death, using machine learning to build genetic signatures that can predict drug sensitivity and guide personalized therapy selection [29, 39]. In genomic and multi-omic interpretation, the validator–integrator role requires the clinician to evaluate variant pathogenicity, distinguish actionable alterations from variants of uncertain significance, interpret AI-derived risk predictions within World Health Organization/International Consensus Classification and European LeukemiaNet frameworks, and integrate molecular findings with morphology, flow cytometry, cytogenetics, clinical history, treatment exposure, and therapeutic intent [40].

To clarify the translational gap between experimental performance and deployable clinical use, Table 1 summarizes the evidence maturity and clinical-readiness considerations for representative AI applications across the leukemia diagnostic workflow.

3.3 | Emerging Computational Frontiers: Quantum-Inspired and Quantum Machine Learning in Leukemia Diagnostics

Beyond conventional machine learning and deep learning, quantum-inspired and quantum machine learning approaches are emerging as experimental computational strategies for hematological image analysis and precision oncology [43]. These methods should be distinguished carefully. Quantum-inspired algorithms are classical computational methods that borrow concepts from quantum mechanics, such as probabilistic state representation or quantum-inspired optimization, but run on classical hardware. In contrast, quantum machine learning algorithms are designed to use quantum circuits or quantum hardware, usually within hybrid quantum–classical pipelines [44]. This distinction is important because many currently reported "quantum" approaches in leukemia detection are quantum-inspired feature-selection methods rather than clinically deployable quantum-computing systems.

In leukemia-related image analysis, the most immediate relevance of these approaches is dimensionality reduction and feature selection. For example, Ahmad et al. proposed a transfer-learning pipeline for leukocyte subtype classification in which deep features extracted from White Blood Cell (WBC) images were reduced using a customized quantum-inspired evolutionary algorithm before classification. The study used a public dataset of 5000 WBC images and reported high classification accuracy with a substantial reduction in feature-vector size, suggesting that quantum-inspired optimization may help address high-dimensional feature spaces in hematological image analysis. However, such approaches remain computational decision-support methods and should not be interpreted as autonomous leukemia diagnostic systems [43].

True quantum machine learning approaches, including variational quantum circuits, quantum neural networks, and quantum kernel methods, are theoretically attractive because they may encode data into high-dimensional quantum feature spaces and potentially improve learning efficiency in settings where annotated datasets are small [45]. This could be relevant to rare leukemia subtypes, limited expert annotations, or multi-omic feature integration. A systematic review of quantum machine learning for digital health found that most studies lacked sufficient technical rigor, few evaluated realistic quantum operating conditions, and performance comparisons did not show a consistent empirical advantage over classical machine learning [46].

The translational barriers are substantial. Current noisy intermediate-scale quantum devices are limited by qubit number, decoherence, gate errors, circuit depth constraints, data-encoding bottlenecks [47], and difficulty scaling from reduced feature vectors to full-resolution clinical images or multi-omic datasets [48]. Many studies rely on simulators or small public datasets, and few include external validation, prospective workflow evaluation, clinically representative cohorts, or direct comparison against optimized classical baselines [45]. Therefore, quantum and quantum-inspired methods should currently be categorized as proof-of-concept or early experimental technologies within leukemia diagnostics [49].

TABLE 1 | Clinical readiness of AI tools in leukemia diagnostics.

Diagnostic domain	AI task	Evidence maturity	Current limitations	Clinical-readiness status	Validator–integrator role
Peripheral blood digital morphology	Automated WBC differential, blast flagging, region-of-interest selection	Strongest for bounded morphology-support tasks; some platforms have prospective or workflow-level evaluation for peripheral blood review [14]	Staining variability, scanner differences, smear quality, artifact detection, and limited rare-subtype representation	Potentially deployable as assistive morphology support when used within a defined intended use and locally validated; not a standalone leukemia diagnostic tool	Verify blast count, review flagged cells/regions, distinguish artifacts/reactive cells from abnormal cells, and correlate with clinical and laboratory data
Bone marrow aspirate and biopsy morphology	Blast estimation, dysplasia recognition, abnormal-cell detection, marrow pattern analysis	Mostly proof-of-concept or retrospective validation; fewer prospective real-world studies [41]	Aspirate quality, hemodilution, crush artifact, biopsy architecture, inconsistent annotation, limited external datasets	Experimental to retrospectively validated; requires external and prospective validation before routine deployment	Confirm cellular identity, assess architecture and dysplasia, reconcile aspirate/biopsy findings with flow and genomics
Flow cytometry	Automated clustering, gating support, and aberrant population detection	Established computational methods; increasing clinical validation for selected AML detection tasks	Antibody panel differences, compensation, instrument calibration, batch effects, subjective ground truth, MRD sensitivity requirements	Locally useful as decision support after analytical validation; clinical readiness depends on intended use and panel harmonization	Review AI-generated clusters/gates, adjudicate aberrant populations, verify MRD-like events, correlate with morphology and treatment status
MRD assessment	Rare-event detection and abnormal-population monitoring	Variable; strongest where standardized panels and reference methods exist	Very low event frequency, false positives from regenerating marrow, therapy effects, sample quality, threshold selection	Requires strict validation, quality control, and human sign-off; not generalizable without standardized protocols	Confirm the clinical significance of flagged rare events and determine whether findings meet reporting thresholds
NGS and molecular interpretation	Variant prioritization, fusion detection support, risk prediction	Retrospective and externally validated models exist for selected tasks; performance varies by platform and cohort [42]	Panel design, sequencing depth, variant-calling pipeline, VUS interpretation, and changing classifications	Decision support only unless validated for the exact intended clinical use	Interpret AI-ranked variants within the WHO/ICC, ELN, cytogenetic, morphologic, and clinical context
Multi-omics integration	Prognostic modeling, subtype discovery, treatment-response prediction	Mostly exploratory or retrospective; limited prospective utility data [38]	Small cohorts, missing data, overfitting, population bias, interpretability, and difficult clinical implementation	Proof-of-concept to retrospectively validated; not routinely deployable without prospective validation	Judge biological plausibility, assess calibration, and integrate predictions with established risk systems and patient-specific management

(Continues)

TABLE 1 | (Continued)

Diagnostic domain	AI task	Evidence maturity	Current limitations	Clinical-readiness status	Validator–integrator role
Quantum-inspired and quantum machine learning	Feature selection, dimensionality reduction, experimental WBC image classification, exploratory multi-omic or drug-response modeling	Mostly proof-of-concept; strongest current evidence is quantum-inspired feature selection on public image datasets; true QML largely simulator-based or early feasibility work [43]	Limited qubits, NISQ noise, decoherence, data-encoding bottlenecks, simulator dependence, small datasets, limited external validation, unclear advantage over optimized classical models	Experimental; not clinically deployable for leukemia diagnosis; useful mainly as a future research frontier	Assess whether feature compression preserves diagnostic meaning, verify outputs against classical baselines and clinical standards, and integrate only within validated intended use

For the pathologist or hematologist, the validator–integrator role remains central even if such approaches mature. The specialist would need to assess whether feature compression or quantum encoding preserves diagnostically meaningful morphology, whether model outputs are robust across staining protocols and instruments, whether predictions generalize across patient populations and leukemia subtypes, and whether any claimed quantum advantage improves clinical decision-making beyond established morphology, flow cytometry, cytogenetic, molecular, WHO/ICC, and ELN frameworks. Thus, quantum machine learning is best framed as a future-facing extension of AI-assisted hematopathology, reinforcing rather than replacing the need for expert validation and clinical integration [50].

3.4 | Translational Gap and Technical Barriers to Clinical Deployment

The domain-specific examples above illustrate a broader implementation problem: clinical deployment requires evidence that models remain reliable outside the curated datasets in which they were developed. In leukemia diagnostics, this gap is not simply a matter of improving algorithmic accuracy; it is a translational challenge involving intended use, external validity, workflow integration, human oversight, regulatory readiness, and post-deployment monitoring. A model that performs well in retrospective image sets, annotated flow cytometry files, or selected molecular cohorts may still be unsuitable for clinical use if its intended role, target population, failure modes, and impact on diagnostic decisions have not been clearly defined [51, 52].

A practical distinction should therefore be made between technical validation, clinical validation, and clinical utility. Technical validation asks whether the model produces reproducible outputs on representative input data. Clinical validation asks whether those outputs remain accurate across independent institutions, patient populations, disease spectra, instruments, software versions, and laboratory workflows. Clinical utility asks whether implementation improves meaningful outcomes, such as diagnostic turnaround time, reproducibility, detection of clinically relevant abnormalities, prioritization of urgent cases, or reduction of avoidable error. These levels of evidence should not be conflated. High AUC, sensitivity, specificity, F1-score, or C-index in a retrospective study may support further evaluation, but such metrics alone do not establish readiness for routine diagnostic reporting [53].

The first requirement for translation is a clearly defined intended use. AI systems used for screening, triage, differential counting, abnormal-population flagging, MRD support, variant prioritization, or integrated risk prediction carry different clinical risks and therefore require different validation thresholds [14, 52]. A low-risk tool that prioritizes slides for expert review is not equivalent to a system whose output influences leukemia classification, MRD interpretation, or treatment stratification. For each intended use, validation should specify the acceptable input material, minimum specimen quality, expected user, reporting context, confidence thresholds, escalation rules, and circumstances in which the model should defer to manual review [51, 52].

A second requirement is representative external validation. Many AI studies in hematopathology are vulnerable to spectrum bias because they are trained and tested on selected cases, enriched disease cohorts, public datasets, or technically optimal specimens [54]. Real-world laboratories encounter borderline cases, mixed phenotypes, post-treatment specimens, therapy-related changes, rare entities, pediatric and adult populations, and variable disease prevalence. Therefore, validation cohorts should include diagnostically difficult and clinically consequential cases, not only clear examples [55]. Performance should also be reported across relevant subgroups and operating conditions because pooled accuracy can conceal underperformance in rare leukemia subtypes, low-prevalence settings, or underrepresented patient groups [56].

A third requirement is evaluation of calibration and decision impact, not only discrimination [57]. For clinical decision support, the question is not merely whether an AI model ranks cases correctly, but whether its probability estimates are trustworthy enough to guide action [58]. Poor calibration can lead to inappropriate confidence in uncertain outputs, especially when disease prevalence differs between the development cohort and the deployment laboratory. Decision thresholds should therefore be justified clinically, and validation should include false-positive and false-negative review, discordance analysis, and assessment of how AI outputs change human decisions [52, 59]. In laboratory hematology, this is particularly important because even small errors in blast recognition, abnormal-population interpretation, or molecular-risk assignment can alter downstream diagnostic classification and management.

A fourth requirement is prospective or workflow-level evaluation. Before full deployment, promising systems can be tested in a silent or shadow mode, in which AI outputs are generated but do not affect reporting. This allows laboratories to measure performance on local cases, identify failure modes, evaluate turnaround time, and determine whether the interface supports or disrupts routine practice [45]. Subsequent controlled implementation should include predefined rules for human review, audit of discordant cases, documentation of overrides, and periodic reassessment. Such staged evaluation aligns with emerging AI reporting and evaluation frameworks, including DECIDE-AI for early-stage clinical evaluation, CONSORT-AI and SPIRIT-AI for clinical trials and protocols, STARD-AI for diagnostic accuracy studies, and TRIPOD+AI for prediction models [51, 52, 55, 60, 61].

A fifth requirement is management of human–AI interaction. AI tools may introduce automation bias, alert fatigue, over-reliance on visually persuasive outputs, or inappropriate dismissal of low-confidence results [62]. Conversely, poorly designed interfaces may increase workload by generating excessive flags or poorly contextualized recommendations [63]. Validation should therefore include the end user, not only the algorithm [64]. The pathologist or hematologist must be able to understand the model's intended use, confidence level, limitations, and reasons for flagging a case [65]. Explainability tools may be helpful, but they should be treated as aids to error analysis and trust calibration rather than proof that a model is correct [66].

A sixth requirement is lifecycle governance. AI systems are not static laboratory instruments; their performance may change with software updates, altered preprocessing pipelines, new scanners or cytometers, revised antibody panels, updated sequencing workflows, changing case mix, or shifts in disease classification. Implementation should therefore include version control, change-management procedures, cybersecurity safeguards, data-quality checks, and post-deployment monitoring for performance drift [67]. Trustworthy AI guidance emphasizes that deployable healthcare AI should be assessed across the full lifecycle, including fairness, universality, traceability, usability, robustness, and explainability [68].

These considerations define the practical translational gap in AI-assisted leukemia diagnostics. The question is not whether AI can identify patterns in hematopathology data; the literature already shows that it can. The clinically relevant question is whether an AI system can perform a defined task safely, reproducibly, transparently, and usefully within a specific laboratory workflow. Until that evidence is available, most AI tools should be regarded as decision-support systems requiring specialist supervision rather than autonomous diagnostic systems. This reinforces the validator–integrator model: the pathologist or hematologist remains responsible for judging technical adequacy, resolving discordance, integrating multimodal evidence, and determining whether an AI-derived output is appropriate for clinical reporting [42, 68, 69].

4 | The Evolving Pathologist/Hematologist in Leukemia Care: Interpreter, Integrator, and Validator

As AI tools demonstrate increasing proficiency, the role of the pathologist/hematologist evolves. AI is becoming an indispensable assistant for high-volume or computationally intensive tasks [4], although successful workflow integration (requiring laboratory information systems compatibility, middleware, data management strategies, etc.) remains a practical hurdle [70]. Automation frees clinicians for higher-level cognitive tasks. The shift in the role is fundamentally toward becoming the chief interpreter, integrator, and validator of all diagnostic information.

Expert oversight and validation are crucial while using these AI tools. Clinicians must thoroughly assess AI outputs, comprehending the specific algorithms employed (e.g., the architecture of a CNN or the parameters of a clustering algorithm), their performance characteristics (sensitivity, specificity, and area under the curve from validation studies), inherent limitations (e.g., performance on out-of-distribution data and sensitivity to artifacts), and compliance with regulatory standards (FDA/EMA requirements) [69]. It is essential to critically evaluate potential biases stemming from training datasets. Addressing the "black box" issue through explainable AI (XAI) methods (e.g., LIME, SHAP, and saliency maps) is vital for building trust, facilitating error analysis, and enabling meaningful assessment of why an AI made a specific prediction [71].

Rigorous validation is paramount because algorithmic errors in leukemia diagnostics carry direct and severe consequences for patient safety. The gap between performance in retrospective

studies and that in real-world clinical practice remains a significant concern, as very few AI-assisted diagnostic tools have been prospectively tested in live clinical environments [72]. A false-negative result, for example, an AI system failing to detect MRD post-treatment, could lead to premature cessation of therapy and subsequent patient relapse. Conversely, a false-positive finding, such as the misclassification of reactive lymphocytes as blasts, could result in unnecessary and toxic chemotherapy, invasive bone marrow biopsies, and immense psychological distress for the patient and their family [73]. Robust mitigation strategies include a mandatory final review and sign-off by a qualified pathologist for all AI-generated results, and the establishment of high-confidence thresholds below which AI findings are flagged for mandatory manual review [11, 50].

Human expertise remains essential for navigating complexity and ambiguity where algorithms might struggle, such as with rare subtypes, subtle morphology, or discordant results. The clinical integration of AI-assisted findings with the complete patient context for accurate classification (WHO/ICC), risk stratification (European LeukemiaNet), and management guidance is a role for humans to perform distinctly [2]. Effectively communicating these integrated findings and overseeing ethical considerations such as privacy, fairness, transparency, and equity are also fundamental responsibilities [73]. The future leukemia specialist will harness specific AI tools while focusing their expertise on critical interpretation, complex problem-solving, holistic integration, communication, and ethical stewardship.

5 | Illustrative Workflow: Human–AI Collaboration in Practice

To conceptualize the synergistic integration of AI and clinical expertise, consider an illustrative workflow for diagnosing acute leukemia (Figure 1). Although specifics will vary depending on institutional resources and case complexity, the following framework exemplifies how AI and human experts could collaborate to enhance diagnostic accuracy and efficiency. Each AI module should be understood as operating behind a validation gate. Before clinical use, the tool must be validated for its intended task, specimen type, instrument or platform, and reporting context. During use, the pathologist/hematologist reviews AI outputs at predefined checkpoints: morphology flags, flow cytometry clusters, MRD-like events, variant-prioritization outputs, and integrated risk predictions. Discordant or low-confidence outputs should trigger manual review, repeat analysis, or multidisciplinary discussion rather than automatic reporting.

The diagnostic process begins with standard sample collection, gathering peripheral blood and bone marrow aspirates, from which morphological smears, flow cytometry samples, and material for molecular and cytogenetic analyses are prepared. Digital WSI transforms conventional microscopy slides into datasets compatible with computational analysis. Concurrently, multi-parametric flow cytometry data and sequencing data from NGS are generated.

In morphological analysis, AI software employing CNNs evaluates digitized slides, automating tasks such as differential cell counting, quantification of blast percentages, and the detection of dysplastic changes or rare suspicious cells. The pathologist or hematologist then reviews the computational results, selectively verifying flagged areas of interest through targeted microscopic evaluation. Here, AI significantly accelerates routine quantification, while the clinician ensures accuracy, clinical relevance, and detailed interpretation.

Flow cytometry data analysis similarly benefits from AI tools, such as automated clustering algorithms that rapidly categorize complex cell populations and highlight potentially aberrant immunophenotypes or MRD signatures. The clinician reviews these automated classifications, scrutinizing proposed gates and populations. In cases of complexities or uncertainties, manual refinement of these automated analyses ensures alignment with morphological findings and clinical context. AI facilitates consistent and rapid initial analysis, while the clinician interprets nuances, validating findings and establishing clinical significance.

AI-assisted bioinformatics pipelines further support the analysis of genomic and cytogenetic data. Machine learning algorithms identify pathogenic mutations, fusion transcripts, and cytogenetic alterations, assisting clinicians by streamlining large-scale genomic data interpretation and initial prognostic stratification. However, the clinician remains integral, evaluating the clinical implications of identified variants, particularly variants of uncertain significance (VUS), and integrating molecular insights with traditional cytogenetic analyses.

6 | Training the Next Generation: Preparing for AI-Assisted Practice

Medical education systems worldwide are recognizing the need to develop the competencies of future pathologists and hematopathologists in AI. In 2025, an international consensus (the DECODE framework) defined broad digital health domains, including health data science and 19 core competencies for physicians, urging medical schools across the world to incorporate these into their curricula [74]. These competencies align with skills crucial for AI-assisted hematopathology, such as data science fundamentals, the ability to interpret AI outputs, an understanding of algorithmic limitations, and ethical/legal literacy in pathology-specific training [75]. A Dutch initiative in 2024 reviewed pathology training entrustable professional activities (EPAs) and concluded that while traditional diagnostic skills remain essential, new competencies in technology evaluation, AI implementation, and awareness of the limits in human–AI interactions are needed [76]. It even proposed a dedicated EPA "using AI in diagnostic pathology practice" with associated competencies to ensure pathologists are "AI-ready" upon completing residency [76].

6.1 | Operationalizing the Validator–Integrator Role in Training

Existing digital health frameworks provide a broad foundation for curriculum design, but AI-assisted hematopathology requires specialty-specific translation into observable diagnostic

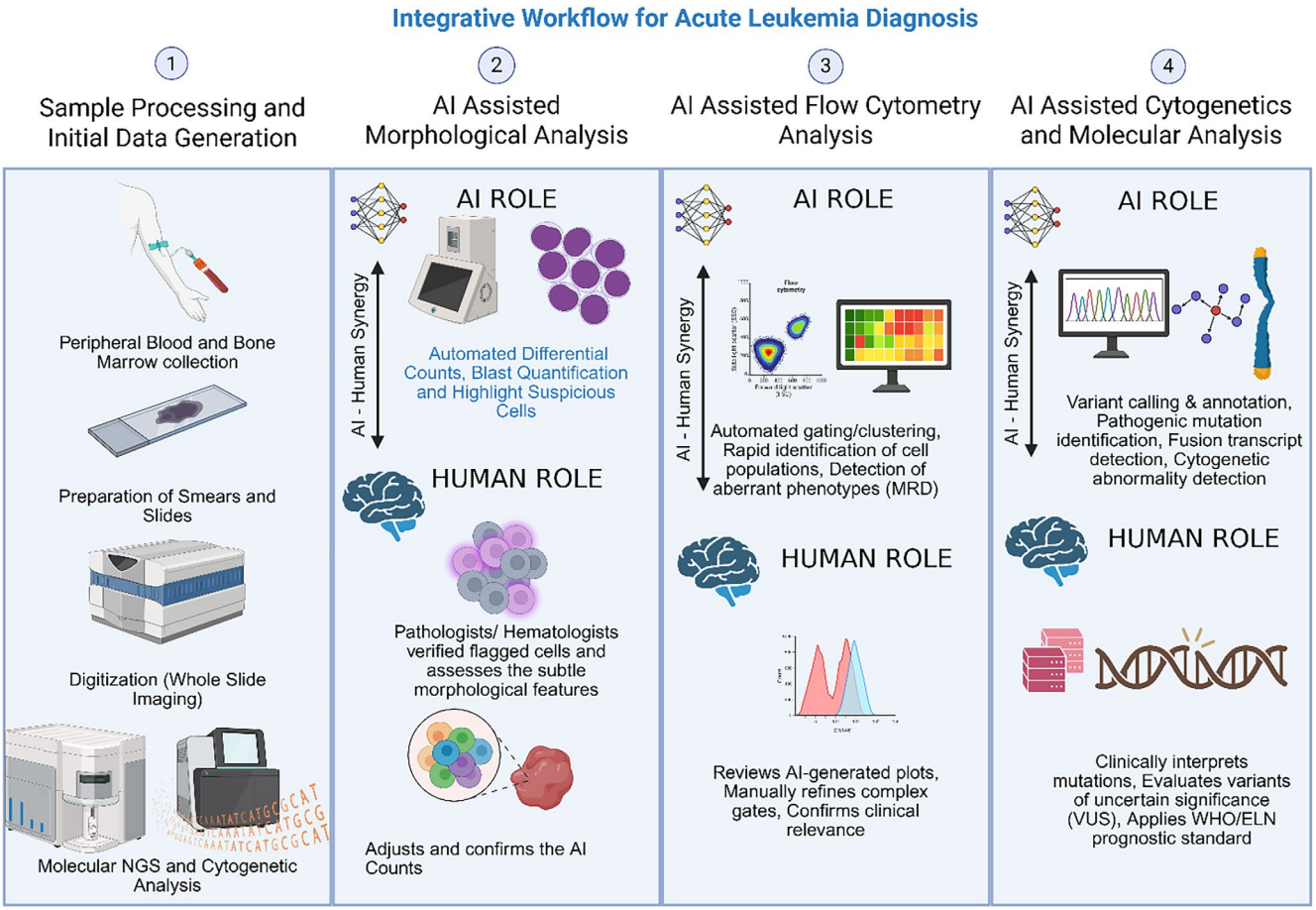

FIGURE 1 | Integrative workflow for acute leukemia diagnosis illustrating synergistic interactions between artificial intelligence (AI) and clinical expertise (Steps 1–4). (Created in BioRender. Das, S. (2025) https://BioRender.com/t4rg3o5). AI, artificial intelligence; ELN, European LeukemiaNet; MRD, minimal residual disease; NGS, next-generation sequencing; WHO, World Health Organization; WSI, whole-slide imaging.

behaviors. The objective is not for every trainee to become an AI developer, but for all trainees to become safe and critical users of AI outputs, while selected trainees may develop more advanced roles as AI "translators" or "developers" who bridge clinical practice, data science, and implementation [77]. The recently proposed pathology-specific EPA for "using AI in diagnostic pathology practice" provides a useful entrustment framework, but in leukemia diagnostics, this EPA should be operationalized around the validator–integrator tasks performed during routine diagnostic work [78]. The Digital Health Competencies in Medical Education framework similarly supports systematic curriculum design, but its broad domains must be adapted to the diagnostic realities of morphology, flow cytometry, molecular pathology, laboratory governance, and multidisciplinary leukemia care [79]. The DECODE framework was developed through an international Delphi process and identified 19 competencies across four digital health domains, making it useful as a global scaffold rather than a leukemia-specific curriculum [74].

In practical terms, trainees should learn a structured sequence: first, define the intended use of the AI tool; second, assess the adequacy of the input data or specimen; third, interpret the AI output and confidence indicators; fourth, compare the output with independent human assessment and other diagnostic modalities; fifth, adjudicate discordant results; and sixth, document the final interpretation, including the role and limitation of AI assistance [41]. This sequence directly mirrors the clinical responsibilities of the hematopathologist as validator–integrator. For example, in digital morphology, a trainee could first perform a manual blast estimate, then review AI-flagged cells or regions, identify false positives due to artifacts or reactive cells, and reconcile the final count with flow cytometry and clinical findings [14]. In flow cytometry, the trainee could review automated clusters or gates, compare them with manual gating, determine whether flagged populations are aberrant, regenerative, therapy-related, or indeterminate, and document the rationale for accepting or rejecting the AI output [40, 80]. In molecular pathology, the trainee could assess whether AI-prioritized variants or risk predictions are consistent with WHO/ICC classification, ELN risk stratification, cytogenetic findings, and the clinical scenario [34, 40].

Training should therefore be embedded into routine diagnostic activities rather than delivered only as lectures. AI-focused cases can be incorporated into morphology sign-out, flow cytometry review, molecular tumor boards, multidisciplinary leukemia meetings, quality assurance conferences, and journal clubs [78]. This approach allows trainees to learn AI supervision in the same clinical contexts in which they will eventually use these tools. Importantly, early exposure should occur in a supervised or simulated environment, including "shadow mode" exercises in which AI outputs are reviewed for educational purposes without influencing patient reports. Such exercises teach trainees not

only when AI is helpful but also when it should be questioned, overridden, or escalated for senior review [52, 77, 79].

6.2 | Assessment and Entrustment of AI Competence

Because the validator–integrator role is a clinical responsibility, assessment should focus on observable performance rather than completion of didactic content alone [81]. A trainee should not be considered competent simply because they can operate AI software or describe machine-learning terminology. Competence should require demonstration that the trainee can interpret AI outputs within the tool's intended use, recognize technical and clinical failure modes, identify when outputs are out-of-distribution, compare AI findings with independent diagnostic evidence, and communicate uncertainty appropriately. This is consistent with the EPA-based approach proposed for AI in diagnostic pathology, where the endpoint is safe participation in AI-assisted practice rather than passive awareness of AI concepts [78]. Vos et al. specifically argue that pathology training requires new competencies in technology evaluation, AI implementation, human–AI interaction limits, ethical/legal awareness, and interpretation of AI results in individual patient context [76].

Assessment can be staged according to the level of training. Junior trainees should demonstrate foundational AI literacy, including understanding supervised versus unsupervised learning, training/validation/test datasets, overfitting, dataset shift, and basic performance metrics [77]. More advanced trainees should be assessed on case-based validation tasks, such as comparing AI-assisted differential counts with manual counts, reviewing automated flow cytometry clusters, evaluating an AI-generated MRD flag, or interpreting AI-prioritized genomic findings [79]. Hematopathology fellows should additionally be expected to design or critique local validation plans, assess subgroup performance, review discordant cases, and explain how AI outputs should be documented in a clinical report [52].

Critical appraisal should be taught explicitly. AI-focused journal clubs should require trainees to evaluate whether a study clearly defines intended use, dataset source and size, ground-truth labeling, separation of training/validation/test cohorts, external validation, subgroup performance, calibration, false-positive and false-negative patterns, and clinical utility [82, 83]. Reporting frameworks such as STARD-AI and TRIPOD+AI can be used as structured appraisal tools for diagnostic accuracy and prediction-model studies, while DECIDE-AI, CONSORT-AI, and SPIRIT-AI can help trainees understand staged clinical evaluation and trial reporting for AI interventions. This ensures that trainees learn not only how AI systems perform under ideal research conditions, but also how to judge whether an algorithm is ready for local clinical use [84].

Entrustment decisions can then be linked to progressively independent responsibilities. At an early level, the trainee observes AI-assisted review and explains the tool's intended use. At an intermediate level, the trainee interprets AI outputs under direct supervision and identifies common failure modes. At an advanced level, the trainee adjudicates discordant AI-human findings, documents limitations, and proposes local validation or quality assurance steps. At the highest level, the trainee can supervise others, contribute to implementation decisions, and participate in post-deployment monitoring. Table 2 translates these expectations into practical competencies, leukemia-specific use cases, assessment methods, and expected trainee levels [85].

To translate the validator–integrator concept into training practice, AI education should be competency-based rather than limited to general awareness. Trainees should learn not only what AI tools can do but also how to validate outputs, recognize failure modes, appraise evidence, document human oversight, and integrate AI-derived findings with morphology, flow cytometry, genomics, and clinical data. These competencies can be implemented without creating an entirely separate curriculum. AI-focused cases can be embedded into routine morphology sign-out, flow cytometry review, molecular pathology conferences, journal clubs, quality assurance meetings, and multidisciplinary leukemia boards. This approach preserves traditional diagnostic training while preparing future specialists to supervise AI-assisted workflows safely.

6.3 | Implementation Strategies in Training Programs

Medical educators worldwide are experimenting with how best to integrate AI training into pathology and hematopathology programs. A common strategy is to embed AI topics into existing curricula through interdisciplinary learning. In North America, for example, the *Pathology Informatics Essentials for Residents* provides a flexible informatics curriculum linked to ACGME milestones and adaptable to different training-program resources [75, 79]. Some residency programs offer focused workshops or electives in computational pathology; institutions like Stanford University and the Mayo Clinic have even organized dedicated AI workshops for pathology trainees, where residents practice interpreting AI-generated analyses and discuss real-world cases of AI application and ethics [75]. These practical sessions function much like simulation exercises, allowing trainees to engage with AI tools in a controlled environment, learning how to integrate AI findings with their own microscopic evaluations without risking patient care.

Pathology training programs in Europe are also incorporating AI through blended learning and simulation. Many departments use virtual microscopy platforms and AI software on digitized blood smear or bone marrow slide cases to simulate how an algorithm might flag blast cells or classify leukemias, thereby teaching hematopathology students how to verify or question an AI's suggestions [86]. Such simulation-based training builds confidence in using AI as a "second reader" in diagnostics. European initiatives also emphasize interdisciplinary collaboration, for instance, pathology residents attending joint seminars with data science students or bioinformatics teams to learn AI development processes, fostering mutual understanding [75]. By learning alongside data experts, pathology trainees gain practical skills in data management (such as curation of digital pathology images) and algorithm validation, which are particularly relevant in hematopathology, where large datasets

TABLE 2 | Entrustable AI competencies for hematopathology training: Leukemia-specific use cases and assessment methods.

Competency	Use case in leukemia diagnostics	Assessment method	Level expected of trainee
Foundational AI literacy	Explain supervised learning, unsupervised clustering, CNNs, training/validation/test sets, overfitting, and dataset shift using examples from morphology and flow cytometry.	Short-answer exercise or oral viva using a published AI study	Core
Critical appraisal of AI studies	Evaluate an AI blast-detection or AML-classification study for dataset size, study design, external validation, confidence intervals, and bias.	AI-focused journal club using a structured checklist such as STARD-AI/TRIPOD + AI principles	Core
AI-assisted morphology validation	Compare the AI-generated blast count or flagged cells with the manual differential count and expert review.	Slide-based simulation using digital peripheral blood or bone marrow cases	Core
Recognition of AI failure modes	Identify false positives from artifacts, smudge cells, reactive lymphocytes, poor staining, or hemodiluted marrow.	Case-based assessment with deliberately challenging images	Core
AI-assisted flow cytometry interpretation	Review automated clusters/gates, identify aberrant populations, and compare AI output with manual gating.	Flow cytometry case simulation with pre/post-AI interpretation	Core
MRD-support interpretation	Assess whether an AI-flagged rare population is compatible with MRD, regeneration, artifact, or nonspecific background.	MRD case conference with threshold and discordance discussion	Advanced/core for hematopathology fellows
Molecular and genomic AI interpretation	Interpret AI-prioritized variants or risk predictions within the WHO/ICC and ELN frameworks.	Molecular tumor board or hematopathology sign-out simulation	Core
Bias and generalizability assessment	Evaluate whether an AI model may underperform across institutions, scanners, staining protocols, antibody panels, sequencing platforms, age groups, ancestry groups, or rare subtypes.	Structured audit of the model validation paper or local validation dataset	Core
Local validation and implementation	Design a local verification plan before using an AI tool for morphology, flow, or NGS decision support.	Resident-led quality improvement or validation protocol assignment	Advanced
Regulatory and governance literacy	Explain intended use, version control, human oversight, change management, post-deployment monitoring, and documentation requirements.	OSCE-style scenario involving the implementation of a new AI tool	Core/advanced
Communication and documentation	Write a report comment describing AI-assisted analysis, human verification, limitations, and final sign-off.	Report-writing exercise using AI-assisted diagnostic cases	Core
Interdisciplinary collaboration	Work with laboratory staff, bioinformaticians, data scientists, and clinicians to resolve AI-output discordance.	Multidisciplinary case presentation	Core

(e.g., flow cytometry and genomic data) increasingly inform diagnoses [87].

Beyond residency, competency-based workshops and continuing classes are provided to keep practicing pathologists up to date. Professional societies in North America and Europe (e.g., the College of American Pathologists and the European Society of Pathology) have begun offering short courses on AI in diagnostics, covering topics such as AI result interpretation, regulatory guidelines, and troubleshooting of AI tools [84]. These efforts ensure that the learning continuum spans from medical school to fellowship and beyond. Notably, many medical schools

are introducing core AI concepts to students early. Surveys indicate, however, that there is still much room for growth—one multi-country study found that while approximately 69% of students were aware of AI in medicine, only 14% learned about it through formal medical school instruction (most learned from independent online sources) [87]. In response, schools in Asia-Pacific and North America have started pilot courses on medical AI for undergraduates, often using case studies in radiology or pathology as teaching examples. For instance, several Chinese medical universities recently added formal AI modules after a nationwide survey showed that over half of the medical students and postgraduates had used AI tools and desired more structured training [88]. Likewise, initiatives are emerging in South Asia and the Middle East. A survey of Jordanian medical students and pathology residents strongly concluded that "AI education should be introduced into medical curricula" to cover applications, pitfalls, and legal aspects [87].

6.4 | Challenges and Considerations in Global Implementation

Implementing AI training in hematopathology on a global scale poses significant challenges. Faculty development and resource limitations are primary concerns. Many regions face a shortage of educators fluent in both pathology and AI, making it difficult to teach new content. According to an international perspective, introducing AI into curricula depends heavily on each institution's resources and faculty expertise; a shortage of trained faculty and high costs can "exacerbate global inconsistency in medical education quality" [89]. Another challenge is regulatory and regional diversity. The rules governing clinical AI use vary widely; for example, the U.S. FDA, European CE marking, and other national regulators have different approval pathways and post-market requirements for AI-assisted diagnostic tools. Hence, training programs must navigate these differences [75]. Modern AI in hematopathology often relies on large datasets of annotated images (e.g., bone marrow aspirate slides and lymph node histology) and other multi-modal data. However, not all regions have the same digital pathology infrastructure or data-sharing practices. Inconsistent slide scanning technologies, variable staining protocols, and fragmented data silos can impede the development and adoption of AI and complicate how we train people to use it. A recent review on preparing pathology data for AI noted that reproducing AI performance across institutions hinges on having *standardized, high-quality data* and consistent annotations, suggesting that "data standardization and representative data across multiple centers" are key to robust AI implementation [84]. In educational terms, this means trainees should learn the importance of data quality and harmonization. Some global initiatives (e.g., cross-border digital pathology networks and the use of common data formats like digital images and communications in medicine (DICOM) for whole-slide images) have begun to address these gaps, but adoption is uneven. Thus, programs must also instill adaptability—pathologists should be prepared to handle AI tools that might work well in one hospital's system but require re-validation in another [90].

Besides these challenges, there are ubiquitous concerns about ethical use, bias, and validation that every program must cover. Ensuring that trainees appreciate issues of algorithmic bias (for instance, how an AI might underperform with underrepresented populations or rare leukemia subtypes) is essential for safe clinical deployment. Accordingly, competency frameworks universally call for teaching the limits of AI—that it is a tool to assist, not replace, the expert judgment of a hematopathologist [76]. Developing this understanding is arguably the capstone of *clinical readiness*.

Key strategies for embedding AI competencies into postgraduate training are summarized in Table 3.

7 | Conclusion and Future Perspectives

The integration of AI in leukemia diagnostics marks a pivotal shift toward a powerful human–AI synergy, not the obsolescence of the hematopathologist. In this model, the specialist remains indispensable as the expert validator, data integrator, and ethical steward who provides ultimate clinical context and decision-making authority. In the future, this synergy will deepen as AI models move beyond initial diagnosis to performing longitudinal data integration for MRD assessment, potentially predicting relapse over time. Furthermore, AI shows immense promise for guiding personalized therapy by using machine learning to predict drug sensitivity and streamline clinical trial matching, translating complex multi-omic data into actionable treatment strategies.

The most immediate future of AI in leukemia diagnostics is therefore not autonomous diagnosis but clinically governed augmentation. AI tools will be most useful when their intended use is clearly defined, their performance is externally and locally validated, their limitations are understood, and their outputs are reviewed by specialists who can integrate morphology, flow cytometry, cytogenetics, genomics, clinical context, and established classification systems. The validator–integrator model provides a

TABLE 3 | Embedding AI competencies into routine hematopathology training.

Integration method	Target competencies	Examples
Dedicated AI Modules	Foundational knowledge, ethical conduct	Interactive lectures, online resources
Integrated Case Discussions	Application, interpretation skills	AI-assisted digital pathology analysis
AI-Focused Journal Clubs	Critical appraisal, foundational knowledge	Structured appraisal exercises
Hands-On Workshops	Application-specific skills, interpretation	AI software practical sessions
Research/Audit Projects	All competencies	Resident-led AI validation projects
Faculty Development	AI teaching and pedagogical strategies	Workshops, train-the-trainer programs

practical framework for this transition: AI can improve efficiency and reproducibility, but the pathologist/hematologist remains responsible for diagnostic validity, clinical interpretation, ethical oversight, and final patient-centered decision-making.

Author Contributions

Kusal K. Das: conceptualization, supervision, validation, and writing – review and editing. **Sayandeep K. Das:** investigation, writing – original draft, and visualization.

Acknowledgements

The authors have nothing to report.

Funding

The authors have nothing to report.

Ethics Statement

The authors have nothing to report.

Consent

The authors have nothing to report.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

References

1. D. A. Arber, A. Orazi, R. P. Hasserjian, et al., "International Consensus Classification of Myeloid Neoplasms and Acute Leukemias: Integrating Morphologic, Clinical, and Genomic Data," *Blood* 140 (2022): 1200–1228, https://doi.org/10.1182/BLOOD.2022015850.

2. H. Döhner, A. H. Wei, F. R. Appelbaum, et al., "Diagnosis and Management of AML in Adults: 2022 Recommendations From an International Expert Panel on Behalf of the ELN," *Blood* 140 (2022): 1345–1377, https://doi.org/10.1182/BLOOD.2022016867.

3. J. D. Khoury, E. Solary, O. Abla, et al., "The 5th Edition of the World Health Organization Classification of Haematolymphoid Tumours: Myeloid and Histiocytic/Dendritic Neoplasms," *Leukemia* 36 (2022): 1703–1719, https://doi.org/10.1038/S41375-022-01613-1.

4. E. J. Topol, "High-Performance Medicine: The Convergence of Human and Artificial Intelligence," *Nature Medicine* 25 (2019): 44–56, https://doi.org/10.1038/s41591-018-0300-7.

5. C. Matek, S. Schwarz, K. Spiekermann, and C. Marr, "Human-Level Recognition of Blast Cells in Acute Myeloid Leukaemia With Convolutional Neural Networks," *Nature Machine Intelligence* 1 (2019): 538–544, https://doi.org/10.1038/s42256-019-0101-9.

6. R. F. Oybek Kizi, T. P. Theodore Armand, and H. C. Kim, "A Review of Deep Learning Techniques for Leukemia Cancer Classification Based on Blood Smear Images," *Applied Biosciences* 4 (2025): 9, https://doi.org/10.3390/APPLBIOSCI4010009.

7. J. Sukhera, "Narrative Reviews: Flexible, Rigorous, and Practical," *Journal of Graduate Medical Education* 14 (2022): 414–417, https://doi.org/10.4300/JGME-D-22-00480.1.

8. C. Baethge, S. Goldbeck-Wood, and S. Mertens, "SANRA—A Scale for the Quality Assessment of Narrative Review Articles," *Research Integrity and Peer Review* 4 (2019): 1–7.

9. M. J. Macawile, V. V. Quiñones, A. Ballado, J. Cruz, and M. V. Caya, "White Blood Cell Classification and Counting Using a Convolutional Neural Network," in *3rd International Conference on Control and Robotics Engineering, ICCRE 2018* (IEEE, 2018), 259–263, https://doi.org/10.1109/ICCRE.2018.8376476.

10. Z. Wang, A. S. Popel, S. Wangsiricharoen, C. Saoud, A. W. James, and J. Sulam, "Label Cleaning Multiple Instance Learning: Refining Coarse Annotations on Single Whole-Slide Images," *IEEE Transactions on Medical Imaging* 41 (2022): 3952–3968, https://doi.org/10.1109/tmi.2022.3202759.

11. H. H. Rashidi, J. Pantanowitz, M. G. Hanna, et al., "Introduction to Artificial Intelligence and Machine Learning in Pathology and Medicine: Generative and Nongenerative Artificial Intelligence Basics," *Modern Pathology* 38 (2025): 100688, https://doi.org/10.1016/J.MODPAT.2024.100688.

12. O. Islam, M. Assaduzzaman, and M. Z. Hasan, "An Explainable AI-Based Blood Cell Classification Using an Optimized Convolutional Neural Network," *Journal of Pathology Informatics* 15 (2024): 100389, https://doi.org/10.1016/J.JPI.2024.100389.

13. B. Z. Katz, M. D. Feldman, M. Tessema, et al., "Evaluation of Scopio Labs X100 Full Field PBS: The First High-Resolution Full Field Viewing of Peripheral Blood Specimens Combined With Artificial Intelligence-Based Morphological Analysis," *International Journal of Laboratory Hematology* 43 (2021): 1408–1416, https://doi.org/10.1111/IJLH.13681.

14. A. Kratz, S. H. Lee, G. Zini, J. A. Riedl, M. Hur, and S. Machin, "Digital Morphology Analyzers in Hematology: ICSH Review and Recommendations," *International Journal of Laboratory Hematology* 41 (2019): 437–447, https://doi.org/10.1111/IJLH.13042.

15. A. E. Aby, S. Salaji, K. K. Anilkumar, and T. Rajan, "A Review on Leukemia Detection and Classification Using Artificial Intelligence-Based Techniques," *Computers and Electrical Engineering* 118 (2024): 109446, https://doi.org/10.1016/J.COMPELECENG.2024.109446.

16. S. Park, Y. H. Park, J. Huh, S. M. Baik, and D. J. Park, "Deep Learning Model for Differentiating Acute Myeloid and Lymphoblastic Leukemia in Peripheral Blood Cell Images via Myeloblast and Lymphoblast Classification," *Digital Health* 10 (2024): 1–11.

17. M. Aria, Z. Javanmard, D. Pishdad, et al., "Towards Diagnostic Intelligent Systems in Leukemia Detection and Classification: A Systematic Review and Meta-Analysis," *Journal of Evidence-Based Medicine* 18 (2025): e70005, https://doi.org/10.1111/JEBM.70005.

18. H. Kim, M. Hur, G. d'Onofrio, and G. Zini, "Real-World Application of Digital Morphology Analyzers: Practical Issues and Challenges in Clinical Laboratories," *Diagnostics* 15 (2025): 677, https://doi.org/10.3390/DIAGNOSTICS15060677.

19. S. Van Gassen, B. Callebaut, M. J. Van Helden, et al., "FlowSOM: Using Self-Organizing Maps for Visualization and Interpretation of Cytometry Data," *Cytometry, Part A* 87 (2015): 636–645, https://doi.org/10.1002/CYTO.A.22625.

20. F. Mair, F. J. Hartmann, D. Mrdjen, V. Tosevski, C. Krieg, and B. Becher, "The End of Gating? An Introduction to Automated Analysis of High-Dimensional Cytometry Data," *European Journal of Immunology* 46 (2016): 34–43, https://doi.org/10.1002/EJI.201545774.

21. Q. Cai, H. Lan, D. Yi, et al., "Flow Cytometry in Acute Myeloid Leukemia and Detection of Minimal Residual Disease," *Clinica Chimica Acta* 564 (2025): 119945, https://doi.org/10.1016/J.CCA.2024.119945.

22. T. Kalina, J. Flores-Montero, V. H. J. Van Der Velden, et al., "EuroFlow Standardization of Flow Cytometer Instrument Settings and Immunophenotyping Protocols," *Leukemia* 26 (2012): 1986–2010, https://doi.org/10.1038/leu.2012.122.

23. J. E. Lewis, L. A. D. Cooper, D. L. Jaye, and O. Pozdnyakova, "Automated Deep Learning-Based Diagnosis and Molecular Characterization of Acute Myeloid Leukemia Using Flow Cytometry," *Modern Pathology* 37 (2024): 37, https://doi.org/10.1016/j.modpat.2023.100373.

24. L. M. Zuromski, J. Durtschi, A. Aziz, et al., "Clinical Validation of a Real-Time Machine Learning-Based System for the Detection of Acute Myeloid Leukemia by Flow Cytometry," *Cytometry Part B, Clinical Cytometry* 108 (2025): 392–403.

25. P. Theunissen, E. Mejstrikova, L. Sedek, et al., "Standardized Flow Cytometry for Highly Sensitive MRD Measurements in B-Cell Acute Lymphoblastic Leukemia," *Blood* 129 (2017): 347–357, https://doi.org/10.1182/BLOOD-2016-07-726307.

26. H. Reza Tizhoosh and L. Pantanowitz, "Artificial Intelligence and Digital Pathology: Challenges and Opportunities," *Journal of Pathology Informatics* 9 (2018): 38, https://doi.org/10.4103/JPI.JPI_53_18.

27. E. Papaemmanuil, M. Gerstung, L. Bullinger, et al., "Genomic Classification and Prognosis in Acute Myeloid Leukemia," *New England Journal of Medicine* 374 (2016): 2209–2221.

28. J. E. Farrar, J. L. Smith, M. Othus, et al., "Long Noncoding RNA Expression Independently Predicts Outcome in Pediatric Acute Myeloid Leukemia," *Journal of Clinical Oncology* 41 (2023): 2949–2962, https://doi.org/10.1200/JCO.22.01114/SUPPL_FILE/DS2_JCO.22.01114.XLSX.

29. P. S. Reel, S. Reel, E. Pearson, E. Trucco, and E. Jefferson, "Using Machine Learning Approaches for Multi-Omics Data Analysis: A Review," *Biotechnology Advances* 49 (2021): 107739, https://doi.org/10.1016/J.BIOTECHADV.2021.107739.

30. A. Curioni-Fontecedro, "A New Era of Oncology Through Artificial Intelligence," *ESMO Open* 2 (2017): e000198, https://doi.org/10.1136/esmoopen-2017-000198.

31. N. Aghaeepour, G. Finak, H. Hoos, et al., "Critical Assessment of Automated Flow Cytometry Data Analysis Techniques," *Nature Methods* 10 (2013): 228–238, https://doi.org/10.1038/nmeth.2365.

32. W. Lotter, M. J. Hassett, N. Schultz, K. L. Kehl, E. M. Van Allen, and E. Cerami, "Artificial Intelligence (AI) in Oncology: Current Landscape, Challenges, and Future Directions," *Cancer Discovery* 14 (2024): 711, https://doi.org/10.1158/2159-8290.CD-23-1199.

33. L. J. Jennings, M. E. Arcila, C. Corless, et al., "Guidelines for Validation of Next-Generation Sequencing–Based Oncology Panels: A Joint Consensus Recommendation of the Association for Molecular Pathology and College of American Pathologists," *Journal of Molecular Diagnostics* 19 (2017): 341–365, https://doi.org/10.1016/j.jmoldx.2017.01.011.

34. S. Roy, C. Coldren, A. Karunamurthy, et al., "Standards and Guidelines for Validating Next-Generation Sequencing Bioinformatics Pipelines: A Joint Recommendation of the Association for Molecular Pathology and the College of American Pathologists," *Journal of Molecular Diagnostics* 20 (2018): 4–27, https://doi.org/10.1016/j.jmoldx.2017.11.003.

35. G. Sirugo, S. M. Williams, and S. A. Tishkoff, "The Missing Diversity in Human Genetic Studies," *Cell* 177 (2019): 26–31, https://doi.org/10.1016/j.cell.2019.02.048.

36. K. G. M. Moons, J. A. A. Damen, T. Kaul, et al., "PROBAST+AI: An Updated Quality, Risk of Bias, and Applicability Assessment Tool for Prediction Models Using Regression or Artificial Intelligence Methods," *BMJ* 388 (2025): e082505, https://doi.org/10.1136/BMJ-2024-082505.

37. F. Zhong, F. Yao, Z. Wang, J. Liu, B. Huang, and X. Wang, "Integrated Multiomics Analysis and Machine Learning Refine Neutrophil Extracellular Trap-Related Molecular Subtypes and Prognostic Models for Acute Myeloid Leukemia," *Frontiers in Immunology* 16 (2025): 1558496, https://doi.org/10.3389/FIMMU.2025.1558496.

38. G. S. Collins, P. Dhiman, J. Ma, et al., "Evaluation of Clinical Prediction Models (Part 1): From Development to External Validation," *BMJ* 384 (2024): e074819, https://doi.org/10.1136/BMJ-2023-074819.

39. A. Kosvyra, A. Karadimitris, M. Papaioannou, and I. Chouvarda, "Machine Learning and Integrative Multi-Omics Network Analysis for Survival Prediction in Acute Myeloid Leukemia," *Computers in Biology and Medicine* 178 (2024): 108735, https://doi.org/10.1016/J.COMPBIOMED.2024.108735.

40. M. M. Li, M. Datto, E. J. Duncavage, et al., "Standards and Guidelines for the Interpretation and Reporting of Sequence Variants in Cancer: A Joint Consensus Recommendation of the Association for Molecular Pathology, American Society of Clinical Oncology, and College of American Pathologists," *Journal of Molecular Diagnostics* 19 (2017): 4–23, https://doi.org/10.1016/j.jmoldx.2016.10.002.

41. K. White, J. Qualtieri, E. L. Courville, et al., "Entrustable Professional Activities in Hematopathology Pathology Fellowship Training: Consensus Design and Proposal," *Academic Pathology* 8 (2021): 8, https://doi.org/10.1177/2374289521990823.

42. E. J. Duncavage, A. Bagg, R. P. Hasserjian, et al., "Genomic Profiling for Clinical Decision Making in Myeloid Neoplasms and Acute Leukemia," *Blood* 140 (2022): 2228–2247, https://doi.org/10.1182/BLOOD.2022015853.

43. R. Ahmad, M. Awais, N. Kausar, U. Tariq, J. H. Cha, and J. Balili, "Leukocytes Classification for Leukemia Detection Using Quantum-Inspired Deep Feature Selection," *Cancers* 15 (2023): 2507, https://doi.org/10.3390/CANCERS15092507.

44. J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, "Quantum Machine Learning," *Nature* 549 (2017): 195–202, https://doi.org/10.1038/nature23474.

45. D. Peral-García, J. Cruz-Benito, and F. J. García-Peñalvo, "Systematic Literature Review: Quantum Machine Learning and Its Applications," *Computer Science Review* 51 (2024): 100619, https://doi.org/10.1016/J.COSREV.2024.100619.

46. R. S. Gupta, C. E. Wood, T. Engstrom, J. D. Pole, and S. Shrapnel, "A Systematic Review of Quantum Machine Learning for Digital Health," *npj Digital Medicine* 8 (2025): 237, https://doi.org/10.1038/s41746-025-01597-z.

47. D. Ranga, A. Rana, S. Prajapat, P. Kumar, K. Kumar, and A. V. Vasilakos, "Quantum Machine Learning: Exploring the Role of Data Encoding Techniques, Challenges, and Future Directions," *Mathematics* 12 (2024): 3318, https://doi.org/10.3390/MATH12213318.

48. K. Bharti, A. Cervera-Lierta, T. H. Kyaw, et al., "Noisy Intermediate-Scale Quantum Algorithms," *Reviews of Modern Physics* 94 (2022): 015004, https://doi.org/10.1103/RevModPhys.94.015004.

49. O. Treeck, B. Krishnamachary, A. Kenzuke Nakamura-García, and J. Espinal-Enríquez, "Pseudogenes in Cancer: State of the Art," *Cancers* 15 (2023): 4024, https://doi.org/10.3390/CANCERS15164024.

50. A. Aggarwal, S. Bharadwaj, G. Corredor, T. Pathak, S. Badve, and A. Madabhushi, "Artificial Intelligence in Digital Pathology—Time for a Reality Check," *Nature Reviews Clinical Oncology* 22 (2025): 283–291, https://doi.org/10.1038/s41571-025-00991-6.

51. X. Liu, S. Cruz Rivera, D. Moher, et al., "Reporting Guidelines for Clinical Trial Reports for Interventions Involving Artificial Intelligence: The CONSORT-AI Extension," *Nature Medicine* 26 (2020): 1364–1374, https://doi.org/10.1038/S41591-020-1034-X.

52. B. Vasey, M. Nagendran, B. Campbell, et al., "Reporting Guideline for the Early-Stage Clinical Evaluation of Decision Support Systems Driven by Artificial Intelligence: DECIDE-AI," *Nature Medicine* 28 (2022): 924–933, https://doi.org/10.1038/s41591-022-01772-9.

53. S. H. Park and K. Han, "Methodologic Guide for Evaluating Clinical Performance and Effect of Artificial Intelligence Technology for

Medical Diagnosis and Prediction," *Radiology* 286 (2018): 800–809, https://doi.org/10.1148/RADIOL.2017171920.

54. M. J. Willemink, W. A. Koszek, C. Hardell, et al., "Preparing Medical Imaging Data for Machine Learning," *Radiology* 295 (2020): 4–15, https://doi.org/10.1148/RADIOL.2020192224.

55. V. Sounderajah, A. Guni, X. Liu, et al., "The STARD-AI Reporting Guideline for Diagnostic Accuracy Studies Using Artificial Intelligence," *Nature Medicine* 31 (2025): 3283–3289, https://doi.org/10.1038/s41591-025-03953-8.

56. L. Seyyed-Kalantari, H. Zhang, M. B. A. McDermott, I. Y. Chen, and M. Ghassemi, "Underdiagnosis Bias of Artificial Intelligence Algorithms Applied to Chest Radiographs in Underserved Patient Populations," *Nature Medicine* 27 (2021): 2176–2182, https://doi.org/10.1038/s41591-021-01595-0.

57. A. C. Alba, T. Agoritsas, M. Walsh, et al., "Discrimination and Calibration of Clinical Prediction Models: Users' Guides to the Medical Literature," *Journal of the American Medical Association* 318 (2017): 1377–1384, https://doi.org/10.1001/JAMA.2017.12126.

58. B. Van Calster, G. S. Collins, A. J. Vickers, et al., "Evaluation of Performance Measures in Predictive Artificial Intelligence Models to Support Medical Decisions: Overview and Guidance," *Lancet Digital Health* 7 (2025): 100916, https://doi.org/10.1016/J.LANDIG.2025.100916.

59. A. J. Vickers and E. B. Elkin, "Decision Curve Analysis: A Novel Method for Evaluating Prediction Models," *Medical Decision Making* 26 (2006): 565–574, https://doi.org/10.1177/0272989X06295361.

60. S. Cruz Rivera, X. Liu, A. W. Chan, et al., "Guidelines for Clinical Trial Protocols for Interventions Involving Artificial Intelligence: The SPIRIT-AI Extension," *Nature Medicine* 26 (2020): 1351–1363, https://doi.org/10.1038/S41591-020-1037-7;SUBJMETA.

61. G. S. Collins, K. G. M. Moons, P. Dhiman, et al., "TRIPOD+AI Statement: Updated Guidance for Reporting Clinical Prediction Models That Use Regression or Machine Learning Methods," *BMJ* 385 (2024): e078378, https://doi.org/10.1136/BMJ-2023-078378.

62. K. Goddard, A. Roudsari, and J. C. Wyatt, "Automation Bias: A Systematic Review of Frequency, Effect Mediators, and Mitigators," *Journal of the American Medical Informatics Association* 19 (2012): 121–127, https://doi.org/10.1136/AMIAJNL-2011-000089.

63. E. Beede, E. Baylor, F. Hersch, et al., "A Human-Centered Evaluation of a Deep Learning System Deployed in Clinics for the Detection of Diabetic Retinopathy," in *Conference on Human Factors in Computing Systems—Proceedings* (Association for Computing Machinery, 2020), https://doi.org/10.1145/3313831.3376718.

64. H. Chen, C. Gomez, C. M. Huang, and M. Unberath, "Explainable Medical Imaging AI Needs Human-Centered Design: Guidelines and Evidence From a Systematic Review," *npj Digital Medicine* 5 (2022): 156, https://doi.org/10.1038/s41746-022-00699-2.

65. M. P. Sendak, W. Ratliff, D. Sarro, et al., "Real-World Integration of a Sepsis Deep Learning Technology Into Routine Clinical Care: Implementation Study," *JMIR Medical Informatics* 8 (2020): e15182, https://doi.org/10.2196/15182.

66. M. Ghassemi, L. Oakden-Rayner, and A. L. Beam, "The False Hope of Current Approaches to Explainable Artificial Intelligence in Health Care," *Lancet Digit Health* 3 (2021): e745–e750, https://doi.org/10.1016/S2589-7500(21)00208-9.

67. S. G. Finlayson, A. Subbaswamy, K. Singh, et al., "The Clinician and Dataset Shift in Artificial Intelligence," *New England Journal of Medicine* 385 (2021): 283–286, https://doi.org/10.1056/NEJMC2104626.

68. K. Lekadir, A. F. Frangi, A. R. Porras, et al., "FUTURE-AI: International Consensus Guideline for Trustworthy and Deployable Artificial Intelligence in Healthcare," *BMJ* 388 (2025): e081554, https://doi.org/10.1136/BMJ-2024-081554.

69. S. L. McNamara, P. H. Yi, and W. Lotter, "The Clinician-AI Interface: Intended Use and Explainability in FDA-Cleared AI Devices for Medical Image Interpretation," *npj Digital Medicine* 7 (2024): 80, https://doi.org/10.1038/s41746-024-01080-1.

70. M. A. Santos-Silva, N. Sousa, and J. C. Sousa, "Artificial Intelligence in Routine Blood Tests," *Frontiers in Medical Engineering* 2 (2024): 1369265, https://doi.org/10.3389/FMEDE.2024.1369265.

71. W. Samek, G. Montavon, S. Lapuschkin, C. J. Anders, and K. R. Müller, "Explaining Deep Neural Networks and Beyond: A Review of Methods and Applications," *Proceedings of the IEEE* 109 (2021): 247–278, https://doi.org/10.1109/JPROC.2021.3060483.

72. V. Rampton, "Artificial Intelligence Versus Clinicians," *BMJ* 369 (2020): m1326, https://doi.org/10.1136/BMJ.M1326.

73. S. Reddy, S. Allan, S. Coghlan, and P. Cooper, "A Governance Model for the Application of AI in Health Care," *Journal of the American Medical Informatics Association* 27 (2020): 491–497, https://doi.org/10.1093/JAMIA/OCZ192.

74. J. Car, Q. C. Ong, T. Erlikh Fox, et al., "The Digital Health Competencies in Medical Education Framework: An International Consensus Statement Based on a Delphi Study," *JAMA Network Open* 8 (2025): e2453131, https://doi.org/10.1001/JAMANETWORKOPEN.2024.53131.

75. H. Gaffney and K. M. Mirza, "Pathology in the Artificial Intelligence Era : Guiding Innovation and Implementation to Preserve Human Insight," *Academic Pathology* 12 (2025): 100166, https://doi.org/10.1016/J.ACPATH.2025.100166.

76. S. Vos, K. Hebeda, M. Milota, et al., "Making Pathologists Ready for the New Artificial Intelligence Era: Changes in Required Competencies," *Modern Pathology* 38 (2025): 100657, https://doi.org/10.1016/J.MODPAT.2024.100657.

77. K. Masters, "Artificial Intelligence in Medical Education," *Medical Teacher* 41 (2019): 976–980, https://doi.org/10.1080/0142159X.2019.1595557.

78. C. B. McCloskey, R. E. Domen, R. M. Conran, et al., "Entrustable Professional Activities for Pathology," *Academic Pathology* 4 (2017): 2374289517714283, https://doi.org/10.1177/2374289517714283.

79. W. H. Henricks, D. S. Karcher, J. H. Harrison, et al., "Pathology Informatics Essentials for Residents: A Flexible Informatics Curriculum Linked to Accreditation Council for Graduate Medical Education Milestones," *Journal of Pathology Informatics* 7 (2016): 27, https://doi.org/10.4103/2153-3539.185673.

80. F. E. Craig and K. A. Foon, "Flow Cytometric Immunophenotyping for Hematologic Neoplasms," *Blood* 111 (2008): 3941–3967, https://doi.org/10.1182/BLOOD-2007-11-120535.

81. O. Ten Cate, "Competency-Based Postgraduate Medical Education: Past, Present, and Future," *GMS Journal for Medical Education* 34 (2017): Doc69, https://doi.org/10.3205/ZMA001146.

82. M. Gottlieb, A. King, R. Byyny, M. Parsons, and J. Bailitz, "Journal Club in Residency Education: An Evidence-Based Guide to Best Practices From the Council of Emergency Medicine Residency Directors," *Western Journal of Emergency Medicine* 19 (2018): 746–755, https://doi.org/10.5811/WESTJEM.2018.4.37507.

83. R. Tolentino, A. Baradaran, G. Gore, P. Pluye, and S. Abbasgholizadeh-Rahimi, "Curriculum Frameworks and Educational Programs in AI for Medical Students, Residents, and Practicing Physicians: Scoping Review," *JMIR Medical Education* 10 (2024): E54793, https://doi.org/10.2196/54793.

84. F. Y. C. Ng, A. J. Thirunavukarasu, H. Cheng, et al., "Artificial Intelligence Education: An Evidence-Based Medicine Approach for Consumers, Translators, and Developers," *Cell Reports Medicine* 4 (2023): 4, https://doi.org/10.1016/j.xcrm.2023.101230.

85. D. Furfaro, L. A. Celi, and R. M. Schwartzstein, "Artificial Intelligence in Medical Education: A Long Way to Go," *Chest* 165 (2024): 771–774, https://doi.org/10.1016/j.chest.2023.11.028.

86. R. J. Christian and M. VanSandt, "Using Dynamic Virtual Microscopy to Train Pathology Residents During the Pandemic: Perspectives on Pathology Education in the Age of COVID-19," *Academic Pathology* 8 (2021): 8, https://doi.org/10.1177/23742895211006819.

87. A. Rjoop, M. Al-Qudah, R. Alkhasawneh, et al., "Awareness and Attitude Toward Artificial Intelligence Among Medical Students and Pathology Trainees: Survey Study," *JMIR Medical Education* 11 (2025): e62669, https://doi.org/10.2196/62669.

88. Q. Li and Y. Qin, "AI in Medical Education: Medical Student Perception, Curriculum Recommendations and Design Suggestions," *BMC Medical Education* 23 (2023): 1–8.

89. Y. Xu, Z. Jiang, D. S. W. Ting, et al., "Medical Education and Physician Training in the Era of Artificial Intelligence," *Singapore Medical Journal* 65 (2024): 159–166, https://doi.org/10.4103/SINGAPOREMEDJ.SMJ-2023-203.

90. C. Eloy, F. Fraggetta, P. J. van Diest, et al., "Digital Transformation of Pathology - the European Society of Pathology Expert Opinion Paper," *Virchows Archiv* 487 (2025): 971–981, https://doi.org/10.1007/s00428-025-04090-w.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Table S1:** Inclusion criteria. **Figure S1:** Study selection flowchart. **Table S2:** Summary of 36 core sources from the initial narrative synthesis. **Table S3:** Additional targeted sources incorporated during revision.