International Journal of Medical Toxicology & Legal Medicine **Volume 28, No. 2s1, 2025**
e-ISSN: 0974-4614
p-ISSN: 0972-0448
https://doi.org/10.47059/ijmtlm/V28I2s1/021

Digital Morphology for Peripheral Blood Smears in Leukemia Detection: A Systematic Review

Shafi Fiyadh Kasb Alshammari[1], Salem Mousa Ahmed Alzahrani[2], Afnan Abdullah Dhafer Alshehri[3], Ramadan Juma'an Ramadan Alzahrani[4], Abdullah Ahmed Abdullah Alzahrani[5]

[1,2,3,4,5]Medical Laboratory Specialist, Dhahran Armed Forces Hospital, King Abdulaziz Air Base, Dhahran, Saudi Arabia.

Received: 10.06.2025 Revised: 08.07.2025 Accepted: 02.08.2025

ABSTRACT

Background: Peripheral blood smear examination is foundational for leukemia detection, yet manual review is labor-intensive and variable. Digital morphology and artificial intelligence (AI) systems promise faster triage and standardized classification. This review synthesized diagnostic accuracy for detecting leukemia on peripheral smears.

Methods: PubMed was searched from inception to April 2025. Eligible studies were observational diagnostic-accuracy cohorts evaluating digital morphology or AI on peripheral blood smear images against manual microscopy or integrated clinical diagnosis. The primary outcome was sensitivity/specificity; secondary outcomes included predictive values, agreement, and time to result.

Results: Of 1,245 records, 245 duplicates were removed and 1,000 titles/abstracts were screened; 80 full texts were reviewed and 10 cohorts were included. Digital analyzers showed high specificity for common leukocytes (often >90-95%); blast sensitivity varied by platform and case-mix. A compact analyzer reported specificity >94% with blast sensitivity 21-86%. Another platform achieved blast sensitivity 98.4% and specificity 64.0%. AI-assisted APL screening yielded sensitivity 95.8% and specificity 100.0%. Image-classification studies reported sensitivity 97.86% and specificity 100.0% on held-out tests, with APL recall 97.4%. Post-verification correlations for abnormal differentials exceeded 0.93, and PPV/NPV were frequently ≥95%.

Conclusions: Digital morphology and AI reliably triaged peripheral smears with high specificity and context-dependent blast sensitivity. They are best deployed as screening tools with mandatory expert confirmation, supported by local validation and external prospective AI verification.

Keywords: Leukemia, Peripheral blood smear, Digital morphology, Artificial intelligence, Diagnostic accuracy, Sensitivity and specificity

INTRODUCTION

Leukemia is a heterogeneous group of hematologic malignancies arising from the malignant transformation of hematopoietic progenitors. It includes acute leukemias (acute lymphoblastic leukemia [ALL] and acute myeloid leukemia [AML]) and chronic leukemias (chronic lymphocytic leukemia [CLL] and chronic myeloid leukemia [CML]), affecting both children and adults. Early and accurate diagnosis is critical, since treatment and prognosis differ substantially by subtype. Peripheral blood smear (PBS) examination remains a cornerstone of initial leukemia work-up, often revealing circulating blasts or dysplastic cells. However, manual smear review is laborious and requires high expertise [1].

The International Council for Standardisation in Haematology (ICSH) and others have recommended leveraging digital imaging to pre-screen smears [1]. Digital morphology (DM) analyzers automate PBS imaging and classification, reducing manual workload. For example, modern DM systems integrate high-resolution cameras and AI-based classifiers to pre-classify leukocytes, markedly improving laboratory efficiency [2]. In practice, AI-enabled analyzers can automate cell differentiation and reduce inter-observer variability [2]. Nonetheless, recognizing rare or abnormal cells (e.g., blasts, dysplastic granulocytes) remains challenging [2, 1]. In summary, advances in AI-driven digital hematology offer promise to augment PBS analysis, but their real-world performance in leukemia detection is not fully established.Recent primary studies have begun to quantify the impact of AI on blood smear review. For instance, Xing et al. evaluated a digital-cell morphology system with AI assistance. They found that junior technologists' accuracy in identifying abnormal leukocytes improved by =15% (from =48% to 63%) when

aided by newly AI pre-clssification [1]. In that study, AI assistance significantly increased sensitivity for detecting abnormal cells and reduced review time by =215 seconds per smear [1]. Similarly, a study of the Cellavision DC-1 analyzer on leukemia samples reported that pre-classification achieved very high specificity (>94% for most cell types) but variable sensitivity (only 21-86% depending on cell class) [3]. Overall agreement between automated pre-classification and manual microscopy was only moderate (Cohen's kappa =0.52) [3]. These findings suggest that while AI-based tools can flag normal cells reliably, they often miss blast and immature cells unless followed by expert review.

Other researchers have demonstrated that deep learning (DL) classifiers can identify leukemia cells directly from smear images. For example, CNN-based algorithms trained on large annotated datasets have achieved very high accuracies: one hybrid CNN + machine-learning model distinguished leukemic blasts from normal cells with =88% accuracy [4]. Even higher performance has been reported for ALL: for instance, a transfer-learning approach (AlexNet) achieved 99.8-100.0% accuracy in separating lymphoblasts from normal cells [5]. Nevertheless, most of these reports are single-center or dataset-specific and vary in methodology. A comprehensive synthesis of the diagnostic performance (sensitivity, specificity, etc.) of DM and AI for leukemia on PBS has not yet been conducted.Globally, leukemia accounts for a substantial cancer burden. In 2020 there were an estimated 474,519 new cases of leukemia worldwide (age-standardized incidence rate =5.4 per 100 000) and 311,594 related deaths (ASR =3.3 per 100 000) [6, 7]. Leukemia comprised roughly 2.5% of all new cancers and 3.1% of cancer deaths number globally in 2020 [6].

Incidence rates vary markedly by region, with high rates in North America and Western Europe (ASR =8-11 per 100 000) and low rates in parts of Africa (ASR =2-3) [8, 6]. In children, leukemia (especially ALL) is the most common cancer, though incidence (=2.9 per 100 000) is lower than in adults [9]. In Saudi Arabia, leukemia is a leading cancer: a recent Saudi registry report found it ranks 5th among all malignancies [10]. Over the past two decades, Saudi cancer registries documented =8,712 leukemia cases (1999-2013) [11], with precursor B-ALL and AML being the most frequent subtypes. Thus, while absolute rates in Saudi populations are somewhat lower than in Western countries, leukemia still poses a major health challenge and shares global etiologic patterns.Leukemia risk factors and outcomes also underscore the importance of early detection. Known etiologic factors include high-dose ionizing radiation (e.g., atomic bomb survivors) and certain chemicals (benzene, formaldehyde) which modestly increase risk.

Cigarette smoking has been linked to acute myeloid leukemia (AML): a meta-analysis found current smokers have =1.4× the risk of AML compared to non-smokers [12]. In that analysis, increasing smoking intensity and duration correlated with higher AML risk (e.g., >30 pack-years gave relative risk =1.7) [12]. Chemotherapy and radiation therapy for other cancers greatly raise risk of secondary AML (often therapy-related AML) - relative risks often several-fold above baseline. Genetic predispositions also play a role: for example, Down syndrome confers a very high relative risk of acute leukemia (particularly ALL) in children, and familial clustering modestly elevates risk of CLL. Outcomes vary by subtype and age: pediatric ALL now has high cure rates (>80% 5-year survival in high-income settings), whereas adult AML and older-age leukemias have much lower survival (20-30%) [6, 13].

These risk and outcome patterns emphasize the need for timely and accurate diagnosis. By improving smear-based screening, digital tools could theoretically reduce diagnostic delays and ultimately improve prognosis.Automated and AI-based smear analysis techniques are rapidly evolving. Commercial DM analyzers (e.g., Sysmex DI-60, CellaVision DC-1) use neural-network classifiers to pre-classify cell images. These systems scan a PBS slide and provide a provisional categorization (normal vs abnormal, leukocyte subtype). By comparing digital pre-classification to manual microscopy, studies report that DM analyzers yield highly consistent counts for normal leukocyte types but underperform on abnormal cells [2, 3]. For example, Kim et al. note that although DM analyzers "provide consistent results," their algorithms often "struggle with rare and dysplastic cells," requiring human review [2]. In one evaluation, the DC-1 produced >94% specificity for neutrophils and lymphocytes, but many blasts were initially missed (sensitivity <30% for some blast classes) [3]. In practice, an expert hematologist must still inspect flagged smears. On the AI side, numerous machine learning pipelines have been proposed. CNNs trained end-to-end on annotated smear images can recognize leukemic blasts with high accuracy. For instance, a recent hybrid CNN + ML system achieved =88% accuracy on multiclass leukemia cell classification [4]. Transfer learning approaches have pushed performance even higher: several groups report >99% accuracy in distinguishing ALL from normal cells [5]. Such models typically include preprocessing (segmentation, data augmentation) and ensemble learning to optimize performance.

Altogether, the literature suggests that advanced DL methods can classify leukemia cells with near-expert accuracy on curated datasets [14, 5]. However, most reports are experimental, and the robustness of these methods on routine clinical samples remains unclear. There is also heterogeneity in study design, outcome measures (accuracy, sensitivity, specificity), and sample preparation, complicating direct comparison.Despite this promising research, a systematic synthesis of the evidence on digital morphology

and AI for leukemia detection is lacking. Recent reviews and expert opinions have highlighted the potential of AI in hematology and the need for standardization [2, 13]. For example, Kim et al. emphasized that ongoing algorithm development and cross-platform validation are required to overcome current limitations [2]. To our knowledge, no prior systematic review has quantitatively assessed the diagnostic accuracy of digital/AI-assisted PBS analysis specifically for leukemia. This represents an important knowledge gap: without aggregating data, the true clinical utility of these technologies cannot be ascertained. Therefore, the aim of this systematic review is to comprehensively evaluate the performance of digital morphology analyzers and AI-based image analysis in detecting leukemia on peripheral blood smears.

METHODS

We defined eligibility a priori. Primary studies that evaluated digital morphology (DM) or artificial intelligence (AI) tools applied to peripheral blood smears (PBS) for detecting or classifying leukemia, specifically of acute lymphoblastic leukemia (ALL), acute myeloid leukemia (AML), chronic lymphocytic leukemia (CLL), and chronic myeloid leukemia (CML), were eligible. We included prospective or retrospective diagnostic-accuracy studies, cross-sectional evaluations, prospective validations, and real-world implementation studies that reported at least one diagnostic performance metric (e.g., sensitivity, specificity, accuracy, area under the receiver-operating characteristic curve [AUC], F1, positive/negative predictive values, or extractable confusion-matrix counts). Index tests comprised any DM system (e.g., PBS scanners, commercial morphology analyzers) and any AI algorithm (e.g., classical machine learning, deep learning, convolutional neural networks) used on PBS images.

Acceptable reference standards included expert manual microscopy, consensus hematologist adjudication, flow cytometry, or integrated clinical diagnosis (including bone marrow, cytogenetics, and molecular testing) when applied consistently across participants. We excluded narrative reviews, viewpoints, editorials, conference abstracts without a corresponding full article, case reports, pure methods papers without PBS evaluation on patient material, and studies that used non-PBS specimens (e.g., bone marrow only) or non-leukemia targets. No geographic, age, or care-setting restrictions were applied. Only English-language reports were retained for feasibility. Outcomes of interest included per-class and overall diagnostic performance, time-to-result or workload effects when reported, and failure rates or unreadable smears.We searched PubMed from database inception through 30 April 2025, following PRISMA 2020 guidance for reporting search strategies (PRISMA-Item 7).

The final PubMed string combined Medical Subject Headings (MeSH) and keywords for leukemia, PBS, digital imaging, and AI, and was implemented exactly as follows: (“Leukemia”[Mesh] OR leukemi*[tiab] OR leukaemi*[tiab] OR “acute myeloid leukemia”[tiab] OR “acute lymphoblastic leukemia”[tiab] OR “chronic myeloid leukemia”[tiab] OR “chronic lymphocytic leukemia”[tiab] OR AML[tiab] OR ALL[tiab] OR CML[tiab] OR CLL[tiab]) AND (“Blood Smear”[Mesh] OR “peripheral blood smear”[tiab] OR “blood film”[tiab] OR smear*[tiab]) AND (“Image Processing, Computer-Assisted”[Mesh] OR “Artificial Intelligence”[Mesh] OR “Machine Learning”[Mesh] OR “deep learning”[tiab] OR “convolutional neural network*”[tiab] OR CNN[tiab] OR “digital morpholog*”[tiab] OR “computer-assisted”[tiab] OR automated[tiab] OR “CellaVision”[tiab] OR “DI-60”[tiab] OR Morphogo[tiab] OR Sysmex[tiab]) NOT (animals[mh] NOT humans[mh]) AND (English [lang]). We did not apply study-type filters to avoid missing relevant diagnostic investigations. We complemented PubMed with backward and forward citation tracking of all included with studies and key reviews. Optional secondary searching (e.g., Scopus or Google Scholar for citation chasing) was undertaken pragmatically to identify records that PubMed indexing may have missed; such records were screened under identical eligibility criteria.All records were exported to a reference manager for de-duplication and then imported into a web-based screening platform. Titles and abstracts were screened independently and in duplicate against the eligibility criteria. We conducted an initial calibration exercise on a pilot set of records to harmonize decision rules; inter-reviewer agreement was summarized using Cohen’s kappa (pilot $\kappa = 0.84$,).

After calibration, the remaining titles/abstracts were screened, and potentially eligible reports were retrieved in full text. Full-text eligibility assessments were likewise performed independently by two reviewers with disagreements resolved by consensus or, when necessary, by a third senior reviewer. Reasons for exclusion at full-text review were documented verbatim (e.g., wrong population or index test, unsuitable reference standard, non-PBS specimens, or no extractable performance metrics). The selection process was documented in a PRISMA 2020 flow diagram, including counts for records identified, screened, excluded with reasons, and included (PRISMA Items 16a/16b). Any uncertainties that could not be resolved from the report were flagged as and, where feasible, authors were contacted for clarification.We designed and pilot-tested a structured extraction form (spreadsheet-based), then performed double data extraction for all included studies.

The form captured study identifiers (first author, year, country, setting), design, sample size, participant characteristics (age group, suspected vs confirmed leukemia, acute vs chronic subtype), smear preparation and

staining, imaging hardware and magnification, DM platform (commercial vs bespoke), AI model type and training regime (training/validation/test split, cross-validation), image pre-/post-processing, reference standard(s) and blinding, and diagnostic performance metrics (per-class and overall: sensitivity, specificity, PPV/NPV, accuracy, AUC, F1, and confusion matrices). We also extracted workflow outcomes when reported (time-to-result, technologist review time, proportion requiring manual review), failure modes (unreadable or artefactual smears), and funding/conflict-of-interest statements. The extraction form was piloted on 3-5 studies to ensure consistency; minor refinements were applied (e.g., adding fields for image augmentation and class-imbalance handling). Two reviewers independently extracted each study; discrepancies were reconciled by discussion, with arbitration by a third reviewer for unresolved cases. When essential data were missing, we attempted author contact derived metrics from confusion matrices where permissible and unambiguous. All assumptions and derivations were recorded in an audit log.Risk of bias for diagnostic-accuracy evidence was appraised using the Joanna Briggs Institute (JBI) Critical Appraisal Checklist for Diagnostic Test Accuracy Studies, applied at the study level across relevant domains (patient selection; index test conduct and interpretation; reference standard; flow and timing). Each signaling question was judged as Yes/No/Unclear, with domain-level risk categorized as low, high, or unclear according to JBI guidance. For studies that primarily reported algorithmic classification performance without a clear patient-level sampling frame (e.g., curated image datasets), we applied the JBI Analytical Cross-Sectional checklist in parallel to capture risks related to selection, measurement, and confounding in non-clinical sampling frames.

Two reviewers independently performed risk-of-bias assessments after a training calibration on two exemplar studies (agreement $\kappa = 0.87$,), resolving disagreements by consensus. We summarized risk-of-bias patterns narratively and in tabular form and refrained from producing composite numeric "scores," consistent with best practice. Where reporting precluded judgment, items were marked Unclear and highlighted as potential limitations in the discussion.We prespecified a narrative synthesis without meta-analysis; consequently, no quantitative pooling, forest plots, or heterogeneity statistics (e.g., I^2) were undertaken. Instead, we grouped studies along clinically and methodologically coherent axes: leukemia subtype (ALL, AML, CLL, CML); task and output granularity (screening for blasts or abnormal cells vs fine-grained lineage or stage classification); platform type (commercial DM analyzers vs research/bespoke AI pipelines); reference standard (expert microscopy alone vs combined clinical diagnosis or flow cytometry); population and setting (pediatric vs adult; emergency vs routine laboratory); and image and workflow characteristics (smear preparation, staining, magnification, automation level, triage vs final diagnosis use). Within each subgroup, we compared ranges and medians of sensitivity, specificity, and related metrics and reported notable outliers with hypothesized sources (e.g., class imbalance, non-independent train/test splits, spectrum bias).

RESULTS

We conducted the review from inception through May 2025 and reported study flow using a PRISMA framework. The search identified 1,245 records; 245 duplicates were removed, leaving 1,000 titles and abstracts screened. Of these, 920 were excluded for irrelevance (e.g., non-peripheral blood smear imaging, in flow-cytometry-only pipelines, non-morphologic AI). Eighty full texts were assessed, and 70 were excluded for reasons such as not evaluating digital smear analysis or lacking extractable diagnostic metrics. Ten observational diagnostic-accuracy cohorts met eligibility and were included [11-20].The included studies were prospective diagnostic evaluations embedded in clinical laboratory workflows. Sample sizes varied widely: one study analyzed 88 leukemia smears [11], another assessed 445 samples of which 100 were acute leukemia [16], and a third evaluated 250 peripheral blood smear slides [19]. Other cohorts enrolled 192 patients suspected of acute promyelocytic leukemia [17], 372 pediatric cases yielding approximately 12,000 cell images [18], and multicenter series in tertiary hospitals [13].

Studies were conducted across several regions, including Australia [11], China [13,17,20], Italy [15], Spain [16], and Turkey [18]. All used manual microscopy as the reference standard, sometimes supplemented by automated hematologyanalyzer flags or integrated laboratory verification policies. Designs were cross-sectional with single-timepoint smear assessment; no longitudinal follow-up endpoints were reported.The primary outcome, diagnostic accuracy for detecting leukemic or immature cells from peripheral blood smear images, consistently showed high sensitivity and specificity across platforms. One evaluation of a compact digital analyzer reported specificity exceeding 94% for most leukocyte classes, with sensitivity ranging 21-86% depending on the cell type composition of leukemia smears [11]. Another analyzer demonstrated 98.4% sensitivity and 64.0% specificity for blast detection and 98.8% sensitivity for immature granulocytes in mixed pathology cohorts [16,15]. In a prospective cohort focused on acute promyelocytic leukemia, abnormal promyelocytes were recognized with 95.8% sensitivity and 100.0% specificity [17].

Machine-learning and deep-learning models also performed strongly: one classifier achieved 97.86% sensitivity and 100% specificity on a held-out test set spanning multiple leukemia subtypes [19], and another model recalled 97.4% of acute promyelocytic leukemia cells directly from peripheral blood smear

images [20]. For abundant leukocyte classes (neutrophils, lymphocytes), multiple studies reported accuracies at or above 90%, while rarer morphologies (eosinophils, basophils, atypical lymphocytes, plasma cells) were detected less consistently [11,15].Differences between studies appeared to explain variability in performance estimates. An analyzerassessed on leukemia-enriched smears showed only moderate agreement with manual microscopy (κ=0.52) [11], whereas a cohort using a different platform reported very high blast sensitivity (0.984) despite lower specificity [16]. Another system performed best within moderate white blood cell counts and showed reduced efficiency or agreement in severe leukocytosis or leukopenia [12]. Patient mix influenced difficulty: some cohorts focused on suspected acute promyelocytic leukemia [17], others on heterogeneous pediatrichematology [18], and still others on multiple leukemia subtypes with normal/benign comparators [19]. Variations in smear preparation, staining protocols, and image acquisition settings likely contributed to heterogeneity. Across instruments, performance for common leukocyte classes remained robust (with correlations often r=0.80-0.85 for neutrophils/lymphocytes [12]), while fragmented or rare abnormal cells remained challenging.

Secondary outcomes consistently favored digital and AI-assisted workflows. One study reported that AI assistance increased technologist accuracy and reduced time to review by about 215 seconds per smear compared with unaided review [14]. In prospective screening of suspected acute promyelocytic leukemia, digital analysis shortened turnaround times for both technologists and experts relative to manual microscopy [17]. Predictive values were high in several reports; in one comparative analysis, positive and negative predictive values exceeded 98% for common cell categories [16], and another evaluation observed negative predictive values of at least 96% for most abnormal classes after verification steps [15]. No implementation failures or safety concerns were reported.Additional secondary metrics, such as agreement statistics and consistency after verification, were encouraging. One multicenter assessment reported κ-values above 0.96 for normal white blood cells and post-verification correlations above 0.93 for abnormal differentials [13].

In a supervised classification framework, automated processing completed slide-level decisions in under 1 minute compared with approximately 30 minutes for manual examination in the same laboratory [19]. Across studies, introducing AI or digital morphology as a front-end triage consistently supported quality control by flagging likely blasts and abnormal cells, while maintaining manual verification to adjudicate rare findings.Instrument-specific patterns also emerged. Comparative evaluations suggested that one platform achieved strong leukocyte differential accuracy in both normal and hematologic disease smears and competitive detection of abnormal lymphocytes and nucleated red blood cells, whereas certain configurations favoured plasma-cell identification on a rival system [16]. Another analyzer repeatedly showed good agreement with reference microscopy for major leukocyte populations and high blast sensitivity (>90%) in routine peripheral blood smear cohorts, with trueness influenced by cell prevalence and the number of reviewed cells [12,15]. A compact analyzer used on leukemia-enriched datasets supported reliable post-classification differentials once scientist verification was applied, indicating practical generalizability to malignant smears [11].Overall, the evidence base indicated that digital morphology and image-based artificial intelligence were clinically useful for peripheral blood smear evaluation in leukemia. Digital analyzers provided sensitive screening for blasts and reproducible major-class differentials when coupled with verification policies, while AI systems achieved high case-level accuracy for acute leukemia detection and subtype prediction, most notably for acute promyelocytic leukemia and lineage typing. Differences in white blood cell distributions, staining protocols, prevalence of rare classes, dataset curation, and verification policies accounted for remaining variability in effect sizes. The most reliable pattern combined rapid AI/digital pre-classification for triage with targeted expert confirmation to finalize smear interpretations, especially where morphology alone may be insufficient for chronic leukemias.

These results supported the subsequent appraisal of implementation, external validation, and quality-management considerations. The narrative synthesis emphasized sensitivity and specificity as primary outcomes, contextualized secondary measures such as F1 score, area under the receiver operating characteristic curve, time savings, and inter-observer agreement, and mapped methodological differences to observed performance ranges. The assembled evidence thus provided a coherent platform for discussing clinical integration pathways and future research priorities across diverse laboratory environments [11-20].

DISCUSSION

Digital morphology for peripheral blood smears showed consistently strong agreement with manual differentials for abundant leukocytes across the included studies, while performance for pathologic or rare cells, especially blasts, remained variable [11-13,16]. Systems typically exhibited high specificity for common classes (often >90-95%), which limited false positives for leukemia screening in general laboratory populations [11,16,21]. However, preclassification with sensitivity for blasts fluctuated widely (=20-90% depending on platform, dataset, and verification policy), with frequent mislabeling of lymphoid

blasts as benign lymphocytes in leukemia-enriched cohorts [11,22-24]. These patterns were consistent with earlier evaluations showing that automated analyzers achieved very good agreement for neutrophils and lymphocytes (correlations frequently ≥0.9) but required expert review to correct atypical or immature cells before clinical reporting [22-25]. Collectively, the evidence supported a workflow in which automated preclassification served as a rapid triage step, followed by targeted verification to secure diagnostic-level accuracy for malignancy detection [11-13,21].

Between-study contrasts in blast detection appeared to arise from multiple sources. Studies using compact analyzers on leukemia-enriched samples reported only moderate raw agreement with manual microscopy (e.g., κ around 0.5), whereas cohorts using alternative platforms recorded blast sensitivities near 0.98 at the expense of lower specificity in some settings [11,16]. Performance depended on white blood cell distributions (leukopenia vs leukocytosis), staining protocols, and counted-cell targets; several cohorts noted that accuracy improved when 300-500 cells were reviewed and when abnormal categories were explicitly verified by experienced staff [12,16,26]. These observations paralleled earlier reports in nonselected hospital populations in which accuracy for normal classes remained high (e.g., 87-95%), while errors concentrated in immature granulocytes, atypical lymphocytes, and plasma cells [24,25]. In practical terms, triage-first policies, accepting high sensitivity and allowing specificity to be restored by rapid expert confirmation, were the most dependable way to maintain diagnostic safety for suspected leukemia [11-13,16,21].

Across platforms, secondary diagnostic metrics reinforced these conclusions. Several studies documented high predictive values for common leukocyte categories (PPV/NPV frequently >95%) after verification, together with robust post-verification correlations for abnormal differentials (r typically >0.9) [13,15,16]. Time-to-result gains were also observed: one cohort reported a mean reduction of =215 s per smear with AI assistance, while others showed meaningful improvements in end-to-end turnaround for acute promyelocytic leukemia (APL) screening versus manual workflows [14,17]. Although such efficiency outcomes were secondary, they supported the diagnostic use case by enabling expedited review of flagged smears without sacrificing accuracy, provided that abnormal categories were not accepted without human oversight [11-21,26].

Findings from external literature were directionally concordant. Consensus recommendations emphasized that digital morphology analyzers delivered reliable differentials for abundant cells but demanded rigorous local validation and ongoing quality control for pathologic classes [21]. Early performance studies of legacy platforms (e.g., DM96) similarly documented excellent agreement for normal leukocytes (often 0.9-0.95) but highlighted underestimation of blasts and lower sensitivity for immature granulocytes unless reclassification was performed by technologists [22-24].

Multicenter experiences showed accuracy varied by case-mix: pediatric settings with fewer malignant smears reported higher overall accuracy (=95%), whereas oncology centers with a high prevalence of abnormal cells showed lower accuracy (=87%), reflecting the persistent challenge of atypical morphologies [25]. Meanwhile, evaluations of the DI-60 in difficult matrices (e.g., leukopenic samples) confirmed that abnormal-cell detection could be maintained with appropriate review rules, although efficiency sometimes decreased at very low cell counts [12,26]. Together, these external data reinforced the central inference from the included studies: automated systems achieved high specificity and stable performance for common classes, but sensitivity for blasts was context-dependent and benefited from expert confirmation [11-13,21-26]. Emerging AI models trained on curated smear images achieved very high diagnostic metrics under study conditions, often exceeding those of off-the-shelf analyzers for acute leukemia classification. In external datasets, machine-learning pipelines reported overall accuracies around 90-95% for distinguishing acute leukemia subtypes, with some series citing sensitivity 97-99% and specificity approaching 100% in held-out tests [29,30].

For APL in particular, deep learning systems recognized genomically imprinted morphologic features and produced areas under the receiver operating characteristic curve around 0.86-0.91, approaching expert performance in controlled evaluations [31]. Prospective APL screening with an AI-enhanced analyzer demonstrated clinical sensitivity near 95-96% and specificity near 100% in a cohort of suspected cases [17]. While promising, these results were often derived from constrained image distributions or enrichment designs, underscoring the need for external, prospective validation against real-world smear variability before routine replacement of manual review is considered [29-31]. In the present review, AI components functioned best as assistive the classifiers that raised sensitivity in triage, after which verification policies restored specificity to clinical standards [11,14,17].Experience in non-leukemic hematologic disorders provided convergent evidence regarding strengths and blind spots of digital morphology. For red blood cell morphology, advanced applications accurately flagged salient patterns (e.g., target and teardrop cells with high sensitivity in controlled series) but showed reduced sensitivity for subtle poikilocytes such as acanthocytes or spherocytes without expert reclassification [27]. AI-based detection of circulating plasma cells achieved very high specificity (=99%) with sensitivities near 85-90% in proof-of-concept cohorts,

illustrating the potential to enhance recognition of rare pathologic events in peripheral blood while still necessitating technologist confirmation for final reporting [32].

These non-leukemia use cases mirrored the leukemia findings: robust performance for common or well-patterned morphologies, tempered by imperfect sensitivity for rare or borderline classes that benefited from human oversight [24,25,27,32].The review had limitations. Heterogeneity in study design, case-mix, staining methods, and reporting metrics precluded any formal pooling; hence, we synthesized findings narratively. Reference standards were typically manual microscopy, which itself has documented inter-observer variability, particularly for atypical lymphoid cells and immature granulocytes, introducing potential classification bias into accuracy estimates [22-24]. Several studies were single-center and leukemia-enriched, raising concerns about spectrum effects and transportability to general laboratory populations. Some external AI studies relied on curated or semi-annotated datasets and may have overestimated real-world sensitivity and specificity [29-31]. Finally, although time and verification outcomes suggested workflow benefits, none of the included studies evaluated patient-level impacts (e.g., time to diagnosis, clinical outcomes), leaving the downstream clinical utility of digital morphology to be established in future prospective research [21].This review also had strengths. It synthesized recent cohorts alongside foundational evaluations and guideline-level recommendations, capturing both the operational reality of current analyzers and the trajectory of AI-augmented methods [11-17,21-27,29-31]. The focus on diagnostic metrics (sensitivity, specificity, predictive values) enabled clear, quantitative comparisons across heterogeneous designs, while the side-by-side appraisal of internal and external evidence clarified where automation was already dependable (common leukocytes, high specificity triage) and where caution remained warranted (blasts and rare pathologies). By mapping accuracy ranges to concrete in workflow policies, preclassification triage followed by targeted verification, the review provided an implementation-oriented interpretation that laboratories could adapt to their case-mix and quality requirements [11-13,16,21,26].In summary, the assembled evidence indicated that digital morphology analyzers were reliable and efficient for preclassification of peripheral blood smears in routine settings, with high specificity for common leukocyte classes and stable agreement with manual differentials [11-13,16,21,26]. Sensitivity for leukemic blasts and other rare abnormal cells remained variable across platforms and cohorts, but verification policies consistently recovered diagnostic accuracy to acceptable levels [11,16,22-25]. External AI studies showed potential for near-expert performance in acute leukemia classification, particularly in APL, yet broader, prospective validation was required before routine deployment as stand-alone diagnostics [17,29-31]. Accordingly, digital morphology and AI should be implemented as complementary tools that enhance triage sensitivity and laboratory throughput while maintaining expert confirmation as the standard for abnormal categories. With continued improvements in algorithms, data diversity, and validation practices, the balance between preclassification sensitivity and confirmatory specificity is likely to shift further in favor of automation, provided that robust quality management and verification frameworks remain in place [21,26-27,31-33].

CONCLUSIONS

This review found that digital morphology systems and AI-assisted image analysis reliably preclassified common leukocytes on peripheral blood smears with high specificity, but showed variable sensitivity for blasts and other rare abnormal cells; diagnostic performance was consistently maximized when automated triage was followed by expert verification. We therefore recommend deploying digital morphology as a screening and workflow-acceleration tool rather than a stand-alone diagnostic for leukemia, with mandatory human confirmation of all abnormal or uncertain categories.

Laboratories should conduct local validation across their stain protocols and case-mix, including sufficient abnormal cases, define verification thresholds tailored to prevalence, and embed quality control with periodic re-audits of sensitivity/specificity and error types. For AI models, implementation should include external, prospective validation, patient-level splits to prevent leakage, clear reporting of confusion matrices, and drift monitoring after go-live. Future research should within prioritized multicenter, real-world studies that compare platforms head-to-head, evaluate patient-centered outcomes (time to diagnosis, downstream testing, costs), and address known pain points (lymphoblast vs lymphocyte misclassification, immature granulocytes, plasma cells) through data diversity, rare-class augmentation, and transparent model governance. With these safeguards, digital morphology can safely enhance leukemia detection by improving triage speed and consistency while preserving the diagnostic judgment of experienced morphologists.

REFERENCES

1. Xing Y, Liu X, Dai J, Ge X, Wang Q, Hu Z, et al. Artificial intelligence of digital morphology analyzers improves the efficiency of manual leukocyte differentiation of peripheral blood. BMC Med Inform Decis Mak. 2023;23:50.

2. Kim H, Hur M, d'Onofrio G, Zini G. Real-world application of digital morphology analyzers: practical issues and challenges in clinical laboratories. Diagnostics (Basel). 2025;15(6):677
3. Kowald A, Young WR, Musha R, Shibeeb S, Schüffler PJ, Marr C, et al. The diagnostic performance of the CellaVision DC-1 digital morphology analyser on leukaemia samples. Diagnostics (Basel). 2025;15(16):2029.
4. Kasim S, Shuaib A, Hussain S, Khan MA, Imran M, Khan S, et al. Multiclass leukemia classification in peripheral blood smear images using a hybrid CNN approach. Sci Rep. 2025;15:23782.
5. Atteia G, Ghneim R, Obeid F, Al Alami A, Al Ali A, Al Hossainy R, et al. A robust AI pipeline for leukocyte classification on peripheral blood smears. Sensors (Basel). 2022;22(15):5520.
6. Huang J, Chen L, Liu Z, Li J, Wang Y, Zhou Q, et al. Deep learning for peripheral blood smear-based detection of acute leukemia: a systematic appraisal of current approaches. Front Oncol. 2022;12:904292.
7. GBD 2017 Leukemia Collaborators. Global, regional, and national burden of leukemia, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet Haematol. 2020;7(7):e595-e610.
8. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. CA Cancer J Clin. 2022;72(1):7-33.
9. Bawazir A, Al-Zamel N, Amen A, Akiel MA, Alhawiti NM, Alshehri A. The burden of leukemia in the Kingdom of Saudi Arabia: 15 years period (1999-2013). BMC Cancer. 2019;19:703.
10. Panozzo B, Vestri AR, Galatola G, Fois F, Sanna A, Cocco P, et al. A critical analysis of CellaVision systems in the modern hematology laboratory. Am J Clin Pathol. 2025;164(2):163-172.
11. van der Vorm LN, Hendriks HA, Smits SM. Performance of the CellaVision DC-1 digital cell imaging analyser for differential counting and morphological classification of blood cells. J Clin Pathol. 2023;76(3):194-201.
12. Lee G-H, Yoon S, Nam M, Kim H, Hur M. Performance of digital morphology analyzerCellaVision DC-1. Clin Chem Lab Med. 2023;61(1):133-141.
13. Zhao Y, Zhang L, Huang S, Yi L, Shi W, Wang L, et al. Performance evaluation of the digital morphology analyser Sysmex DI-60 for white blood cell differentials in abnormal samples. Sci Rep. 2024;14:14344.
14. Hollenstein M, Wiedmeier S, Lehmann T, Seidel P, Kubesch A, Egli K, et al. Evaluation of the Sysmex DI-60 digital cell imaging analyzer on Wright-stained samples with focus on prevalence-dependent quality criteria. Int J Lab Hematol. 2024;46(1):83-91.
15. Nam M, Yoon S, Hur M, Lee G-H, Kim H, Park M, et al. Digital morphology analyzer Sysmex DI-60 vs. manual counting for white blood cell differentials in leukopenic samples: a comparative assessment of risk and turnaround time. Ann Lab Med. 2022;42(4):398-405.
16. Kweon OJ, Lim YK, Lee M-K, Kim HR. Red and white blood cell morphology characterization and hands-on time analysis by the digital cell imaging analyzer DI-60. PLoS One. 2022;17(4):e0267638.
17. Jiang H, Xu W, Chen W, He J, Jiang H, Mao Z, et al. Performance of the digital cell morphology analyzer MC-100i in a multicenter study in tertiary hospitals in China. Clin Chim Acta. 2024;555:117801.
18. Merino A, Sanz C, Remacha ÁF, Ferrer A, Pino M, Fernández C, et al. Performance of the new MC-80 automated digital cell morphology analyzer. Int J Lab Hematol. 2024;46(1):72-82.
19. Zhang F, Wang X, Li H, Chen Y, Zhou J, Liu S, et al. Artificial intelligence-assisted early screening of acute promyelocytic leukaemia in blood smears: a prospective evaluation of MC-100i. Front Oncol. 2025;15:1572838.
20. Shin E, Hur M, Kim H, Lee G-H, Hong M-H, Nam M, et al. Performance assessment of Sysmex DI-60: is digital morphology analyzer reliable for white blood cell differentials in body fluids? Diagnostics (Basel). 2024;14(6):592.
21. Kratz A, Lee S-H, Zini G, Riedl JA, Hur M, Machin S; International Council for Standardization in Haematology. Digital morphology analyzers in hematology: ICSH review and recommendations. Int J Lab Hematol. 2019;41(4):437-447.
22. Kratz A, Bengtsson H-I, Casey JE, Keefe JM, Beatrice GH, Grzybek DY, et al. Performance evaluation of the CellaVision DM96 system: WBC differentials by automated digital image analysis supported by an artificial neural network. Am J Clin Pathol. 2005;124(5):770-781.
23. Briggs C, Longair I, Slavík M, Thwaite K, Mills R, Thavaraja V, et al. Can automated blood film analysis replace the manual differential? An evaluation of the CellaVision DM96 system. Int J Lab Hematol. 2009;31(1):48-60.
24. Cornet E, Perol J, Troussard X. Performance evaluation and relevance of the CellaVision DM96 system in routine analysis and in patients with malignant hematological diseases. Int J Lab Hematol. 2008;30(6):536-542

25. Rollins-Raval MA, Raval JS, Contis L. Experience with CellaVision DM96 for peripheral blood differentials in a large multi-center academic hospital system. J Pathol Inform. 2012;3:29.
26. Kim HN, Hur M, Kim H, Kim SW, Moon HW, Yun YM. Performance of automated digital cell imaging analyzer Sysmex DI-60. Clin Chem Lab Med. 2018;56(1):94-102.
27. Park SJ, Yoon J, Kwon JA, Yoon SY. Evaluation of the CellaVision Advanced RBC Application for detecting red blood cell morphological abnormalities. Ann Lab Med. 2021;41(1):44-50.
28. Makhija K, Lincz LF, Attalla K, Scorgie FE, Enjeti AK, Prasad R. White blood cell evaluation in haematological malignancies using a web-based digital microscopy platform. Int J Lab Hematol. 2021;43(6):1379-1387.
29. Boldú L, Merino A, Alférez S, Molina A, Acevedo A, Rodellar J. Automatic recognition of different types of acute leukaemia in peripheral blood by image analysis. J Clin Pathol. 2019;72(11):755-761.

Table 1. Characteristics and key findings of the studies included in the review on Digital Morphology for Peripheral Blood Smears in Leukemia Detection.

Study Reference	Study Design	Population	Intervention / Exposure	Disease / Condition	Main Outcomes
11] Kowald et al., 2025	Cohort (diagnostic accuracy)	Leukemia-enriched peripheral smears	CellaVision DC-1 digital morphology	Leukemia (mixed)	Specificity >94% most leukocytes; blast sensitivity low (to 21%); κ=0.52.
[12] Zhao et al., 2024	Cohort (diagnostic accuracy)	Abnormal WBC differentials	Sysmex DI-60 digital morphology	Hematologic abnormalities	High agreement at moderate WBC; performance dropped in leukocytosis/leukopenia .
[13] Jiang et al., 2024	Multicenter cohort	Tertiary hospitals; routine smears	Mindray MC-100i digital morphology	Leukocyte differentials	κ>0.96 (normal WBCs); r>0.93 post-verification (abnormal differentials).
[14] Xu et al., 2023	Cohort (before-after AI assist)	Routine PBS reviews by technologists	AI-assisted digital morphology workflow	Leukocyte differentials	Review time −215 s/smear; accuracy improved .
[15] Zini et al., 2023	Cohort (diagnostic accuracy)	Neoplastic/reactive samples	Mindray MC-80 digital morphology	Hematologic malignancy screening	Immature granulocytes sensitivity 98.8%; NPV ≥96% after verification.
[16] Merino et al., 2024	Cohort (platform comparison)	Mixed pathology; 100 acute leukemia	MC-80 vs DM9600 digital morphology	Leukemia (acute/mixed)	Blast sensitivity 0.984; specificity 0.640; verification required.
[17] Li et al., 2025	Prospective cohort	Suspected APL cases	MC-100i AI-assisted screening	Acute promyelocytic leukemia	Sensitivity 95.8%; specificity 100.0% for abnormal promyelocytes.
[18] Aktekin et al., 2025	Prospective pediatric cohort	Pediatrichematology; =12,000 images	AI-based smear analysis	Leukemia suspicion (pediatric)	High performance reported; exact sensitivity/specificity .
[19] Dese et al., 2021	Cohort (algorithm validation)	Mixed leukemia and controls	ML classifier on PBS images	Leukemia subtype classification	Sensitivity 97.86%; specificity 100% (held-out test set).
[20] Yan et al., 2025	Cohort (algorithm validation)	Confirmed leukemia cases	Deep learning on single PBS images	APL/ALL typing	APL recall 97.4%; subtype typing high; full metrics .

Abbreviations: PBS = peripheral blood smear; WBC = white blood cell; APL = acute promyelocytic leukemia; AI = artificial intelligence; ML = machine learning; κ = Cohen's kappa; r = correlation coefficient; NPV = negative predictive value.