



Chinese Pharmaceutical Association
Institute of Materia Medica, Chinese Academy of Medical Sciences

Acta Pharmaceutica Sinica B

www.elsevier.com/locate/apsb
www.sciencedirect.com



REVIEW

Application of artificial intelligence in laboratory hematology: Advances, challenges, and prospects



Hongyan Liao^{a,b,c,†}, Feng Zhang^{a,b,†}, Fengyu Chen^{a,b}, Yifei Li^d,
Yanrui Sun^d, Darcée D. Sloboda^c, Qin Zheng^{a,b}, Binwu Ying^{a,b,*},
Tony Hu^{c,*}

^aDepartment of Laboratory Medicine, Clinical Laboratory Medicine Research Center, West China Hospital, Sichuan University, Chengdu 610041, China

^bSichuan Clinical Research Center for Laboratory Medicine, Chengdu 610041, China

^cCenter for Cellular and Molecular Diagnostics and Department of Biochemistry and Molecular Biology, Tulane University School of Medicine, New Orleans, LA 70112, USA

^dWest China School of Medicine, Sichuan University, Chengdu 610041, China

Received 7 December 2024; received in revised form 15 February 2025; accepted 28 March 2025

KEY WORDS

Artificial intelligence;
Machine learning;
Deep learning;
Laboratory hematology;
Diagnosis;
Prognosis;
Clinical workflow;
MICM classification

Abstract The diagnosis of hematological disorders is currently established from the combined results of different tests, including those assessing morphology (M), immunophenotype (I), cytogenetics (C), and molecular biology (M) (collectively known as the MICM classification). In this workflow, most of the results are interpreted manually (*i.e.*, by a human, without automation), which is expertise-dependent, labor-intensive, time-consuming, and with inherent interobserver variability. Also, with advances in instruments and technologies, the data is gaining higher dimensionality and throughput, making additional challenges for manual analysis. Recently, artificial intelligence (AI) has emerged as a promising tool in clinical hematology to ensure timely diagnosis, precise risk stratification, and treatment success. In this review, we summarize the current advances, limitations, and challenges of AI models and raise potential strategies for improving their performance in each sector of the MICM pipeline. Finally, we share perspectives, highlight future directions, and call for extensive interdisciplinary cooperation to perfect AI with wise human-level strategies and promote its integration into the clinical workflow.

*Corresponding authors.

E-mail addresses: yingbinwu@scu.edu.cn (Binwu Ying), tonyhu@tulane.edu (Tony Hu).

[†]These authors made equal contributions to this work.

Peer review under the responsibility of Chinese Pharmaceutical Association and Institute of Materia Medica, Chinese Academy of Medical Sciences.

<https://doi.org/10.1016/j.apsb.2025.05.036>

2211-3835 © 2025 The Authors. Published by Elsevier B.V. on behalf of Chinese Pharmaceutical Association and Institute of Materia Medica, Chinese Academy of Medical Sciences. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Laboratory hematology is an important branch of medical science that provides information about hematologic (blood) disorders, thereby assisting in the diagnosis, differential diagnosis, and prognosis of these disorders, as well as treatment planning and treatment efficacy monitoring. This branch of science uses laboratory testing methods to analyze cellular, molecular, and biochemical components in various biospecimens, such as peripheral blood (PB), bone marrow (BM), urine, or pleural effusion. Laboratory hematology plays a critical role in diagnosing and monitoring a wide range of hematologic disorders. Among these, hematologic neoplasms are a group of heterogeneous neoplastic disorders marked by clonal expansion of abnormal cells in the BM, lymph nodes, and/or blood^{1–4}. For these malignancies, timely diagnosis is vital for successful treatment. In addition to hematologic neoplasms, laboratory hematology is also essential for diagnosing and managing non-neoplastic disorders such as anemia, infections, acute hemorrhagic states, allergies, and immunodeficiencies, which are equally critical in clinical practice. Following the guidelines of the World Health Organization (WHO), clinical hematology laboratories use a conventional approach for diagnosis, evaluating a patient's blood parameters, BM aspirate, or the results of tissue biopsies to test for tumor cells. The combined results of different tests, which include those assessing morphology (M), immunophenotype (I), cytogenetics (C), and molecular biology (M), are collectively known as the MICM classification. This workflow may also provide evidence for differential diagnosis, risk stratification, and targeted therapy.

To conduct this workflow, the clinical hematology laboratory is incorporated with multi-disciplinary diagnostic tools and methods⁵. This requires hematopathologists or technicians with outstanding expertise and experience—both of which are a lacking resource in healthcare centers. Despite the emergence of cutting-edge diagnostic technologies, most of the results still have to be manually interpreted (*i.e.*, by a human, without automation), which renders a long turnaround time (TAT). Moreover, the advent of new technologies and instruments has brought about data of increasing dimensionality and throughput, such as from the electronic health record (which includes laboratory results, imaging studies, and diagnosis codes) and genetic testing⁶. And not only have hematologic neoplasms increased in prevalence but global pandemics showing hematological alterations have emerged (*e.g.*, the COVID-19 pandemic), causing more complexity in the clinical hematology laboratory⁷. Therefore, establishing a more efficient, accurate, and traceable approach, an approach that also requires less human guidance in laboratory hematology, is mandatory to tackle these emerging health concerns.

Artificial intelligence (AI) is a branch of computer science that is usually aimed at both understanding and building intelligent entities *via* software programs⁸. AI can assist in big data analysis, disease pattern identification, early screening and confirmative diagnosis, and patient survival prediction in healthcare settings. Generally recognized as a subbranch of AI, machine learning (ML) is widely defined as a computational strategy with an array

of algorithms to achieve AI. Basic learning methods of ML include supervised learning, unsupervised learning, and reinforced learning, from which more hybrid methods can be derived. The workflow of ML generally comprises study design, data acquisition, data cleaning and preprocessing, model construction and training, optimization of the model (adding more steps for processing the data based on continuous performance monitoring), and model deployment (Fig. 1). To obtain robust performance, ML models must be trained on plenty of high-quality data.

ML classifiers have created new opportunities for accurate, data-intensive science across multiple disciplines, and their integration into the medical care field has expanded to several areas, including clinical trials and research; disease diagnosis, prediction, and treatment; smart patient management systems; and AI-powered chatbots and telemedicine. As many firmly believe, AI and ML have already transformed clinical medicine by improving the quality of care, and they have great potential to further improve medical care in the near future and beyond¹⁰.

In clinical hematology, substantial progress has been made in the utilization of AI for diagnostic purposes, in areas including instrument research and development, workflow improvement, and result interpretation. Proof-of-principle studies have been conducted for multiple tests, for those assessing morphology, immunophenotype, cytogenetics, and molecular biology (the MICM classification) in addition to other auxiliary techniques. These promising studies suggest that clinical hematology—traditionally associated with numerous specimens, long TAT, tedious manual intervention, and challenging comprehensive result interpretation—will be revolutionized with AI.

In this review, we aim to introduce the trending AI models and focus on how they would influence laboratory hematology. Specifically, we highlight how AI models influence primary tests or techniques associated with hematology diagnosis by providing examples associated with specific hematological disorders and briefly mention how AI has helped advance prognosis evaluation and treatment decision-making. We also discuss opportunities and challenges to give a comprehensive perspective of future AI-assisted laboratory hematology, which aims to accurately diagnose hematological disorders in a timely manner.

2. Complete blood count (CBC)

2.1. CBC and its related challenges

The complete blood count (CBC) is one of the most commonly ordered laboratory tests and usually the first test ordered for disease screening and checkups, initiating practical diagnostic algorithms and thereby assisting in diagnosis¹¹. CBCs are ordered for a variety of conditions, the most common being hematological abnormalities including anemia, thrombocytopenia, leukopenia, polycythemia, thrombocytosis, and leukocytosis. Thus, they are helpful for diagnosing anemia, certain cancers (hematological malignancies), infections, acute hemorrhagic states, allergies, and immunodeficiencies; for treatment monitoring; and for routine

checkups. Leveraging the CBC in a cost-effective manner has helped speed diagnosis, track disease progression, determine therapeutic response, and adjust treatment strategy.

In the context of CBC analysis, several challenges remain that may lead to delays of CBC reports, or even worse, misdiagnoses. One of these challenges is the need for recognizing abnormal data effectively and precisely. CBC data is complex; it includes counts of red blood cells (RBCs), white blood cells (WBCs), and platelets (PLTs) as well as the WBC differential and the volume distribution of RBCs. The data often give the first clue of hematologic malignancy—an abnormal cell count or abnormal morphology¹². According to the “41 rules,” a consensus guideline published by the International Society for Laboratory Hematology (ISLH), slide review is the first action suggested when encountering an abnormal parameter or flag¹³. But assessing cell morphology by manual microscopy, the long-standing gold standard for CBC review, is inefficient and requires laboratory hematologists with accumulated experience. Addressing these issues, an automated hematological analyzer is now widely accessible that provides not only indices but also flags, histograms, and scattergrams, which may be helpful for accurate diagnosis. Using these types of tools, physicians or laboratory technologists can look beyond the common indices or the next prescribed “rule” and interpret the CBC outputs in a comprehensive manner. Another challenge is the need for an efficient processing framework to reduce TAT despite the increasing number of specimens. This challenge could be addressed by point-of-care (POC) testing, which could help obtain investigative results without delay¹⁴, and there is an increasing demand for the development of POC CBC analyzers with easier operation, higher throughput, and greater accuracy. AI has showed outstanding performance in extracting explicative features and hidden patterns from complex medical data¹⁵. Recently, efforts have been made to incorporate AI into the analysis of parameters related to hematologic neoplasms, and the relevant literature is summarized in Table 1¹⁶⁻²⁶.

2.2. Screening, diagnosis, and prediction of hematologic malignancies by AI trained on CBC results

Timely diagnosis and treatment is vitally important for a better outcome of leukemia, especially in cases of acute promyeloblast leukemia (APL). In pursuit of an early detection method for APL, Haider et al.^{16,17} developed a CBC-parameter-driven artificial neural network (ANN) predictive model based on data collected from a massive number of study subjects. They found that using the CBC digit parameters, including PLT count, immature fraction of PLTs (IPF), and DNA/RNA content-related neutrophils could assist in the early detection of APL¹⁷. More recently, Alcazer et al.¹⁸ developed an extreme gradient boosting (XGBoost) model aimed at predicting 3 main acute leukemia subtypes based on CBC and routine biochemical parameters. This model demonstrated high area under the (receiver operating characteristic) curve (AUC) and accuracy in the diagnosis of APL, acute myeloid leukemia (AML), and acute lymphoblastic leukemia (ALL), having the potential to assist in guiding initial treatment decisions, particularly in situations where cytological expertise is limited¹⁸. AI has also been used to predict chronic myelogenous leukemia (CML) from CBC results. CML often presents with left shift and unexplained leukocytosis, but almost half of diagnosed CML cases begin asymptotically, prompting researchers to explore whether blood cell counts prior to diagnosis could predict CML. One group, Hauser et al.¹⁹, used XGBoost and the least absolute

shrinkage and selection operator (LASSO) to construct an ML model based on historical CBC results from 1623 patients. They concluded that blood cell counts collected up to 5 years prior to the diagnostic workup of CML successfully predicted the BCR-ABL1 test result¹⁹. Thus, CBC results have clear potential for predicting CML status, with the counts of basophils, leukocytes, and neutrophils being especially important. While these researchers utilized the digit parameters in ML model training, our group reported an innovative CBC framework consisting of a convolutional neural network (CNN) model, ResNet-18, that quantitatively mapped CBC scattergrams to quickly and robustly indicate probable susceptibility to APL and, like the ANN and ML models, showed great sensitivity, specificity, and precision²⁰. Thus, for diseases that require rapid and prompt diagnosis, or that show prominent alterations in PB, building AI models based on CBC outputs shows clinical promise.

In addition to aiding early diagnosis, AI models trained on CBC results have been used for differential diagnosis, when 2 or more conditions share similar signs or symptoms. For example, parameters of cell population data, including morphological parameters and those related to the immature leukocyte fraction, have been used to train AI models to differentiate acute from chronic—as well as myeloid from lymphoid—leukemias¹⁶, which increases one’s confidence that a CBC result can show the “disease fingerprint” of leukemia. In the case of myelodysplastic syndromes (MDS), hallmarks of hematologic complications generally include macrocytic anemia and ineffective or dysplastic hematopoiesis. But because there are numerous other neoplastic or nonneoplastic causes of cytopenia, the diagnosis of MDS is particularly challenging. Blood smear examinations recommended by the ISLH¹³ have not been satisfactory, possibly because a high percentage of slides are skipped from recheck rules. Thus, incorporating cell population data may add value to the laboratory diagnostic work-up of MDS. Supporting this idea is the seminal work describing the “MDS-CBC score,” a promising tool incorporating 3 parameters: absolute neutrophil count (ANC), structural neutrophil dispersion (Ne-WX), and mean corpuscular volume (MCV). The score was shown to exclude or suggest MDS in patients with cytopenia from unknown causes at the time of CBC assessment²⁷. Inspired by this, Zhu et al.²¹ used an ML approach named Breiman’s random forest (RF) to predict MDS and highlighted Ne-WX and IPF as the 2 strongest discriminatory predictors of MDS among patients with cytopenia.

2.3. Risk stratification, prognosis prediction, and treatment monitoring by AI using CBC parameters

For patients with confirmed hematological disorders, CBCs are routinely ordered for treatment monitoring. These CBC results may serve as a valuable data source for AI algorithms, which can integrate these results with others to identify risk factors and underlying causes of disease. For example, Mahmood et al.²² examined the blood biochemistry (including CBC results), phenotypic data, and environmental risk factors of pediatric patients with ALL to reveal associations and enable prediction of the disease. In their study, 4 supervised ML algorithms (classification and regression trees (CART), RF, gradient boosted machine (GBM) and C5.0) were evaluated, and CART provided the best fit for the entire data set with high accuracy. And notably, PLT abnormality was demonstrated as the main attribute in predicting pediatric ALL²².

Of note, ML algorithms have also demonstrated a comprehensive contribution to the prognosis and management of hematologic neoplasms. One example is LASSO for the prognosis of diffuse large B-cell lymphoma (a subtype of heterogeneous non-Hodgkin lymphoma), the course of which had been challenging to predict. In a retrospective multicenter study²³, results indicated that, compared to widely used existing prognostic models, LASSO was more accurate for risk stratification of patients with diffuse large B-cell lymphoma. In addition to age, gender, stage of disease, and central nervous system involvement, LASSO identified WBC count and hemoglobin level as new risk factors for this form of lymphoma²³. More recently, 9 ML models were used to predict the complete remission and 2-year overall survival of AML with multimodal clinical, laboratory, cytogenetic, and molecular genetic data²⁴. Notably, hemoglobin at initial diagnosis was predictive of both complete remission and 2-year overall survival²⁴. The researchers proposed to incorporate 9 ML classifiers (instead of 1) into the pipeline to facilitate greater transferability of the methodology²⁴. Although the models from these studies were limited by their retrospective nature, they have been externally validated, providing proof of concept that ML algorithms based on CBC parameters can be used for a decision support system in PB.

2.4. Integration of AI into POC devices for CBC analysis

Despite the CBC ranking among the most ordered tests, POC devices capable of performing CBCs have lagged²⁸, largely because automating the morphology check has been challenging. This has fostered novel hematological platforms integrated with AI for POC CBC tests. One example is the Sight OLO, a POC CBC test based on recent advances in AI and computerized image analysis (computer vision), designed and validated in a multicenter setting²⁵. Requiring only 27 μ L of whole blood, the Sight OLO hematology analyzer provides a 19-parameter, 5-part differential CBC. It shows strong concordance to the Sysmex XN-1000 and supports fingerprick capillary blood analysis, but challenges remain in ensuring consistent performance across all parameters and managing user errors²⁵. Another POC CBC test, the Hilab System, uses microscopy and chromatography techniques to analyze blood cells and hematimetric parameters and combines these techniques with AI, ML, and deep learning (DL) to provide the main parameters evaluated in the CBC²⁶. Similar to the OLO system, this analyzer shows a strong correlation with the Sysmex XE-2100 for most parameters ($r \geq 0.9$) and offers similar advantages: it requires a small sample volume (90 μ L) and supports both venous and fingerprick blood samples. Overall, both devices provide CBC results with accuracy, repeatability, and flagging capabilities comparable to that of conventional automated hematological analyzers, but work faster, require smaller sample volumes, and allow for less invasive sample collection^{25,29}. Thus, AI integration may facilitate the development of POC CBC platforms, overcoming the challenges attributed to their larger relatives: requirements for liquid reagent replacement, washout and calibration procedures, and frequent quality control processes.

2.5. Challenges and future directions of AI-assisted CBC analysis

One of the primary challenges in AI-assisted CBC analysis is the reliance on extensive, high-quality training data to ensure robust ML models. The current models are usually based on retrospective

data, and lack validation in large and diverse cohorts from external centers, which limits their broad utility. To address this, hematologic professionals must continue to validate these models through microscopic evaluation and data interpretation across varied populations. Beyond data requirements, integrating AI into POC devices for CBC analysis faces several technical and practical hurdles. These include the complexity of technical integration, the absence of standardized data formats, the need for more generalizable models, and the critical task of addressing data bias. Additionally, cost constraints and data privacy concerns further complicate the widespread adoption of AI in clinical settings.

Despite these challenges, AI has demonstrated significant promise in areas such as diagnosis, risk assessment, prognosis prediction, and POC test development. To fully realize this potential, future efforts should focus on developing disease-specific and mixed-disease AI models, validating them with larger, multicenter datasets, and conducting prospective studies. These steps are essential to bridge the gap between proof-of-concept results and their translation into routine clinical practice.

3. Cytomorphology

3.1. Cytomorphology and the benefits of AI

Cytomorphology examination refers to the use of a microscope to interpret and analyze the types, quantities, and morphological changes of blood cells to quickly assess smears of PB or BM³⁰. Considering the vast majority of specimens reflect normal, reactive/non-neoplastic, or other pathological conditions, the main task of PB cytomorphology examination is to screen for potential cases of hematological neoplasms and subject the specimens to further analysis. In contrast, BM cytomorphology often confirms a diagnosis, at least for certain disease types such as acute leukemia according to the French–American–British (FAB) classification³¹. Both PB and BM cytomorphology methods have drawbacks; they are labor-intensive and prone to error, and they may result in significant interobserver variability, especially when the workload is high³². Moreover, the training of qualified technicians takes a long time, and senior personnel are lacking.

For these reasons, the examination of cytomorphology has benefitted from ML, with numerous models demonstrating outstanding performance, as shown in Table 2³³⁻⁵⁶. ML algorithms are usually trained with massive amounts of data annotated by experts. These algorithms perform tasks automatically, requiring little human instruction or intervention, and they can handle a high workload, one beyond human capability. Accordingly, the incorporation of AI into the clinical workflow of cytomorphology has been initiated early and successfully.

3.2. PB morphology recognition by automated digital imaging instruments

To capture high-resolution images of cells and convert them into a digital format for storage and analysis, digital microscopes have been developed that integrate advanced image processing software or algorithms. One example is the CellaVision digital microscope series, which employs ML algorithms such as ANN and DL models to efficiently and accurately classify WBCs and RBCs for PB smear analysis, reducing the total analysis time per smear by approximately 50% and achieving an overall pre-classification

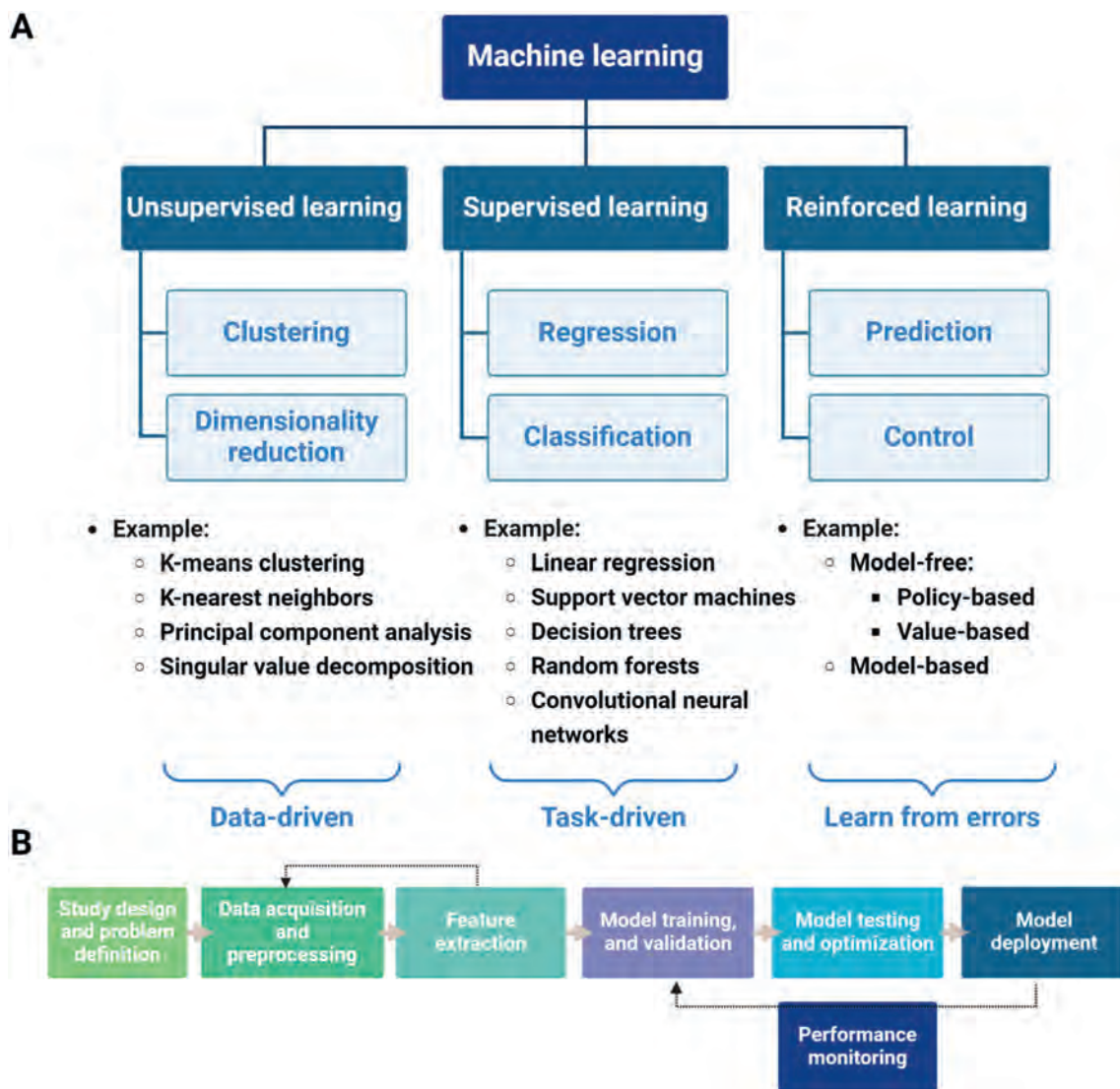


Figure 1 Overview of the machine learning (ML) method categories, applications, and corresponding algorithm examples (A) and ML workflow for effective data processing and model building (B). (A) Unsupervised learning is a form of data-driven ML that finds patterns in unlabeled input data, attempting to make sense of the data by extracting useful features and patterns on its own. Unsupervised ML is often used for clustering, dimensionality reduction and association tasks. Common unsupervised ML methods include k-means and k-nearest neighbor for clustering, and principal component analysis for dimensionality reduction. Supervised learning is a form of task-driven ML in which models are typically trained with a fully labeled data set. Supervised learning is used for classification tasks, which involve mapping inputs to a discrete value (*i.e.*, class label), and regression tasks, which involve mapping inputs to a continuous value. Examples of supervised ML methods include linear and logistic regression for regression tasks and support vector machines, decision trees, and random forests for classification tasks. Reinforced learning is a different form of ML—the model isn’t trained with a sample data set. Instead, it is a feedback-based ML, which involves learning from mistakes through trial and error. Reinforced learning is often used for prediction and control in problem domains where time and event sequences matter, feedback may be delayed, and actions have consequences. In reinforced learning, prediction involves predicting the performance of some policy, whereas control involves determining the optimal policy that yields the best performance⁹. (B) A generalized ML pipeline is shown. The first step is study design and problem definition, which requires thoroughly understanding the problem to be solved by gaining enough domain knowledge to make informed decisions for the remaining steps in the ML pipeline. Next, sufficient and representative data relevant to the problem are gathered. In the data preprocessing stage, the raw data are cleaned and converted to a machine-readable format in preparation for feature extraction. Then, the data are transformed for feature extraction, which ensures only the most salient information relevant to the problem will be used for training the ML model. Common feature extraction techniques include feature selection (*e.g.*, linear discriminant analysis), dimensionality reduction (*e.g.*, principal component analysis), and computer vision techniques for input images (*e.g.*, color, shape, and texture feature extraction methods). During model training and validation, the extracted features are used to train the ML model. Validation is a critical step in which a separate dataset (not used for training) is employed to evaluate the model’s performance during the training process. This helps to detect issues such as overfitting, where the model performs well on training data but poorly on unseen data. After validation, the model is further evaluated with an independent testing dataset that has not been used in either training or validation. The type of ML model(s) to use will depend on a variety of factors, such as the nature and amount of input data, format of the desired output, and the given problem’s associated ML

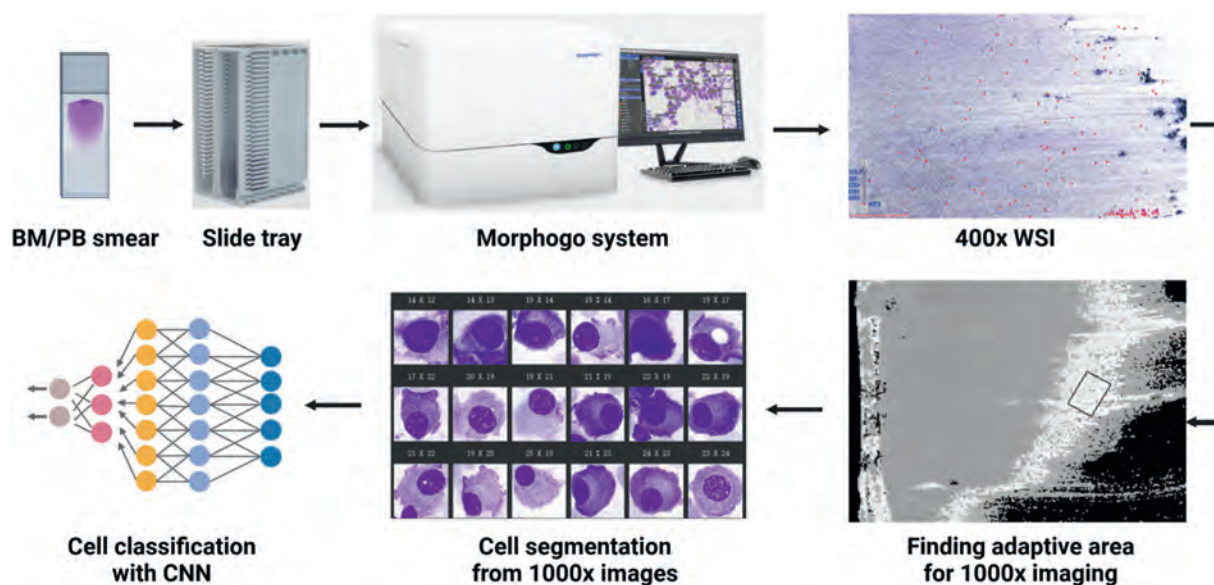


Figure 2 Workflow of Morpho to digitize smears and analyze nucleated cells. Adapted from Ref. 42 under a Creative Commons license (CC BY-NC).

accuracy of 92% compared to manual slide review^{57,58}. Moreover, this digital microscope series showed acceptable agreement with manual light microscopy for determining most erythrocyte morphologies⁵⁹. Another CNN-based predictive model, DysplasiaNet, was specifically designed to distinguish dysplastic neutrophils with fewer granules, one of the most challenging morphological abnormalities to detect in PB smears. When tested in a validation dataset consisting of dysplastic (3071) and normal (4237) neutrophil images selected by clinical pathologists, this model achieved a sensitivity above 95% and a specificity above 94%, with a global accuracy of 94.85%³⁴. Subsequently, the Sysmex DI-60, a fully automated digital morphology analyzer based on an ANN model, demonstrated acceptable accuracy—not only in identifying cells such as polychromatophilic erythrocytes, target cells, and ovalocytes but also in preclassifying WBCs from healthy specimens—while significantly reducing the hands-on time compared to that of manual slide review (reducing 144.1 s/slide for abnormal samples and 144.6 s/slide for normal samples)³⁵. The Sysmex XN series, which incorporates the Sysmex DI-60, achieved complete automation of the WBC classification process, enabling automated CBC, slide preparation and staining, and digital scanning with preclassification of cells³⁶. Additionally, the core performance of the MC-100i, another automatic digital cell morphology analyzer based on a CNN model, was evaluated in a multicenter study conducted across 11 tertiary hospitals in China³⁷. The study assessed key performance metrics of the analyzer, including its ability to classify normal and abnormal WBCs, RBCs, PLTs, and PLT clumps, as well as within-run precision and reading time. The MC-100i demonstrated excellent performance in the classification of all these blood cell types, achieving >90% sensitivity and specificity for RBCs and >97%

accuracy for normal WBCs compared with manual differential results³⁷. The abovementioned and similar products have been commercialized, revolutionizing clinical hematology laboratories routinely facing an extremely large and growing number of samples on a daily basis. Such automated cell morphology analysis systems are being progressively developed to streamline the process of locating, capturing, and analyzing cells on PB smears, providing classification recommendations and cell count results. Following this automated analysis, laboratory staff can review the results and manually correct any misclassified cells.

3.3. AI-assisted methods for PB and BM cytomorphology

AI technologies have been incorporated into the entire spectrum of applications in hematological cell morphology analysis, including automated cell detection, segmentation, feature extraction, classification, image enhancement, data augmentation, and leukemia subtype prediction. For instance, Alf  rez et al.³⁸ developed an automated system for detecting and segmenting different types of lymphoid cells in PB, achieving a classification accuracy of 97.67% for normal, reactive, and abnormal cells. Pohlkamp et al.³⁹ used ML to combine cell detection, segmentation, and feature extraction, achieving an overall accuracy of 91.7% and 93.4% for classifying critical pathological cell types. In the area of feature extraction and classification, Matek et al.⁴⁰ used CNNs to identify blast cells in AML, achieving a high classification accuracy. The system was able to differentiate between various subtypes of leukemia with a high degree of accuracy, and it successfully identified blast cells in PB smears⁴⁰. Chen et al.⁴¹ introduced a multi-level feature fusion method combined with deformable self-attention mechanisms, significantly enhancing the

task. In the next stage, model testing and optimization, the model is tuned to ensure readiness for deployment. Usually, testing evaluates the trained model on unseen data, while validation uses different data to evaluate the model during the training process. Such separate training, validation, and testing data ensures the model is not overtrained (*i.e.*, memorizing, rather than actually learning). After the ML model achieves an optimized performance on testing data and has been verified, the model is ready for deployment in the field. It may be necessary to continuously monitor its performance, which may lead to collecting additional data, and refining or retraining the model.

feature extraction process. Their model improved classification performance by addressing scale disparity and extracting global features, achieving superior results compared to other leukocyte detection models. As image enhancement and data augmentation are key techniques for improving model performance, Boldú et al.⁵⁵ used image enhancement methods along with color clustering and mathematical morphology to process cell images, extracting features and performing classification. This approach improved the recognition accuracy of different types of acute leukemia cells, reaching an overall classification accuracy of 94% when compared with the true confirmed diagnosis. Fu et al.⁵⁴, meanwhile, developed the Morphogo system, which reached an accuracy of over 85.7% in classifying hematopoietic lineage cells, contributing to BM analysis. AI technologies have also been successfully applied in predicting leukemia subtypes by combining the multiple tasks mentioned above. A recent study by Lincz et al.⁶⁰ compared three versions of AI algorithms for classifying acute leukemia in abnormal PB films. Their results demonstrated significant improvements in AI performance over time, with sensitivity for blast identification increasing from 97%

to 100%, and the specificity for blast identification improving from 24% (AI1) to 12% (AI3)⁶⁰.

While cell recognition is important for disease diagnosis, it also has prognostic value. For example, increased levels of circulating plasma cells have been identified as a reliable prognostic factor for the progression of multiple myeloma, as well as for patient survival, and the number of these cells helps assess measurable/minimal residual disease (MRD)⁶¹. In a recent study⁴², researchers used a commercial digital imaging system, Morphogo, to capture thousands of nucleated cells in each PB smear *via* whole slide imaging. They built a CNN model with EfficientNet as its backbone to refine the capacity of Morphogo to identify circulating plasma cells, positioning it as a potential method for effective circulating plasma cell screening *via* PB smears (Fig. 2)⁴². When identifying circulating plasma cells, Morphogo showed a sensitivity of 89.0%, a specificity of 99.7%, and an accuracy of 99.7%, and its performance in PB smear detection of circulating plasma cells was always superior to pathologists at different sensitivity thresholds (0.600 and 0.870).

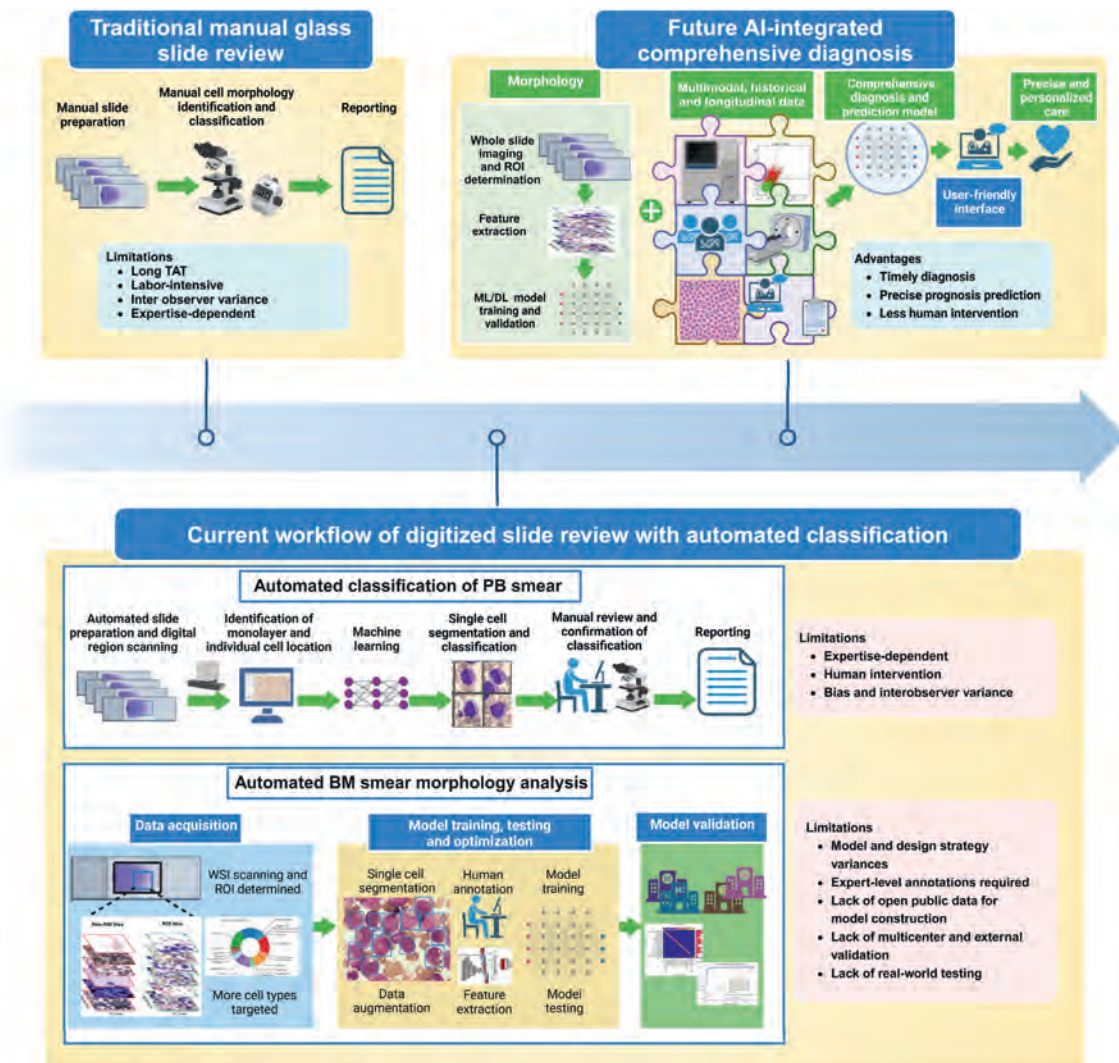


Figure 3 Schematic overview of the development of AI-assisted cytomorphology analysis and AI integration into a comprehensive diagnostic workflow for hematological disorders. BM: bone marrow; PB: peripheral blood; ROI: region of interest; TAT: turnaround time; WSI: whole slide imaging.

Compared to assessing PB morphology, assessing BM morphology has been more difficult to automate. BM is complex, with mixed cell classes having different maturation stages. Additionally, the quality of BM sample aspiration, slide-making, and staining can vary. And malignancy-related pathological alterations such as dysplasia and fibrosis can further affect cell recognition and classification. Furthermore, many of the automated systems or algorithms currently available have been trained primarily on physiological cell types or PB smears and may not be directly applicable to BM morphology.

Despite these challenges, automating the examination of BM morphology has seen some progress. In early studies, hematopathologists extracted single-cell features from digitized images and used them to classify the cells in question. For example, Krappe et al.⁴³ developed an ML classifier by using the hierarchical tree model for 16 cell types from BM and reported an overall accuracy of 66.3% with 46,189 cells. In a more recent study⁴⁶, researchers applied a CNN based on a large data set of 171,374 expert-annotated single-cell images of BM to 945 patients diagnosed with a variety of hematological diseases and remarkably, 22 classes of BM leukocytes could be automatically classified. As the authors of these studies point out, enlarged datasets with expert annotation could be an important resource, not only for educational purposes but also for the future development of automated image-based BM cell classification systems, should these datasets become available to be public.

Another promising direction for BM automation involves the use of whole slide images digitized by computer-assisted image analysis systems. This approach offers high-resolution images that capture the entire BM smear, which is advantageous for comprehensive analysis. In an initial study, Wang et al.⁶² proposed a hierarchical patch-based DL framework with two Cascade R-CNN DL models, adopting a coarse-to-fine strategy. This framework rapidly located BM particles and cellular trails, identified BM cells in whole slide images, and subsequently identified multiple cell types, including megakaryocytes and mitotic cells, and detailed 4 stages of erythroblasts⁶². In both cross-validation and generalization tests, the accuracy of this framework for BM whole slide image analysis was higher than 98.8%, and in addition, the analysis time was greatly reduced, superior to existing small image-based benchmark methods. A consistent result was found by a contemporary study⁶³, which identified regions of interest first. This study utilized the You Only Look Once (YOLO) approach for model training to generate a “cytological fingerprint” called the Histogram of Cell Types, achieving high accuracy in region detection (97% accuracy) and cell detection and cell classification (75% mean AP)⁶³. Both studies were based on data from healthy people and from patients with different hematological disorders including MDS, acute leukemia, and plasma cell neoplasm. Hu et al.⁶⁴ also proposed a 2 stage-detection framework, this time from the whole slide image level to the patch level, which outperformed other models in almost all evaluation criteria

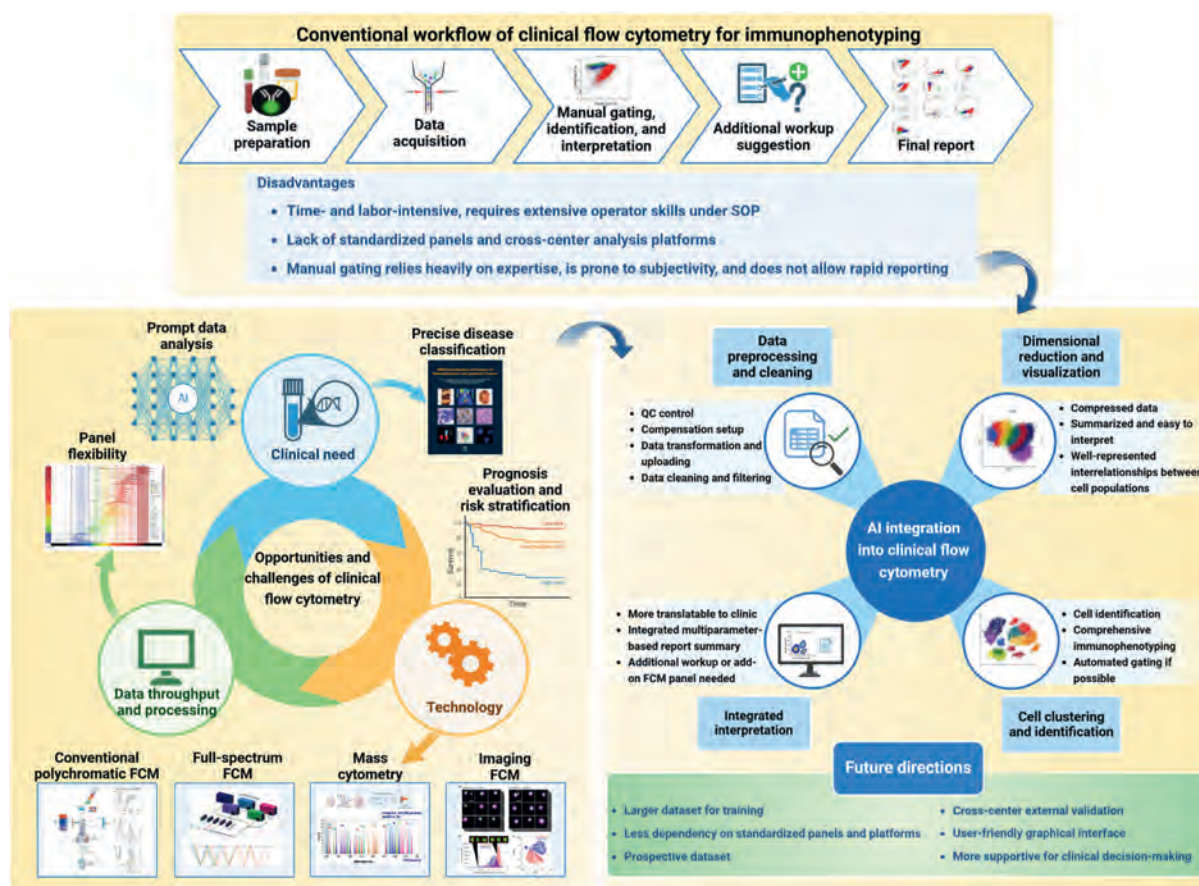


Figure 4 Schematic overview of current immunophenotyping workflow, opportunities and challenges of clinical flow cytometry, and AI integration into the immunophenotyping of hematological disorders, with prospects for future improvement. SOP: standard operating procedures; FCM: flow cytometry.

when used to evaluate binuclear cells from multicenter datasets. For example, when using Deep Layer Aggregation (DLA) as the backbone in a multicenter hybrid dataset, this framework has higher AP (0.798 vs. 0.781) and AP₅₀ (0.959 vs. 0.956) than the CircleNet model. Although this coarse-to-fine strategy was originally designed for binuclear cells in PB, it may offer a new method for BM-cell detection.

The integration of AI in BM analysis has achieved substantial advancements in single-cell feature extraction and whole-slide image analysis. Single-cell feature extraction analysis, providing in-depth insights at the cellular level, is of great significance for understanding cell heterogeneity and complexity. This technology can identify and analyze different cell subsets in the BM, including rare cell types, which is particularly important for studying the spatial anatomy of BM hematopoiesis. Additionally, single-cell feature extraction analysis leverages ML classifiers to classify BM cells with high accuracy, provided that large, well-annotated datasets are available. When available, these datasets serve as invaluable resources, not only for educational purposes but also for the iterative development of more sophisticated automated classification systems. However, these data sets are not always available, making the dependence on extensive datasets and the initial training costs significant limitations of single-cell feature extraction. In contrast to AI-enhanced single-cell analysis, AI-enhanced whole-slide image analysis provides a more comprehensive view of the BM smear, enabling detailed examination and classification of BM cells with high accuracy and efficiency. Also, it can achieve remarkable performance in detecting and classifying various BM cell types. However, the challenge of handling large-scale image data can hinder the capture of fine-grained cellular details, while also being affected by high resolution and artifact-related issues⁶⁵. In summary, both AI-assisted single-cell feature extraction and AI-enhanced whole-slide image analysis present unique strengths and limitations. The strategic combination of these approaches could harness their complementary advantages, fostering a more robust, efficient, and comprehensive BM analysis pipeline⁶⁶. As research continues to evolve, the integration of AI in BM analysis promises to revolutionize hematological research and diagnostics.

3.4. AI-assisted methods for hematologic malignancy diagnosis based on cytomorphology

Though there have been achievements in single-cell morphology recognition, what's most essential is a comprehensive diagnostic system to inform clinical decision-making. To this end, Kimura et al.⁴⁷ were the first to develop an automated image diagnostic system for MDS using PB smears. By combining a CNN-based image recognition system with an XGBoost decision-making system, the diagnostic system differentiated MDS from aplastic anemia with high sensitivity (96.2%) and specificity (100%), with an AUC of 0.990. Beyond the myeloid malignancies, efforts have been made in the diagnosis of ALL. When diagnosing this disease, features from both the cell nucleus and the cytoplasm may help. Assuming this, Neoh et al.⁵⁰ designed a novel clustering algorithm with stimulating discriminant measures of both within- and between-cluster scatter variances to first segment the nucleus and cytoplasm of lymphocytes/lymphoblasts and then consider features from both for WBC recognition. Different from previous algorithms that mainly focus on nuclei, this novel algorithm achieved superior recognition rates of 96.72% and 96.67%⁵⁰. Another DL classifier directly took in the values from image pixels and, through the use of multilayer architecture, slowly constructed useful features which were then used to recognize the patterns relevant to the disease background⁵¹. This DL approach has been applied to recognize lymphocytes and distinguish pre-T cells from pre-B cells. Its accuracy is comparable to the dominant approach of hand-crafted feature engineering with classification algorithms such as support vector machine, k-nearest neighbors, and multi-layer perceptron, so it still has great potential for lymphoblast cell image classification⁵¹.

Recent progress has also been made on differentiation diagnosis of mixed disease subtypes related to different genetic backgrounds. For example, one recently published DL-based pipeline, AMLnet, could discriminate not only between AML patients and healthy individuals but also between different AML subtypes based on BM images⁵³. The model comprised a variable output number of deep CNN modules to process images and a voting module to transform the image level to the patient level,

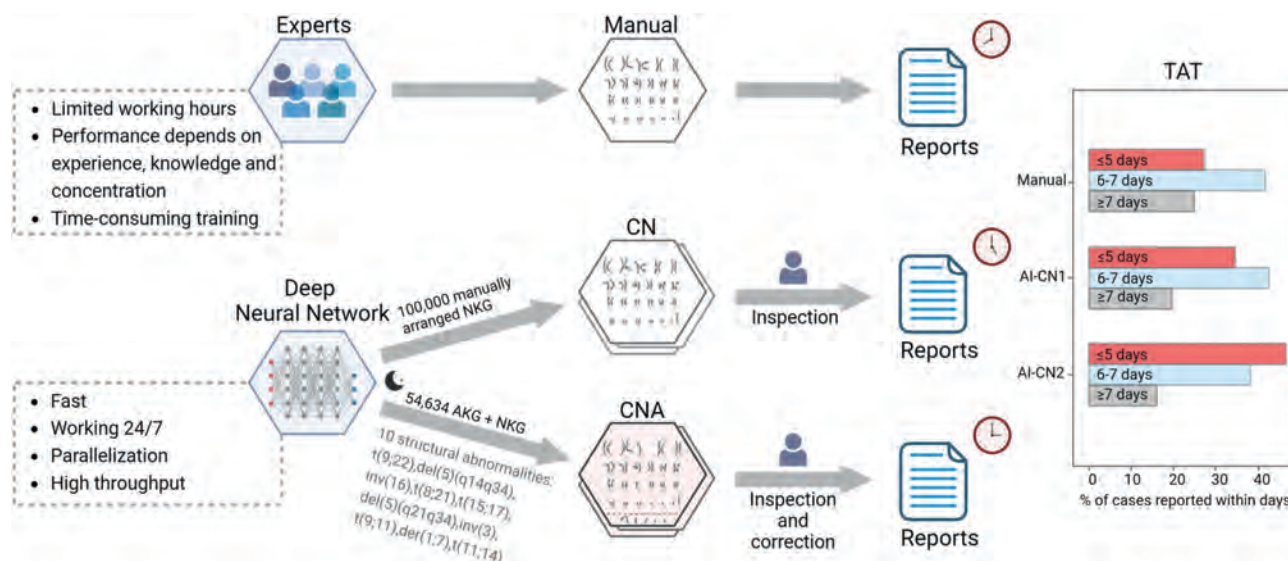


Figure 5 AI-enhanced chromosome banding analysis. Adapted with permission from Ref. 118. Copyright © 2020 American Society of Hematology.

Table 1 Summary of studies of AI in CBC.

ML model	Dataset	Application	Performance	Validation	Ref.
ANN: prediction/differentiation and classification	CBC and CPD parameters of 1577 hematological neoplasm subjects Various dataset partitions of 50, 60, 70 and 50, 40, 30 for training and testing	Differentiation of various types of leukemias	Accuracy: 83.1% (train sets) 89.4% (test sets) - AUC: 0.79–0.94	–	16
ANN: prediction and classification	CBC and CPD parameters of 1067 hematological neoplasm subjects	Early flagging of APL cases	Accuracy 95.7% (train sets) 97.7% (test sets) - AUC: 89.4%	–	17
AI-PAL (XGBoost): predicting subtypes	CBC and biochemical parameters from 6 French university hospital databases	Accurate diagnosis of AML, APL, and ALL	Accuracy (validation sets): 99.5% (ALL) 98.8% (AML) 99.7% (APL) AUC (validation sets): 0.97 (APL) 0.90 (ALL) 0.89 (AML) AUC: 0.59–0.92	External validation	18
- XGBoost: Model optimization - LASSO: Variable selection	CBC parameters of CML (BCR-ABL1) patients - Train set: 80% - Test set: 20%	Predicting the future diagnosis of CML		–	19
ResNet-18 (CNN): Mapping scattergrams	CBC scattergrams of 242 AL patients and 384 HCs - Train set: 80% - Test set: 20%	Quantitatively mapping a CBC scattergram to indicate the susceptibility to APL	Precision: >0.99 Sensitivity: 95% AUC: >0.99	External validation	20
- Breiman's RF: parameter identification - CART: Model optimization	CBC and morphological parameters of 525 individuals with cytopenias	Incorporating platelet-derived CPD parameters (Ne-WX and IPF) to improve MDS-CBC score	Sensitivity: 96% (RF) 84.5% (CART) Specificity: 87% (RF) 97.8% (CART) PPV: 94.7% (CART) NPV: 93.1% (CART)	Bicentric study	21
- CART: Attribute splitting - RF: feature importance evaluation - GM: Model fitting - C5.0 decision tree: decision making	Clinical (including CBC) and phenotypic data of 50 ALL patients	Finding the most discriminative attributes for predicting the risk of ALL	Accuracy: 99.83% (CART) 98.6% (C5.0) 94.44% (RF) 95.61% (GM)	–	22
- LASSO: prediction - RF: prognostic factor identification	Clinical and CBC data of 1211 DLBCL patients - Train set: 70% - Validation set: 30%	Identifying prognostic factors of DLBCL	AUC: 75.8% (LASSO) 71.6% (RF)	–	23

(continued on next page)

Table 1 (continued)

ML model	Dataset	Application	Performance	Validation	Ref.
9 ML models: prediction	Clinical, laboratory, cytogenetic and molecular genetic data of 1383 AML patients - Train set: 90% - Test set: 10%	Predicting CR and 2-year OS of AML	CR prediction AUC: $\sim 0.77-0.86$ (test set) $\sim 0.71-0.80$ (validation cohort) 2-year OS prediction AUC: $\sim 0.63-0.74$ (test set) $\sim 0.65-0.75$ (validation cohort)	External, multicenter validation	24
- Sight OLO platform - CNN: characterizing RBC - ML: classification - Hilab system (ML, DL): classification, imaging, and result processing	679 whole blood samples	Automating the analysis of various blood parameters for rapid CBC reports and POCT	Sensitivity: 93.0% Specificity: 80.6%	Multicenter, clinical laboratory validation	25
	CBC data of 450 blood samples	- Analyzing blood cells and hematimetric parameters - POCT	Accuracy: >80.0% Sensitivity: >89.0% Specificity: >88.5% Kappa coefficient: >81.0%	Extensive clinical validation	26

–, not applicable; AI-PAL, artificial intelligence prediction of acute leukemia; AL, acute leukemia; ALL, acute lymphoblastic leukemia; AML, acute myeloid leukemia; ANN, artificial neural network; APL, acute promyelocytic leukemia; AUC, area under the curve; CART, classification and regression trees; CBC, complete blood cell count; CML, chronic myelogenous leukemia; CNN, convolutional neural network; CPD, cell population data parameter; CR, complete remission; DL, deep learning; DLBCL, diffuse large B-cell lymphoma; GM, gradient boosted machine; HCs, healthy controls; IPF, immature fraction of platelet; LASSO, least absolute shrinkage and selection operator; MDS, myelodysplastic syndromes; ML, machine learning; Ne-WX, neutrophil dispersion; OS, overall survival; POCT, point-of-care testing; RBC, red blood cell; RF, random forest; XGBoost, extreme gradient boosting.

thus giving a probability for the predicted subtype classification and also indicating the most significant areas for manual review⁵³. Such an approach may serve as a rapid prescreening and decision support tool for cytomorphological pathologists, which would especially help those in low-resource areas.

Although various models have claimed validated performance in cell identification, additional improvements are needed to construct a comprehensive AI-aided morphologic diagnostic framework with multimodal data inputs (in addition to BM-cell morphology) and training in more disease subtypes. For example, cytomorphological diagnosis of CML would be more robust by simultaneously considering aspects of BM: BM-cell conformation, the degree of hyperplasia, and the alkaline phosphatase score. Incorporating these 3 aspects, Li et al.⁵⁶ reported an AI-aided diagnostic framework (virtual hematological morphologist) for the precise diagnosis of chronic phase CML on the basis of BM cells with a balance accuracy of 99.23%, a sensitivity of 97.96%, and a specificity of 100%. The results also suggested that integrating explicit medical expertise into a comprehensive AI-aided workflow (similar to the clinical diagnostic procedure) is remarkably valuable, outperforming widely used end-to-end frameworks in both testing accuracy (96.88% vs. 68.75%) and generalization ability (97.11% vs. 68.75%) when distinguishing between normal and abnormal cases⁵⁶.

3.5. The success of AI-assisted cytomorphology recognition

In summary, AI-assisted PB cytomorphology analysis has significantly reduced the workload of laboratory personnel and has thus achieved revolutionary success in the translation to clinical settings compared to that of other techniques in the hematology laboratory. This example of AI integration has largely paralleled the incorporation of AI into image processing in other medical fields (e.g., pathology, radiology)^{67,68}. In hematology, the automated systems have improved efficiency and reduced the TAT for routine analysis, significantly advancing hematologic malignancy diagnosis and decreasing interobserver variances. But issues remain, including the limited number of “expert-annotated” cell morphologies, especially for rare cell types, and the lack of public databases. Also, while automated systems enhance efficiency, human oversight is still essential for ensuring diagnostic accuracy, particularly in complex cases. To optimize the utility of AI in hematology, future research must improve rare and atypical cell detection, especially for BM examination. Despite these issues, the success in AI-assisted cytomorphology encourages the hematology field to accept and utilize AI to a greater degree. In the future, integrating AI seamlessly into clinical workflows may require multimodal, historical, and longitudinal data to be taken into consideration to ensure comprehensive, accurate, and efficient diagnostic and prediction models for precise and personalized medicine (Fig. 3).

4. Immunophenotyping

4.1. Essential role of immunophenotyping in hematopathology

Immunophenotyping for lineage assignment is an essential step for confirming a diagnosis and subcategorizing hematology lymphoid malignancies according to the WHO classifications^{3,69}. The primary method for immunophenotyping is multiparametric flow cytometry (FCM), which sensitively detects antigen expression on or within

Table 2 Studies of AI in cell morphology.

ML model	Dataset	Application	Performance	Validation	Ref.
• CNN: differentiation	600 images of neutrophils having 5 different nuclear lobulations from PB smears • Train set: 500 images • Test set: 100 images	• Automatically recognizing and categorizing neutrophil types having different nuclear lobulations	Best accuracy: 70.0%–99.0% Average precision: 70.2%–99.0% Average recall: 70.0%–99.0% Average F-measure: 69.9%–99.0%	–	33
• DysplasiaNet (CNN): recognition	20,670 images of hypogranulated and normal neutrophils • Train set: 4810 images • Validation set: 1000 images	• Automatically distinguishing dysplastic neutrophils with fewer granules	Sensitivity: 95.5% Specificity: 94.3% Precision: 94% Overall accuracy: 94.85%	–	34
• Sysmex DI-60 (ANN): locating, identifying, and pre-classifying cells	822 PB smears for routine CBC analysis	RBC morphology characterization and WBC differential counts	Sensitivity: 9.1%–100.0% Specificity: 6.3%–100.0%	–	35
• Sysmex XN-Series equipped with the DI-60: Automating the WBC differential process	232 blood samples from patients for morphological analysis and 2000 blood samples for routine analysis	• Fully automating CBC, slide preparation/staining, and digital scanning with pre-classification of cells	Overall accuracy: 88.4%	–	36
• MC-100i (CNN): classifying cells	600, 382, and 445 blood samples for WBC, RBC, and PLT classification analysis, respectively	• Automating the analysis of blood smears by classifying WBC, RBC, and PLT	Accuracy: 97.8% for normal WBC 89.95% for abnormal WBC Sensitivity: >96% for normal WBC >87% for abnormal WBC >90% for RBC Specificity: 99.40% for blast cells >90% for RBC	–	37
• SVM with a radial basis function kernel: processing images	4389 lymphoid images from 105 patients	• Classifying normal, reactive lymphocytes and follicular lymphoma cells	Accuracy: 91.23% for 3 groups 97.67% for 5 groups	–	38
• ImageNet-pretrained Xception model: identification and prediction	• Train set: 82,974 images from 514,183 cells • Test set: 8297 cells	• Capturing and flagging suspicious cells from 21 classes of hematological disorders	Reproducibility: >90.0% Accuracy: >90.0% Sensitivity: >90.0% Specificity: >90.0%	–	39
• ResNeXt (CNN): differentiation	18,365 cell images from 200 individuals • Train set: 80% • Test set: 20%	• Classifying the most important leukocyte types	Accuracy: 91.7% Precision: >90% Sensitivity: >90%	–	40

(continued on next page)

Table 2 (continued)

ML model	Dataset	Application	Performance	Validation	Ref.
MFDS-DETR:	- WBCDD datasets encompassing 684 samples: - Train set: 540 samples - Test set: 144 samples - Train set: 305,019 cell images from 137 BM smears and 72 PB smears of patients with MM - Test set: 151,807 cell images from 184 PB smears	Automatically locating and counting leukocyte types	AP: 79.7% AP ₅₀ : 97.2%	External validation in 2 public databases	41
Morphogo (CNN): Automating CPC detection	- Train set: 305,019 cell images from 137 BM smears and 72 PB smears of patients with MM - Test set: 151,807 cell images from 184 PB smears	CPC identification <i>via</i> PB smears	Accuracy: 99.66% Sensitivity: 89.03% Specificity: 99.68% PPV: 55.96% NPV: 99.96%	—	42
- SVM: feature selection - Hierarchical Tree Classifier: structured decision-making - DL (ANN): classification (localization, segmentation, recognition) and differential cell count - MLFL-net: cell classification - YOLOX-s: cell localization	- Train set: 600,000 cell images - Validation set: 30,867 cell images 11,788 fully annotated micrographs from 728 smears and 131,300 expert-annotated single cell images - Train set: 80% - Validation set: 20% 171,374 images from BM smears of 945 patients - Train set: 80% - Test set: 20% 703,970 digitalized cell images from 3261 PB smears	- Automatically classifying BM cells into 16 different classes - Automatically classifying marrow cells	Accuracy: 52.1% (hierarchical Tree Classifier) 51.5% (SVM) Accuracy: 90.1% ICC: ≥ 0.883	—	43
ResNet 50 (CNN), sequential model: classification	- Train set: 600,000 cell images - Validation set: 30,867 cell images 11,788 fully annotated micrographs from 728 smears and 131,300 expert-annotated single cell images - Train set: 80% - Validation set: 20% 171,374 images from BM smears of 945 patients - Train set: 80% - Test set: 20% 703,970 digitalized cell images from 3261 PB smears	Automatically classifying BM leukocytes	Accuracy: 89.53% AUC: 0.98	Preliminary validation in a hospital	44
- CNN: recognition and classification - XGBoost: diagnostic optimization	- Train set: 600,000 cell images - Validation set: 30,867 cell images 11,788 fully annotated micrographs from 728 smears and 131,300 expert-annotated single cell images - Train set: 80% - Validation set: 20% 171,374 images from BM smears of 945 patients - Train set: 80% - Test set: 20% 703,970 digitalized cell images from 3261 PB smears	Automatically classifying BM leukocytes	Precision and recall across multiple cell types are high	External validation	46
ResNet 50 (CNN): classification and differentiation	- Train set: 600,000 cell images - Validation set: 30,867 cell images 11,788 fully annotated micrographs from 728 smears and 131,300 expert-annotated single cell images - Train set: 80% - Validation set: 20% 171,374 images from BM smears of 945 patients - Train set: 80% - Test set: 20% 703,970 digitalized cell images from 3261 PB smears	- Differentiating from AA - Diagnosing MDS and differentiating between AA, MDS, and AML	Image-recognition system: Sensitivity: 93.5% Specificity: 96.0% MDS diagnostic system: Sensitivity: 96.2% Specificity: 100% AUC: 0.990	—	47
ResNet 50 (CNN): classification and differentiation	- Train set: 600,000 cell images - Validation set: 30,867 cell images 11,788 fully annotated micrographs from 728 smears and 131,300 expert-annotated single cell images - Train set: 80% - Validation set: 20% 171,374 images from BM smears of 945 patients - Train set: 80% - Test set: 20% 703,970 digitalized cell images from 3261 PB smears	Automatic recognition of AA, MDS, and AML	2-classification model: AUC: 0.985 Accuracy: 0.914 Sensitivity: 0.992 3-classification model: AUC: 0.97 Accuracy: 0.93 Sensitivity: 0.86	External validation	48

• ML: Automated morphology analysis	131 samples from BM trephines	• Identifying and subclassifying megakaryocytes from BM trephine samples • Identifying ALL	Precision: 0.86	—	49
• Clustering algorithm: segmentation • SVM, MLP: recognition • GA: segmentation enhancement • Ensemble classifiers: classification improvement • CNN: classification • SVM-GA: feature extraction and parameter optimization • AlexNet: detection and classification	Data from a public ALL image dataset: ALL-IDB2 • Train set: 90% • Test set: 10%		Accuracy: 96.67% Recognition rates: 96.72%	Evaluation using open databases	50
	• Train set: 2420 augmented data of 121 images • Test set: 121 images	• Detection and subclassification of ALL	Accuracy: >80% Sensitivity: >68%	—	51
	• Test set: 121 images • 108 and 260 microscopic blood images from ALL-IDB1 and ALL-IDB2, respectively • Train set: 60% • Evaluation set: 40%	• Detection and subclassification of ALL	For detection: Sensitivity: 100% Specificity: 98.11% Accuracy: 99.50% For subclassification: Sensitivity: 96.74% Specificity: 99.03% Accuracy: 96.06%	—	52
• AMLnet (DL): identification and classification	8245 BM smear images from 651 patients • Train set: 60% • Evaluation set: 40%	• Discriminating AML patients and healthy individuals • Identifying AML subtypes	AUC: 0.89 (image level) AUC: 0.92 (patient level)	External validation at 2 centers	53
• MorphoGo (CNN): cell classification	65,986 nucleated cells from 230 BM smears	• Automating classification of cell morphology and diagnostic prediction in mixed blood malignancies	Accuracy: >85.7% Sensitivity: 69.4% Specificity: 97.2%	—	54
• ML: segmentation, feature extraction and classification	442 PB smears from 206 patients • Train set: 75% • Test set: 25% An image dataset and a case dataset	• Automatically identifying and classifying different types of acute leukemia smears • Distinguishing normal and abnormal cases <i>via</i>	Accuracy: 85.8% (for 6 cell groups) 94.0% (for individual smears) Accuracy: >97% Sensitivity: 97%	—	55
				—	56

(continued on next page)

Table 2 (continued)

ML model	Dataset	Application	Performance	Validation	Ref.
• VHM (faster R-CNN, SVM): feature extraction and case identification	• Train set: 80% • Test set: 20%	multimodal image recognition and case identification	Specificity: >92% AUC: >0.95		
–, not applicable; AA, aplastic anemia; ALL, Acute lymphoblastic leukemia; ALL-IDB, acute lymphoblastic leukemia image databases; AML, acute myeloid leukemia; ANN, artificial neural network; AP, average precision; AUC, area under the curve; BMSNet, bone marrow smears network; CBC, complete blood cell count; CNN, convolutional neural network; CPC, circulating plasma cell; DL, deep learning; GA, genetic algorithm; ICC, reliability coefficient; MDS, myelodysplastic syndromes; MFDS-DETR, multi-level feature fusion and deformable self-attention DETR; MLFL-Net, multi-level feature learning network; ML, machine learning; MLP, multi-layer perceptron; MM, multiple myeloma; NPV, negative predictive value; PB, peripheral blood; PLT, platelet; PPV, positive predictive value; RBC, red blood cell; WBC, white blood cell; ResNet, residual neural network; ResNeXt, residual networks with next-generation aggregated transformations; SVM, support vector machine; VHM, virtual hematological morphologist; WBCDD, Leukocyte Detection Dataset; XGBoost, extreme gradient boosting; YOLOX-s, You Only Look Once X-s.					

individual cells in a semiquantitative manner⁷⁰. Thus, the FCM identification of abnormal cells plays an indispensable role, both in the initial confirmative diagnosis and in MRD monitoring.

4.2. The challenges of manual gating and the opportunities for AI in the FCM workflow

Compared to other techniques in the clinical hematology laboratory, FCM has been lagging in automation. Interpreting FCM data largely relies on the manual gating of bivariate plots and organizing the dot plots or histograms in a sequential manner. This process involves either hierarchical gating or nonhierarchical gating. Hierarchical gating systematically evaluates bivariate plots using rigid threshold values to create successive subpopulations from the previous gated ones (*i.e.*, “child gates” from “parent gates”), whereas nonhierarchical gating uses Boolean operators AND, OR, and NOT to create nonrigid and hand-drawn gates to label cell populations of interest. Of note, both methods rely heavily on the expertise of the operator, which can introduce bias, and thus these methods tend to be labor-intensive, subjective, and difficult to reproduce. Moreover, compared to FCM in immune monitoring, FCM in immunophenotyping can be more challenging. This is largely due to the individualized panel of antibodies required for diagnostic purposes and the essential role of FCM in clinical decision-making. Meanwhile, with the development of novel fluorophores and laser technology, cytometry data can be highly dimensional, with an increasing number of parameters simultaneously measured per cell. For example, spectral FCM and mass cytometry (*i.e.*, cytometry by time-of-flight), which have shown potential in both clinical and research scenarios^{71,72}, allow measurement of over 40 markers per cell in single-tube assays⁷³. This enables more flexibility in the design of antibody panels but can add substantial difficulty in manual gating and data interpretation. Therefore, hematology cytometry data analysis should always be conducted by experienced hematopathologists capable of expert-level analysis who can communicate extensively with physicians for guidance regarding additional workup.

To advance the development of computational methods for the identification of cell populations of interest in FCM data, members of 3 communities—algorithm developers, FCM users, and software and instrument vendors—initiated the Flow Cytometry: Critical Assessment of Population Identification Methods (FlowCAP) project⁷⁴. The FlowCAP project has provided well-annotated FCM datasets for comparing the performance of automated cluster algorithms⁷⁴. Although none of the individual FlowCAP methods have provided perfect results for all use cases and sample sets, automated methods may be practical for many FCM use cases, especially the task of cell population identification⁷⁴. Thus, incorporating AI may increase the efficiency, sensitivity, accuracy, and standardization of the FCM workflow in the clinical hematology laboratory⁷⁵.

Both supervised and unsupervised learning methods have been utilized for the automated analysis of FCM data⁷⁶. ML algorithms have been primarily designed for dimensionality reduction, cell population identification, and sample prediction. Alternatively, the reported computational/bioinformatics approaches can generally be grouped as quality control (*e.g.*, flowCore⁷⁷), clustering-based automated gating (*e.g.*, flowClust⁷⁸), visualization (*e.g.*, flow data analysis using self-organizing maps^{79,80}), t-distributed stochastic neighbor embedding (t-SNE), and spanning tree progression of density normalized events⁸¹. The representative models published to date have been summarized in Table 3^{82–102}.

Table 3 Studies of AI in flow cytometry.

ML model	Dataset	Application	Performance	Validation	Ref.
• viSNE: dimensionality reduction	10,000 cells from a complete dataset of healthy human bone marrows	• Mapping high-dimensional cytometry data onto 2 dimensions • Comparing leukemia diagnosis and relapse samples • Identifying a rare leukemia population reminiscent of minimal residual disease	Consistent and reproducible	–	82
• PhenoGraph: defining phenotypes • t-SNE: dimensionality reduction	3 different mass cytometry datasets of healthy human BM	• Defining phenotypes in high-dimensional single-cell data • Revealing the heterogeneity of intracellular state	–	–	83
• GMM: cellular distribution capture • Fisher kernel methods: feature embedding • SVM: classification	Data from 531 patients who underwent evaluation for cytopenias and/or AL • Train set: 80% • Test set: 20%	• Classifying acute leukemias and distinguishing them from non-neoplastic cytopenias • Identifying key FCM parameters that contribute to the model's performance	Accuracy: 94.2% AUC: 99.5%	–	84
• SOM: dimensionality reduction and data clustering • CNN: pattern recognition	MFC data from 20,622 samples of suspected B-NHL patients • Train set: 18,274 cases • Test set: 2348 cases	• Transforming MFC raw data into a multicolor 2D image • Distinguishing samples from healthy individuals and those with disease • Differentiating 7 subtypes of mature B-cell neoplasm	F1 score: 0.94 Accuracy: 0.830	–	85
• UMAP: feature extraction and dimensionality reduction • RF: classification	3,417 cases of PB data • Train set: 80% • Test set: 20%	• Classified cases with a B-cell malignancy and screened out normal cases under a basal B-cell panel, and identified cases requiring add-on studies under this pipeline	Accuracy: >95% Sensitivity: 100% (with 14% specificity)	–	86
• EnsembleCNN Classifier: decision automation • RF: integration	Samples from 9635 patients • Train set: 80% • Validation set: 20%	• Flagging suspected cases that require additional antibody panel to distinguish CLL from MCL	Accuracy: 94% AUROC: 89%	–	87
• viSNE: dimensionality reduction • PhenoGraph: cellular clustering • Barnes-Hut SNE: dimensionality reduction	10,000 mononuclear cells from 10 PB specimens 72 routine laboratory samples	• Identifying abnormal T-cell populations <i>via</i> on a single 8-color, 10-parameter antibody combination • Clustering CLL MRD separately from normal B cells	Correlation Coefficient: 0.99	–	88
			–	–	89

(continued on next page)

Table 3 (continued)

ML model	Dataset	Application	Performance	Validation	Ref.
- PCA: data exploration and analysis - UMAP and t-SNE: dimensionality reduction - FlowSOM and Phenograph: clustering - DeepFlow: clustering and cell population identification	High-dimensional cytometry data	Enhancing the analysis and interpretation of high-dimensional cytometric data	—	—	90
- Clustering algorithm: cell identification - CytoNorm: data normalization - UMAP: dimensionality reduction - t-SNE: data structure discovery	113 CLL MRD FCS files - Train set: 41 files - Validation set: 72 files Cytometry data of whole blood samples from 27 patients	Cell classification and quantification of CLL MRD	Pearson correlation coefficient: 0.8650 LOA: 95%	—	91
	~1.3 million cells from mass cytometry and single-cell RNA sequencing datasets	Normalizing for batch effect removal that enables mass cytometry analysis of large clinical cohorts	—	—	92
MetaCyt: Automated meta-analysis	Train set: 50,000 cells Test set: 50,000 cells Cytometry data from over 30 different datasets	Improving visualization and interpretation of single-cell data via dimensionality reduction	Accuracy: ~95%	External validation	93
	Single-cell mass cytometry data from multiple clinical samples	- Identifying common cell subsets from heterogeneous cytometry data - Automatic meta-analysis of cytometry data - Revealing differences in immune cells between ethnic groups	—	—	94
- Citrus: cluster identification, characterization and regression - PCA: dimensionality reduction - FloReMi: preprocessing, feature extraction and selection, prediction	Train set: 2/3 Test set: 1/3 Train set: FCM data from 191 patients Test set: FCM data from 192 patients	Automating identification of stratifying signatures in cellular subpopulations	AUC: 0.69 (train set) 0.80 (test set)	—	95
CellCnn: Enhanced detection and classification	A benchmark dataset based on samples from specific patients	- Predicting survival times of HIV patients by analyzing FCM data - Detecting rare cell subsets associated with disease by using high-dimensional single-cell measurements	—	—	96
			Achieving a high level of precision at lower recall levels	—	97

• Deep CNN and CellCNN: data analysis • RF with FlowSOM as input: cell clustering • scvis: Visualization • PCA: preprocessing • t-SNE: dimensionality reduction • Probabilistic generative model: Uncertainty quantification • LDA: classification	CyTOF data from 472 PB samples	• Identifying biomarkers associated with latent CMV infection and other diseases	AUC: 0.93–0.97	–	98
• Single-cell RNA-seq data and 2 CyTOF datasets • Train set: 80% • Validation/test set: 20%		• Capture and visualization of low-dimensional structures in single-cell gene expression data	Accuracy: >81%	–	99
• UMAP and HDBSCAN: dimension reduction • Graphic clustering: cell identification • DL: prediction	Cell populations in CyTOF data from 4 benchmark datasets ~433,000 cells from two benchmark CyTOF datasets • Train set: 50% • Validation set: 50%	• Predicting cell populations in single-cell mass cytometry data • Predicting cell types of single-cell mass cytometry data	Accuracy: ~98% F1-score: 0.99 F-scores: 0.9921 (CyTOF1) 0.9992 (CyTOF2) Accuracy: ~71.9–99.4% (CyTOF1) ~96.5–99.8% (CyTOF2) AUC: ~0.94–1.00 (CyTOF1) 1.00 (CyTOF2)	Multicenter validation External validation	100 101
• DeepCyTOF (DL): Automatic gating and classification • Domain adaptation: calibrates datasets	FCM and CyTOF data of blood samples from 14 patients and 34 healthy subjects	• Automating the gating process in cytometry by accurately classifying cell types across varied datasets without recalibration	F-measure value: 0.985	Multicenter validation	102

–, not applicable; AUC, area under the curve; AUROC, area under the ROC curve; BM, bone marrow; B-NHL, B-cell non-Hodgkin lymphoma; CLL, chronic lymphocytic leukemia; CMV, cytomegalovirus; CNN, convolutional neural network; CyTOF, cytometry by time-of-flight mass spectrometry or mass cytometry; DL, deep learning; FCM, flow cytometry; FCS, flow cytometry standard; FloReMi, flow density survival regression using minimal feature redundancy; GMM, gaussian mixture model; HDBSCAN, hierarchical density-based spatial clustering of applications with noise; LDA, linear discriminant analysis; LOA, limit of agreement; MCL, mantle cell lymphoma; MFC, multiparameter flow cytometry; ML, machine learning; MRD, measurable/minimal residual disease; PB, peripheral blood; PCA, principal component analysis; RF, random forest; RNA-seq, RNA sequencing; SNE, stochastic neighbor embedding; SOM, self-organizing map; SVM, support vector machine; t-SNE, t-distributed stochastic neighbor embedding; UMAP, uniform manifold approximation and projection; viSNE, visual interactive stochastic neighbor embedding.

4.3. ML models for the diagnosis and differentiation of leukemia

As immunophenotyping by FCM plays an essential role in the diagnosis and differentiation of acute leukemia, the integration of AI-assisted methods seems feasible and could provide a fast, automated, data-driven overview of each suspected case. Early advances in the computer-driven analysis of FCM data used cell clustering techniques (*e.g.*, flow data analysis using self-organizing maps) in combination with ML techniques (*e.g.*, support vector machine and RF) and showed increased diagnostic accuracy in various hematologic malignancies¹⁰³. And the dimensional reduction tool based on stochastic neighbor embedding, viSNE, consistently discriminated between BM samples from healthy individuals and individuals with cancer (*i.e.*, AML, ALL) and further discriminated between samples from individuals with newly diagnosed leukemia and relapsed leukemia⁸². Cellular subpopulations may overlap visually when reduced to 2 dimensions, however, which is an inevitable limitation of methods aimed at dimensional reduction. This issue led to the subsequent development of PhenoGraph, which defined phenotypes *via* high-dimensional, single-cell data and revealed the heterogeneity of their intracellular state—likely a new pathophysiology of pediatric AML⁸³. Using data from 531 patients who underwent evaluation for cytopenia and/or acute leukemia, Monaghan et al.⁸⁴ developed an ML model to rapidly distinguish among APL, non-APL AML, ALL, and nonneoplastic cytopenia. Interestingly, adding more FCM parameters, light scatter properties, and CD117 did not significantly improve the model further, indicating both the robustness of their ML model and the possibility of using a simplified panel for malignancy screening. In a recent cross-panel model aimed at automated AML *versus* nonneoplastic classification, using a fraction of the parameters within the whole panel did not seem to compromise the performance¹⁰⁴, suggesting that when developing future algorithms, researchers can use multicenter datasets based on different antibody panels.

The diagnosis of chronic lymphocytic leukemia (CLL) involves PB immunophenotyping by FCM¹⁰⁵, and ML has been applied to the classification and diagnosis of CLL^{16,106}. One example is a model built by Zhao et al.⁸⁵, which classified 9 diagnostic classes of mature B-cell neoplasms (*e.g.*, CLL) and healthy controls *via* a self-organizing map algorithm using a relatively large cohort of 20,622 patients for model training and testing. The results of the model were then utilized as input for a CNN classifier, which showed 83% accuracy in classifying CLL cases. Later, an RF classifier, built from a dataset of 3417 PB cases, prospectively classified cases with a B-cell malignancy, screened out normal cases by a basal B-cell panel, and identified cases requiring add-on studies using this pipeline⁸⁶.

Finally, complications arise when attempting to accurately gate out neoplastic T cells, as they show greater variance in both phenotype and light scatter properties¹⁰⁷. Combining viSNE and PhenoGraph was able to identify abnormal T-cell populations based on a single 8-color, 10-parameter antibody combination from routine clinical FCM data⁸⁸.

4.4. ML models for MRD assessment in hematological malignancies

MRD is a reliable and sensitive prognostic marker for acute leukemia and lymphoma. Among the methods assessing MRD, FCM is faster, more widely applicable, and less expensive. But because of the rarity of leukemia or lymphoma cells and FCM's

dependence on expert-level experience, using this method to distinguish positive events could be extremely challenging, laborious, and time-consuming. In addition, higher-dimensional data have been generated to increase the resolution over that of conventional FCM (measuring 8 to 10 colors) or full-spectrum FCM; one example is the recently published 19-color single-tube panel¹⁰⁸. Thus, it seems that traditional gating strategies do not meet the requirements of shorter TAT and lower risk of missing important features.

AI-assisted tools have been employed to analyze clinical MRD results by correctly classifying rare cells in certain disease types. For example, owing to an aberrant phenotypic “fingerprint,” researchers were able to utilize viSNE to identify a rare leukemia population reminiscent of MRD by using synthetic MRD samples created by spiking a healthy sample with ALL cells⁸². And viSNE was used as a pre-gating strategy on CD19-positive events to decrease the computational burden, identifying a putative MRD population. Similarly, with the help of an additional manual gating step, Barned-Hutt SNE was able to cluster CLL MRD separately from normal B cells⁸⁹. In another study⁹⁰, the authors developed a hybrid deep neural network approach for measuring CLL MRD, demonstrating a high accuracy of 97.1% and an excellent correlation with expert analysis. In a recent study⁹¹, DeepFlow was used to estimate the level of MRD in 113 randomly selected CLL FCM analysis files (in.fcs format) and showed a highly consistent output with the expert analysis obtained manually. The advantages of DeepFlow are its simple graphical user interface and quick analysis with minimal batch effects, in contrast to other AI-assisted algorithms such as t-SNE and Uniform Manifold Approximation and Projection⁹⁰.

4.5. The prospects of AI-assisted FCM data analysis

There is enormous potential for AI integration into additional steps of the clinical workflow; AI could be used, for example, to monitor cytometer performance, to enhance or improve existing data cleaning strategies, and to perform compensation or data transformation¹⁰⁹. As a proof of concept, an AI-assisted diagnostic tool in the clinical FCM laboratory could prime the cytometry data analysis pipeline by reliably screening out abnormal cases without further manual review. This tool could provide both an initial differential diagnosis to trigger the appropriate confirmatory studies and treatment monitoring of certain hematological malignancies.

Several issues should be addressed to promote translation of such tools, currently being developed in the research lab, into clinical practice. First, the models require larger datasets for training the models and multicenter and external validation. Second, most studies have reported the accuracy and sensitivity of their ML models based on retrospective data, but prospective studies should also be conducted. Third, the derived ML models are mostly panel- and disease-specific, which limits their broad applicability to the complex disease types observed in real practice. Fourth, even though there are consensus recommendations for both the initial diagnosis and the MRD measurement of hematological disorders, variations in instrumentation, reagents, antibody panels, and compensation settings hinder standardization globally, as do differences in analysis, interpretation, and reporting practices. Fifth, tools with a user-friendly graphical interface, rather than tools requiring programming expertise to fully utilize them, are a crucial need. Finally, a versatile model may simultaneously allow for early screening, differential diagnosis,

suggestive diagnosis, confirmative diagnosis, and prognosis evaluation, which all support clinical decision-making. Thus, future models may require more intensive and extensive interdisciplinary cooperation, including lab–clinical–technological communication (Fig. 4).

5. Cytogenetics

5.1. The importance of cytogenetic analysis and associated challenges

Cytogenetic analysis, the detection of chromosome abnormalities associated with copy numbers, structural alterations, and karyotype complexities, is vital in the clinical evaluation of nearly every form of hematologic malignancy. For example, detection of fusion transcripts $t(8;21)(q22;q22.1)/RUNX1-RUNX1T1$, $inv(16)(p13.1q22)$, $t(16;16)(p13.1;q22)/CBFB-MYH11$, or $t(15;17)(q22;q12)/PML-RARA$ allow for the diagnosis of AML even when the blast count is less than 20%¹¹⁰. Conventional cytogenetic assays include karyotyping by chromosomal banding analysis, fluorescence *in situ* hybridization (FISH), and chromosomal microarray¹¹¹. They all show limitations—low resolution, few analyzable metaphase cells, and constraints in detecting balanced chromosomal rearrangements¹¹²—but are continuously evolving as genomic evaluations of cancer increase and molecular diagnostic technologies improve. New techniques, or “next-generation cytogenetics,” have also emerged, such as optical genome mapping (OGM)¹¹³. The following paragraphs will discuss the status of AI integration into the analysis of cytogenetics, with the examples summarized in Table 4^{114–137}.

5.2. Karyotyping by chromosomal banding analysis

A karyotype is a photograph of the chromosomes, usually ordered in pairs by size and banding patterns. Karyotype analysis is the mainstream approach for detecting and investigating chromosomal abnormalities, including structural chromosomal rearrangements and abnormal chromosome number (*i.e.*, aneuploidy), in cytogenetics, genetics, and clinical diagnostics of hematologic disorders. Karyotyping by chromosomal banding analysis is performed by culturing cells, separating the chromosomes from the nucleus during metaphase, and staining them on a slide to allow for microphotography. Accurately segmenting chromosomes is of paramount importance because it directly impacts the accuracy of the subsequent classification process. Similar to the manual analysis of cytomorphology, the traditional manual analysis of chromosome images is labor-intensive, time-consuming, and nontrivial, requiring expertise from rigid training and accumulated clinical experience. As a result, early endeavors attempted to automate manual cytogenetic analyses, for automatic segmentation of chromosomes¹³⁸, detection of CML¹³⁹, leukemia diagnosis¹⁴⁰, and karyotyping of deformed chromosomes from BM cells¹⁴¹. But they were unanimously challenged by poor-quality images featuring bent, curved, distorted, overlapped, and blurred chromosomes. Therefore, novel AI methods are greatly needed to enhance the quality of karyogram images for correct diagnosis.

In pursuit of this, Sharma et al.¹¹⁴ first used a CNN to classify chromosomes, in which 400 images were fed into the crowdsourcing platform CrowdFlower to annotate the chromosome boundaries. Then, features were extracted and fed into a deep neural network for classification. Despite straightening the bent

chromosomes before classification, the model showed low accuracy, possibly due to the limited size of its training dataset¹¹⁴. To improve the performance, another group developed mCNN_GO, a multi-input CNN, which instantly classified chromosomes with higher accuracy¹¹⁷. Both studies attempted to address bent, distorted, and curved chromosomes, the former requiring a pre-processing step¹¹⁴, the latter integrating a straightening algorithm based on the medial axis of the chromosome¹¹⁷. To enhance the quality of karyotype images, Bokhari et al.¹¹⁹ developed a method based on a cycle generative adversarial network, ChromoEnhancer, that did not necessitate a training set. This model transformed poor-quality BM cell karyograms characterized by distorted, overlapped, and blurred chromosomes into good-quality images, with an average peak signal-to-noise ratio of 40.795 and a structural index similarity of 0.988. More recently, a novel CNN-based end-to-end chromosome segmentation network, ChroSegNet, was published¹¹⁵. This model seems to outcompete the others; it was built upon a large dataset, containing 13,096 pairs of chromosomes, and it effectively extracted key features by putting more emphasis on spatial information, making it better suited for dealing with complex chromosome structures and changing positions¹¹⁵. These seminal studies, however, were not done in the real context of hematological malignancy.

To improve the TAT and quality of karyotyping by chromosomal banding analysis, Haferlach et al.¹¹⁸ designed a fully automated workflow using a deep neural network classifier (Fig. 5). After validation by experienced staff, the classifier demonstrated high accuracy in assigning chromosomes to groups of class and orientation, achieving 98.6% accuracy for 23,000 chromosomes with a normal karyotype according to routine diagnostic methods. Remarkably, this AI-based method has been incorporated into the routine workflow (including ISO 15189), representing a big step toward real clinical integration and even in a cloud-based manner. More recently, Vajen et al.¹¹⁶ developed a CNN to predict both the chromosome class and orientation, identifying the correct class for 98.8% of chromosomes from patients with AML, MDS, and chronic myeloproliferative disorder and leading to a time savings of 42% for the karyotyping workflow. For translation to clinical settings, having access to a reference dataset with an abundance of abnormal karyotype images is crucial, especially for rare hematologic disorders. With this in mind, Deng et al.¹⁴² successfully employed ML to manually build a dataset containing abnormal karyotype images of $t(9;22)(q34;q11)$, the cytogenetic hallmark of CML. This pioneering endeavor serves as a proof of concept for establishing a clinical reference of abnormal karyotypes in the context of hematologic malignancies.

5.3. FISH

FISH is a technique that employs fluorescent probes to specifically bind a target genome sequence. FISH has greater sensitivity than karyotyping by chromosomal banding analysis, and it can be an alternative diagnostic method in samples lacking metaphases. But standard clinical manual scoring for FISH is still labor-intensive, time-consuming, and subjective¹⁴³, and a few studies have attempted to overcome these shortcomings. Using a confocal whole slide imaging scanner, Frankenstein et al.¹²² established automated 3D FISH scoring of z-stack images of lymphomas and solid tumors from 10 patients. Such 3D gene positioning allowed for precise localization, resulting in the identification of more variations in nuclear patterns associated with chromosome

Table 4 Studies of ML in cytogenetics and molecular biology.

ML model	Dataset	Application	Performance	Validation	Ref.
• Crowd sourcing: segmentation • CNN: classification	400 healthy patient images of chromosomes • Train and validation set: 200 images • Test set: 200 images	• Segmenting and classifying chromosomes	Accuracy: >80%	—	114
• ChroSegNet (CNN): chromosome segmentation	13,096 pairs of chromosome data from original microscopic images of chromosomes • Train set: 10,477 pairs • Test set: 2619 pairs	• Automating chromosome karyotype segmentation and classification • Dealing with complex chromosome structures and changing positions	Accuracy: 93.31% F1 value: 92.99%	—	115
• CNN: chromosome classification and rotation	330,131 normal karyograms and images • Train set: 297,119 karyograms • Validation set: 16,506 karyograms • Test set: 16,506 karyograms	• Predicting the type and correct orientation of fluorescent R-Band metaphase chromosomes • Identifying normal and abnormal chromosomes in patients with AML, MDS and CMPD • Automatically segmenting and classifying chromosomes	Accuracy: 98.8%	Integrated into clinical practice	116
• mCNN_GO (CNN): classification • Mask R-CNN: segmentation • Geometric optimization: preprocessing chromosome • AI-CN (DNN): Automatic karyotyping (class/number and orientation)	5000 sub-images from individual chromosomes • Train set: 100,000 karyograms from routine diagnostics • Validation set: 500 normal karyotypes • Train set: 1056 poor-quality images and 1727 excellent-quality images • Test set: 690 poor-quality karyogram images • Train set: DNA FISH images from 28 wells of a 384-well plate • Test set: DNA FISH images from 48 wells	• Automatically identifying the class/number and orientation of chromosomes with a normal karyotype • Transforming poor-quality BM cell karyograms into good-quality images • Detecting and counting DNA FISH signals	Accuracy: 0.9570 F1 score: >0.92	—	117
• CycleGAN: image enhancement	5000 sub-images from individual chromosomes • Train set: 100,000 karyograms from routine diagnostics • Validation set: 500 normal karyotypes • Train set: 1056 poor-quality images and 1727 excellent-quality images • Test set: 690 poor-quality karyogram images • Train set: DNA FISH images from 28 wells of a 384-well plate • Test set: DNA FISH images from 48 wells	• Automatically identifying the class/number and orientation of chromosomes with a normal karyotype • Transforming poor-quality BM cell karyograms into good-quality images • Detecting and counting DNA FISH signals	Accuracy: 98.6%	Integrated into routine clinical workflow	118
• SpotLearn (CNN): RF: spot filtering • U-net: spot detection	5000 sub-images from individual chromosomes • Train set: 100,000 karyograms from routine diagnostics • Validation set: 500 normal karyotypes • Train set: 1056 poor-quality images and 1727 excellent-quality images • Test set: 690 poor-quality karyogram images • Train set: DNA FISH images from 28 wells of a 384-well plate • Test set: DNA FISH images from 48 wells	• Automatically identifying the class/number and orientation of chromosomes with a normal karyotype • Transforming poor-quality BM cell karyograms into good-quality images • Detecting and counting DNA FISH signals	PSNR: 40.795 SSIM: 0.988	—	119
• SpotLearn (CNN): RF: spot filtering • U-net: spot detection	5000 sub-images from individual chromosomes • Train set: 100,000 karyograms from routine diagnostics • Validation set: 500 normal karyotypes • Train set: 1056 poor-quality images and 1727 excellent-quality images • Test set: 690 poor-quality karyogram images • Train set: DNA FISH images from 28 wells of a 384-well plate • Test set: DNA FISH images from 48 wells	• Automatically identifying the class/number and orientation of chromosomes with a normal karyotype • Transforming poor-quality BM cell karyograms into good-quality images • Detecting and counting DNA FISH signals	Accuracy: >94.22% Precision: >94.21% Recall: >99.4% F-score: >97.02%	—	120

RetinaNet-based localization and classification	CNNs:	299 FISH images and 301 images of individual nuclei	Detecting, localizing, and classifying nuclei and FISH signals without the need for segmentation	Accuracy: 96%	—	121
SHIMARIS Automated 3D scoring	PAFQ:	3D FISH images from confocal whole slide imaging scans	Automatically localizing and identifying variations in nuclear patterns associated with rearrangement	—	—	122
XGBoost: prediction of telomeric outcomes		128,800 data points of telomere length measurements - Train set: 80% - Test set: 20%	Predicting treatment responses and prognoses in patients with <i>POT1</i> mutations	R^2 : 0.931	—	123
- RF and GBMs: identification and classification - DL: Outcome prediction		500,000 adult patients with any of 6 chronic diseases - Train set: 70% - Test set: 15% - Validation set: 15%	Predicting which patients might benefit from genetic test associated with genetic diseases	Accuracy: >0.90	—	124
ML: predictive modeling		Mutation data from BM samples of 868 patients - Train set: 80% - Validation set: 20%	Predicting the clinical phenotype and outcomes of patients with various myeloid leukemias	Accuracy: 63%–88%	External validation	125
DeepVariant (DL): Variant calling		High-throughput NGS data available from the CEPH female sample NA12878	Calling SNP and indel mutations from NGS data	SNP F1 score: 99.95% Indel F1 score: 98.98%	External validation	126
ML: data analysis and relationship mapping		Clinical and commonly mutated gene data of 2697 patients with myeloid malignancies	Revealing associations between clinical characteristics and specific genomic alterations	—	A multicenter study	127
GBM (ML): diagnostic modeling		Genomic and clinical data from 2697 individuals with myeloid malignancies - Train set: 80% - Test set: 20%	- Analyzing 24 commonly mutated genes in MDS - Distinguishing MDS from other myeloid malignancies - Elucidating clinicogenomic relationships	Diagnostic AUC: 0.951 Other AUC: >0.675	External validation	128
EVE (deep generative models): Variant effect prediction		~250 million protein sequence data - Train set: 90% - Test set: 10%	Predicting variant pathogenicity without relying on labels	Accuracy: ~90%	—	129
		71 million missense variants		AUC: >0.809	—	130

(continued on next page)

Table 4 (continued)

ML model	Dataset	Application	Performance	Validation	Ref.
AlphaMissense (fine-tuning AlphaFold): Variant effect prediction	across the human proteome - Validation set: 2526 ClinVar variants - Evaluation set: 1263 pathogenic and 1263 benign variants	Predicting the pathogenicity of missense variants			
StrVCTVRE (RF): Variant pathogenicity prediction	- Train set: rare SVs from ClinVar, gnomAD and a recent great ape sequencing study - Validation/test set: rare SVs from DECIPHER and the 1000 genomes project phase 3	Distinguishing pathogenic SVs from benign	Accuracy: 0.83 Sensitivity: 90%	Multicenter evaluation	131
Cue (DL): localization and genomic mapping	13,504 SVs of various types and sizes	Calling and genotyping SVs of diverse sizes and types	Precision: >85% Recall: >87% F1 score: >86%	External validation	132
CNV-P (RF): classification and prediction	- Train set: data from 6 individuals - Test set: data from 3 individuals and 2 independent validation datasets - Validation set: NA12878 and HG002 datasets	Calling and classifying CNVs from genome sequencing data	Precision: >0.900 Recall rate: 0.850	External validation	133
GECNVNet (DL): feature embedding and interaction	CNV data from 752 MM patients - Train set: 80% - Test set: 20%	Facilitating accurate detection of chromothripsis	AUC: 0.8309	External validation	134
X-CNV (XGBoost): classification and diagnosis	- Train set: 5315 pathogenic and 14,260 benign CNVs - Validation set: 4893 pathogenic and 4073 benign CNVs	- Predicting the pathogenicity of CNVs - Quantitatively measuring the pathogenic effect	Accuracy: 0.83 F1 score: 0.84 Sensitivity: 0.85 Specificity: 0.8	—	135

• RF: classification	429 annotated cases of different histology • Train set: 283 cases • Validation set: 146 cases	• Identifying the subtypes of B-NHLs using ligation-dependent RT-PCR and NGS data	Concordance: 0.8–1.0	–	136
• Bayesian latent Class analysis: cluster identification • ML: Genomic subclassification	Cytogenetic and gene sequencing data from 6788 AML patients • Train set: 80% • Test set: 20% • Validation set: 203 AML patients	• Integrating cytogenetic and gene sequencing data to subclassify AML and predict clinical outcomes	Accuracy: 97%	External validation	137

–, not applicable; AI-CN, AI based classifier normal; AML, acute myeloid leukemia; AUC, area under the curve; BM, bone marrow; B-NHLs, B-cell non-Hodgkin lymphomas; CEPH, Centre d'Etude du Polymorphisme Humain; CMPD, chronic myeloproliferative disorder; CNN, convolutional neural network; CNV, copy number variation; CNV-P, copy number variation prediction; CycleGANs, the cycle generative adversarial network; DECIPHER, database of chromosomal imbalance and phenotype in humans using ensemble resources; DL, deep learning; DNN, deep neural network; EVE, evolutionary model of variant effect; FISH, fluorescence *in situ* hybridization; GBM, gradient boosted machine; GECNVNet, graph embedding copy number variation network; gnomAD, The Genome Aggregation Database; indels, insertions and deletions; MDS, myelodysplastic syndromes; ML, machine learning; MM, multiple myeloma; NGS, next-generation sequencing; PSNR, peak signal-to-noise ratio; RF, random forest; RT-PCR, reverse transcription polymerase chain reaction; SNP, single nucleotide polymorphism; SSIM, structural index similarity; StrVCTVRE, structural variant classifier trained on variants rare and exonic; SV, structural variation; U-net, upsampling in the network; X-CNV, X-copy number variations; XGBoost, extreme gradient boosting.

rearrangement. Another ML-based workflow, SpotLearn, was developed for the segmentation of chromosomes in multi-channel FISH images with low signal-to-noise ratios, which allowed for a fully automated, single-allele analysis of 3D genome organization in a high-throughput fashion¹²⁰. Building on this, another group introduced an automatic DL-based pipeline, which enabled automated analysis on images captured by a single channel and localized and classified nuclei and FISH signals without the need for segmentation, as in SpotLearn¹²¹.

Since AI has been employed for FISH analysis of various cancers, including breast and thyroid cancer¹⁴⁴, it is believed that similar AI models or workflows will be developed for the diagnosis and treatment monitoring of hematological neoplasms.

5.4. Chromosomal microarray

In contrast to conventional karyotyping and FISH techniques, the chromosomal microarray stands out as a potent genome-wide technology, known for its impressive high-resolution capability. The chromosomal microarray is predominantly employed to detect and pinpoint various pathogenic structural variations (SVs) and copy number variations (CNVs)¹⁴⁵. With respect to its limitations, the method is unable to identify balanced chromosomal rearrangements¹¹², which are common in many hematologic malignancies, and has a relatively high cost. Additionally, the integration of AI into chromosomal microarray analysis has not been reported, with the exception of one study; this study used chromosomal microarray test results from electronic health records to help build an ML model to predict who should receive a disease-related genetic test¹²⁴.

5.5. OGM as an emerging cytogenetic analysis method

OGM is a new technology rapidly being adopted in clinical genetics laboratories because of its ability to detect SVs and CNVs. It involves imaging very long and linear single DNA molecules (median size >250 kb) that have been labeled at specific sites¹¹³. OGM combines microfluidics, high-resolution microscopy, and automated image analysis to facilitate high-throughput whole-genome imaging and de novo assembly of sequencing reads. It can dramatically simplify the laboratory workflow by replacing multiple tests (conventional karyotype, FISH, and chromosomal microarray) with 1 test.

Because of its higher resolution and shorter TAT, OGM is expected to be a supplement and potential alternative to standard cytogenetics¹⁴⁶. Recently, OGM has been used to efficiently detect a wide range of chromosomal anomalies or SVs in hematologic malignancies (*e.g.*, AML, B-cell ALL, CLL)^{147–149}. OGM results require rapid and accurate mapping of DNA fragment images to images of a reference genome, and AI can potentially help with this process. But the computational methods currently available for OGM fall short in terms of accuracy and computational speed, presenting an avenue for future research.

5.6. Future directions of AI-assisted cytogenetics analysis

The application of AI to cytogenetics has significant potential for improving the accuracy and efficiency of diagnostic procedures; however, further research is needed to address specific challenges associated with this approach. These challenges include the need for extensive, high-quality datasets to construct robust AI models and more advanced AI classifiers, ones that don't rely on

histological classification or medical evaluation. Specific strategies include fostering interdisciplinary collaboration among computational technologists, geneticists, and hematopathologists to design AI tools tailored to cytogenetic diagnostics. Additionally, efforts should focus on creating large, standardized datasets through partnerships between research institutions and hospitals, while ensuring data privacy and security. Advanced AI techniques, including DL, transfer learning, and explainable AI, should be leveraged to improve model accuracy and interpretability. Rigorous validation using independent datasets and the establishment of standardized protocols will further ensure the reliability and clinical applicability of these tools. By addressing these challenges through targeted strategies, AI can significantly enhance the diagnosis and treatment of hematologic malignancies, thus paving the way for more accurate, efficient, and rapid cytogenetic analysis.

6. Molecular biology

6.1. Growing support from molecular biology in clinical hematology

In regard to using genetic analyses for diagnosis, classification, risk assessment, prognostication, and therapeutic decision-making, hematologic malignancies have historically been the vanguard among cancers. With the development of high-throughput sequencing techniques coupled with the analysis of extensive gene panels, molecular genetics has entered the era of big data¹⁵⁰. Advanced molecular technologies, including mate pair sequencing and next-generation sequencing (NGS) (*e.g.*, whole-exome sequencing, whole-genome sequencing, and whole-transcriptome sequencing or RNA sequencing) are employed either individually or in combination to identify genomic features driven by single nucleotide variants (SNVs), insertions and deletions (indels), SVs, CNVs, expressed gene fusions, and rearrangements¹¹². As genomic features obtained from these advanced sequencing technologies can offer a more comprehensive understanding of clinical and pathological characteristics, both the 5th edition of the WHO classification and the 2022 International Consensus Classification of Myeloid Neoplasms and Acute Leukemias rely heavily on genomic data^{1–3}. But analyzing and interpreting genomic data from hematologic malignancies often requires expertise, which can introduce human error and bias. Therefore, AI can help reduce the dependency on expert knowledge and potentially increase the consistency of data interpretation. Many studies have applied ML methodologies to detect various types of genetic variants (Table 4), enhancing the precision of diagnosis and classification of hematologic malignancies.

6.2. Single gene mutations

Gene mutations play a pivotal role in the pathogenesis and progression of hematological malignancies, contributing to the molecular landscape of abnormalities observed in these disorders. Single gene mutations contribute to dysregulated signaling pathways, abnormal cellular interactions, and uncontrolled growth, ultimately driving the initiation and progression of hematological disorders. Examples include *FLT3*-, *NPM1*-, and *KRAS*-mutant AML, *IGHV*-mutant CLL, and *DDX41*-mutant MDS. The identification and characterization of specific gene mutations have significantly advanced our understanding of hematological

malignancies, providing valuable insights into disease etiology and potential therapeutic targets. Advances in genomic technologies have enabled comprehensive profiling of the mutational landscape, revealing recurrent genetic alterations that can potentially serve as therapeutic targets or prognostic markers.

6.2.1. SNVs

An SNV is the substitution of a single nucleotide for another at a specific position in the genome. While “variant” may be used as a general term for any single nucleotide change in a DNA sequence, SNVs specifically refer to both common single nucleotide polymorphisms (SNPs) and rare mutations, whether germline or somatic¹⁵¹. When the variant presents in at least 1% of the population, it is called an SNP. Identifying disease-related SNPs has significantly improved our understanding of disease pathogenesis and molecular pathways and has facilitated the development of better treatments¹⁵². The genome-wide distribution of SNPs and their easy adaptability for high-throughput analysis make them the target of choice for identifying genomic abnormalities in hematological malignancies and other cancers.

Despite rapid advances in sequencing technologies, accurately distinguishing genetic variants present in an individual genome from billions of short, error-prone outputs of sequence reads from NGS remains challenging. An ML approach called DeepVariant, based on a CNN model, effectively identified SNP and indel mutations from NGS data by learning statistical relationships between images of read pileups around putative variant and true genotype calls¹²⁶. Strikingly, the model could even identify variants from a variety of sequencing data by different technologies and experimental designs, including deep whole genomes from 10 × Genomics and exomes from Ion Ampliseq, highlighting the benefit of using more automated and generalizable techniques for variant identification¹²⁶. In another study¹²⁵, a total of 1372 SNVs and 939 indels in 87 genes were used to train an ML-based model to predict the clinical phenotype and outcomes of patients with panmyeloid leukemias, including AML, MDS, chronic myelomonocytic leukemia, and myeloproliferative neoplasms. By analyzing 295 cancer genes or whole exome sequencing data from BM samples of these patients, the ML-based model predicted phenotypes with an accuracy varying from 63% to 88% across different myeloid leukemias¹²⁵. Similarly, the analysis of indel mutation data has been enhanced by the 2 AI models mentioned above.

6.2.2. Point mutations

The term “SNV” has also been used to refer to point mutations found in cancer cells¹⁵³. A point mutation is a genetic mutation in which a single nucleotide is changed, inserted, or deleted from a DNA sequence within an organism’s genome. Some point mutations in hematological malignancies are considered disease-defining and can significantly impact clinical characteristics and disease phenotypes. Accordingly, Radakovich et al.¹²⁷ used ML models to analyze 20 commonly mutated genes in myeloid malignancies, revealing associations between clinical characteristics and specific genomic alterations, such as mutations in *TP53*, *ASXL1*, and *KRAS*. Additionally, 24 commonly mutated genes in MDS, including *SF3B1*, *TET2*, and *ASXL1*, among others, were analyzed by an interpretable ML model¹²⁸. Leveraging both clinical and NGS data, this model demonstrated remarkable precision in distinguishing MDS from other myeloid malignancies, with confidence intervals of nearly 95%¹²⁸.

In addition to linking specific point mutations with specific diseases, AI can also help predict the pathogenicity of these mutations, which if adopted, would significantly impact clinical decisions. In 2020, deep generative models of evolutionary data were developed to predict variant pathogenicity without relying on labels¹²⁹. Subsequently, Cheng et al.¹³⁰ introduced a model named AlphaMissense, which was specifically designed for predicting the pathogenicity of missense variants, a subset of point mutations. The group provided a database containing predictions for all possible single amino acid substitutions in humans, classifying 89% of 71 million possible missense variants as either likely benign or likely pathogenic.

6.3. Genomic structural variations

Genomic SVs are alterations in the structure of an organism's chromosomes involving sequence lengths >50 bp. While numerous SVs are linked to genetic diseases, many are not. SVs of uncertain significance are part of the reason why some clinical cases remain unresolved after whole-genome sequencing. Therefore, methods to predict the pathogenicity of these SVs are essential for fully harnessing the diagnostic capabilities of long-read sequencing. To help clinicians and researchers resolve cases and understand new disease mechanisms, a publicly available classifier named StrVCTVRE, which utilizes an RF algorithm, was developed to distinguish pathogenic SVs from benign SVs that overlap exons¹³¹. Furthermore, Popic et al.¹³² proposed an extensible DL framework, Cue, to call and genotype SVs of different sizes and types. But according to a study that systematically evaluated the performance of 69 SV-detection algorithms, accurate SV calls depend more on the selection and combination of algorithms than on the methods used in the algorithms¹⁵⁴.

CNVs are a subset of SVs that involve changes in the genome's copy number through duplication or deletion of large DNA segments, significantly affecting the genome and contributing to the development and progression of hematologic malignancies. For example, TP53 copy number loss is more common in AML patients with harmful TP53 missense mutations, and this is associated with cytogenetic risk groups, particularly those with a complex karyotype¹⁵⁵. Accurate CNV detection and interpretation are vital for the diagnosis and prognostic monitoring of hematologic malignancies. But detecting CNVs can be challenging, as they can be difficult to distinguish from noise in whole-genome—sequencing data. This is where AI has come into play, assisting in the detection, interpretation, and prediction of CNV pathogenicity. For example, Zhang et al.¹³⁵ developed a computational framework called X—CNV to predict the pathogenicity of CNVs by integrating more than 30 predictive features on a genome-wide basis; the pathogenic effect was quantitatively measured by a meta-voting prediction score based on the probabilistic value generated from the XGBoost algorithm. The X—CNV model achieved an AUC value of 0.96 in the training set and 0.94 in the validation set¹³⁵. More recently, Yu et al.¹³⁴ designed a DL algorithm to facilitate the accurate detection of chromothripsis, which can predict poor clinical outcomes in patients with multiple myeloma. Their algorithm leverages the intrinsic relationships between various CNV features. Building on this, researchers have integrated CNV and SNV signals to further improve detection accuracy and sensitivity, providing a more comprehensive tumor genomic profile. Recently, a research team developed MRD-EDGE, a ML-guided platform for plasma whole-

genome sequencing, which combines SNV and CNV signals to enable ultra-sensitive circulating tumor DNA detection and tumor burden monitoring. Together, these advances illustrate how AI-driven tools are transforming the detection and interpretation of CNVs, optimizing cancer diagnosis, prognosis, and treatment strategies¹⁵⁶.

6.4. Gene fusion and rearrangement

One critical aspect contributing to the complexity of hematological malignancies is the presence of genetic aberrations in the form of fusion genes, resulting from the fusion of 2 separate genes. Fusion genes arise from chromosomal translocations, inversions, or other genomic rearrangements, leading to the juxtaposition of previously independent genetic elements. The identification and characterization of fusion genes has offered valuable insights into the underlying molecular mechanisms governing hematopoietic cell differentiation, proliferation, and survival. Moreover, fusion genes have emerged as crucial diagnostic and prognostic markers for guiding clinical decision-making in the management of hematological malignancies. Data from different types of mutations can be integrated to collectively suggest disease diagnosis, classification, and risk prediction. But manual detection is limited by its time-intensive nature, subjectivity in interpretation, tolerance for noise, scalability challenges, and high costs. Therefore, incorporating AI could revolutionize fusion gene analysis, refine diagnosis, and improve prognostic accuracy in hematologic cancers.

To prevent misclassification in low-grade lymphomas and to retrieve clinically important characteristics in high-grade lymphomas, one research group developed a platform that combines a middle-throughput gene expression assay and ML to identify the subtypes of B-cell non-Hodgkin lymphomas¹³⁶. Using ligation-dependent real-time polymerase chain reaction and NGS data, including SNVs, gene fusions, and other genetic markers, they developed an RF classifier to discriminate the 7 most frequent categories of B-cell non-Hodgkin lymphomas, showing about 80%–100% concordance with previous classification results. Another study by Awada et al.¹³⁷ introduced an AML molecular subclassification method through the integration of cytogenetic and gene sequencing data. Incorporating ML-driven genomic signatures for AML, they successfully revealed novel genomic AML subclasses. Notably, this comprehensive model displayed an impressive 97% accuracy in efficiently categorizing various AML subtypes and predicting their clinical outcomes.

6.5. AI-assisted prognosis marker identification

In the field of hematology disorders, precisely predicting a patient's prognosis hinges significantly on molecular typing. The transition from knowing diseases only by phenotype to also by genotype has led to new insights into pathogenesis. The availability of molecular information also allows for the identification of more diagnostic and prognostic markers, personalized treatment to address interpatient variability, and targeted therapy. Using clinical genomic data, researchers have established scoring systems and constructed prognostic prediction models that seem to enhance the precision and efficiency of patient diagnosis, treatment response prediction, and prognosis prediction. Creating a distinct scoring system by identifying specific genomic

aberrations, exemplified in NPM1-mutated AML¹⁵⁷, aids in categorizing new prognostic subgroups¹⁵⁸.

With a broader scope in mind, Fleming et al.¹⁵⁹ employed recursive partitioning and RF to construct a hierarchical prognostic risk model, which incorporated cytogenetic and molecular factors from 2074 non-APL AML patients into groupings, resulting in a lower error rate than that of the European Leukemia Net–based classification and cytogenetic analysis alone. Shreve et al.¹⁶⁰ crafted an ML model using genomic and clinical data from 3421 AML patients to predict individualized patient outcomes, which demonstrated superior performance compared to that of the European Leukemia Net classification. Additionally, Bersanelli et al.¹⁶¹ integrated 63 clinical and genomic variables from 2043 MDS patients to construct a prognostic model using Bayesian network analysis and hierarchical Dirichlet processes to delineate genomic associations and subgroups, thereby defining distinct subgroups correlated with specific clinical features. The predictive accuracy of these features has been corroborated by a dynamic ML-based general clinical model¹⁶², which achieved concordant indices of 0.74 for overall survival and 0.81 for leukemic transformation, surpassing those of models previously described by the same group¹⁶³.

6.6. Future directions of AI in molecular genetics

AI in the molecular genetics of hematological malignancies promises enhanced diagnostic precision and classification by identifying diverse genetic variants. Despite its potential to reduce reliance on specialized knowledge and improve data consistency, AI faces several challenges, including the need for extensive, high-quality datasets for training, the risk of bias in algorithm development, and the complexity of integrating AI into clinical workflows. To address these challenges, specific strategies can be implemented. First, collaborative efforts among institutions should be encouraged to create large, diverse, and well-annotated datasets, which are essential for training robust and generalizable AI models. Second, transparent and standardized protocols for algorithm development and validation should be established to minimize bias and ensure reproducibility. Third, interdisciplinary collaboration between AI experts, clinicians, and pathologists is crucial to design user-friendly AI tools that seamlessly integrate into existing clinical practices. Additionally, ongoing education and training for healthcare professionals will be key to fostering trust and adoption of AI technologies. By addressing these challenges through targeted strategies, the integration of AI into the molecular genetics of hematological malignancies can be optimized, ultimately improving diagnostic accuracy and patient outcomes.

7. Concluding remarks and prospects

Along with advancements in technology, AI has great potential to improve the field of medicine⁶⁷. We foresee a pivotal role for AI in the future of diagnostic hematology, as it has the potential to revolutionize disease detection, classification, and monitoring through enhanced precision, efficiency, and scalability. By integrating AI into clinical workflows, we can automate complex tasks such as cell morphology analysis, immunophenotyping, and genetic variant identification, thereby reducing diagnostic TATs and alleviating the burden on laboratory personnel. Furthermore, AI-driven tools can uncover subtle patterns in large datasets,

enabling earlier detection of hematologic malignancies and facilitating personalized treatment strategies. Despite these promising applications, challenges remain regarding the high frequency of cases requiring manual intervention. Mitigating biases and human errors, desirable AI models could enable medical professionals to evaluate or verify individual cases more efficiently while “safely” skipping common and routine cases. As no model is perfect (and data quality is the key determinant), integrating all diagnostic modalities through a combination of various AI techniques would offer a rapid, automated, data-driven overview of each suspected case of hematologic malignancy. The development of AI techniques will have to incorporate data from patients with rare conditions to prevent the oversight of these cases, which remains a challenge.

In clinical settings, AI-driven tools are alleviating the workload of laboratory personnel by automating tasks like cell morphology analysis. These systems can quickly detect abnormal cells, minimizing human error and inter-observer variability, thus enhancing the reliability and timeliness of diagnostic workflows. By automating the interpretation of flow cytometry and genetic data, AI uncovers subtle patterns and mutations that might otherwise be overlooked, improving diagnostic accuracy and enabling more informed treatment decisions. This ability to detect critical details facilitates better risk stratification, ultimately guiding more effective and personalized treatment plans. Furthermore, AI’s integration in analyzing genomic data helps clinicians identify specific mutations and molecular markers, allowing for a more tailored approach to therapy that not only increases treatment efficacy but also reduces unnecessary side effects for patients. While seminal studies have led to breakthroughs, AI integration into real clinical settings has faced several barriers, including but not limited to technical, ethical, and clinical barriers. The truth is that even the most robust ML model follows the maxim of “garbage in, garbage out” in algorithm development¹⁶⁴. The performance and robustness of different AI methods are largely dependent on the availability of extensive standardized digital data, which may greatly eliminate potential bias when training the algorithms. Furthermore, a global agreement on “ethical AI” emerges around 5 principles—transparency, justice and fairness, nonmaleficence, responsibility, and privacy^{165,166}—highlighting the importance of integrating guideline-development efforts with substantive ethical analysis and adequate implementation strategies. And since the diagnosis of hematolymphoid diseases relies on multifaceted empirical evidence including clinical features, leveraging historical, longitudinal, and multimodal data is critical. Last but not least, clinical trials and multicenter studies are prerequisites for any AI model to be applied in a clinical workflow. Tackling these barriers will require interdisciplinary cooperation of computer scientists, health policy makers, hospital managing personnel, and, especially, diagnostic experts. Frontline hematopathologists and technicians should learn to embrace and utilize the assistance from AI. Importantly, they can also help to improve AI with wise human-level strategies.

Acknowledgments

The work was supported by research funding provided by the National Cancer Institute (U01CA252965, USA), the Eunice Kennedy Shriver National Institute of Child Health and Human Development (R01HD090927 and R01HD103511, USA), the National Institute of Allergy and Infectious Diseases

(R01AI144168, R01AI175618, R01AI173021, R01AI174964, R01AI177986, and R01AI179714, USA), the U.S. Department of Defense (W8IXWH1910926, USA), the National Institute of Neurological Disorders and Stroke (R21NS130542, USA), the Science and Technology Project of Sichuan Province of China (Grant No.2023NSFSC1484, China), and the 1·3·5 Project for Disciplines of Excellence, West China Hospital, Sichuan University (Grant Nos. 2023HXFH034 and 25HXJS035, China).

Author contributions

Hongyan Liao, Feng Zhang, Binwu Ying, and Tony Hu contributed to the conceptualization. Hongyan Liao, Feng Zhang, Yifei Li, Yanrui Sun, and Qin Zheng drafted the manuscript. Hongyan Liao, Feng Zhang, Fengyu Chen, Darcée D. Sloboda, Qin Zheng, Binwu Ying, and Tony Hu provided critical revisions. Hongyan Liao and Tony Hu provided the funding support. Binwu Ying and Tony Hu were responsible for the decision to submit the manuscript.

Conflicts of interest

The authors have no conflicts of interest to declare.

References

- Arber DA, Orazi A, Hasserjian RP, Borowitz MJ, Calvo KR, Kvasnicka HM, et al. International consensus classification of myeloid neoplasms and acute leukemias: integrating morphologic, clinical, and genomic data. *Blood* 2022;**140**:1200–28.
- Campo E, Jaffe ES, Cook JR, Quintanilla-Martinez L, Swerdlow SH, Anderson KC, et al. The international consensus classification of mature lymphoid neoplasms: a report from the Clinical Advisory Committee. *Blood* 2022;**140**:1229–53.
- Alaggio R, Amador C, Anagnostopoulos I, Attygalle AD, Araujo IBdO, Berti E, et al. The 5th edition of the World Health Organization classification of haematolymphoid tumours: lymphoid neoplasms. *Leukemia* 2022;**36**:1720–48.
- Swerdlow SH CE, Harris NL, Jaffe ES, Pileri SA, Stein H, Thiele J. WHO classification of tumours of haematopoietic and lymphoid tissues, revised 4th ed., Volume vol. 2. Lyon: IARC Press; 2017.
- Taylor J, Xiao WB, Abdel-Wahab O. Diagnosis and classification of hematologic malignancies on the basis of genetics. *Blood* 2017;**130**:410–23.
- Weber GM, Mandl KD, Kohane IS. Finding the missing link for big biomedical data. *JAMA* 2014;**311**:2479–80.
- Frater JL, Zini G, d'Onofrio G, Rogers HJ. COVID-19 and the clinical hematology laboratory. *Int J Lab Hematol* 2020;**42**:11–8.
- Russell SJ, Norvig P. *Artificial intelligence: a modern approach*. Boston: Pearson; 2016.
- Alpaydin E. *Introduction to machine learning*. Cambridge: MIT press; 2020.
- Haug CJ, Drazen JM. Artificial intelligence and machine learning in clinical medicine, 2023. *N Engl J Med* 2023;**388**:1201–8.
- Tefferi A, Hanson CA, Inwards DJ. How to interpret and pursue an abnormal complete blood count in adults. *Mayo Clin Proc* 2005;**80**:923–36.
- Ratley A, Minj J, Patre P. Leukemia disease detection and classification using machine learning approaches: a review. In: *Proceedings of the 2020 first International Conference on power, control and computing technologies (ICPC2T)*. New York: IEEE; 2020 Jan 1–3. Raipur, India.
- Barnes PW, McFadden SL, Machin SJ, Simson E, et al. The international consensus group for hematology review: suggested criteria for action following automated CBC and WBC differential analysis. *Lab Hematol* 2005;**11**:83–90.
- Junker R, Schlebusch H, Lupp PB. Point-of-care testing in hospitals and primary care. *Dtsch Arztebl Int* 2010;**107**:561–7.
- El Alaoui Y, Elomri A, Qaraqe M, Padmanabhan R, Yasin Taha R, El Omri H, et al. A review of artificial intelligence applications in hematology management: current practices and future prospects. *J Med Internet Res* 2022;**24**:e36490.
- Haider RZ, Ujjan IU, Khan NA, Urrechaga E, Shamsi TS. Beyond the in-practice CBC: the research CBC parameters-driven machine learning predictive modeling for early differentiation among leukemias. *Diagnostics (Basel)* 2022;**12**:138.
- Haider RZ, Ujjan IU, Shamsi TS. Cell population data-driven acute promyelocytic leukemia flagging through artificial neural network predictive modeling. *Transl Oncol* 2020;**13**:11–6.
- Alcazer V, Le Meur G, Roccon M, Barriere S, Le Calvez B, Badaoui B, et al. Evaluation of a machine-learning model based on laboratory parameters for the prediction of acute leukaemia subtypes: a multicentre model development and validation study in France. *Lancet Digit Health* 2024;**6**:e323–33.
- Hauser RG, Esserman D, Beste LA, Ong SY, Colomb Jr DG, Bhargava A, et al. A machine learning model to successfully predict future diagnosis of chronic myelogenous leukemia with retrospective electronic health records data. *Am J Clin Pathol* 2021;**156**:1142–8.
- Liao HY, Xu YX, Meng Q, Qiao YF, Liu Y, et al. A convolutional neural network-based, quantitative complete blood count scattergram-mapping framework promptly screens acute promyelocytic leukemia with high sensitivity. *Cancer* 2023;**129**:2986–98.
- Zhu JJ, Lemaire P, Mathis S, Ronez E, Clauser S, Jondeau K, et al. Machine learning-based improvement of MDS-CBC score brings platelets into the limelight to optimize smear review in the hematology laboratory. *BMC Cancer* 2022;**22**:972.
- Mahmood N, Shahid S, Bakhshi T, Riaz S, Ghufuran H, Yaqoob M. Identification of significant risks in pediatric acute lymphoblastic leukemia (ALL) through machine learning (ML) approach. *Med Biol Eng Comput* 2020;**58**:2631–40.
- Shen ZY, Zhang S, Jiao Y, Shi YX, Zhang H, Wang F, et al. LASSO model better predicted the prognosis of DLBCL than random forest model: a retrospective multicenter analysis of HHLWG. *J Oncol* 2022;**2022**:1618272.
- Eckardt JN, Röhlrig C, Metzeler K, Kramer M, Stasik S, Georgi JA, et al. Prediction of complete remission and survival in acute myeloid leukemia using supervised machine learning. *Haematologica* 2023;**108**:690–704.
- Bachar N, Benbassat D, Brailovsky D, Eshel Y, Gluck D, Levner D, et al. An artificial intelligence-assisted diagnostic platform for rapid near-patient hematology. *Am J Hematol* 2021;**96**:1264–74.
- Gasparin AT, Araujo CIF, Schmitt P, Cardoso MR, Perussolo MC, de Jesus TCS, et al. Hilab system, a new point-of-care hematology analyzer supported by the internet of things and artificial intelligence. *Sci Rep* 2022;**12**:10409.
- Boutault JN, Peterlin P, Boubaya M, Sockel K, Chevallier P, Garnier A, et al. A novel complete blood count-based score to screen for myelodysplastic syndrome in cytopenic patients. *Br J Haematol* 2018;**183**:736–46.
- Hassan U, Reddy Jr B, Damhorst G, Sonoiki O, Ghonge T, Yang C, et al. A microfluidic biochip for complete blood cell counts at the point-of-care. *Technology (Singap World Sci)* 2015;**3**:201–13.
- Gasparin AT, Araujo CIF, Cardoso MR, Schmitt P, Godoy JB, Reichert ES, et al. Hilab system device in an oncological hospital: a new clinical approach for point of care CBC test, supported by the internet of things and machine learning. *Diagnostics (Basel)* 2023;**13**:10.
- Bain BJ. Diagnosis from the blood smear. *N Engl J Med* 2005;**353**:498–507.
- Bennett JM, Catovsky D, Daniel MT, Flandrin G, Galton DA, Gralnick HR, et al. Proposals for the classification of the acute

- leukaemias. French-American-British (FAB) co-operative group. *Br J Haematol* 1976;**33**:451–8.
32. Sasada K, Yamamoto N, Masuda H, Tanaka Y, Ishihara A, Takamatsu Y, et al. Inter-observer variance and the need for standardization in the morphological classification of myelodysplastic syndrome. *Leuk Res* 2018;**69**:54–9.
 33. Yabuta M, Nakamura I, Ida H, Masauzi H, Okada K, Kaga S, et al. Deep learning-based nuclear lobe count method for differential count of neutrophils. *Tohoku J Exp Med* 2021;**254**:199–206.
 34. Acevedo A, Merino A, Boldu L, Molina A, Alferez S, Rodellar J. A new convolutional neural network predictive model for the automatic recognition of hypogranulated neutrophils in myelodysplastic syndromes. *Comput Biol Med* 2021;**134**:104479.
 35. Kweon OJ, Lim YK, Lee MK, 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**:e0267638.
 36. Tabe Y, Yamamoto T, Maenou I, Nakai R, Idei M, Horii T, et al. Performance evaluation of the digital cell imaging analyzer DI-60 integrated into the fully automated Sysmex XN hematology analyzer system. *Clin Chem Lab Med* 2015;**53**:281–9.
 37. Jiang H, Xu W, Chen W, He J, Jiang HQ, Mao ZG, 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.
 38. Alferez S, Merino A, Bigorra L, Rodellar J. Characterization and automatic screening of reactive and abnormal neoplastic B lymphoid cells from peripheral blood. *Int J Lab Hematol* 2016;**38**:209–19.
 39. Pohlkamp C, Jhalani K, Nadarajah N, Heo I, Wetton W, Drescher R, et al. Machine learning (ML) can successfully support microscopic differential counts of peripheral blood smears in a high throughput hematology laboratory. *Blood* 2020;**136**:45–6.
 40. Matek C, Schwarz S, Spiekermann K, Marr C. Human-level recognition of blast cells in acute myeloid leukaemia with convolutional neural networks. *Nat Mach Intell* 2019;**1**:538–44.
 41. Chen YF, Zhang CY, Chen B, Huang YY, Sun YF, Wang CM, et al. Accurate leukocyte detection based on deformable-DETR and multi-level feature fusion for aiding diagnosis of blood diseases. *Comput Biol Med* 2024;**170**:107917.
 42. Chen P, Zhang L, Cao XY, Jin XY, Chen N, Zhang L, et al. Detection of circulating plasma cells in peripheral blood using deep learning-based morphological analysis. *Cancer* 2024;**130**:1884–93.
 43. Krappe S, Wittenberg T, Haferlach T, Münzenmayer C. Automated morphological analysis of bone marrow cells in microscopic images for diagnosis of leukemia: nucleus-plasma separation and cell classification using a hierarchical tree model of hematopoiesis. In: *Proceedings of the SPIE medical imaging 2016 Conference*. San Diego, CA: Bellingham, WA: SPIE; 2016 Mar 01-05. 2016 Mar 1–5.
 44. Jin H, Fu XY, Cao XY, Sun MX, Wang XF, Zhong YH, et al. Developing and preliminary validating an automatic cell classification system for bone marrow smears: a pilot study. *J Med Syst* 2020;**44**:1–10.
 45. Wang WN, Luo MG, Guo PR, Wei Y, Tan Y, Shi HX. Artificial intelligence-assisted diagnosis of hematologic diseases based on bone marrow smears using deep neural networks. *Comput Methods Progr Biomed* 2023;**231**:107343.
 46. Matek C, Krappe S, Münzenmayer C, Haferlach T, Marr C. Highly accurate differentiation of bone marrow cell morphologies using deep neural networks on a large image data set. *Blood* 2021;**138**:1917–27.
 47. Kimura K, Tabe Y, Ai T, Matsuzaki A, Nishibe K, Ebihara M, et al. A novel automated image analysis system using deep convolutional neural networks to diagnose MDS. *Blood* 2019;**134**:4670.
 48. Wang MF, Dong CX, Gao Y, Li JL, Han MR, Wang LJ. A deep learning model for the automatic recognition of aplastic anemia, myelodysplastic syndromes, and acute myeloid leukemia based on bone marrow smear. *Front Oncol* 2022;**12**:844978.
 49. Sirinukunwattana K, Aberdeen A, Theissen H, Sousos N, Psaila B, Mead AJ, et al. Artificial intelligence-based morphological fingerprinting of megakaryocytes: a new tool for assessing disease in MPN patients. *Blood Adv* 2020;**4**:3284–94.
 50. Neoh SC, Srisukkharn W, Todryk LZ, Greystoke B, Lim CP, Hossain MA, et al. An intelligent decision support system for leukaemia diagnosis using microscopic blood images. *Sci Rep* 2015;**5**:14938.
 51. Pansombut T, Wikaisuksakul S, Khongkrapan K, Phon-On A. Convolutional neural networks for recognition of lymphoblast cell images. *Comput Intell Neurosci* 2019;**2019**:7519603.
 52. Shafique S, Tehsin S. Acute lymphoblastic leukemia detection and classification of its subtypes using pretrained deep convolutional neural networks. *Technol Cancer Res Treat* 2018;**17**:1533033818802789.
 53. Yu ZB, Li JH, Wen X, Han YL, Jiang PL, Zhu M, et al. AMLnet, A deep-learning pipeline for the differential diagnosis of acute myeloid leukemia from bone marrow smears. *J Hematol Oncol* 2023;**16**:27.
 54. Fu XY, Fu M, Li Q, Peng XG, Lu J, Fang FQ, et al. Morphogo: an automatic bone marrow cell classification system on digital images analyzed by artificial intelligence. *Acta Cytol* 2020;**64**:588–96.
 55. Boldú L, Merino A, Alferez 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**:755–61.
 56. Li N, Fan LQ, Xu H, Zhang XW, Bai ZZ, Li MH, et al. An AI-aided diagnostic framework for hematologic neoplasms based on morphologic features and medical expertise. *Lab Invest* 2023;**103**:100055.
 57. Ceelie H, Dinkelaar RB, van Gelder W. Examination of peripheral blood films using automated microscopy; evaluation of Diffmaster Octavia and Cellavision DM96. *J Clin Pathol* 2007;**60**:72–9.
 58. Saad Albichr I, Sottiaux JY, Hotton J, De Laveleye M, Dupret P, Detry G. Cross-evaluation of five slidemakers and three automated image analysis systems: the pitfalls of automation?. *Int J Lab Hematol* 2020;**42**:573–80.
 59. Nakashima MO, Doyle TJ, Phelan-Lewin K, Summers JM, Bena J, Hsi ED. Assessment of semi-quantitative grading of red blood cell abnormalities utilizing images from the CellaVision DM96 compared to manual light microscopy. *Int J Lab Hematol* 2017;**39**:e110–2.
 60. Lincz LF, Makhija K, Attalla K, Scorgie FE, Enjeti AK, Prasad R. A comparative evaluation of three consecutive artificial intelligence algorithms released by Techcyte for identification of blasts and white blood cells in abnormal peripheral blood films. *Int J Lab Hematol* 2024;**46**:92–8.
 61. Granell M, Calvo X, Garcia-Guiñón A, Escoda L, Abella E, Martínez CM, et al. Prognostic impact of circulating plasma cells in patients with multiple myeloma: implications for plasma cell leukemia definition. *Haematologica* 2017;**102**:1099.
 62. Wang CW, Huang SC, Lee YC, Shen YJ, Meng SI, Gaol JL. Deep learning for bone marrow cell detection and classification on whole-slide images. *Med Image Anal* 2022;**75**:102270.
 63. Tayebi RM, Mu Y, Dehkharghanian T, Ross C, Sur M, Foley R, et al. Automated bone marrow cytology using deep learning to generate a histogram of cell types. *Commun Med* 2022;**2**:45.
 64. Hu G, Wang BM, Hu BX, Chen D, Hu LH, Li C, et al. From WSI-level to patch-level: structure prior-guided binuclear cell fine-grained detection. *Med Image Anal* 2023;**89**:102931.
 65. Dimitriou N, Arandjelovic O, Caie PD. Deep learning for whole slide image analysis: an overview. *Front Med (Lausanne)* 2019;**6**:264.
 66. Sarachakov A, Varlamova A, Svekolkin V, Polyakova M, Valencia I, Unkenholz C, et al. Spatial mapping of human hematopoiesis at single-cell resolution reveals aging-associated topographic remodeling. *Blood* 2023;**142**:2282–95.
 67. Yu KH, Beam AL, Kohane IS. Artificial intelligence in healthcare. *Nat Biomed Eng* 2018;**2**:719–31.
 68. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med* 2019;**25**:44–56.
 69. Arber DA, Orazi A, Hasserjian R, Thiele J, Borowitz MJ, Le Beau MM, et al. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood* 2016;**127**:2391–405.

70. Dunphy CH. Applications of flow cytometry and immunohistochemistry to diagnostic hematopathology. *Arch Pathol Lab Med* 2004;**128**:1004–22.
71. Koladiya A, Davis KL. Advances in clinical mass cytometry. *Clin Lab Med* 2023;**43**:507–19.
72. Brestoff JR. Full spectrum flow cytometry in the clinical laboratory. *Int J Lab Hematol* 2023;**45**:44–9.
73. Spitzer Matthew H, Nolan Garry P. Mass cytometry: single cells, many features. *Cell* 2016;**165**:780–91.
74. Aghaeepour N, Finak G, Consortium F, Consortium D, Hoos H, Mosmann TR, et al. Critical assessment of automated flow cytometry data analysis techniques. *Nat Methods* 2013;**10**:228–38.
75. Fuda F, Chen MY, Chen WN, Cox A. Artificial intelligence in clinical multiparameter flow cytometry and mass cytometry-key tools and progress. *Semin Diagn Pathol* 2023;**40**:120–8.
76. Mair F, Hartmann FJ, Mrdjen D, Tosevski V, Krieg C, Becher B. The end of gating?. An introduction to automated analysis of high dimensional cytometry data. *Eur J Immunol* 2016;**46**:34–43.
77. Hahne F, LeMeur N, Brinkman RR, Ellis B, Haaland P, Sarkar D, et al. FlowCore: a bioconductor package for high throughput flow cytometry. *BMC Bioinf* 2009;**10**:1–8.
78. Lo K, Brinkman RR, Gottardo R. Automated gating of flow cytometry data via robust model-based clustering. *Cytometry A* 2008;**73**:321–32.
79. Van Gassen S, Callebaut B, Van Helden MJ, Lambrecht BN, Demeester P, Dhaene T, et al. FlowSOM: using self-organizing maps for visualization and interpretation of cytometry data. *Cytometry A* 2015;**87**:636–45.
80. Quintelier K, Couckuyt A, Emmaneel A, Aerts J, Saeys Y, Van Gassen S. Analyzing high-dimensional cytometry data using FlowSOM. *Nat Protoc* 2021;**16**:3775–801.
81. Qiu P, Simonds EF, Bendall SC, Gibbs Jr KD, Bruggner RV, Linderman MD, et al. Extracting a cellular hierarchy from high-dimensional cytometry data with SPADE. *Nat Biotechnol* 2011;**29**:886–91.
82. Amir E-aD, Davis KL, Tadmor MD, Simonds EF, Levine JH, Bendall SC, et al. viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia. *Nat Biotechnol* 2013;**31**:545–52.
83. Levine JH, Simonds EF, Bendall SC, Davis KL, El-ad DA, Tadmor MD, et al. Data-driven phenotypic dissection of AML reveals progenitor-like cells that correlate with prognosis. *Cell* 2015;**162**:184–97.
84. Monaghan SA, Li JL, Liu YC, Ko MY, Boyiadzis M, Chang TY, et al. A machine learning approach to the classification of acute leukemias and distinction from nonneoplastic cytopenias using flow cytometry data. *Am J Clin Pathol* 2021;**157**:546–53.
85. Zhao M, Mallesh N, Höllein A, Schabath R, Haferlach C, Haferlach T, et al. Hematologist-level classification of mature B-cell neoplasm using deep learning on multiparameter flow cytometry data. *Cytometry A* 2020;**97**:1073–80.
86. Ng DP, Zuromski LM. Augmented human intelligence and automated diagnosis in flow cytometry for hematologic malignancies. *Am J Clin Pathol* 2021;**155**:597–605.
87. Simonson PD, Lee AY, Wu D. Potential for process improvement of clinical flow cytometry by incorporating real-time automated screening of data to expedite addition of antibody panels. *Am J Clin Pathol* 2022;**157**:443–50.
88. DiGiuseppe JA, Cardinali JL, Rezuze WN, Pe'er D. PhenoGraph and viSNE facilitate the identification of abnormal T-cell populations in routine clinical flow cytometric data. *Cytometry B* 2018;**94**:744–57.
89. Goshaw JM, Gao Q, Wardrope J, Dogan A, Roshal M. 14-Color single tube for flow cytometric characterization of CD5+ B-LPDs and high sensitivity automated minimal residual disease quantitation of CLL/SLL. *Cytometry B* 2021;**100**:509–18.
90. Liechti T, Weber LM, Ashhurst TM, Stanley N, Prlc M, Van Gassen S, et al. An updated guide for the perplexed: cytometry in the high-dimensional era. *Nat Immunol* 2021;**22**:1190–7.
91. Bazinet A, Wang A, Li X, Jia F, Mo H, Wang W, et al. Automated quantification of measurable residual disease in chronic lymphocytic leukemia using an artificial intelligence-assisted workflow. *Cytometry B Clin Cytom* 2024;**106**:264–71.
92. Van Gassen S, Gaudilliere B, Angst MS, Saeys Y, Aghaeepour N. CytoNorm: a normalization algorithm for cytometry data. *Cytometry A* 2020;**97**:268–78.
93. Becht E, McInnes L, Healy J, Dutertre CA, Kwok IWH, Ng LG, et al. Dimensionality reduction for visualizing single-cell data using UMAP. *Nat Biotechnol* 2018;**37**:38–44.
94. Hu Z, Juijavarapu C, Hughey JJ, Andorf S, Lee HC, Gherardini PF, et al. MetaCyto: a tool for automated meta-analysis of mass and flow cytometry data. *Cell Rep* 2018;**24**:1377–88.
95. Bruggner RV, Bodenmiller B, Dill DL, Tibshirani RJ, Nolan GP. Automated identification of stratifying signatures in cellular subpopulations. *Proc Natl Acad Sci U S A* 2014;**111**:E2770–7.
96. Van Gassen S, Vens C, Dhaene T, Lambrecht BN, Saeys Y. FloReMi: flow density survival regression using minimal feature redundancy. *Cytometry A* 2016;**89**:22–9.
97. Arvaniti E, Claassen M. Sensitive detection of rare disease-associated cell subsets via representation learning. *Nat Commun* 2017;**8**:14825.
98. Hu ZC, Tang A, Singh J, Bhattacharya S, Butte AJ. A robust and interpretable end-to-end deep learning model for cytometry data. *Proc Natl Acad Sci U S A* 2020;**117**:21373–80.
99. Ding JR, Condon A, Shah SP. Interpretable dimensionality reduction of single cell transcriptome data with deep generative models. *Nat Commun* 2018;**9**:2002.
100. Abdelaal T, van Unen V, Holtt T, Koning F, Reinders MJT, Mahfouz A. Predicting cell populations in single cell mass cytometry data. *Cytometry A* 2019;**95**:769–81.
101. Cheng LJ, Karkhanis P, Gokbag B, Liu YZ, Li L. DGCyTOF: deep learning with graphic cluster visualization to predict cell types of single cell mass cytometry data. *PLoS Comput Biol* 2022;**18**:e1008885.
102. Li HM, Shaham U, Stanton KP, Yao Y, Montgomery RR, Kluger Y. Gating mass cytometry data by deep learning. *Bioinformatics* 2017;**33**:3423–30.
103. Duetz C, Bachas C, Westers TM, van de Loosdrecht AA. Computational analysis of flow cytometry data in hematological malignancies: future clinical practice?. *Curr Opin Oncol* 2020;**32**:162–9.
104. Wang YF, Li JL, Lee CC, Wallace PK, Ko BS. Using artificial intelligence to interpret clinical flow cytometry datasets for automated disease diagnosis and/or monitoring. In: Hawley Teresa S Hawley, Robert G, editors. *Methods in molecular biology (Clifton, NJ)*. New York: Springer; 2024. p. 353–67.
105. Shadman M. Diagnosis and treatment of chronic lymphocytic leukemia: a review. *JAMA* 2023;**329**:918–32.
106. Elhadary M, Elshoeibi AM, Badr A, Elsayed B, Metwally O, Elshoeibi AM, et al. Revolutionizing chronic lymphocytic leukemia diagnosis: a deep dive into the diverse applications of machine learning. *Blood Rev* 2023;**62**:101134.
107. Hunt AM, Shallenberger W, Ten Eyck SP, Craig FE. Use of internal control T-cell populations in the flow cytometric evaluation for T-cell neoplasms. *Cytometry B* 2016;**90**:404–14.
108. Fokken H, Waclawski J, Kattre N, Kloos A, Müller S, Ettinger M, et al. A 19-color single-tube full spectrum flow cytometry assay for the detection of measurable residual disease in acute myeloid leukemia. *Cytometry A* 2024;**105**:181–95.
109. College of American Pathologists [Internet]. Northfield, IL: College of American Pathologists; c2023 [cited 2025 Apr 18]. Is it magic? Artificial intelligence integration into clinical cytometry workflows. Available from: <https://www.cap.org/member-resources/articles/is-it-magic-artificial-intelligence-integration-into-clinical-cytometry-workflows>.
110. Döhner H, Estey E, Grimwade D, Amadori S, Appelbaum FR, Büchner T, et al. Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. *Blood* 2017;**129**:424–47.

111. Granada I, Palomo L, Ruiz-Xivillé N, Mallo M, Solé F. Cytogenetics in the genomic era. *Best Pract Res Clin Haematol* 2020;**33**:101196.
112. Akkari YMN, Baughn LB, Dubuc AM, Smith AC, Mallo M, Dal Cin P, et al. Guiding the global evolution of cytogenetic testing for hematologic malignancies. *Blood* 2022;**139**:2273–84.
113. Mantere T, Neveling K, Pebrel-Richard C, Benoist M, van der Zande G, Kater-Baats E, et al. Optical genome mapping enables constitutional chromosomal aberration detection. *Am J Hum Genet* 2021;**108**:1409–22.
114. Sharma M, Saha O, Sriraman A, Hebbalaguppe R, Vig L, Karande S. Crowdsourcing for chromosome segmentation and deep classification. In: *Proceedings of the 2017 IEEE Conference on computer vision and pattern recognition Workshops; 2017 Jul 21-26; Honolulu, HI, USA*. New York: IEEE; 2017.
115. Chen XY, Cai Q, Ma N, Li HS. ChroSegNet: an attention-based model for chromosome segmentation with enhanced processing. *Appl Sci* 2023;**13**:1–5.
116. Vajen B, Hänselmann S, Lutterloh F, Käfer S, Espenkötter J, Beening A, et al. Classification of fluorescent R-Band metaphase chromosomes using a convolutional neural network is precise and fast in generating karyograms of hematologic neoplastic cells. *Cancer Genet* 2022;**260–261**:23–9.
117. Xie N, Li X, Li K, Yang Y, Shen HT. Statistical karyotype analysis using CNN and geometric optimization. *IEEE Access* 2019;**7**:179445–53.
118. Haferlach C, Hänselmann S, Walter W, Volkert S, Zenger M, Kern W, et al. Artificial intelligence substantially supports chromosome banding analysis maintaining its strengths in hematologic diagnostics even in the era of newer technologies. *Blood* 2020;**136**:47–8.
119. Bokhari Y, Alhareeri A, Aljouie A, Alkhalidi A, Rashid M, Alawad M, et al. ChromoEnhancer: an artificial-intelligence-based tool to enhance neoplastic karyograms as an aid for effective analysis. *Cells* 2022;**11**:2244.
120. Gudla PR, Nakayama K, Pegoraro G, Misteli T. SpotLearn: convolutional neural network for detection of fluorescence *in situ* hybridization (FISH) signals in high-throughput imaging approaches. *Cold Spring Harb Symp Quant Biol* 2017;**82**:57–70.
121. Zakrzewski F, de Back W, Weigert M, Wenke T, Zeugner S, Mantey R, et al. Automated detection of the HER2 gene amplification status in fluorescence *in situ* hybridization images for the diagnostics of cancer tissues. *Sci Rep* 2019;**9**:8231.
122. Frankenstein Z, Uraoka N, Aypar U, Aryeequaye R, Rao M, Hameed M, et al. Automated 3D scoring of fluorescence *in situ* hybridization (FISH) using a confocal whole slide imaging scanner. *Appl Microsc* 2021;**51**:4.
123. Luxton JJ, McKenna MJ, Lewis AM, Taylor LE, Jhavar SG, Swanson GP, et al. Telomere length dynamics and chromosomal instability for predicting individual radiosensitivity and risk via machine learning. *J Personalized Med* 2021;**11**:188.
124. Morley TJ, Han L, Castro VM, Morra J, Perlis RH, Cox NJ, et al. Phenotypic signatures in clinical data enable systematic identification of patients for genetic testing. *Nat Med* 2021;**27**:1097–104.
125. Morita K, Wang F, Makishima H, Yan Y, Yoshizato T, Yoshida K, et al. Pan-myeloid leukemia analysis: machine learning-based approach to predict phenotype and clinical outcomes using mutation data. *Blood* 2018;**132**:1801.
126. Poplin R, Chang PC, Alexander D, Schwartz S, Colthurst T, Ku A, et al. A universal SNP and small-indel variant caller using deep neural networks. *Nat Biotechnol* 2018;**36**:983–7.
127. Radakovich N, Malcovati L, Meggendorfer M, Sekeres MA, Shreve J, Hilton CB, et al. Genotype-phenotype correlations in patients with myeloid malignancies using explainable artificial intelligence. *Blood* 2020;**136**:31–2.
128. Radakovich N, Meggendorfer M, Malcovati L, Hilton CB, Sekeres MA, Shreve J, et al. A geno-clinical decision model for the diagnosis of myelodysplastic syndromes. *Blood Adv* 2021;**5**:4361–9.
129. Frazer J, Notin P, Dias M, Gomez A, Min JK, Brock K, et al. Disease variant prediction with deep generative models of evolutionary data. *Nature* 2021;**599**:91–5.
130. Cheng J, Novati G, Pan J, Bycroft C, Zemgulyte A, Applebaum T, et al. Accurate proteome-wide missense variant effect prediction with AlphaMissense. *Sci Technol Humanit* 2023;**381**:eadg7492.
131. Sharo AG, Hu Z, Sunyaev SR, Brenner SE. StrVCTVRE: a supervised learning method to predict the pathogenicity of human genome structural variants. *Am J Hum Genet* 2022;**109**:195–209.
132. Popic V, Rohlicek C, Cunial F, Hajirasouliha I, Meleshko D, Garimella K, et al. Cue: a deep-learning framework for structural variant discovery and genotyping. *Nat Methods* 2023;**20**:559–68.
133. Wang TF, Sun JH, Zhang XQ, Wang WJ, Zhou Q. CNV-P: a machine-learning framework for predicting high confident copy number variations. *PeerJ* 2021;**9**:e12564.
134. Yu J, Chen N, Zheng Z, Gao M, Liang N, Wong KC. Chromothripsis detection with multiple myeloma patients based on deep graph learning. *Bioinformatics* 2023;**39**:7.
135. Zhang L, Shi JR, Ouyang J, Zhang RQ, Tao YR, Yuan DS, et al. X-CNV: genome-wide prediction of the pathogenicity of copy number variations. *Genome Med* 2021;**13**:132.
136. Bobee V, Drieux F, Marchand V, Sater V, Veresezan L, Picqueton JM, et al. Combining gene expression profiling and machine learning to diagnose B-cell non-Hodgkin lymphoma. *Blood Cancer J* 2020;**10**:59.
137. Awada H, Durmaz A, Gurnari C, Kishtagari A, Meggendorfer M, Kerr CM, et al. Machine learning integrates genomic signatures for subclassification beyond primary and secondary acute myeloid leukemia. *Blood* 2021;**138**:1885–95.
138. Grisan E, Poletti E, Tomelleri C, Ruggeri A. Automatic segmentation of chromosomes in Q-band images. *Annu Int Conf IEEE Eng Med Biol Soc* 2007;**2007**:5513–6.
139. Wang XW, Zheng B, Li SB, Mulvihill JJ, Chen XD, Liu H. Automated identification of abnormal metaphase chromosome cells for the detection of chronic myeloid leukemia using microscopic images. *J Biomed Opt* 2010;**15**:046026–12.
140. Khmelinskii A, Ventura R, Sanches J. A novel metric for bone marrow cells chromosome pairing. *IEEE Trans Biomed Eng* 2010;**57**:1420–9.
141. Khan S, Dsouza A, Sanches J, Ventura R. Geometric correction of deformed chromosomes for automatic karyotyping. In: *Proceedings of the Annual International Conference of IEEE engineering in medicine and biology Society*. San Diego, CA. New York: IEEE; 2012 Aug 28 – Sep 1. 2012.
142. Deng JH, Peng WX, Lu QY, Wang Z, Fu Q, Zhou XG, et al. Manually-established abnormal karyotype dataset based on normal chromosomes effectively train artificial intelligence model for better cytogenetic abnormalities prediction. *Research Square [Preprint]* 2023. <https://doi.org/10.21203/rs.3.rs-2913988/v1> [cited 2025 Mar 23].
143. Huber D, Voith von Voithenberg L, Kaigala GV. Fluorescence *in situ* hybridization (FISH): history, limitations and what to expect from micro-scale FISH?. *Micro Nano Eng* 2018;**1**:15–24.
144. Wang CW, Lin KY, Lin YJ, Khalil MA, Chu KL, Chao TK. A soft label deep learning to assist breast cancer target therapy and thyroid cancer diagnosis. *Cancers (Basel)* 2022;**14**:5312.
145. He J, Yi K, Zhang YJ, Zhang H, Hou J, Li R. A novel human multiple myeloma cell line with a 1q21 gain genetic abnormality and CKS1B overexpression. *Ann Transl Med* 2023;**11**:126.
146. Duncavage EJ, Bagg A, Hasserjian RP, DiNardo CD, Godley LA, Iacobucci I, et al. Genomic profiling for clinical decision making in myeloid neoplasms and acute leukemia. *Blood* 2022;**140**:2228–47.
147. Valkama A, Vorimo S, Kumpula TA, Rasanen H, Savolainen ER, Pylkas K, et al. Optical genome mapping as an alternative to FISH-based cytogenetic assessment in chronic lymphocytic leukemia. *Cancers (Basel)* 2023;**15**:1294.
148. Liu M, Zhang Y, Chen X, Wang F, Chen JQ, Cao PX, et al. Deciphering the structural variants in B cell acute lymphoblastic leukemia by optical genome mapping. *Blood* 2023;**142**:1610.
149. Levy B, Baughn LB, Akkari Y, Chartrand S, LaBarge B, Claxton D, et al. Optical genome mapping in acute myeloid leukemia: a multi-center evaluation. *Blood Adv* 2023;**7**:1297–307.

150. Haferlach T, Walter W. Challenging gold standard hematology diagnostics through the introduction of whole genome sequencing and artificial intelligence. *Int J Lab Hematol* 2023;**45**:156–62.
151. Goya R, Sun MG, Morin RD, Leung G, Ha G, Wiegand KC, et al. SNVMix: predicting single nucleotide variants from next-generation sequencing of tumors. *Bioinformatics* 2010;**26**:730–6.
152. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, et al. 10 Years of GWAS discovery: biology, function, and translation. *Am J Hum Genet* 2017;**101**:5–22.
153. Khurana E, Fu Y, Chakravarty D, Demichelis F, Rubin MA, Gerstein M. Role of non-coding sequence variants in cancer. *Nat Rev Genet* 2016;**17**:93–108.
154. Kosugi S, Momozawa Y, Liu X, Terao C, Kubo M, Kamatani Y. Comprehensive evaluation of structural variation detection algorithms for whole genome sequencing. *Genome Biol* 2019;**20**:117.
155. Tashakori M, Kadia T, Loghavi S, Daver N, Kanagal-Shamanna R, Pierce S, et al. TP53 copy number and protein expression inform mutation status across risk categories in acute myeloid leukemia. *Blood* 2022;**140**:58–72.
156. Widman AJ, Shah M, Frydendahl A, Halmos D, Khamnei CC, Ogaard N, et al. Ultrasensitive plasma-based monitoring of tumor burden using machine-learning-guided signal enrichment. *Nat Med* 2024;**30**:1655–66.
157. Heath EM, Chan SM, Minden MD, Murphy T, Shlush LI, Schimmer AD. Biological and clinical consequences of NPM1 mutations in AML. *Leukemia* 2017;**31**:798–807.
158. Patkar N, Shaikh AF, Kakirde C, Nathany S, Ramesh H, Bhanshe P, et al. A novel machine-learning-derived genetic score correlates with measurable residual disease and is highly predictive of outcome in acute myeloid leukemia with mutated NPM1. *Blood Cancer J* 2019;**9**:79.
159. Fleming S, Tsai C-H, Döhner H, Döhner K, Papaemmanuil E, Tien H-F, et al. Use of machine learning in 2074 cases of acute myeloid leukemia for genetic risk profiling. *Blood* 2019;**134**:1392.
160. Shreve J, Meggendorfer M, Awada H, Mukherjee S, Walter W, Hutter S, et al. A personalized prediction model to risk stratify patients with acute myeloid leukemia (AML) using artificial intelligence. *Blood* 2019;**134**:2091.
161. Bersanelli M, Travaglini E, Meggendorfer M, Matteuzzi T, Sala C, Mosca E, et al. Classification and personalized prognostic assessment on the basis of clinical and genomic features in myelodysplastic syndromes. *J Clin Oncol* 2021;**39**:1223–33.
162. Nazha A, Komrokji R, Meggendorfer M, Jia X, Radakovich N, Shreve J, et al. Personalized prediction model to risk stratify patients with myelodysplastic syndromes. *J Clin Oncol* 2021;**39**:3737–46.
163. Nazha A, Narkhede M, Radivoyevitch T, Seastone DJ, Patel BJ, Gerds AT, et al. Incorporation of molecular data into the revised international prognostic scoring system in treated patients with myelodysplastic syndromes. *Leukemia* 2016;**30**:2214–20.
164. Beam AL, Kohane IS. Big data and machine learning in health care. *JAMA* 2018;**319**:1317–8.
165. Jobin A, Ienca M, Vayena E. The global landscape of AI ethics guidelines. *Nat Mach Intell* 2019;**1**:389–99.
166. World Health Organization. *Ethics and governance of artificial intelligence for health: guidance on large multi-modal models*. Geneva, Switzerland: World Health Organization; 2024. Available from: <https://www.who.int/publications/i/item/9789240084759>.