



How artificial intelligence might disrupt diagnostics in hematology in the near future

Wencke Walter¹ · Claudia Haferlach¹ · Niroshan Nadarajah¹ · Ines Schmidts¹ · Constanze Kühn¹ · Wolfgang Kern¹ · Torsten Haferlach¹ 

Received: 12 March 2021 / Revised: 11 May 2021 / Accepted: 24 May 2021 / Published online: 8 June 2021
© The Author(s) 2021. This article is published with open access

Abstract

Artificial intelligence (AI) is about to make itself indispensable in the health care sector. Examples of successful applications or promising approaches range from the application of pattern recognition software to pre-process and analyze digital medical images, to deep learning algorithms for subtype or disease classification, and digital twin technology and in silico clinical trials. Moreover, machine-learning techniques are used to identify patterns and anomalies in electronic health records and to perform ad-hoc evaluations of gathered data from wearable health tracking devices for deep longitudinal phenotyping. In the last years, substantial progress has been made in automated image classification, reaching even superhuman level in some instances. Despite the increasing awareness of the importance of the genetic context, the diagnosis in hematology is still mainly based on the evaluation of the phenotype. Either by the analysis of microscopic images of cells in cytomorphology or by the analysis of cell populations in bidimensional plots obtained by flow cytometry. Here, AI algorithms not only spot details that might escape the human eye, but might also identify entirely new ways of interpreting these images. With the introduction of high-throughput next-generation sequencing in molecular genetics, the amount of available information is increasing exponentially, priming the field for the application of machine learning approaches. The goal of all the approaches is to allow personalized and informed interventions, to enhance treatment success, to improve the timeliness and accuracy of diagnoses, and to minimize technically induced misclassifications. The potential of AI-based applications is virtually endless but where do we stand in hematology and how far can we go?

Introduction

Over the last 15 years, comprehensive diagnostics in leukemia and lymphoma has become increasingly challenging. In order to follow the guidelines of the World Health Organization (WHO) classification, the results from different fields, including cytomorphology, cytogenetics, immunophenotyping, and molecular genetics, have to be combined to establish a diagnosis. Gains in throughput from the introduction of next-generation sequencing (NGS) technologies and the accompanied broadening of the analytical spectrum in molecular genetics have boosted the value of molecular genetic results for diagnostics, as

indicated by the revision of the WHO classification of leukemias and lymphomas in 2017 (Ref. [1]).

The plethora of available molecular information has broadened the landscape in leukemia and lymphoma diagnostics and has led to new insights in the underlying biology of the respective diseases, provoking a shift in diagnostics from phenotype to genotype. Moreover, the identification of an increasing list of diagnostic and prognostic markers, the refined estimate of inter-individual variability, and the ongoing effort to establish correlations between different layers of information that might eventually lead to improved targeted therapy options, are paving the way for personalized medicine. In parallel, it is indisputable that the data collection process goes digital, allowing the automated integration of different test results and easy access for all involved stakeholders. This journey also offers the opportunity to share information between clinical and genomic experts from multiple institutions, facilitating the assignment of patients to specific clinical trials or targeted treatment

✉ Torsten Haferlach
torsten.haferlach@mll.com

¹ MLL Munich Leukemia Laboratory, Munich, Germany

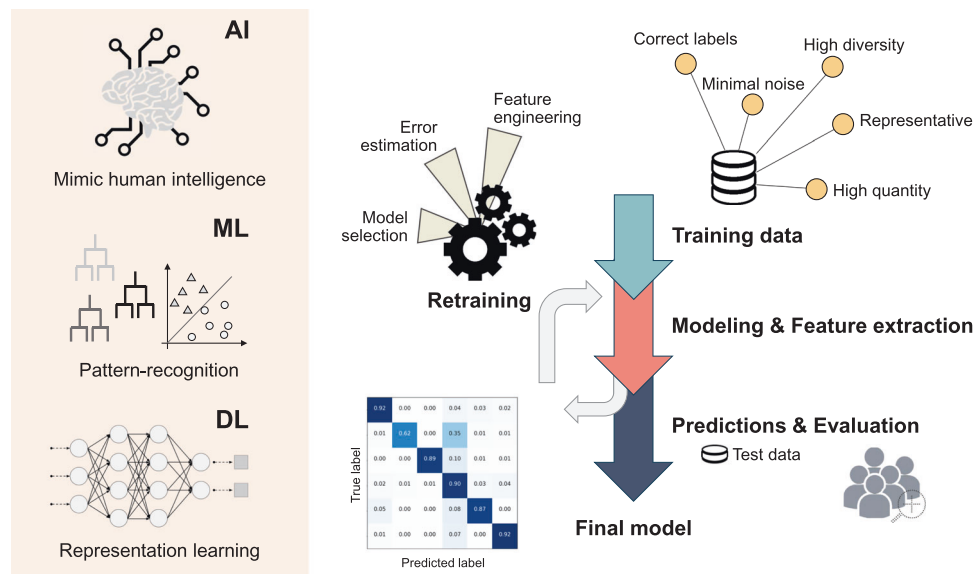


Fig. 1 Overview of the different domains and the process of supervised learning. The left side represents the different domains of supervised learning going from artificial intelligence to machine learning and finally deep learning. The right side depicts the process of supervised learning. At the top right corner, the requirements for a good training data set are listed. The data is used for automated feature

extraction, leading to the generation of a model, the performance of which is evaluated by its capability to correctly predict the labels of unseen instances (= test data). Based on the evaluation outcome, the model is retrained to refine the selected features and to optimize the model. After several rounds of retraining the final model emerges. AI artificial intelligence, DL deep learning, ML machine learning.

options [2]. Hence, the journey goes from analogous to digital and from phenotype to genotype.

Digital data is also a basic prerequisite for the application of emerging artificial intelligence (AI) techniques. Together with deep learning (DL) and machine learning (ML), AI is currently a buzzword across almost all scientific disciplines and has the potential to revolutionize diagnostic approaches in hematology. With the dramatic performance improvements in the last years, AI is at the brink to be introduced into routine diagnostics to enhance diagnostic methods but even more to facilitate disease classification and guidance of treatment. One exciting prospect is the development of digital twins to forecast cancer trajectories and to predict the potential impact of different therapeutic strategies *in silico*. The evaluation of these simulations might help to select the most promising interventions for each individual patient, minimizing side effects and the risk of complications [2, 3].

Here, we try to speculate what will happen in the next five years, how the landscape of leukemia diagnostics will be influenced by ML technologies, and how the future integration of AI-based methods will shape routine diagnostics in hematology. It is not our intention, to comprehensively review recent advances of the last five years but we like to highlight ML applications that are already being used, at least in research. Five years from now, it might be interesting to see where we were too optimistic and whether what we currently anticipate becomes a reality. We might even have AI-based methods so advanced, the complexity and capability of which, we are currently incapable of contemplating.

Quick introduction to the principles of machine learning

Due to the widespread interest and success of AI-based applications, the terms: artificial intelligence and machine learning, resound throughout the varying scientific disciplines, while often being used interchangeably in medicine. However, whereas AI strives to simulate human behavior and intelligence, ML, as a subdomain of AI, refers to the automatic detection of patterns and associations within the data (Fig. 1). DL, as a subfield of ML, allows layered neural networks to learn an abstract representation of often very complex data sets. AI and ML are not new and already in the early years the potential, risks and limits of AI have been hotly debated [4].

The remarkable improvement in technology in the biological field, especially for high-throughput methods, such as next-generation sequencing (NGS), has led to the faster generation of high quality data for a fraction of the original costs, resulting in the increased availability of digital data (=big data). Nonetheless, medical image classification, as an example for supervised learning, has benefitted the most from the introduction of ML methods to medicine so far [5, 6]. The advantage is the model's adaptability and the fact that ML performs a task automatically on experience without getting explicit instructions, on a scale exceeding the capacity of the human brain. The algorithm is trained with a massive amount of data, requiring little human intervention, except for providing the correct class label to

each image, and then left to extract relevant features and to draw its own connections, ultimately developing a set of rules and associations. The performance of the method is evaluated by its capability to predict the correct labels for a set of images not used in the training process (Fig. 1). The different types and techniques of ML in hematology have recently been reviewed [7].

The selection of an adequate training and test data set, in terms of quantity and quality, is crucial to obtain reliable results that are reproducible in real world scenarios and agnostic to the location and patient population (Fig. 1). In medicine, the training data sets are inevitably unbalanced because of the different prevalence of certain phenotypes and diseases in general but also in different subpopulations. The data sets on which the algorithms are trained, can be biased by a multitude of confounding factors that are not always obvious to the developers. Hence, it is essential to capture as much variability as possible by collecting a very diverse training set, reducing the risk of overfitting and increasing the likelihood of strong performances even for unseen instances. Ideally, the developed ML method should be transferable between hospitals and/or laboratories without loss of quality. However, due to the current lack in standardization for the operating procedures, minor adjustments are most likely unavoidable.

Current applications and advances

Cytomorphology

For more than 150 years, cytomorphology has been the backbone of hematological diagnostics, which is still true today. If an aberrant blood count is detected, cytomorphologic examinations are performed first, providing an initial diagnosis and guiding other diagnostic methods such as cytogenetics, immunophenotyping or molecular genetics to substantiate the result. However, the obtained preliminary diagnosis is solely based on the phenotype, and its correctness depends on pre-analytic procedures, as well as on the experience and capabilities of the personnel to accurately detect the aberrant cells, even very rare types. Hence, an automated pre-processing and evaluation of the digital microscopic images could benefit the reproducibility of results and would allow the hematologists and pathologists to focus on edge cases that do not fit the standard pattern, reducing the overall workload.

In the era of digital microscopic imaging and ML technologies, automated image processing, data analysis, and classification have become feasible. Initial attempts started with the segmentation of pre-processed images, followed by object identification, feature extraction, and lastly the classification of the different cell types. For the classification, different techniques have been applied

ranging from support vector machines, to random forests, and artificial neural networks (ANN). With the increasing availability of digital images and the introduction of DL, the steps are less strictly separated, relying on the algorithm to differentiate between artefacts and informative material, as well as to extract relevant and meaningful features with limited human interference.

With respect to peripheral blood cells, the spectrum of successfully applied approaches ranges from the automated counting of white and red blood cells [8], to the differentiation between myeloblasts and lymphoblasts [9], and the simultaneous classification of different lymphoid cell types [10–12], as well as 17 (ref. [13]) and 21 (ref. [14]) cell types of different lineages and maturation states, including rare and malignant leukemia cells, to name a few. The automated identification and annotation of individual cells also forms the basis for the classification of different types of AML [15], the discrimination of reactive and MPN samples [16], and the differentiation of malignant and healthy cells for ALL diagnosis [17, 18] and ALL subtype classification [19, 20]. The quality of the results of each approach depends heavily on standardized pre-analytic, analytic and post-analytic parameters.

However, while in peripheral blood smears the cell density is sufficiently low to readily identify individual cells, the interpretation of bone marrow smears is much more difficult. Therefore, a pre-screening of the microscopic image is necessary to identify areas of high quality and single cell resolution. Here it is important that the areas are selected from different parts of the image to ensure the detection of all malignant cells, even if this means compromising on quality. Due to the increased complexity of the task, it is not surprising that even DL models for assisted interpretation of bone marrow smears [21, 22] could only yield moderate results so far (Table 1). Recently, digital pathology, which encompasses the digitization of histology glass slides, has emerged as a powerful tool for cancer diagnostics in general but also for diagnostics in hematology [23], largely benefitting from the introduction of DL for whole slide image analysis [24].

Cytogenetics

Another diagnostic method in hematology that relies on the analysis and interpretation of morphological features is cytogenetics. Chromosome banding analysis has long been used in hematology and is the gold standard to identify cytogenetic abnormalities that allow the stratification of patients into disease subtypes with distinct prognosis. The patient-specific information is retrieved from the classification of chromosomes by size and banding as displayed in a karyogram. However, the generation of an accurate karyogram strongly depends on the quality of the captured

Table 1 Overview of the different diagnostic tests, the current challenges and known confunders for the clinical implementation of AI-based methods, and the requirements for a successful implementation.

	Cytomorphology	Cytogenetics	Immunophenotyping	Molecular genetics
Method	Microscopy	Chromosome banding analysis	Multiparameter flow cytometry	Genomic analysis
Aim	Identification and characterization of cell populations based on morphology	Identification of cytogenetic abnormalities	Identification and characterization of cell populations based on light-scattering properties and antigen expression patterns	Identification of individual molecular profiles
Challenges	Differentiation between artefacts and informative material (=cells) Correct identification of borders (very dense regions with overlapping cells)	Identification and selection of individual chromosomes Correct identification of structural abnormalities	Accurate representation and transformation of the raw data	Data matrix usually sparse and informative signals might be lost in noise Meaningful combination of various data types and differentiation between absence of information (e.g., insufficient coverage) and true negative results Lack of knowledge for annotation and interpretation of variants in coding and especially non-coding regions
Confounders	Extraction of features that allow the differentiation of maturation states of the same cell type Resolution, image capturing, image cropping are not standardized between laboratories	Different banding and staining methods	Different methods for data pre-processing and data/image transformation	Plethora of methods for the identification of features (CNV, SV, SNV, Fusions, etc), data transformation, and dimensionality reduction with limited concordance and individual biases
Requirements for final implementation	Unbalanced data sets for training with rare cell types being underrepresented Time and cost efficient digitization of glass slides Standardized, automated systems for the recording of digital microscopic images Balanced training data capturing as much biological and technical variety as possible	Unbalanced data sets for training with rare structural abnormalities being underrepresented Time and cost efficient digitization of glass slides	Harmonization and standardization of used antibodies	Harmonization of gene panels Standardization of analysis pipelines and variant interpretation

metaphases and chromosomes, for which viable cells have to be grown and arrested in the metaphase stage of cell division. Subsequent banding and staining of the chromosomes is essential to highlight the details of diagnostic importance and to identify normal and abnormal chromosomes. However, karyotyping is a very time-consuming and complex task and a high degree of automation is desirable. For more than 30 years, various analysis systems for automated metaphase capturing and semi-automatic and/or interactive karyotyping have been available. These systems have been increasingly useful in classical human genetics, especially in prenatal diagnostics.

The challenges for automated karyotyping are multifarious and not yet fully resolved. First, individual chromosomes have to be identified and selected, excluding artefacts and overlapping or touching chromosomes from the downstream analysis. Already in 2007, an automated workflow was proposed [25] but most of the procedures still require some manual curation to avoid extensive discard and the artificial creation of abnormal chromosomes due to unfortunate cutting of overlapping chromosomes [26] (Table 1). It follows the labeling of the separated chromosomes and their assignment to the respective position in the karyogram. For an automated procedure, the identification of an optimal and small chromosome feature set is key for accurate performance and robustness. Common features considered for labeling include shape and size of the chromosome, the centromere location, and the unique banding pattern profile. Especially the banding pattern has been researched intensively to efficiently compute the profiles as a prerequisite for chromosome classification.

Over the last 20 years, various chromosome classifiers have been developed, ranging from ANN [27, 28], to competitive neural network teams [26], wavelet transform based linear discriminant analysis [29], and different versions of (deep) convoluted neural networks [30–34], achieving an accuracy from 85.2% to 98.6%. The classification often improved by correct alignment and orientation of the chromosomes along the vertical axis as a pre-processing step. Misclassification usually involved chromosomes, very similar in size, shape, and appearance and, hence, are challenging to differentiate, even for human professionals.

In tumor cytogenetics, chromosome anomalies, including numerical and structural abnormalities, are quite common and pose a further challenge for automated approaches. Numerical abnormalities usually involve normal chromosomes and, hence, most methods can be extended quite easily for this task. Structural abnormalities on the other hand are more challenging, due to the huge variety of possibilities and the sometimes limited available training material. However, promising early results show that also structurally abnormal karyograms might be detectable in an automated fashion in the future [34].

Immunophenotyping

Besides cytomorphology, multiparameter flow cytometry (MFC, immunophenotyping) is the central method for the diagnosis of leukemias and lymphomas. MFC uses fluorescence dye-conjugated monoclonal antibodies, targeting diagnostically relevant antigens, to analyze cell populations based on their light-scattering properties and antigen expression patterns. Specific software automatically measures and captures the expression of the respective fluorescent dyes. Subsequently, human experts apply a sequential gating procedure to large sets of bidimensional plots to identify and label cell populations of interest. Although this method seems much less subjective than cytomorphology or histology, all the steps are error prone, influencing or even biasing the later interpretation. While standardized procedures are in place to control quality of sample preparation and measurement, the interpretation still relies on expert knowledge with inherited inter-observer variability. Thus, to reduce the dependency on expert knowledge and to increase reproducibility of data interpretation the implementation of automated procedures is desirable (Table 1).

An attempt in this field was conducted by Zhao et al. [35], who used self-organizing maps of light emission profiles as an input for a deep convolutional neural network to differentiate between healthy and neoplastic samples, as well as classification of mature B-neoplasm subtypes. Different clustering and ML techniques have been applied for joint cell clustering and identification of anomalous sample phenotypes for various hematologic malignancies [36], including AML samples [37, 38] and lymphomas [39–41]. Angeletti [42] applied a genetic algorithm to differentiate between AML and control samples and Bigorra et al. [43] could show that neural network approaches yield the highest accuracy for the differentiation between healthy controls, virus-infected samples and CLL patients. Only few approaches [44, 45] have attempted to use flow cytometry data for the classification without preceding image transformation. While Biehl et al. [44] used generalized matrix relevance learning vector quantification to separate AML patients from healthy controls, Müller et al. [45] applied a XGBoost model to assign lymphoma samples to their respective subtype.

Moreover, AI-based methods have been applied to accelerate, harmonize, and standardize interpretation of minimal residual disease (MRD) measurements obtained by flow cytometry, predicting the outcome of AML and MDS patients with high accuracy [46, 47].

Molecular genetics

While the other fields integrate AI technologies to mimic human intelligence and to reproduce the knowledge of experienced diagnosticians, clinical molecular genetics aims

to implement ML-based methods to perform tasks that are impractical for humans to do. With the introduction of high-throughput sequencing techniques and the accompanied analysis of large gene panels or even whole genomes and transcriptomes, molecular genetics has entered the realm of big data (Table 1). A single human genome contains $2 \times \sim 3.2$ billion nucleotides worth of information and mining the data to obtain clinically relevant insights soon becomes cumbersome.

Basically every step in clinical genomic analysis could benefit from the integration of ML and DL methods [48, 49], including, but not limited to, the recognition of DNA sequence patterns [50, 51], variant calling [52], and variant effect prediction [53, 54] and classification [55]. Especially variant interpretation becomes more and more important with the increase in analyzed genes. Prioritization of causal and clinically actionable genetic variants is required for clinical decision-making and forms the basis for automated disease classification.

For some entities, the current WHO classification only mentions molecular genetic markers as a footnote so far but there is increasing recognition of the diagnostic relevance of the broader genetic context. Sequencing of a predefined set of gene regions associated with the suspected disease is currently favored, as opposed to whole exome sequencing, owing to decreased costs, to reduce turn-around times, and to limit an overburden of information that is not clinically actionable. The obtained mutational profile can be enriched by integration with phenotypic changes and clinical data as recently done by Nagata et al. who applied Bayesian ML techniques to identify diagnostically and prognostically relevant associations between genetic variants and cytomorphological changes in myelodysplastic syndromes (MDS). ML algorithms have also been applied to integrate mutational data, peripheral blood values, and clinical data into a genoclinical model to differentiate various bone marrow disorders [56, 57]. However, the long-term goal might be to solely rely on molecular genetic data for disease and subgroup classification as previously done by different groups [58, 59].

Applying ML-based methods to molecular genetic data is not only relevant for clinical diagnosis but also for prognostication and the prediction of drug-responses. Wagner et al. [61] combined the molecular results of different databases in an ANN to identify a prognostic 3-gene signature that separated AML patients of European LeukemiaNet (ELN) strata in subgroups with different survival probabilities. A supervised ML approach applied to an *NPM1*^{mut} AML cohort identified clinically important mutations which were combined to a genetic score to pinpoint patients who are at high risk of relapse [62]. Supervised machine learning identified features that reliably assigned AML patients with *RUNX1-RUNX1T1* to favorable and poor risk classes [63]. In an MDS cohort an a priori

market basket analysis algorithm was used to identify molecular signatures strongly associated with response to hypomethylating agents [64].

Although, the clinical application of transcriptome analysis is usually limited to the quantification of expression of a handful of genes by qPCR, different studies have also demonstrated the combined strength of larger gene panels and DL-based approaches for patient classification [60], biomarker detection [65], and predicting clinical response to anti-cancer drugs [66]. To overcome the difficulties presented by bulk gene expression data, DL has been applied to estimate cell type compositions from tissue expression profiles [67].

Current challenges for the clinical implementation of AI-based methods

As outlined in the previous chapters, hematology could benefit substantially from the application of AI-based approaches to reduce the workload, to merge the knowledge of different experts, to decrease turn-around times, and to minimize technically induced misclassifications. The accurate interpretation of results depends a lot on a person's level of experience and reliable AI-based clinical decision-support systems could install confidence in unexperienced clinicians, especially for morphology-based diagnostics. Different ML-based applications address these points, with the automated classification of cells and chromosomes from digital microscopic images nearly reaching expert clinical accuracy.

However, the performance of the algorithms largely depends on the availability of extensive standardized digital data to train the algorithms. The degree of workflow automation and hence, the homogeneity and reproducibility of the obtained data, varies between the different fields, potentially biasing the results (Table 1). Moreover, various parameters, such as the selected staining technique, the used antibodies or the selected genes and hotspots of a testing panel, often differ between different institutions, laboratories, and hospitals, impeding a generalization of the developed methods (Table 1). Hence, interlaboratory validation studies are necessary for a successful application and transferability of ML models and comprehensive guidelines have to be established to ensure standardized method outputs that can subsequently be fed into the ML models.

For some instances, the available training data is also limited in its complexity, capturing only part of the biological variability (Table 1). In cytomorphology, for example, it is challenging to obtain sufficient training material of rare cells, restricting the algorithm's capability to extract meaningful features and to reliably identify the phenotype. The same applies to immunophenotyping and to rare chromosomal abnormalities in tumor cytogenetics. In molecular genetics,

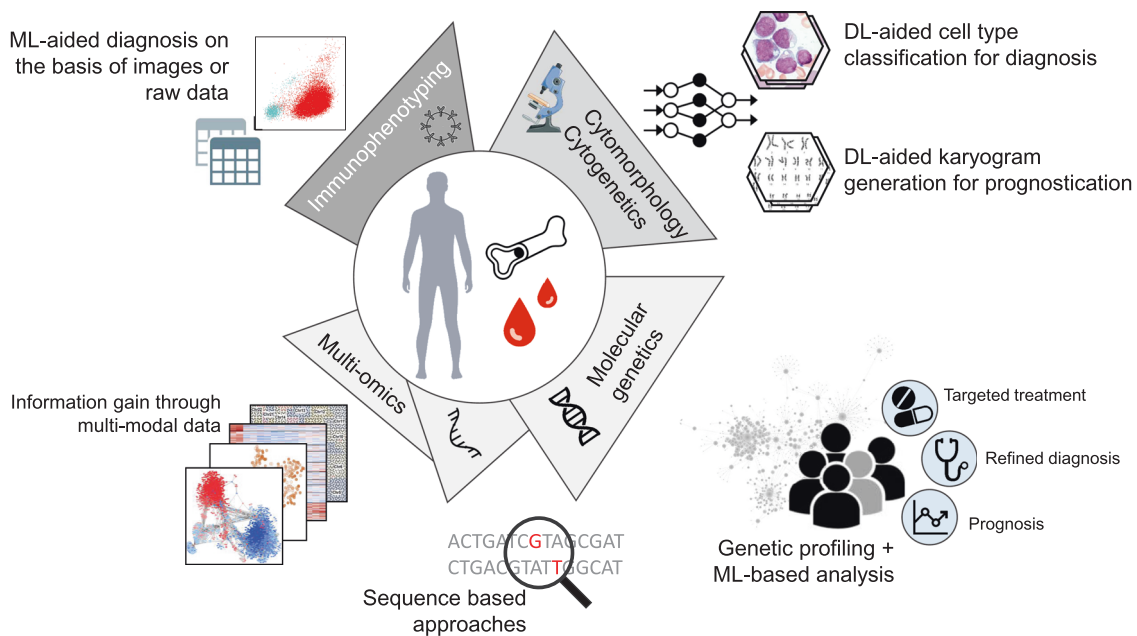


Fig. 2 Overview of implemented and potential ML applications in hematology. The central part of the figure displays current and future diagnostic tests and methods, while the outer part illustrates the

various data types and the potential clinical impact of ML-based applications and analyses. DL deep learning, ML machine learning.

analyzed gene panels are often inconsistent among patients, depending on the suspected diagnosis and associated genes. Due to the sometimes-limited training material, there is a high risk of overfitting and the accidental fitting of confounders, resulting in significantly worse performances for unseen data. In addition, it is necessary to pay attention to avoid unintended discriminatory bias and failed generalization to new populations.

The benefit of AI technologies is the potential to improve the performance by constant learning, raising at the same time the unprecedented question about how to regulate such a machine. Initial guidelines have already been created around the globe [68]. However, one as yet unanswered question is the question of responsibility if an algorithm gets a diagnosis wrong. Most likely, there will not be an all-encompassing answer but each occurrence will have to be evaluated independently, taking into account the exact circumstances. Due to the initial poor results of AI-based approaches, many potential users are sceptical and successful, robust implementations, validated by clinical studies, will be necessary to eliminate the doubts. Furthermore, there is the ‘black box’ or ‘explainability’ problem of DL methods, allowing humans to comprehend the decisions of the algorithms only to a limited extent, undermining the scientific value of the method. The shortcomings of current ML-based methods in healthcare and potential solutions have been widely discussed [69–71]. Due to recent efforts, various methods exist to improve the transparency of an applied DL model [72], creating the neologism of ‘explainable artificial intelligence’ and paving the way for the clinical use.

Opportunities and outlook

As indicated before, the best opportunities for a fast implementation of ML-based technologies into routine diagnostic workflows offer cytomorphology and cytogenetics (Fig. 2). With the recent integration of DL methods, the accuracies of the different methods are close to expert level, offering the possibility of faster and more accurate sample processing, bringing expertise to the fingertips of less experienced hematologists. Although the current AI-based applications in flow cytometry are less extensive and have not reached clinically acceptable accuracies in all domains, an automated workflow could potentially lead to more standardized and reproducible results (Fig. 2).

In contrast to the other fields, the opportunities and possibilities in molecular genetics do not solely rely on the optimization of AI-based methods or an increase in standardized training data sets but also depend on the creation of comprehensive data collections and large-scale data mining efforts to improve the understanding of individual sequence variants and complex molecular constellations (Fig. 2). Comprehensive assays such as whole genome sequencing (WGS) and whole transcriptome sequencing (WTS) have great and so far mostly unexplored potential to revolutionize diagnostics of hematologic malignancies. Due to the high dimensionality and multi-modality, genomic data sets are the perfect candidate for DL explorations and can be used to refine diagnosis, further develop classification systems, identify prognostic factors, and to provide

targets for individualized therapy (Fig. 2). Especially the identification of significant interactions and regulatory mechanisms are of tremendous interest. Dynamic mechanistic models can be used to analyze signaling regulations and even predict the effects of targeted therapies for personalized treatments [73]. Here, AI-based methods can also help to filter the clinically most relevant pathways [74]. Moreover, integrating molecular profiles and networks with deep, longitudinal physiological data and AI methods opens up the possibility for early detection of disease transition, prediction of clinical outcome, and the design of personalized treatment strategies [75, 76]. Another proposed solution to the challenge of efficiently linking an individual's molecular profile to a compatible drug treatment is based on constructing Digital Twins [3]. The idea behind a digital twin in medicine is to build a computational model of a patient, which can be modified *in silico*, testing different treatments more quickly, economically, and safely than is possible in real life. The concept could be extended to *in silico* clinical trials, testing new drugs for a fraction of the current cost without putting patients at risk. The accuracy of such models could further be improved by adding multi-omics data (Fig. 2). Different ML-based methods have been developed for multi-omics analysis to conquer challenges such as data integration, biomarker discovery, and the identification of therapeutic compounds [77, 78]. The integration of multi-omics data opens up the opportunity for a more comprehensive diagnosis with a tailored treatment based on an individual's genetic make-up.

Conclusions

AI-based technologies continue to transform our everyday life and increasingly also different sectors of health care, including hematology. On one hand, the implementation of ML methods will aid clinicians in their analysis and interpretation of the data, increase objectivity and accuracy in the work-up, while on the other, the collected and the integrated knowledge will aid less experienced doctors in guiding their decision-making process. Here, regulated integration of new applications is important to ensure that patients are neither exposed to flawed interventions with potentially harmful effects nor denied access to beneficial innovations. Furthermore, in the future, molecular data, in combination with AI-based methods, might be superior to the phenotype-based standard methods in hematology, potentially defining new therapeutic approaches, yielding informed treatment decisions. Over the next five years, we will see an increasing need for AI approaches to integrate multi-modal patient data and treatment options as it becomes impossible for human intelligence to capture all this. However, in any field the prerequisite for the integration of ML-based methods is a

reproducible and accredited data generation workflow. Only standardized outputs can be used for automated processing and provide results that, in combination with clinical judgment, can be used for diagnostics, for prognosis, and for therapy management in the context of precision medicine.

Despite the recent success of AI-based technologies in medicine, no system is infallible and results always have to be examined critically. It is not the intention to replace doctors with AI technologies but it would be negligent to ignore their potential. We need the support of AI to fulfill our primary goal: to provide patients with the best possible care. It is our duty to use all the means at our disposal to achieve this.

Compliance with ethical standards

Conflict of interest TH, CH and WK declare part ownership of Munich Leukemia Laboratory (MLL). WW, NN, IS and CK are employed by the MLL.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Swerdlow SH, Campo E, Harris NL, Jaffe ES, Pileri SA, Stein H et al. editors. WHO classification of tumours of haematopoietic and lymphoid tissues. Revised 4th edn. Lyon: International Agency for Research on Cancer; 2017.
2. Madhavan S, Beckman RA, McCoy MD, Pishvaian MJ, Brody JR, Macklin P. Envisioning the future of precision oncology trials. *Nat Cancer*. 2021;2:9–11.
3. Björnsson B, Borrebaeck C, Elander N, Gasslander T, Gawel DR, Gustafsson M, et al. Digital twins to personalize medicine. *Genome Med*. 2019;12:4.
4. Fjelland R. Why general artificial intelligence will not be realized. *Humanit Soc Sci Commun*. 2020;7:10.
5. Wu JT, Wong KCL, Gur Y, Ansari N, Karargyris A, Sharma A, et al. Comparison of chest radiograph interpretations by artificial intelligence algorithm vs radiology residents. *JAMA Netw Open*. 2020;3:e2022779.
6. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017;542:115–8.
7. Shouval R, Fein JA, Savani B, Mohty M, Nagler A. Machine learning and artificial intelligence in haematology. *Br J Haematol*. 2021;192:239–50.

8. Alomari YM, Sheikh Abdullah SNH, Zaharatul Azma R, Omar K. Automatic detection and quantification of WBCs and RBCs using iterative structured circle detection algorithm. *Comput Math Methods Med*. 2014;2014:979302.
9. Bigorra L, Merino A, Alférez S, Rodellar J. Feature analysis and automatic identification of leukemic lineage blast cells and reactive lymphoid cells from peripheral blood cell images. *J Clin Lab Anal*. 2017;31:e22024.
10. Alférez S, Merino A, Mujica LE, Ruiz M, Bigorra L, Rodellar J. Automatic classification of atypical lymphoid B cells using digital blood image processing. *Int J Lab Hematol*. 2014;36:472–80.
11. Alférez S, Merino A, Bigorra L, Mujica L, Ruiz M, Rodellar J. Automatic recognition of atypical lymphoid cells from peripheral blood by digital image analysis. *Am J Clin Pathol*. 2015;143:168–76. quiz 305
12. Alférez 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.
13. Kimura K, Tabe Y, Ai T, Takehara I, Fukuda H, Takahashi H, et al. A novel automated image analysis system using deep convolutional neural networks can assist to differentiate MDS and AA. *Sci Rep*. 2019;9:13385.
14. 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.
15. Boldú L, Merino A, Alférez S, Molina A, Acevedo A, Rodellar J. Automatic recognition of different types of acute leukaemia in peripheral blood by image analysis. *J Clin Pathol*. 2019;72:755–61.
16. 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.
17. Moshavash Z, Danyali H, Helfroush MS. An automatic and robust decision support system for accurate acute leukemia diagnosis from blood microscopic images. *J Digit Imaging*. 2018;31:702–17.
18. Chin Neoh S, Srisukkharn W, Zhang L, Todryk S, Greystoke B, Peng Lim C, et al. An intelligent decision support system for leukaemia diagnosis using microscopic blood images. *Sci Rep*. 2015;5:14938.
19. Pansombut T, Wikaisuksakul S, Khongkrapan K, Phon-On A. Convolutional neural networks for recognition of lymphoblast cell images. *Comput Intell Neurosci*. 2019;2019:7519603.
20. 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.
21. Wu Y-Y, Huang T-C, Ye R-H, Fang W-H, Lai S-W, Chang P-Y, et al. A hematologist-level deep learning algorithm (BMSNet) for assessing the morphologies of single nuclear balls in bone marrow smears: algorithm development. *JMIR Med Inform*. 2020;8:e15963.
22. Fu X, Fu M, Li Q, Peng X, Lu J, Fang F, et al. Morphogor: an automatic bone marrow cell classification system on digital images analyzed by artificial intelligence. *Acta Cytol*. 2020;64:588–96.
23. El Achi H, Khoury JD. Artificial intelligence and digital microscopy applications in diagnostic hematopathology. *Cancers*. 2020;12:797.
24. Dimitriou N, Arandjelović O, Caie PD. Deep learning for whole slide image analysis: an overview. *Front Med*. 2019;6:264.
25. 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.
26. Gagula-Palalic S, Can M. Human chromosome classification using Competitive Neural Network Teams (CNNT) and Nearest Neighbor. In: *IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*. IEEE. 2014; 626–9.
27. Delshadpour S. Reduced size multi layer perceptron neural network for human chromosome classification. In: *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE Cat. No.03CH37439)*. IEEE. 2004; 3: 2249–52.
28. Oskouei BC, Shanbehzadeh J. Chromosome classification based on wavelet neural network. In: *2010 International Conference on Digital Image Computing: Techniques and Applications*. IEEE. 2010; 605–10.
29. Roshkhar MJ, Setarehdan SK. Linear Discriminant Analysis of the wavelet domain features for automatic classification of human chromosomes. In: *2008 9th International Conference on Signal Processing*. IEEE. 2008; 849–52.
30. Qiu Y, Lu X, Yan S, Tan M, Cheng S, Li S et al. Applying deep learning technology to automatically identify metaphase chromosomes using scanning microscopic images: an initial investigation. In: *Chen WR (ed). Biophotonics and Immune Responses XI*. SPIE, 2016.
31. Sharma M, Saha O, Sriraman A, Hebbalaguppe R, Vig L, Karande S. Crowdsourcing for chromosome segmentation and deep classification. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE. 2017; 786–93.
32. Swati, Gupta G, Yadav M, Sharma M, Vig L. Siamese networks for chromosome classification. In: *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*. IEEE. 2017; 72–81.
33. Qin Y, Wen J, Zheng H, Huang X, Yang J, Song N, et al. Varifocal-net: a chromosome classification approach using deep convolutional networks. *IEEE Trans Med Imaging*. 2019;38:2569–81.
34. 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.
35. 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. *Cytom A*. 2020;97:1073–80.
36. 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.
37. Dundar M, Akova F, Yerebakan HZ, Rajwa B. A non-parametric Bayesian model for joint cell clustering and cluster matching: identification of anomalous sample phenotypes with random effects. *BMC Bioinforma*. 2014;15:314.
38. Manninen T, Huttunen H, Ruusuvoori P, Nykter M. Leukemia prediction using sparse logistic regression. *PLoS ONE*. 2013;8:e72932
39. Lakoumentas J, Drakos J, Karakantza M, Nikiforidis GC, Sakellaropoulos GC. Bayesian clustering of flow cytometry data for the diagnosis of B-chronic lymphocytic leukemia. *J Biomed Inform*. 2009;42:251–61.
40. Zare H, Bashashati A, Kridel R, Aghaeepour N, Haffari G, Connors JM, et al. Automated analysis of multidimensional flow cytometry data improves diagnostic accuracy between mantle cell lymphoma and small lymphocytic lymphoma. *Am J Clin Pathol*. 2012;137:75–85.
41. Pouyan MB, Jindal V, Birjandtalab J, Nourani M. Single and multi-subject clustering of flow cytometry data for cell-type identification and anomaly detection. *BMC Med Genomics*. 2016;9:41.
42. Angeletti C. A method for the interpretation of flow cytometry data using genetic algorithms. *J Pathol Inf*. 2018;9:16.
43. Bigorra L, Larriba I, Gutiérrez-Gallego R. Machine learning algorithms for accurate differential diagnosis of lymphocytosis based on cell population data. *Br J Haematol*. 2019;184:1035–7.

44. Biehl M, Bunte K, Schneider P. Analysis of flow cytometry data by matrix relevance learning vector quantization. *PLoS ONE*. 2013;8:e59401
45. Müller M-L, Nadarajah N, Jhalani K, Heo I, Wetton W, Haferlach C, et al. Employment of machine learning models yields highly accurate hematological disease prediction from raw flow cytometry matrix data without the need for visualization or human intervention. *Blood*. 2020;136:11.
46. Ko B-S, Wang Y-F, Li J-L, Li C-C, Weng P-F, Hsu S-C, et al. Clinically validated machine learning algorithm for detecting residual diseases with multicolor flow cytometry analysis in acute myeloid leukemia and myelodysplastic syndrome. *EBioMedicine*. 2018;37:91–100.
47. Licandro R, Reiter M, Diem M, Dworzak M, Schumich A, Kämpel M. Application of machine learning for automatic MRD assessment in paediatric acute myeloid leukaemia. In: *Proceedings of the 7th International Conference on Pattern Recognition Applications and Methods*. SCITEPRESS - Science and Technology Publications. 2018.
48. Dias R, Torkamani A. Artificial intelligence in clinical and genomic diagnostics. *Genome Med*. 2019;11:70.
49. Zou J, Huss M, Abid A, Mohammadi P, Torkamani A, Telenti A. A primer on deep learning in genomics. *Nat Genet*. 2019;51:12–8.
50. Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF, Darbandi SF, Knowles D, Li YI, et al. Predicting splicing from primary sequence with deep learning. *Cell*. 2019;176:535–48. e24
51. Albaradei S, Magana-Mora A, Thafar M, Uludag M, Bajic VB, Gojobori T, et al. Splice2Deep: an ensemble of deep convolutional neural networks for improved splice site prediction in genomic DNA. *Gene X*. 2020;5:100035.
52. Poplin R, Chang P-C, 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.
53. Hoffman GE, Bendl J, Girdhar K, Schadt EE, Roussos P. Functional interpretation of genetic variants using deep learning predicts impact on chromatin accessibility and histone modification. *Nucleic Acids Res*. 2019;47:10597–611.
54. Lai C, Zimmer AD, O'Connor R, Kim S, Chan R, van den Akker J, et al. LEAP: Using machine learning to support variant classification in a clinical setting. *Hum Mutat*. 2020;41:1079–90.
55. Quang D, Chen Y, Xie X. DANN: a deep learning approach for annotating the pathogenicity of genetic variants. *Bioinformatics*. 2015;31:761–3.
56. Hilton CB, Meggendorfer M, Sekeres MA, Shreve J, Radakovich N, Roupail Y, et al. Geno-clinical model for the diagnosis of bone marrow myeloid neoplasms. *Blood*. 2019;134:4238.
57. Radakovich N, Meggendorfer M, Malcovati L, Sekeres MA, Shreve J, Hilton CB, et al. A personalized clinical-decision tool to improve the diagnostic accuracy of myelodysplastic syndromes. *Blood*. 2020;136:33–5.
58. Parida L, Haferlach C, Rhrissorakrai K, Utro F, Levovitz C, Kern W, et al. Dark-matter matters: discriminating subtle blood cancers using the darkest DNA. *PLoS Comput Biol*. 2019;15:e1007332.
59. Meggendorfer M, Walter W, Haferlach C, Kern W, Haferlach T. Challenging blast counts by machine learning techniques and genome sequencing for discriminating AML and MDS. *Blood*. 2019;134:4663.
60. Mallick PK, Mohapatra SK, Chae G-S, Mohanty MN. Convergent learning-based model for leukemia classification from gene expression. *Pers Ubiquitous Comput*. 2020. <https://doi.org/10.1007/s00779-020-01467-3>. (e-pub ahead of print 16 October 2020).
61. Wagner S, Vadakekolathu J, Tasian SK, Altmann H, Bornhäuser M, Pockley AG, et al. A parsimonious 3-gene signature predicts clinical outcomes in an acute myeloid leukemia multicohort study. *Blood Adv*. 2019;3:1330–46.
62. 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.
63. Shaikh AF, Kakirde C, Dhamne C, Bhanshe P, Joshi S, Chaudhary S, et al. Machine learning derived genomics driven prognostication for acute myeloid leukemia with RUNX1-RUNX1T1. *Leuk Lymphoma*. 2020;61:3154–60.
64. Nazha A, Sekeres MA, Bejar R, Rauh MJ, Othus M, Komrokji RS, et al. Genomic biomarkers to predict resistance to hypomethylating agents in patients with myelodysplastic syndromes using artificial intelligence. *JCO Precis Oncol*. 2019;3:1–11.
65. Castillo D, Galvez JM, Herrera LJ, Rojas F, Valenzuela O, Caba O, et al. Leukemia multiclass assessment and classification from Microarray and RNA-seq technologies integration at gene expression level. *PLoS ONE*. 2019;14:e0212127
66. Sakellariopoulos T, Vougas K, Narang S, Koinis F, Kotsinas A, Polyzos A, et al. A deep learning framework for predicting response to therapy in cancer. *Cell Rep*. 2019;29:3367–3373. e4
67. Menden K, Marouf M, Oller S, Dalmia A, Magruder DS, Kloiber K, et al. Deep learning-based cell composition analysis from tissue expression profiles. *Sci Adv*. 2020;6:eaba2619.
68. Jobin A, Ienca M, Vayena E. The global landscape of AI ethics guidelines. *Nat Mach Intell*. 2019;1:389–99.
69. Holzinger A, Langs G, Denk H, Zatloukal K, Müller H. Causability and explainability of artificial intelligence in medicine. *Wiley Interdiscip Rev Data Min Knowl Disco*. 2019;9:e1312.
70. Amann J, Blasimme A, Vayena E, Frey D, Madai VI. Precise4Q consortium. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. *BMC Med Inf Decis Mak*. 2020;20:310.
71. Cuttillo CM, Sharma KR, Foschini L, Kundu S, Mackintosh M, Mandl KD, et al. Machine intelligence in healthcare-perspectives on trustworthiness, explainability, usability, and transparency. *NPJ Digit Med*. 2020;3:47.
72. Samek W, Wiegand T, Müller K-R. Explainable artificial intelligence: understanding, visualizing and interpreting deep learning models. *arXiv*. 2017. <http://arxiv.org/abs/1708.08296>.
73. Saez-Rodriguez J, Blüthgen N. Personalized signaling models for personalized treatments. *Mol Syst Biol*. 2020;16:e9042.
74. Ai C, Kong L. CGPS: a machine learning-based approach integrating multiple gene set analysis tools for better prioritization of biologically relevant pathways. *J Genet Genom*. 2018;45:489–504.
75. Yurkovich JT, Tian Q, Price ND, Hood L. A systems approach to clinical oncology uses deep phenotyping to deliver personalized care. *Nat Rev Clin Oncol*. 2020;17:183–94.
76. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med*. 2019;25:44–56.
77. Nicora G, Vitali F, Dagliati A, Geifman N, Bellazzi R. Integrated multi-omics analyses in oncology: a review of machine learning methods and tools. *Front Oncol*. 2020;10:1030.
78. Patel-Murray NL, Adam M, Huynh N, Wassie BT, Milani P, Fraenkel E. A multi-omics interpretable machine learning model reveals modes of action of small molecules. *Sci Rep*. 2020;10:954.