



Artificial Intelligence in Hematology: Current Challenges and Opportunities

Nathan Radakovich¹ · Matthew Nagy¹ · Aziz Nazha^{2,3}

Published online: 1 April 2020

© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

Purpose of Review Artificial intelligence (AI), and in particular its subcategory machine learning, is finding an increasing number of applications in medicine, driven in large part by an abundance of data and powerful, accessible tools that have made AI accessible to a larger circle of investigators.

Recent Findings AI has been employed in the analysis of hematopathological, radiographic, laboratory, genomic, pharmacological, and chemical data to better inform diagnosis, prognosis, treatment planning, and foundational knowledge related to benign and malignant hematology. As more widespread implementation of clinical AI nears, attention has also turned to the effects this will have on other areas in medicine.

Summary AI offers many promising tools to clinicians broadly, and specifically in the practice of hematology. Ongoing research into its various applications will likely result in an increasing utilization of AI by a broader swath of clinicians.

Keywords Machine learning · Artificial intelligence · Hematology · Deep learning

Introduction

Artificial intelligence (AI)—the use of computer algorithms to perform tasks or solve problems that were traditionally associated with the human capacity for flexible thinking and adaptivity—has undergone explosive progress in the last decade thanks to improvements in computing power and access to ever-increasing repositories of data [1–3]. AI now plays a role in several commonplace technologies, including voice recognition or targeted advertising, and is poised to soon

become involved in others, such as autonomously driving vehicles [1].

Given the portion of the gross domestic product that healthcare represents in developed countries, and given the enormous quantities of data generated both from clinical medicine and biomedical research, it is unsurprising that healthcare has been a hotbed for recent AI research. Indeed, as of 2019 in radiology, the medical field where AI has arguably made the most progress to date, 90% of practicing radiologists anticipate that AI will be incorporated into their future practice.

Simultaneously, as the volume of medical data available grows at an exponential pace, investigators and clinicians are faced with the “data rich, information poor” problem, wherein the quantity—and complexity—of the data generated exceed our ability to make use of it; other commentators have noted that merely possessing increasing amounts of data is no guarantee that more knowledge, or better patient care, will follow [3]. While still nascent, AI holds similar promise for medicine broadly, and for the practice of hematology. This review provides a description of relevant concepts in AI for those unfamiliar with it, applications where AI shows utility in hematology, and future challenges pertaining to AI’s integration into clinical practice.

This article is part of the Topical Collection on *Social Media Impact of Hematologic Malignancies*

✉ Aziz Nazha
nazhaa@ccf.org

¹ Cleveland Clinic Lerner College of Medicine of Case Western Reserve University, Cleveland, OH, USA

² Center for Clinical Artificial Intelligence, Cleveland Clinic, Cleveland, OH, USA

³ Department of Hematology and Medical Oncology, Taussig Cancer Institute, Cleveland Clinic, Desk R35 9500 Euclid Ave., Cleveland, OH 44195, USA

Machine Learning: Basic Concepts

The vast majority of healthcare-oriented AI falls under the subheading of machine learning (ML) (Fig. 1). Broadly, ML refers to any algorithm that, rather than being programmed directly to complete some objective, is instead designed to take source data and find the most effective way to use those data to complete that objective [4]. Among ML tasks, most applications are considered “supervised learning,” meaning that a specific outcome of interest (called a “label”) is known and the algorithm in question is designed to find the best way to predict that outcome [5]. This typically consists of either classification or regression problems. “Unsupervised” applications, where a prespecified outcome is not known, also exist and are used to identify previously unknown structures within data, for example by clustering different samples according to similar genomic data [4, 5].

Myriad algorithms exist under the umbrella of machine learning. Of these, neural network methods are some of the most commonly used in recent years. As suggested by their name, neural networks draw inspiration from the structure of the nervous system; the basic units of communication in a neural network, called nodes or neurons, are arrayed in sequential layers, with connections of varying strength between each layer [1, 6]. The initial layer of a neural network receives input data such as an array of pixels, a series of words, or an array of categorical data. These data are relayed through intermediate, “hidden” layers, which can create representations of relationships or higher-level concepts within data, for example progressing from a set of individual genes to gene networks, or from an array of pixels to a collection of objects like faces and bodies [1, 6]. Hidden layers ultimately feed into an output layer, which can perform standard classification/regression tasks or create more sophisticated outputs, such as novel images, sound, or text. “Deep neural networks” or “deep learning” refers to any neural network with more than one hidden

layer; most neural networks employed in healthcare are deep networks [2, 6].

Neural networks can be built in a variety of configurations that allow them to excel at various functions. Convolutional neural networks (CNNs) are highly successful models designed primarily for handling visual data. CNNs work by filtering small pieces of an image (typically three to seven pixels) in order to detect simple elements of an image such as edges, curves, or changes in color; the output of these filters is passed through subsequent layers of the network, which in turn encode increasingly high-level features such as shapes, textures, and eventually entire objects [1, 7].

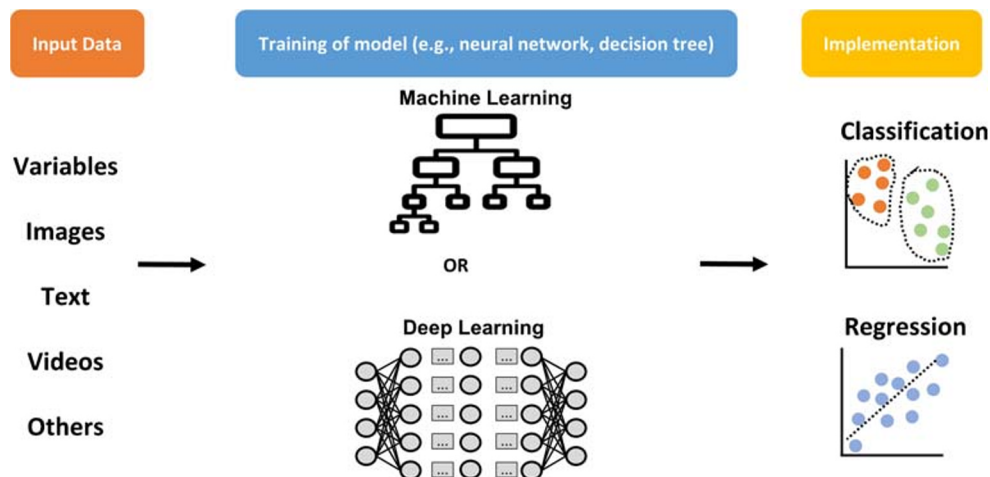
CNNs’ architecture gives them remarkable flexibility, and since their inception, they have significantly outperformed previous approaches [8]. CNNs carry the additional benefit of requiring less domain-specific knowledge to construct than other methods, and because their initial layers encode basic shapes rather than specific objects, highly accurate existing networks can be repurposed for other uses in a process called “transfer learning,” which is a crucial ability when, as in many healthcare settings, datasets number in the hundreds of samples rather than the millions used to develop the original network [6]. Some CNN-based models have even achieved human-level accuracy in diverse image processing tasks such as screening for diabetic retinopathy, correctly classifying histopathologic data, and interpreting screening mammograms, highlighting both the effectiveness of CNNs and their applicability in diverse applications [9–12].

Image Analysis

Pathology

As with other medical applications, AI-based image processing in hematology has leapt forward with the advent of CNNs.

Fig. 1 Overview of machine learning implementation. Source data of varying types are collected for a given dataset, then fed into a machine learning model (decision tree in this diagram, on the upper part) or deep neural network (lower part) which is trained to make prediction of two types of problems. Classification problem (defining a class or multi-class prediction) and a regression problem (continuous number prediction). For example, a model can be fed in images of dogs and cats and trained on a deep neural network to predict each class



CNN-based models can now discriminate between different leukocytes on peripheral smears with area under the receiver operating characteristic (AUROC) scores exceeding 95%, demonstrating promising proof of principle for greater automation in routine pathology practice [13–15]. Similar work is under way in the automated interpretation of bone marrow specimens [16]. CNNs have also demonstrated utility in describing qualitative and quantitative differences within individual lineages of cells, such as erythrocyte morphology and textural changes in sickle cell disease [17]. These successes extend to the differential diagnosis of disease, where models have demonstrated the ability to diagnose acute myeloid leukemia, differentiate between causes of bone marrow failure, and, in resource-constrained settings, serve as a point of care screening tool for lymphoma [14, 18, 19].

The above developments suggest several possible practical uses for AI. Pathology results often play a major role in informing the treatment of malignancy, but significant variation may exist among observers, a fact that can be especially true in the case of uncommon diseases, and one which carries significant clinical ramifications [20–22]. In such settings, AI models could serve as a consistent reference standard that could serve either to support diagnoses or to prompt review by another individual. In the case of uncommon conditions or resource-limited environments, AI systems could serve as a means of effectively triaging patients for referral to specialist care.

Radiology

Radiology has enjoyed similar benefits from advancements in computer vision. Detection of marrow involvement and bony lesions in patients based on PET or CT data [23, 24]. Similar methods have been used to define regions of marrow involvement in recurrent acute myeloid leukemia (AML) [25]. CNN-based methods have also been employed in image segmentation, a process that involves highlighting the boundaries of different structures in an image and is laborious for human operators to perform [26]. Beyond diagnosis and anatomic measurement, AI can extract additional information from radiographic data, such as more effective risk stratification of patients with Hodgkin lymphoma, which could in turn affect treatment decisions [27, 28]. As with pathology, AI has several potential roles to play in radiology, including use as a screening tool, decision aid, or prognostic model.

Laboratory and EHR Data

Beyond image processing, several other data sources, either individually or in concert with one another, provide valuable substrate for AI models and can augment clinical care in the setting of diagnosis, prognosis, and diagnosis.

Diagnosis

AI has been used in several avenues to improve the reliability, convenience, or efficiency of diagnoses. CNN-based approaches have been demonstrated to effectively diagnose multiple myeloma based exclusively on mass spectrometry data from peripheral blood [29]. In the case of difficult-to-differentiate conditions, such as various causes of bone marrow failure, personalized models have demonstrated high diagnostic ability by integrating patient demographics, laboratory data, and basic genetic information [30]. Similar approaches have also been employed for differential diagnosis of peripheral leukemia versus lymphoma [31]. As with image analysis, these advances open the door to AI's use as a means to reduce the amount of resources required for diagnostic studies and to standardize their interpretation.

Prognosis and Risk Stratification

Prognosis is a notoriously difficult task, and even in widely used clinical prognostication tools significant variability exists within risk strata [32]. AI, which is well equipped to handle nonlinear, complex data, has the potential to provide more refined, personalized prognoses. Such approaches have been used in the benign hematology setting to refine risk scores for central catheter thrombosis, identifying low-risk individuals with a 95% negative predictive value [33]. For patients receiving hematopoietic stem cell transplants, AI has been used to stratify individuals at low versus high risk for acute graft-versus-host disease, with implications for decision-making about immunosuppression of such individuals [34]. Similar efforts have been undertaken in autologous transplant for multiple myeloma [35]. In malignant hematology, AI has been used to improve upfront risk stratification for AML/MDS [36, 37]. In the post-treatment setting, where the presence of minimal residual disease (MRD) is an adverse prognostic factor, AI has demonstrated the ability to achieve human-level performance at MRD detection via flow cytometry and mass cytometry, something which could streamline and standardize the processing of such data [38–40].

Beyond providing valuable information to patients, advances in prognostic ability may better inform clinicians' treatment decisions by better assessing risk in heterogeneous risk strata. Sasaki et al. (2019) describe the use of a decision tree-based approach to CML treatment, and in retrospective data demonstrate that ML-informed treatment resulted in longer survival for patients compared with usual care [41]. While prospective validation is needed, the ability to apply risk-stratification to treatment planning is appealing, and in CML, it could provide guidance for clinicians as they seek to balance drug tolerability and efficaciousness.

Genomics and Response Prediction

In the case of hematologic malignancy, patients' disease biology is, with a few notable exceptions, driven by heterogeneous and complex genetic factors that are difficult to elucidate. The advent of accurate, relatively inexpensive next-generation sequencing (NGS) technologies makes it more feasible to enumerate the genetic alterations present in individual patients' malignancies. In lymphoma, NGS in tandem with AI has been used, with or without other data such as *in vitro* drug sensitivities, to predict response to chemotherapy [42, 43]. In MDS, this has been used to predict response to lenalidomide or hypomethylating agents using a recommender system inspired by targeted advertising algorithms [44, 45]. Similar approaches have been employed when considering the treatment of AML, where the generation of the BeatAML data repository has additionally provided an example of the utility of creating large, well-curated datasets available to the research community at large [46].

As seen in larger pan-cancer cohorts, adequately sized datasets also provide the ability to apply ML to the pursuit of basic science insights; recent work by Rheinbay et al. demonstrates the use of unsupervised methods to augment the discovery of novel non-coding driver mutations in a pan-cancer cohort [47]. Developing both the techniques and the data collections to better elucidate disease biology will, hopefully, in turn better inform novel, biologically sound therapies that are better suited to patients' unique disease biology.

Novel Therapeutics and Trial Design

Devising, refining, and testing new therapies is an expensive, time-consuming task with a low success rate. AI-based strategies may aid in developing more rational and efficient pipelines for drug development. This includes models designed to integrate *in vitro* sensitivity data from drug screens with genomic information about the cell lines used in order to more accurately predict the response to new agents and to investigate alternative uses of existing compounds [42]. From the perspective of medicinal chemistry, neural networks have demonstrated the ability to closely approximate the performance of more computationally expensive computational techniques for modeling protein-drug interactions, which may potentially allow researchers to take on more computationally demanding tasks [48].

Beyond basic and translational science, personalized care informed by AI has the potential to affect how clinical trials are conducted. Low response rates confer the need for increased size and expense in clinical trials, and strategies to effectively select the patients most likely to benefit from an intervention can lower costs and increase the likelihood of finding use for new therapies [49, 50]. Conversely, methods

that effectively identify patients unlikely to benefit from a treatment spare them futile treatment and needless toxicities, and open up the possibility for accessing other avenues of treatment. In the setting of malignancy, this could mean that patients unlikely to benefit from the current standard of care could receive investigational agents without being subjected to agents unlikely to benefit from them.

Future Challenges

AI has several hurdles to clear before realizing its potential in medicine. It is likely, however, to increase in prominence and, ultimately, to foster changes and evolution in clinical practice. Considering this, an understanding of AI's basics will need to become part of physicians' statistical literacy, and they should be aware of some of the challenges that accompany AI's emergence. These include logistical questions related to its implementation, the adequacy of data used to develop AI models, and the adoption of clinically meaningful standards for AI development. While there has been some speculation that AI will displace human pathologists or radiologists, these views are more commonly held by those less familiar with those fields and those less familiar with AI; domain experts instead envision complementary roles for AI and physicians [51–53].

Implementation

Bringing AI into the clinic represents a distinct challenge from the initial development of AI systems. Given its need for access to large amounts of patient data, AI systems will likely have to be implemented into EHR, posing potential challenges regarding security and data ownership. The prospect of an EHR-based system also raises questions about the effect on clinicians. EHR is frequently cited as a contributing factor in physician burnout, and potential unintended consequences such as perceived autonomy or an increased documentation burden need to be considered.

Practically speaking, AI also needs to be engineered in a manner that makes its use convenient in the setting of a busy practice. Systems that require extra work to manually enter patient data or access models are less likely to see uptake regardless of how well they perform. In cases such as pathology where physical specimens are used, the logistics of specimen preparation and analysis bear consideration. For implementation-focused research, e.g., recently published work in AI-augmented microscopy, the outcomes of interest include not only the discriminative ability of the system but also its ability to perform in real time and without substantially affecting existing workflows [54] (Table 1). Such factors directly influence a technology's ability to be incorporated into the practice of medicine, and as such should be considered endpoints in similar work.

Table 1 Representative AI publications in hematology. AA, aplastic anemia; MDS, myelodysplastic syndrome; CNN, convolutional neural network; *sens.*, sensitivity; *spec.*, specificity; AML, acute myeloid

leukemia; AUROC, area under the receiver operating characteristic curve; SUV, standardized uptake volume; MRD, minimal residual disease; SVM, support vector machine; NGS, next-generation sequencing

First author	Application	Method	Results
Kimura, 2019	Leukocyte classification and differential diagnosis of AA and MDS	CNN	94%/96% <i>sens./spec.</i> for leukocyte classification, 96%/100% <i>sens./spec.</i> for differentiating AA/MDS
Li, 2019	PET-CT to detect AML bone marrow involvement	PyRadiomics (manual feature engineering)	Outperformance of human radiologists; 88%/90% <i>sens./spec.</i> for bone marrow involvement
Milgrom, 2019	PET-CT prediction of refractory Hodgkin lymphoma via flow cytometry	CNN	AUROC 0.95 for refractory disease compared to 0.78 and 0.65 for feature-based predictions (volume, SUV)
Moraes L, 2019	Flow cytometry for differential diagnosis	Decision tree	66% top DDX; 95% inclusion in DDX
Ni W, 2016	AML MRD detection by flow cytometry	SVM	Equivalent performance to human data analysis ($c = 0.986$)
Arai Y, 2019	Acute leukemia relapse prediction after HSCT	Decision tree	0.75 AUROC
Goswami C 2019	Risk stratification for multiple myeloma autologous stem cell transplant	Decision tree	Statistically significant differentiation into risk strata; identification of high-risk patients
Nazha A, 2019	Use of NGS panels to predict hypomethylating agent resistance in MDS	Recommender algorithm	Statistically significant risk stratification by hypomethylating agent resistance likelihood
Gal O, 2019	Pediatric AML response likelihood based on NGS	KNN	0.85 AUROC for induction response

Appropriate Training Data

As with any other statistical model, AI models' performance hinges on the data used to develop them. Watson for Oncology (WFO), an IBM initiative designed to process, integrate, and implement large volumes of both patient data and medical literature to help oncologists choose appropriate options for patients, is an illustrative case. Enthusiasm for WFO has waned substantially, in large part because WFO's treatment recommendations often varied from those of clinicians, sometimes to an alarming degree [55, 56]. Studies of WFO's recommendations found such discrepancies more often in intermediate-stage malignancy or in patients unfit for standard therapies [57, 58]. WFO's example highlights both the inherent challenge of delivering truly personalized care and the significance of adequate source data. As with other clinical research, source data need to be considered when applying AI models in new settings, whether that is in a new health system or in a population different from the one used to develop the model.

The data used to generate models can also precipitate ethical challenges. Language models derived from repositories of free text from the internet display a clear tendency to associate male individuals more closely with leadership roles and professional accomplishment than their female counterparts [59, 60]. Similarly, programs designed to predict crime or approve loans have been observed to exhibit bias against racial minorities [61, 62]. These cases highlight the downside of AI's ability to learn from large datasets; if the data themselves reflect problematic outcomes or practices, then models run the risk of internalizing them.

Making AI Clinically Meaningful

Algorithms can only be as clinically meaningful as the outcomes that they are designed to predict. This requires, as with

other clinical research, the use of appropriately patient-oriented endpoints. On a more technical level, it also requires the use of appropriate metrics to evaluate model performance. For instance, the use of accuracy as a metric is of dubious use in medical settings because it assigns equal value to true positive and true negative results; in healthcare settings, where the outcome of interest is often only present in a minority of the population, accuracy will overstate the utility of a model due to a high proportion of true negative results, despite poor performance in identifying the true outcome of interest. In order to best align algorithms' performance with patients' interests, metrics should be selected and reported with an eye to the outcomes that clinicians and patients care about (e.g., explicitly describing the burden of false positive and negative results alongside model accuracy).

Beyond reliable performance, AI models' decisions need to be interpretable. Knowing not only *what* an algorithm predicts, but *why* it does so is critical when faced with clinical ambiguity, when discussing decision-making with patients, or in instances where there is disagreement between human and machine predictions. One of the ways to accomplish this is through the use of separate algorithms designed to identify how individual variables contribute to a model's output, an approach that has been developed for ML more broadly but has also studied specifically in the context of healthcare [63, 64]. Such approaches have been used in hematologic malignancy to highlight the most salient features of individual patients' diseases, something that can prove useful both in the context of explaining diagnoses and in making predictions about responsiveness to treatment [30, 37, 65].

In settings outside of hematology, there has also been success in designing neural networks whose architecture contains intrinsic mechanisms for explainability, such as highlighting regions of interest on pathology slides or identifying the phrases in a patient's medical record that most strongly suggest a particular diagnosis or prognosis [66]. Such measures

lend accountability and credibility to predictions, and facilitate AI models' integration into the larger picture of clinical deliberation rather than serving as stand-alone decision points.

Conclusion

Hematology as a field stands to benefit significantly from contemporary AI, both across a spectrum of different types of data and across a spectrum of patient care, including diagnosis, prognosis, and more effective management of hematologic disorders. As AI continues to gain a foothold in the management of pathologic, radiologic, genomic, and EHR data, attention needs to be paid to its effective implementation into clinical practice and the development of AI systems with an eye to how they will ultimately impact patient care.

Compliance with Ethical Standards

Conflict of Interest The authors report no conflicts of interest and have no financial disclosures related to this work.

Ethical Statement All reported studies/experiments with human or animal subjects performed by the authors have been previously published and complied with all applicable ethical standards (including the Helsinki declaration and its amendments, institutional/national research committee standards, and international/national/institutional guidelines).

Human and Animal Rights and Informed Consent This article does not contain any studies with human or animal subjects performed by any of the authors.

References

- LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015 May;521(7553):436–44.
- Miotto R, Wang F, Wang S, Jiang X, Dudley JT. Deep learning for healthcare: review, opportunities and challenges. *Brief Bioinform*. 2018 Nov 27;19(6):1236–46.
- Alsuliman T, Humaidan D, Sliman L. Machine learning and artificial intelligence in the service of medicine: necessity or potentiality? *Curr Res Transl Med* 2020;S2452318620300192.
- Deo RC. Machine learning in medicine. *Circulation*. 2015;132(20):1920–30.
- Buch VH, Ahmed I, Maruthappu M. Artificial intelligence in medicine: current trends and future possibilities. *Br J Gen Pract*. 2018 Mar;68(668):143–4.
- Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, et al. A guide to deep learning in healthcare. *Nat Med*. 2019 Jan;25(1):24–9.
- Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL. Artificial intelligence in radiology. *Nat Rev Cancer*. 2018 Aug;18(8):500–10.
- Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Commun ACM*. 2017 May 24;60(6):84–90.
- Keel S, Lee PY, Scheetz J, Li Z, Kotowicz MA, MacIsaac RJ, et al. Feasibility and patient acceptability of a novel artificial intelligence-based screening model for diabetic retinopathy at endocrinology outpatient services: a pilot study. *Sci Rep*. 2018 Dec;8(1):4330.
- Abramoff MD, Lavin PT, Birch M, Shah N, Folk JC. Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. *Npj Digit Med*. 2018 Dec;1(1):39.
- Coudray N, Ocampo PS, Sakellaropoulos T, Narula N, Snuderl M, Fenyo D, et al. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nat Med*. 2018;24(10):1559–67.
- Shen L, Margolies LR, Rothstein JH, Fluder E, McBride R, Sieh W. Deep learning to improve breast cancer detection on screening mammography. *Sci Rep*. 2019 Dec;9(1):12495.
- Hegde RB, Prasad K, Hebbar H, Singh BMK. Comparison of traditional image processing and deep learning approaches for classification of white blood cells in peripheral blood smear images. *Biocybern Biomed Eng*. 2019 Apr;39(2):382–92.
- 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 Dec;9(1):13385.
- Wang Q, Bi S, Sun M, Wang Y, Wang D, Yang S. Deep learning approach to peripheral leukocyte recognition. Zhang J, editor. *PLOS ONE*. 2019 Jun 25;14(6):e0218808.
- Chandradevan R, Aljudi AA, Drumheller BR, Kunanantaseelan N, Amgad M, Gutman DA, et al. Machine-based detection and classification for bone marrow aspirate differential counts: initial development focusing on nonneoplastic cells. *Lab Invest*. 2020 Jan;100(1):98–109.
- Xu M, Papageorgiou DP, Abidi SZ, Dao M, Zhao H, Kamiadakis GE. A deep convolutional neural network for classification of red blood cells in sickle cell anemia. Nie Q, editor. *PLOS Comput Biol*. 2017 Oct 19;13(10):e1005746.
- Alsalem MA, Zaidan AA, Zaidan BB, Hashim M, Madhloom HT, Azeed ND, et al. A review of the automated detection and classification of acute leukaemia: coherent taxonomy, datasets, validation and performance measurements, motivation, open challenges and recommendations. *Comput Methods Prog Biomed*. 2018 May;158:93–112.
- Im H, Pathania D, McFarland PJ, Sohani AR, Degani I, Allen M, et al. Design and clinical validation of a point-of-care device for the diagnosis of lymphoma via contrast-enhanced microholography and machine learning. *Nat Biomed Eng*. 2018 Sep;2(9):666–74.
- Elmore JG, Longton GM, Carney PA, Geller BM, Omega T, Tosteson ANA, et al. Diagnostic concordance among pathologists interpreting breast biopsy specimens. *JAMA*. 2015 Mar 17;313(11):1122–32.
- Fadi B, Luciana S, Epstein Jonathan I. The value of mandatory second opinion pathology review of prostate needle biopsy interpretation before radical prostatectomy. *J Urol*. 2010 Jul 1;184(1):126–30.
- Naqvi K, Jabbour E, Bueso-Ramos C, Pierce S, Borthakur G, Estrov Z, et al. Implications of discrepancy in morphologic diagnosis of myelodysplastic syndrome between referral and tertiary care centers. *Blood*. 2011 Oct 27;118(17):4690–3.
- Xu L, Tetteh G, Lipkova J, Zhao Y, Li H, Christ P, et al. Automated whole-body bone lesion detection for multiple myeloma on ⁶⁸ Ga-Pentixafor PET/CT imaging using deep learning methods. *Contrast Media Mol Imaging*. 2018;2018:1–11.
- Martinez-Martinez F, Kybic J, Lambert L, Mecková Z. Fully automated classification of bone marrow infiltration in low-dose CT of patients with multiple myeloma based on probabilistic density model and supervised learning. *Comput Biol Med*. 2016 Apr;71:57–66.
- Li H, Xu C, Xin B, Zheng C, Zhao Y, Hao K, et al. ¹⁸F-FDG PET/CT radiomic analysis with machine learning for identifying bone marrow involvement in the patients with suspected relapsed acute leukemia. *Theranostics*. 2019;9(16):4730–9.

26. Moon H, Huo Y, Abramson RG, Peters RA, Assad A, Moyo TK, et al. Acceleration of spleen segmentation with end-to-end deep learning method and automated pipeline. *Comput Biol Med.* 2019 Apr;107:109–17.
27. Milgrom SA, Elhalawani H, Lee J, Wang Q, Mohamed ASR, Dabaja BS, et al. A PET radiomics model to predict refractory mediastinal Hodgkin lymphoma. *Sci Rep.* 2019 Dec;9(1):1322.
28. Guo B, Tan X, Ke Q, Cen H. Prognostic value of baseline metabolic tumor volume and total lesion glycolysis in patients with lymphoma: a meta-analysis. *PLoS One.* 2019;14(1):e0210224.
29. Deulofeu M, Kolářová L, Salvadó V, María Peña-Méndez E, Almáši M, Štork M, et al. Rapid discrimination of multiple myeloma patients by artificial neural networks coupled with mass spectrometry of peripheral blood plasma. *Sci Rep.* 2019 Dec;9(1):7975.
30. Hilton C. Geno-Clinical Model for the diagnosis of bone marrow myeloid neoplasms. In *ASH; 2019* [cited 2019 Nov 14]. **Available from:** <https://ash.confex.com/ash/2019/webprogram/Paper126967.html>
31. Moraes LO, Pedreira CE, Barrena S, Lopez A, Orfao A. A decision-tree approach for the differential diagnosis of chronic lymphoid leukemias and peripheral B-cell lymphomas. *Comput Methods Prog Biomed.* 2019 Sep;178:85–90.
32. Patel SS, Sekeres MA, Nazha A. Prognostic models in predicting outcomes in myelodysplastic syndromes after hypomethylating agent failure. *Leuk Lymphoma.* 2017 Nov 2;58(11):2532–9.
33. Liu S, Zhang F, Xie L, Wang Y, Xiang Q, Yue Z, et al. Machine learning approaches for risk assessment of peripherally inserted central catheter-related vein thrombosis in hospitalized patients with cancer. *Int J Med Inf.* 2019 Sep;129:175–83.
34. Arai Y, Kondo T, Fuse K, Shibasaki Y, Masuko M, Sugita J, et al. Using a machine learning algorithm to predict acute graft-versus-host disease following allogeneic transplantation. *Blood Adv.* 2019 Nov 26;3(22):3626–34.
35. Goswami C, Poonia S, Kumar L, Sengupta D. Staging system to predict the risk of relapse in multiple myeloma patients undergoing autologous stem cell transplantation. *Front Oncol.* 2019 Jul 12;9:633.
36. Nazha A, Komrokji RS, Meggendorfer M, Mukherjee S, Al Ali N, Walter W, et al. A personalized prediction model to risk stratify patients with myelodysplastic syndromes. *Blood.* 2018 Nov 29;132(Supplement 1):793–793.
37. Shreve J. A personalized prediction model to risk stratify patients with acute myeloid leukemia (AML) using artificial intelligence. In *ASH; 2019* [cited 2019 Nov 14]. **Available from:** <https://ash.confex.com/ash/2019/webprogram/Paper128066.html>
38. Ni W, Hu B, Zheng C, Tong Y, Wang L, Li Q, et al. Automated analysis of acute myeloid leukemia minimal residual disease using a support vector machine. *Oncotarget* [Internet]. 2016 Nov 1 [cited 2019 Nov 7];7(44). **Available from:** <http://www.oncotarget.com/fulltext/12430>
39. 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 Nov;37:91–100.
40. Arvaniti E, Claassen M. Sensitive detection of rare disease-associated cell subsets via representation learning. *Nat Commun.* 2017 Apr;8(1):14825.
41. Sasaki K, Kantarjian HM, Jabbour E, Ravandi F, Konopleva MY, Borthakur GM, et al. The impact of treatment recommendation by Leukemia Artificial Intelligence Program (LEAP) on survival in patients with chronic myeloid leukemia in chronic phase (CML-CP). *Blood.* 2019 Nov 13;134(Supplement 1):1642–1642.
42. Chang Y, Park H, Yang H-J, Lee S, Lee K-Y, Kim TS, et al. Cancer drug response profile scan (CDRscan): a deep learning model that predicts drug effectiveness from cancer genomic signature. *Sci Rep.* 2018 Dec;8(1):8857.
43. Mani-Vamosfaderani A, Neiband MS, Benvidi A. Identification of molecular features necessary for selective inhibition of B cell lymphoma proteins using machine learning techniques. *Mol Divers.* 2019 Feb;23(1):55–73.
44. 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 Sep;3:1–11.
45. Madanat YF, Sekeres MA, Mukherjee S, Hirsch CM, Guan Y, Nagata Y, et al. Genomic biomarkers predict response/resistance to lenalidomide in non-Del(5q) myelodysplastic syndromes. *Blood.* 2018 Nov 29;132(Supplement 1):1797–1797.
46. Lee S-I, Celik S, Logsdon BA, Lundberg SM, Martins TJ, Oehler VG, et al. A machine learning approach to integrate big data for precision medicine in acute myeloid leukemia. *Nat Commun.* 2018 Dec;9(1):42.
47. Rheinbay E. Analyses of non-coding somatic drivers in 2,658 cancer whole genomes. *Nature.* 2020;578:102–11.
48. Yakovenko O, Jones SJM. Modern drug design: the implication of using artificial neuronal networks and multiple molecular dynamic simulations. *J Comput Aided Mol Des.* 2018 Jan;32(1):299–311.
49. Harrer S, Shah P, Antony B, Hu J. Artificial intelligence for clinical trial design. *Trends Pharmacol Sci.* 2019 Aug;40(8):577–91.
50. Woo M. An AI boost for clinical trials. *Nature.* 2019 Sep 25;573: S100–2.
51. Recht M, Bryan RN. Artificial intelligence: threat or boon to radiologists? *J Am Coll Radiol.* 2017 Nov;14(11):1476–80.
52. Sharma G, Carter A. Artificial intelligence and the pathologist: future frenemies? *Arch Pathol Lab Med.* 2017 May;141(5):622–3.
53. Pinto dos Santos D, Giese D, Brodehl S, Chon SH, Staab W, Kleinert R, et al. Medical students' attitude towards artificial intelligence: a multicentre survey. *Eur Radiol.* 2019 Apr;29(4):1640–6.
54. Chen P-HC, Gadepalli K, MacDonald R, Liu Y, Kadowaki S, Nagpal K, et al. An augmented reality microscope with real-time artificial intelligence integration for cancer diagnosis. *Nat Med.* 2019 Sep;25(9):1453–7.
55. Ross C, Swetlitz I. IBM pitch its Watson supercomputer as a revolution in cancer care. It's nowhere close. *Stat news* [Internet]. 2017 Sep 5; **Available from:** <https://www.statnews.com/2017/09/05/watson-ibm-cancer/>
56. Schmidt C. M. D. Anderson Breaks With IBM Watson, Raising questions about artificial intelligence in oncology. *JNCI J Natl Cancer Inst* [Internet]. 2017 May [cited 2019 Nov 10];109(5). **Available from:** <https://academic.oup.com/jnci/article-lookup/doi/10.1093/jnci/djx113>
57. Choi YI, Chung J, Kim KO, Kwon KA, Kim YJ, Park DK, et al. Concordance rate between clinicians and Watson for oncology among patients with advanced gastric cancer: early, real-world experience in Korea. *Can J Gastroenterol Hepatol.* 2019 Feb 3;2019: 1–6.
58. Kim EJ, Woo HS, Cho JH, Sym SJ, Baek J-H, Lee W-S, et al. Early experience with Watson for oncology in Korean patients with colorectal cancer. Orzechowski P, editor. *PLOS ONE.* 2019 Mar 25;14(3):e0213640.
59. Machines taught by photos learn a sexist view of women. *Wired* [Internet]. [cited 2019 Nov 26]; **Available from:** <https://www.wired.com/story/machines-taught-by-photos-learn-a-sexist-view-of-women/>
60. Zou J, Schiebinger L. AI can be sexist and racist — it's time to make it fair. *Nature.* 2018 Jul;559(7714):324–6.
61. Courtland R. Bias detectives: the researchers striving to make algorithms fair. *Nature.* 2018 Jun 20;558:357–60.
62. Hague DC. Benefits, pitfalls, and potential bias in health care AI. *N C Med J.* 2019 Jul;80(4):219–23.
63. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. In: Guyon I, Luxburg UV, Bengio S, Wallach H, Fergus

- R, Vishwanathan S, et al., editors. Advances in neural information processing systems 30 [Internet]. Curran Associates, Inc.; 2017 [cited 2019 Nov 17]. p. 4765–4774. **Available from:** <http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf>
64. Lundberg SM, Nair B, Vavilala MS, Horibe M, Eisses MJ, Adams T, et al. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nat Biomed Eng.* 2018 Oct;2(10):749–60.
65. Radakovich N. Predicting Response to hypomethylating agents in patients with myelodysplastic syndromes (MDS) using artificial intelligence (AI). In ASH; 2019 [cited 2019 Nov 14]. **Available from:** <https://ash.confex.com/ash/2019/webprogram/Paper126275.html>
66. Tomašev N, Glorot X, Rae JW, Zielinski M, Askham H, Saraiva A, et al. A clinically applicable approach to continuous prediction of future acute kidney injury. *Nature.* 2019 Aug;572(7767):116–9.

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