HHS Public Access

Author manuscript

Best Pract Res Clin Haematol. Author manuscript; available in PMC 2020 October 12.

Published in final edited form as:

Best Pract Res Clin Haematol. 2020 September ; 33(3): 101192. doi:10.1016/j.beha.2020.101192.

Acute myeloid leukemia and artificial intelligence, algorithms and new scores

Nathan Radakovich[a], **Matthew Cortese**[b], **Aziz Nazha**[a,b,c,*]

[a]Cleveland Clinic Lerner College of Medicine of Case Western Reserve University, United States

[b]Department of Hematology and Medical Oncology, Cleveland Clinic, United States

[c]Center for Clinical Artificial Intelligence, Cleveland Clinic, United States

Abstract

Artificial intelligence, and more narrowly machine-learning, is beginning to expand humanity's capacity to analyze increasingly large and complex datasets. Advances in computer hardware and software have led to breakthroughs in multiple sectors of our society, including a burgeoning role in medical research and clinical practice. As the volume of medical data grows at an apparently exponential rate, particularly since the human genome project laid the foundation for modern genetic inquiry, informatics tools like machine learning are becoming crucial in analyzing these data to provide meaningful tools for diagnostic, prognostic, and therapeutic purposes. Within medicine, hematologic diseases can be particularly challenging to understand and treat given the increasingly complex and intercalated genetic, epigenetic, immunologic, and regulatory pathways that must be understood to optimize patient outcomes. In acute myeloid leukemia (AML), new developments in machine learning algorithms have enabled a deeper understanding of disease biology and the development of better prognostic and predictive tools. Ongoing work in the field brings these developments incrementally closer to clinical implementation.

Keywords

Acute myeloid leukemia; Artificial intelligence; Machine learning; Malignant hematology; Genomics; Multi-omics; Risk stratification

Introduction

Acute myeloid leukemia (AML) contains substantial clinical and genomic heterogeneity. There is a commensurately broad spectrum of prognoses; therefore, risk-stratification at diagnosis plays a critical role in treatment selection. While the decision of whether to pursue an anthracycline- and cytarabine-based induction regimen is, with few exceptions, largely

*Corresponding author. Center for Clinical Artificial Intelligence, Cleveland Clinic, Lerner College of Medicine, Department of Hematology and Medical Oncology, Taussig Cancer Institute, Cleveland Clinic, Desk R35 9500 Euclid Ave, Cleveland, OH, 44195, United States. nazhaa@ccf.org (A. Nazha).

Declaration of competing interest

AN reports research support from Jazz Pharmaceuticals, is part of the speaker bureau for Incyte and Novartis, is a member of the data monitoring committee of MEI Pharma, and is an advisory board member and provides consulting for Karyopharma, Abbvie, and Daiichi Sankyo. The other authors declare no competing interests.

determined by non-genetic factors such as patient performance status and comorbidites, genetics play a significant role in decision-making about if—and when—to pursue subsequent hematopoietic stem cell transplant (HCST) [1]. Increasingly, they also inform selection of targeted therapies, both in the upfront and relapse setting, as well as in patients whose performance status precludes induction therapy [2,3].

Currently, risk stratification and treatment planning in AML are informed by a relatively limited number of genetic criteria. The 2017 European LeukemiaNet (ELN) risk stratification criteria rely primarily on cytogenetic abnormalities, with the exception of select somatic mutations [4]. Although informative, scoring systems such as ELN are intrinsically limited by significant intra-stratum heterogeneity. The advent of increasingly inexpensive technologies such as next-generation sequencing now makes it feasible to capture the genetic changes that underlie that heterogeneity. Thus, the challenge that faces clinicians as they try to better tailor therapies to individual patients is how to best make sense of the surfeit of biological data available to them.

Machine learning (ML), a branch of computer science that makes predictions based on patterns that it has learned from training data rather than explicit *a priori* design, is an appealing tool to help address this challenge. Briefly, ML falls into two major categories: the first, supervised learning, identifies patterns in data (e.g., a set of patient characteristics) that are most predictive of a specified outcome (e.g., response or non-response to an intervention); in the second category, unsupervised learning, algorithms identify potentially meaningful structures in data (e.g., distinct phenotypes within a disease) that are then appraised by investigators for potential clinical/biological significance [5].

Because of its ability to effectively handle nonlinear data with complex internal relationships and because of the wealth of data available in healthcare, interest in medical ML has grown exponentially in recent years across medical specialties. While still a developing field, ML holds significant promise for altering how physicians are able to classify and treat disease. This review summarizes work done to date applying ML to the study of AML (Table 1), and highlights potential future applications in risk stratification, prognosis, and treatment decisions (Fig. 1) (an in-depth introduction to ML is beyond the scope of this article, but several excellent overviews of the topic exist [5–7]).

Deciphering-Omics data

Technologies such as next-generation sequencing and proteomics make it increasingly feasible to query the biological underpinnings of malignancy, and are increasingly used in the study of AML. While ML and ML-adjacent techniques such as clustering and dimensionality reduction are employed widely throughout such research, this review focuses more narrowly on the use of supervised learning approaches.

Diagnosis and sub-classification of AML is challenging and often hinges on expert interpretation of several diagnostic modalities including morphology, immunophenotype, cytogenetics, and sequencing. Warnat-Herresthal et al. (2020) describe the use of LASSO regression, an ML method for automatically selecting the most predictive features from a

dataset and performing regression on them, in the automated diagnosis and classification of AML based on based on transcriptomic data pooled from 105 studies comprising 12,029 AML, MDS, ALL, and healthy patients [8]. Without ancillary data or expert input, the authors' model identified AML based on transcriptomic data with > 99% accuracy, and distinguished between M3 AML and other subtypes with 97.6% sensitivity and 99.5% specificity. By including detailed observations of model performance in the differential diagnosis of other hematologic malignancies, the authors also illustrate their model's performance for a variety of distinct clinical scenarios. Interestingly, model performance was not significantly impacted by the exclusion of known AML-related genes, suggesting a complementary role for ML and other, more widely used genomic analysis methods such as differential gene expression [8].

Work by Shreve, et all demonstrates the ability of genomic data to more effectively risk-stratify patients with AML in recently presented work employing gradient boosting trees (GBT), an ML approach that generates predictions based on the collective input of hundreds or thousands of small decision trees [9]. When trained on data from 3241 unique patients, a GBT model achieved a c-index for overall survival ranging from 0.80 to 0.85 across five datasets compared to a c-index of 0.59 when using the 2017 ELN criteria, demonstrating not only ML's increased prognostic power but robustness across varied patient populations and clear explanations for individual risk [9].

Gerstung et al. (2017) highlight the use of genomic data in conjunction with clinical information to inform treatment decisions in AML [10]. In their paper the authors describe the use of a modified regression-based method for estimating the likelihood of survival based on whether a patient received HCST in their first remission (CR1) after induction for AML, or after their first relapse. Using data from 1540 AML patients the authors were able to predict survival with a c-index of 0.72. Based on retrospective analysis using their model they estimate that they would be able to reduce HSCT rates without adversely affecting mortality by identifying patients with favorable subtypes (e.g., inv(16) or t(8:21) AML) who would, if relapsed, still face a good prognosis with HCST [10]. The authors' work highlights the potential for ML strategies to tailor the aggressiveness of therapies, although given the novelty of such approaches prospective validation is needed to demonstrate their efficacy and safety.

Lee et al. (2018) highlight another ML approach to genomic data in a paper describing a novel method for predicting drug sensitivities [11]. The authors describe an approach that incorporates information learned from the Tumor Cancer Genome Atlas (TCGA) to help inform the likelihood that alterations in a given gene will have therapeutic significance, thereby mitigating the risk of identifying chance patterns without a strong biological underpinning. Using this approach, the authors attained state-of the art performance compared to models that did not incorporate such background information. Although such approaches are still removed from the point of clinical implementation, they highlight the potential for a given ML model to incorporate biological insights from external datasets in order to refine its predictive ability.

Work in other hematologic malignancies highlights potential alternative methods for handling genomic data. A 2019 paper describes the use of a recommender algorithm, which functions similarly to those used to generate shopping or media suggestions, to help identify patients at risk for resistance to hypomethylating agents in patients with myelodysplastic syndromes (MDS) [12]. By applying the same paradigm to the relationship between gene mutations and drug responsiveness, the model, which was based on 433 patients from 3 institutions, was able create meaningful risk strata among patients with at least 3 mutations identified, with low-risk patients surviving 22.8 months and high-risk patients surviving 14.6 months [12]. Such algorithms are designed to handle sparse data in other settings (such as recommending new purchases when based of the purchase of a few dozen items out of millions of potential options), and thus may be useful frameworks for evaluating gene alterations, which are similarly sparse. Similarly, the associations between genes and outcomes identified by the algorithm play a potentially informative role in elucidating novel genomic interactions.

Greater personalization from clinical data

Non-genomic features still provide a valuable substrate for clinical predictions, and may serve as the basis for models either independently or in conjunction with other forms of data. Indeed, even relatively simple models can yield informative predictions, such as a regression-based model described by Krug et al. (2010) that achieved an area under the receiver operating characteristic curve (AUROC) of 0.72 when combining patient characteristics, laboratory values, and known genetic risk factors [13]. Subsequently, new ML strategies have emerged, with the potential to enhance clinicians' predictive abilities. In contrast to simple regression-based methods, newer algorithms are capable of capturing interactions and hierarchies within variables. For instance, by creating trees with one mutation or clinical characteristic only considered in the context of an antecedent variable, decision tree based methods such as random forests and gradient boosted trees can capture more "context" for a variable if informative. Notably, this ability also increases the attendant risk of overfitting, a phenomenon wherein a model is trained based on features specific to its training data, which in turn reduces its external validity.

Work in other hematologic malignancies highlights the utility of these advances. Recent work has employed GBT-based strategies to refine the diagnosis of bone marrow failure syndromes such as MDS, with AUROC > 0.95 based primarily on peripheral smears, clinical data, and a handful of single mutations from 2471 patients, demonstrating the potential for some clinical questions to be solved highly effectively by ML approaches [14]. Importantly, the model can also highlight which factors, for example particular lab or genomic findings, most contribute to an individual's diagnosis, thereby providing an intelligible description of how it makes predictions. In the setting of elevated hemoglobin and a known JAK2 mutation, a patient's monocyte count may not be relevant to the diagnostic process, however it might be pivotal in distinguishing MDS and CMML. Methods such as GBT can capture this complexity, and thus make more sophisticated predictions (it bears repeating that this ability also increases the risk of overfitting—for example, if every patient named "John" in a training set happens to have CMML, models given that information might incorporate that decidedly non-generalizable fact into their decision-making process rather than correctly

dismissing it as a product of random chance). ML may also be used to adaptively inform response to treatment, as demonstrated in our work using serial blood count measurements to make expedited predictions about success or failure of hypomethylating agents in MDS, and are in the process of applying similar strategies to evaluating the success of induction therapy for AML [15].

Unsupervised methods have also been employed to directly observe clinical data. Gandelman et al. (2019) highlight an unsupervised, clustering-based approach in their use of t-distributed stochastic neighbor embeddings (tSNE), a dimensionality reduction method designed to create two-dimensional representations of high-dimension data, to propose clinically distinct subtypes of graft-versus-host disease (GVHD) [16]. Through inspection of seven different clusters identified by the tSNE method, the authors describe seven distinct clinical phenotypes of GVHD based on the severity of involvement in eye, liver, joints, fascia, mouth, and gastro-intestinal tract. The authors then used frequentist statistics to examine clinical differences between the groups, thereby identifying distinct risk strata. Such experiments demonstrate the potential for unsupervised methods to yield novel, clinically meaningful insights into how to approach diagnosis and risk stratification. They also serve as a means of hypothesis generation for corollary work investigating whether observed clinical subtypes reflect any underlying differences in disease biology. Such results should, however, be interpreted with caution in the absence of external validation. In this case the authors examine the reproducibility of their findings by applying a separate algorithm, self-organizing maps, to their *t*-SNE results, in turn finding that five of seven phenotypes were reproducible. Approaches such as this may be beneficial not only in risk stratification, but in hypothesis generation about biologically distinct disease subtypes.

Image analysis

Image analysis in the medical community at large has been transformed by the advent of convolutional neural networks (CNNs), a subtype of neural network that first rose to prominence in 2012 [17]. Briefly, CNNs function by mimicking the underlying mechanisms of biological vision, using a series of filters that progress from elementary shape detection to abstract, higher-level representation of objects [18]. While large datasets are ordinarily required to train highly accurate CNNs, excellent performance on smaller datasets can be achieved by repurposing other models. This process, called transfer learning, entails retaining the superficial shape- and texture-recognition layers of a pre-trained CNN while altering the deeper layers in order to tune it for a new task. This approach has enabled several high-profile advances in interpreting histopathological, radiological, and photographic medical data [19–22].

Currently, the majority of AI-based image analysis in hematologic malignancy has focused on classification tasks. Chandradevan et al. (2019) describe classification of cells on bone marrow aspirates based on extensive expert labeling [23]. Similar work in AML-adjacent fields (MDS and ALL) has also been undertaken, in each case classifying diseases of interest with sensitivity and specificity > 95% [24,25]. Despite the remarkable performance of CNNs, the process of creating adequately labeled datasets for model training is laborious

and time-consuming. Methods that require a lesser degree of image annotation are an ongoing area of research, and their development may facilitate further progress in this field.

Beyond identification and classification tasks, image analysis may be able to yield novel information from pathology slides. Studies in solid tumors have demonstrated CNNs' ability to predict clinically meaningful single mutations and microsatellite instability, and even to accurately risk-stratify patients purely on the basis of tissue architecture in colon cancer [26–28]. It is possible that there are corollary data to be gleaned from hematopathology, however this remains an area in need of further study. The ability to glean clinically meaningful information directly from histopathology slides carries significant potential advantages given the ubiquity, relatively rapid turnaround time, and low cost of obtaining histopathologic data.

Challenges, lessons learned, and the path to clincical implementation

Shouval et al. highlight potential limitations of ML nicely in a 2016 study attempting to use clinical data to predict mortality from a dataset 26,266 patients undergoing HSCT for AML [29]. After employing several contemporary models, the authors observed that despite an enviable sample size, model performance consistently plateaued above training sizes of 6000–8000 patients, and that, for all models tested, performance similarly plateaued after the inclusion of merely six variables; they conclude that improvements in predictive ability require the identification of additional prognostic features. Needless to say, the ceiling for predictive ability is unique to each machine learning task and is also influenced by the type of machine learning strategy employed, but as noted by other commentators, successful ML approaches to some clinical problems may ultimately require multimodal approaches capable of integrating clinical, genomic, and other data in order to attain clinically meaningful levels of performance [30].

As with any other statistical method, care needs to be taken when devising new scores and predictive models. As demonstrated by Wang et al. (2017), existing tools such as the ELN criteria vary in their prognostic utility when applied to novel cohorts of patients [31]. The studies in this review that document reliable performance across different patient cohorts are likely able to do so in part due to the inclusion of multi-institutional data in the training process [8,9]. Similar approaches will be important to ensure the same external validity in future work. As illustrated by Warnat-Herresthal et al., the test characteristics of ML models vary based on the population in which they are applied, so care needs to be taken to ensure that the data used to develop and validate ML models reflect the setting in which they will ultimately be used [8]. Prospective validation, when feasible, also provides additional evidence for models' utility, and will become increasingly important as ML approaches widespread clinical implementation.

Widespread clinical acceptance of ML will also hinge on its ability to generate transparent explanations for its predictions. Interpretability is an active area in machine learning research, and work in recent years has yielded several flexible solutions for interpreting a diverse array of models; for instance, the SHAP (SHapely Additive exPlanations) approach described by Lundberg et al. (2017) can summarize how particular variables in an ML model contribute to individual-level predictions, as well as what variables are of most significance

to the model globally [32]. We and others have used SHAP to illustrate variable importance with patient-level resolution for several clinical applications including interpreting NGS data, differential diagnosis, response prediction based on time series analysis, and generating treatment recommendations [9,14,15,33].

How will ML affect current paradigms in AML management as it edges toward clinical utility? ML-based diagnostic strategies may be useful in clarifying ambiguous data, streamlining the workflow of diagnosticians, and providing consistent performance in resource-constrained settings. The ability to more thoroughly examine patient-specific disease biology may elucidate the mechanisms that cause different treatments to succeed or fail in particular patients, and thus inform decision-making around treatments. From the perspective of prognosis and risk stratification, ML strategies offer a more granular, individualized approach; while scores and strata are useful heuristics that reflect clinically meaningful distinctions, their simplicity may place an upper limit on their predictive ability. With sufficiently accurate ML models, it may be possible to adopt a prognostic approach that directly estimates probabilities (along with the degree of uncertainty about those estimates) and provides clinician-accessible explanations for its predictions. Although aspirational at this point in time, the eventual adoption of such a paradigm would facilitate treatment approaches that more closely reflect patient-specific disease factors.

Conclusion

AML's heterogeneous nature makes the development of truly personalized care challenging; current standards for attempting to understand disease biology are clearly imperfect in their ability to fully appraise individual risk, and thus to best inform treatment. Emerging progress in the medical use of ML, both in AML and other malignancies, heralds new, potentially transformative means for approaching its management.

Practice points

- Machine learning provides a useful tool for making sense of the expanding depth and volume of data available in AML treatment
- Machine learning models provide a diverse and flexible framework for handling multi-omic, clinical, and laboratory data
- Machine learning research for AML remains pre-clinical; multi-center data collection and prospective model validation are key to realizing machine learning's clinical potential

Research agenda

Multi-omic data

- Identification of novel, clinically meaningful disease subtypes
- Correlation of disease biology to clinical outcomes
- Integration of large, multi-omic datasets

Leveraging clinical data

- Increasing personalization of risk stratification and treatment recommendations
- Development of interpretable, clinician-friendly models

Image analysis (histopathology and radiology)

- Automated/assisted diagnosis based on bone marrow and peripheral blood images
- Acquisition of novel clinical data from images using deep learning
- Correlation of imaging findings to prognosis/treatment approach

Bringing ML from research to practice

- Development of multi-institutional databases for robust model generation
- Prospective validation of clinical ML models
- Incorporation of ML into clinical trials and research
- Practitioner education and awareness about machine learning research and best practices

References

[1]. Estey EH. Acute myeloid leukemia: 2019 update on risk-stratification and management. Am J Hematol 2018;93(10):1267–91. [PubMed: 30328165]

[2]. Patel SA, Gerber JM. A user's guide to novel therapies for acute myeloid leukemia. Clin Lymphoma, Myeloma & Leukemia 2020 1 30.

[3]. Versluis J, Cornelissen JJ, Craddock C, Sanz MÁ, Canaani J, Nagler A. Acute myeloid leukemia in adults [Internet] In: Carreras E, Dufour C, Mohty M, Kröger N, editors. The EBMT handbook: hematopoietic stem cell transplantation and cellular therapiesseventh ed.Cham (CH): Springer; 2019 [cited 2020 Mar 4]. Available from:. http://www.ncbi.nlm.nih.gov/books/NBK554029/.

[4]. Döhner H, Estey E, Grimwade D, Amadori S, Appelbaum FR, Büchner T, et al. Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. Blood 2017 1 26;129(4):424–47. [PubMed: 27895058]

[5]. Liu Y, Chen P-HC, Krause J, Peng L. How to read articles that use machine learning: users' guides to the medical literature. J Am Med Assoc 2019 11 12;322(18):1806–16.

[6]. Deo RC. Machine learning in medicine. Circulation 2015 11 17;132(20):1920–30. [PubMed: 26572668]

[7]. Sivapalaratnam S Artificial intelligence and machine learning in haematology. Br J Haematol 2019 4;185(2):207–8. [PubMed: 30729496]

[8]. Warnat-Herresthal S, Perrakis K, Taschler B, Becker M, Baßler K, Beyer M, et al. Scalable prediction of acute myeloid leukemia using high-dimensional machine learning and blood transcriptomics [Internet]. iScience 2019 12 18 [cited 2020 Mar 4];23(1). Available from:. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6992905/.

[9]. Shreve J A personalized prediction model to risk stratify patients with acute myeloid leukemia (AML) using artificial intelligence. ASH 2019 [cited 2019 Nov 14]. Available from:. https://ash.confex.com/ash/2019/webprogram/Paper128066.html.

[10]. Gerstung M, Papaemmanuil E, Martincorena I, Bullinger L, Gaidzik VI, Paschka P, et al. Precision oncology for acute myeloid leukemia using a knowledge bank approach. Nat Genet 2017 3;49(3):332–40. [PubMed: 28092685]

[11]. 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 12;9(1):42. [PubMed: 29298978]

[12]. 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 9(3):1–11.

[13]. Krug U, Röllig C, Koschmieder A, Heinecke A, Sauerland MC, Schaich M, et al. Complete remission and early death after intensive chemotherapy in patients aged 60 years or older with acute myeloid leukaemia: a web-based application for prediction of outcomes. Lancet Lond. Engl 2010 12 11;376(9757):2000–8.

[14]. Hilton C Geno-clinical model for the diagnosis of bone marrow myeloid neoplasms. ASH 2019 [cited 2019 Nov 14]. Available from:. https://ash.confex.com/ash/2019/webprogram/Paper126967.html.

[15]. Radakovich N Predicting response to hypomethylating agents in patients with myelodysplastic syndromes (MDS) using artificial intelligence (AI). Ash. 2019 [cited 2019 Nov 14]. Available from:. https://ash.confex.com/ash/2019/webprogram/Paper126275.html.

[16]. Gandelman JS, Byrne MT, Mistry AM, Polikowsky HG, Diggins KE, Chen H, et al. Machine learning reveals chronic graft-versus-host disease phenotypes and stratifies survival after stem cell transplant for hematologic malignancies. Haematologica 2019 1;104(1):189–96. [PubMed: 30237265]

[17]. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Commun ACM 2017 5 24;60(6):84–90.

[18]. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015 5;521(7553):436–44. [PubMed: 26017442]

[19]. Yu K-H, Zhang C, Berry GJ, Altman RB, Ré C, Rubin DL, et al. Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features. Nat Commun 2016 11;7(1):12474. [PubMed: 27527408]

[20]. McKinney SM, Sieniek M, Godbole V, Godwin J, Antropova N, Ashrafian H, et al. International evaluation of an AI system for breast cancer screening. Nature 2020 1;577(7788):89–94. [PubMed: 31894144]

[21]. Roffman D, Hart G, Girardi M, Ko CJ, Deng J. Predicting non-melanoma skin cancer via a multi-parameterized artificial neural network. Sci Rep 2018 12;8(1):1701. [PubMed: 29374196]

[22]. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. J Am Med Assoc 2016 12 13;316(22):2402–10.

[23]. Chandradevan R, Aljudi AA, Drumheller BR, Kunananthaseelan 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 2019 9 30 [Internet]. http://www.nature.com/articles/s41374-019-0325-7 [cited 2019 Nov 2]; Available from:.

[24]. 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 12;9(1):13385. [PubMed: 31527646]

[25]. Shafique S, Tehsin S. Acute lymphoblastic leukemia detection and classification of its subtypes using pretrained deep convolutional neural networks. Technol Canc Res Treat 2018 1;17 1533033818802789.

[26]. Kather JN, Pearson AT, Halama N, Jäger D, Krause J, Loosen SH, et al. Deep learning can predict microsatellite instability directly from histology in gastro-intestinal cancer. Nat Med 2019;25(7):1054–6. [PubMed: 31160815]

[27]. Coudray N, Ocampo PS, Sakellaropoulos T, Narula N, Snuderl M, Fenyö 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. [PubMed: 30224757]

[28]. Kather JN, Krisam J, Charoentong P, Luedde T, Herpel E, Weis C-A, et al. Predicting survival from colorectal cancer histology slides using deep learning: a retrospective multicenter study In: Butte AJ, editor. PLoS Med 16. 2019 1 24:e1002730. 1.

[29]. Shouval R, Labopin M, Unger R, Giebel S, Ciceri F, Schmid C, et al. Prediction of hematopoietic stem cell transplantation related mortality- lessons learned from the in-silico approach: a European society for blood and marrow transplantation acute leukemia working party data mining study [Internet]. PloS One 2016 3 4 [cited 2020 Mar 4];11(3). Available from:. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4778768/.

[30]. Haferlach T, Schmidts I. The power and potential of integrated diagnostics in acute myeloid leukaemia. Br J Haematol 2020 1;188(1):36–48. [PubMed: 31808952]

[31]. Wang M, Lindberg J, Klevebring D, Nilsson C, Mer AS, Rantalainen M, et al. Validation of risk stratification models in acute myeloid leukemia using sequencing-based molecular profiling. Leukemia 2017 10;31(10):2029–36. [PubMed: 28167833]

[32]. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions [Internet] Guyon I, Luxburg UV, Bengio S, Wallach H, Fergus R, Vishwanathan S, editors. Advances in neural information processing systems, vol. 30 Curran Associates, Inc.; 2017 [cited 2019 Nov 17]. pp. 4765–4774. Available from:. http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf.

[33]. 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 11 13;134(Supplement_1). 1642–1642.

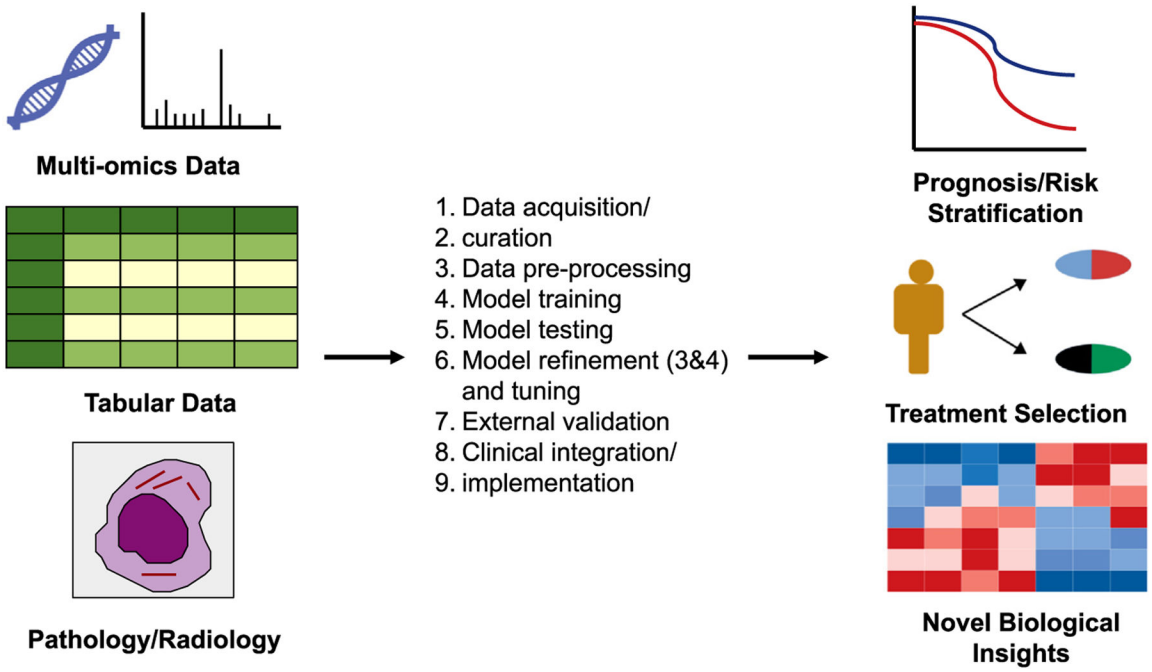

Fig. 1.
Example implementations of machine learning in oncologic practice.

Table 1

Survey of recent publications employing machine learning in the study of AML.

Author, Year	Study Size	Application	ML Model	Results
Warnat-Herresthal, 2019	12,029 AML, leukemia, control patients; 102 studies	Transcriptomics-based AML diagnosis and subtype classification	LASSO regression	98% accuracy for AML DDx: ≥ 97% accuracy for AML subclassification (except FAB M7 leukemia)
Shreve, 2019	3421 AML patients	Prediction of OS based on 44-gene NGS panel	GBT	C-index of 0.80 compared to ELN C-index of 0.59; individual-level prediction explanations
Lee, 2018	30 patients, 160 drugs	Multi-omic approach to predicting drug sensitivities	Novel model (MERGE)	Identification of gene hubs/networks from TCGA; reproduction of known genomic markers; identification of putative novel drug susceptibilities
Nazha, 2019	433 MDS patients	NGS panel based prediction of HMA resistance	Recommender algorithm	Stratification of patients with ≥ 3 mutations into high (OS = 14.6 months) and low risk (OS = 22.8 months)
Krug, 2010	2208 AML patients	CR and mortality prediction from clinical data	Multiple regression	AUROC 0.72 with cytogenetic risk factors; AUROC 0.65 with clinical data alone
Gerstung	1540	Risk stratification for HSCT in AML	Novel regression method	C-index of 0.72 for overall survival, individualized risk prediction for HSCT in CR1 versus relapse
Kimura, 2019	3261 peripheral smears	Differentiation between MDS/aplastic anemia	CNN	96.2% sensitivity, 100% specificity, 0.99 AUROC for distinguishing MDS/aplastic anemia
Chandradevan, 2019	10,000 annotated cells	Classification of cells on bone marrow aspirate	CNN	0.98 AUROC for overall classification on non-neoplastic tissue
Shouval, 2016	26,266 AML + HCST patients	Prediction of mortality s/p HCST	Multiple models*	AUROC up to 0.72; identified key

LASSO: least absolute shrinkage and selection operator, DDx: differential diagnosis, NGS: next-generation sequencing, GBT: gradient boosted tree, ELN: European Leukemia Net, TCGA: The Cancer Genome Atlas, HMA: hypomethylating agent, AUROC: area under the receiver operating characteristic curve, CNN: convolutional neural network.