Using machine learning to unravel the intricacy of acute myeloid leukemia

by Luca Guarnera and Valeria Visconte

Received: September 25, 2023.
Accepted: October 2, 2023.

Citation: Luca Guarnera and Valeria Visconte.
Using machine learning to unravel the intricacy of acute myeloid leukemia.

Publisher's Disclaimer.
E-publishing ahead of print is increasingly important for the rapid dissemination of science. Haematologica is, therefore, E-publishing PDF files of an early version of manuscripts that have completed a regular peer review and have been accepted for publication. E-publishing of this PDF file has been approved by the authors. After having E-published Ahead of Print, manuscripts will then undergo technical and English editing, typesetting, proof correction and be presented for the authors’ final approval; the final version of the manuscript will then appear in a regular issue of the journal. All legal disclaimers that apply to the journal also pertain to this production process.
Using machine learning to unravel the intricacy of acute myeloid leukemia

Luca Guarnera¹,², Valeria Visconte¹

¹Department of Translational Hematology & Oncology Research, Taussig Cancer Institute, Cleveland Clinic, Cleveland OH, 44114, USA
²Department of Biomedicine and Prevention, University of Rome Tor Vergata, Rome, Italy

Keywords: acute myeloid leukemia, machine learning, genetic risk stratification

Correspondence to:
Valeria Visconte, Ph.D.
Department of Translational Hematology and Oncology Research, Taussig Cancer Institute
9620 Carnegie Ave n building, Building NE6-250, Cleveland, OH, USA 44106
ORCID ID: 0000-0002-2993-1509
E-mail: visconv@ccf.org

Text word count: 656

Disclosures
No conflicts of interest to disclose.

Contributions
LG and VV wrote the manuscript.
The paper by Park et al.\textsuperscript{1} published in Haematologica proposes the application of machine learning (ML) algorithms to refine cluster signatures characterized by cytogenetic and mutational features common to patients with acute myeloid leukemia (AML). Such effort is inspired by the goal of defining similar clusters possibly informing survival outcomes and response or refractoriness to conventional therapies (intensive chemotherapy [IC], hypomethylating agents [HMA], and HMA plus venetoclax [VEN]). The study cohort comprised 279 patients who underwent IC (n=131), HMA (n=76), and HMA/VEN (n=72) in a time span of almost 4 years. The focus of the study is to validate ELN 2022 classification in older patients and for such aim a cohort of patients $\geq$60 years was needed. The study also expands on the investigation of ELN 2022 in patients for whom IC is not appropriate.

Using unsupervised hierarchical clustering the authors were able to merge features according to similarities pointing out the heterogeneity of the disease with the identification of 9 genomic clusters characterized by diverse survival outcomes based on treatment. Some clusters were associated to better outcomes in one or another treatment group. For instance, cluster 4 was enriched in core-binding factor-AML [96% CBF-AML] and associated with better prognosis in the IC group, which reflected the choice of IC in older patients with CBF-AML. One of the major additions of this study to the generalized concept of using ML to measure the effects of combinatorial gene mutations, was the incorporation of treatment data. However, although ML was able to distinguish cluster-types associated with treatment, given the small sample size per treatment group, a definitive conclusion is difficult to be reached.

Having said that, this study complements other key results achieved through ML in the field of AML in recent years. The interconnection of several variables in large cohorts of patients allowed to explore, through ML, different patients’ stratifications\textsuperscript{2,3}, integrated prognostic algorithms\textsuperscript{4}, identified biomarkers\textsuperscript{5} and supported cytomorphological diagnosis.\textsuperscript{6} This large amount of data offered insights on different aspects of AML management, respecting the granularities of disease features and suggesting the possibilities of adding new factors or classifiers to consider in the tailoring of treatment strategy.
Thus, in the near future, one could envision a role of ML into the refinement of the disease classifications and as a useful guide to a proper integration of emerging strategies such as immunotherapies, results from clinical trials, and maintenance treatments.

Furthermore, the paper by Park et al.\textsuperscript{1} offers the opportunity to reflect on the discrepancies among the studies using ML clustering in AML. In 2021, Awada et al.\textsuperscript{2} applied standard and ML driven analysis to 6,788 AML cases and defined a genomic 4-tiered model, challenging the conventional dichotomy between de novo and secondary AML. Recently, 4 clusters were also identified in a large European cohort analyzed by Eckardt et al.\textsuperscript{7} by re-stratifying patients in comparison to ELN 2017 criteria. Ultimately, the current study identified 9 genomic clusters by incorporating treatment data. These differences can be attributed to several biases determined by unavailable/ different choice of data, small sample sizes (a limitation pointed out also by Park et al.\textsuperscript{1} on their study), short patient follow-up, exclusion/ inclusion of clinical data and misclassifications. Of note, the comparison among studies can be challenging and misleading, especially when using unsupervised learning approaches in which, lack of prediction, might be applicable only to a specific context. The context-dependent interpretation of the data underpins, in fact, one of the most important pitfall and concern of ML which is the restriction of an algorithm to a single specific use.

In summary, this article together with others demonstrate the utility of ML algorithms in resolving intricate molecular relationships and their impact on clinical outcomes. More importantly, the future holds the promise of dissecting genomic interplay guiding precision medicine. In line with several new tools being tested, large studies, standardization, validation cohorts, and uniformity of pipelines across studies could be the keys to unlock the full potential of ML.
References