

Using machine learning to unravel the intricacy of acute myeloid leukemia

Luca Guarnera^{1,2} and Valeria Visconte¹

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

Correspondence: V. Visconte

visconv@ccf.org


Received: September 25, 2023.

Accepted: October 2, 2023.

Early view: October 12, 2023.

<https://doi.org/10.3324/haematol.2023.284085>

©2024 Ferrata Storti Foundation

Published under a CC BY license 

The paper by Park *et al.*¹ published in this issue of *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 on 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 four years. The focus of the study is to validate European LeukemiaNet (ELN) 2022 classification in older patients, and for this, a cohort of patients ≥ 60 years was analyzed. 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 with 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, it is difficult to reach a definitive conclusion.

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 has allowed researchers to explore, through ML, different patient stratifications^{2,3} and integrated prog-

nostic algorithms,⁴ to identify biomarkers,⁵ and support cytomorphological diagnosis.⁶ This huge amount of data has offered insights into 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 for ML in 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.*¹ offers the opportunity to reflect on the discrepancies among the studies using ML clustering in AML. In 2021, Awada *et al.*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.*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 also pointed out by Park *et al.*¹ with regards to 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 inputs might be applicable only to a specific context. In fact, the context-dependent interpretation of the data underpins one of the most important pitfalls and concerns of ML: the restriction of an algorithm to a single specific use.

In summary, this article contributes to the literature demonstrating 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.

Disclosures

No conflicts of interest to disclose.

Contributions

Both authors contributed equally.

References

1. Park S, Kim TY, Cho B-S, et al. Prognostic value of European Leukemia Net 2022 criteria and genomic clusters using machine learning in older adults with acute myeloid leukemia. *Haematologica*. 2024;109(4):1095-1106.
2. Awada H, Durmaz A, Gurnari C, et al. Machine learning integrates genomic signatures for subclassification beyond primary and secondary acute myeloid leukemia. *Blood*. 2021;138(19):1885-1895.
3. Kewan T, Durmaz A, Bahaj W, et al. Molecular patterns identify distinct subclasses of myeloid neoplasia. *Nat Commun*. 2023;14(1):3136.
4. Eckardt JN, Röllig C, Metzeler K, et al. Prediction of complete remission and survival in acute myeloid leukemia using supervised machine learning. *Haematologica*. 2023;108(3):690-704.
5. Zhang Y, Liu D, Li F, et al. Identification of biomarkers for acute leukemia via machine learning-based stemness index. *Gene*. 2021;804:145903.
6. Eckardt JN, Schmittmann T, Riechert S, et al. Deep learning identifies acute promyelocytic leukemia in bone marrow smears. *BMC Cancer*. 2022;22(1):201.
7. Eckardt JN, Röllig C, Metzeler K, et al. Unsupervised meta-clustering identifies risk clusters in acute myeloid leukemia based on clinical and genetic profiles. *Commun Med (Lond)*. 2023;3(1):68.