SOHO ANNUAL MEETING 2024

Enhancing Prognostic Models in Acute Lymphoblastic Leukemia Using Machine Learning: A Multivariate Approach

S. Sulley, Y. Zhou

1. Boston University, Department of Pathology & Laboratory Medicine

Add your logo here

INTRODUCTION

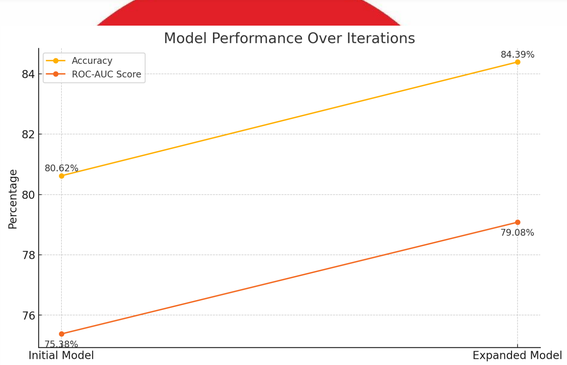
Acute Lymphoblastic Leukemia (ALL) is a critical hematological malignancy with diverse outcomes based on a variety of clinical and genetic factors. The integration of machine learning (ML) in hematopathology promises to enhance prognostic accuracy and personalize treatment strategies, leveraging complex datasets to predict patient outcomes.

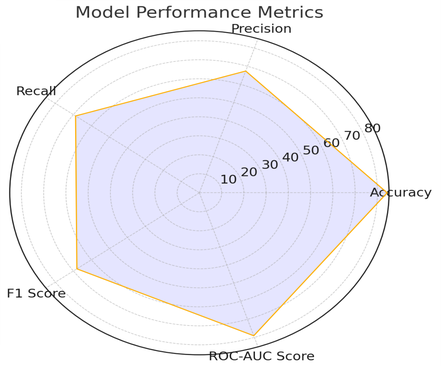
AIM

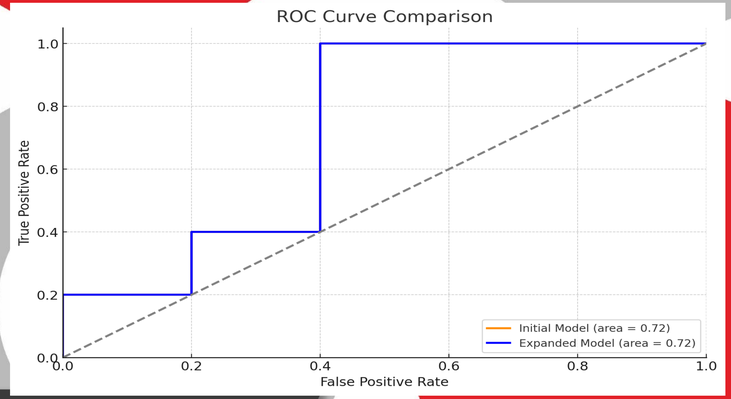
To develop and evaluate a machine learning model that predicts overall survival status in ALL patients using clinical and genetic features.

METHOD

- This study utilized a retrospective cohort dataset comprising clinical and genomic data from three studies on pediatric Acute Lymphoblastic Leukemia1. Initial model development focused on two clinical features: Diagnosis Age and White Blood Cell (WBC) count. The model was later expanded to include additional features (Sex, MLL Status, DNA Index) to investigate their impact on predictive performance.
- The dataset combined information from the Pediatric Acute Lymphoid Leukemia - Phase II (TARGET, 2018), and Acute Lymphoblastic Leukemia studies (St. Jude, Nat Genet 2015 & 2016), offering a comprehensive view of clinical and genetic characteristics influencing ALL outcomes.
- The primary outcome was overall survival status, categorized as living or deceased. Model performance was evaluated using accuracy, precision, recall, F1 score, and ROC-AUC score metrics.







RESULTS

The initial model, utilizing Diagnosis Age and WBC, yielded an accuracy of 80.62% and a ROC-AUC score of 75.38%. Incorporating additional features improved accuracy to 84.39% and the ROC-AUC score to 79.08%. These enhancements suggest the added features' significant role in accurately predicting survival outcomes.

CONCLUSIONS

Machine learning models can effectively predict overall survival status in ALL patients using a combination of clinical and genetic features. The inclusion of broader feature sets significantly enhances model performance, underscoring the importance of integrating diverse data types in prognostic modeling. This approach holds potential for improving clinical decision-making and personalized patient care in hematopathology..

REFERENCES

- Cerami et al. The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. Cancer Discovery. May 2012 2; 401. PubMed.
- **Hunger, S. P., & Mullighan, C. G. (2015).** Acute Lymphoblastic Leukemia in Children. *New England Journal of Medicine*, 373(16), 1541-1552.
- **Pui, C. H., & Evans, W. E. (2006).** Treatment of Acute Lymphoblastic Leukemia. *New England Journal of Medicine*, 354(2), 166-178.
- **Zhang, J., Ding, L., Holmfeldt, L., Wu, G., Heatley, S. L., Payne-Turner, D., ... & Downing, J. R. (2012).** The genetic basis of early T-cell precursor acute lymphoblastic leukaemia. *Nature*, 481(7380), 157-163.
- **Roberts, K. G., Gu, Z., Payne-Turner, D., McCastlain, K., Harvey, R. C., Chen, I. M., ... & Mullighan, C. G. (2017).** High frequency and poor outcome of Philadelphia chromosome-like acute lymphoblastic leukemia in adults. *Journal of Clinical Oncology*, 35(4), 394-401.

ACKNOWLEDGEMENTS

Patients, Researchers and contributors to dataset such as CBioPortal

CONTACT INFORMATION

sulleys@bu.edu

X: @saaniesulley