The 65th ASH Annual Meeting Abstracts

ORAL ABSTRACTS

613.ACUTE MYELOID LEUKEMIAS: CLINICAL AND EPIDEMIOLOGICAL

Harnessing Artificial Intelligence for Risk Stratification in Acute Myeloid Leukemia (AML): Evaluating the Utility of Longitudinal Electronic Health Record (EHR) Data Via Graph Neural Networks

Riya Sinha[1], *Matthew Schwede, MD*[2,3], *Ben Viggiano, MS*[2], *David Kuo, MD*[2], *Solomon Henry*[4], *Douglas Wood*[4], *Gabriel Mannis, MD*[5], *Ravindra Majeti, MD PhD*[3,6,7], *Jonathan Chen, MD PhD*[8,9,10], *Tian Y. Zhang, MD PhD*[5,11]

[1] Department of Biomedical Data Science, Stanford University, Menlo Park, CA
[2] Department of Biomedical Data Science, Stanford University, Stanford, CA
[3] Division of Hematology, Department of Medicine, Stanford University, Stanford, CA
[4] Technology & Digital Solutions (TDS), Research Technology, and Research Data Services, Stanford Health Care and School of Medicine, Stanford, CA
[5] Department of Medicine, Division of Hematology, Stanford University School of Medicine, Stanford, CA
[6] Institute for Stem Cell Biology and Regenerative Medicine, Stanford University School of Medicine, Stanford, CA
[7] Stanford Cancer Institute, Stanford, CA
[8] Clinical Excellence Research Center, Stanford University, Stanford, CA
[9] Division of Hospital Medicine, Department of Medicine, Stanford University School of Medicine, Stanford, CA
[10] Stanford Center for Biomedical Informatics Research, Stanford University, Stanford, CA
[11] Stanford Cancer Institute, Stanford University School of Medicine, Stanford, CA

Background:
AML is a life-threatening disease, and to determine which patients need allogeneic stem cell transplantation, hematologists risk-stratify each case. However, standard risk stratification using the European LeukemiaNet (ELN) criteria is focused on baseline mutations and chromosomal aberrations, and the risk estimate is not updated during a patient's course. In other blood cancers, recalculating the risk with treatment response data can help guide the need for more intensive therapy (Kurtz, et al, *Cell*, 2019). Furthermore, deep learning graph neural networks (GNN) applied to EHR data have strong predictive power in a hematology context (Fouladvand, et al, *J Biomed Inform*, 2023). Thus, we evaluated the power of a GNN to predict survival in AML using longitudinal EHR data, specifically with labs and histological features that are not included in the ELN but may capture the treatment response.

Methods:
Patients who were seen at the Stanford Cancer Institute, had EHR data available within six months of diagnosis, and were diagnosed with AML between June 1998 and January 2021 were included in this retrospective analysis. The GNN was trained to predict survival at two years from diagnosis using the first six months of clinical data. Patients were excluded if they were lost to follow-up before two years or died before six months. Data were collected from structured databases associated with Stanford's EHR, except that diagnosis dates were from Stanford's Cancer Registry, and survival data was supplemented with other databases including the Social Security Death Index. Dysplasia, bone marrow cellularity, and bone marrow blast percentages from pathology reports ("pathology report data") were extracted using text processing algorithms and weakly supervised machine learning (Ratner, et al, *ArXiv*, 2017).

To represent time series information, we framed each patient's timeline as a network (or "graph") of events. The primary GNN model was a heterogenous graph transformer classifier with two node types: complete blood count (CBC) data and pathology report data (Hu, et al, *ArXiv*, 2020). Data from the same week were assumed to be from the same timeframe and connected with bidirectional edges. Data separated by longer time periods were connected with unidirectional edges of a separate edge type. The independent test dataset consisted of patients whose ELN 2022 classification was available, and to train the model, the remaining data were divided into train/validation splits of 0.9/0.1.

Results:
Of the 2,535 patients with survival data, 1,029 met inclusion criteria. Table 1 summarizes the data available in the EHR for each variable, and nearly all patients had CBC and pathology report data. The area under the receiver operating characteristic (AUROC) using the ELN 2022 criteria for predicting survival in the test dataset was 0.79. The AUROC curve for the GNN model

was comparable at 0.76, despite not using any variables from the ELN criteria, and the model effectively stratified patients' disease into high- and low-risk in the independent test dataset (hazard ratio [HR] 3.0, log-rank p = 0.0009). Interestingly, despite not having access to mutation or cytogenetic data, the high-risk cases were enriched in known high-risk mutations, like *TP53* and *RUNX1*, and in high-risk chromosomal aberrations, like 5q deletion (Table 1). Although the model predictions correlated with the ELN criteria in some ways, they also stratified the ELN intermediate-risk AML cases into high and low risk (HR 6.1 for model-predicted high risk among ELN intermediate cases, p = 0.07).

Conclusions:

Risk stratification using artificial intelligence and longitudinal data from the EHR performed comparably to the ELN 2022 criteria and has the potential to further stratify the ELN categories. The model performed well despite only using histological features and lab values, which are more readily available and more frequently updated than next-generation sequencing results. In the future, this approach may further improve with a larger sample size and additional variables, such as measurable residual disease and treatment information. Given the heterogeneity and increasing complexity of AML classification, leveraging artificial intelligence to assist with classification will be crucial, and these results are a step towards a future where data are automatically extracted from the EHR and used for continuously updated risk stratification.
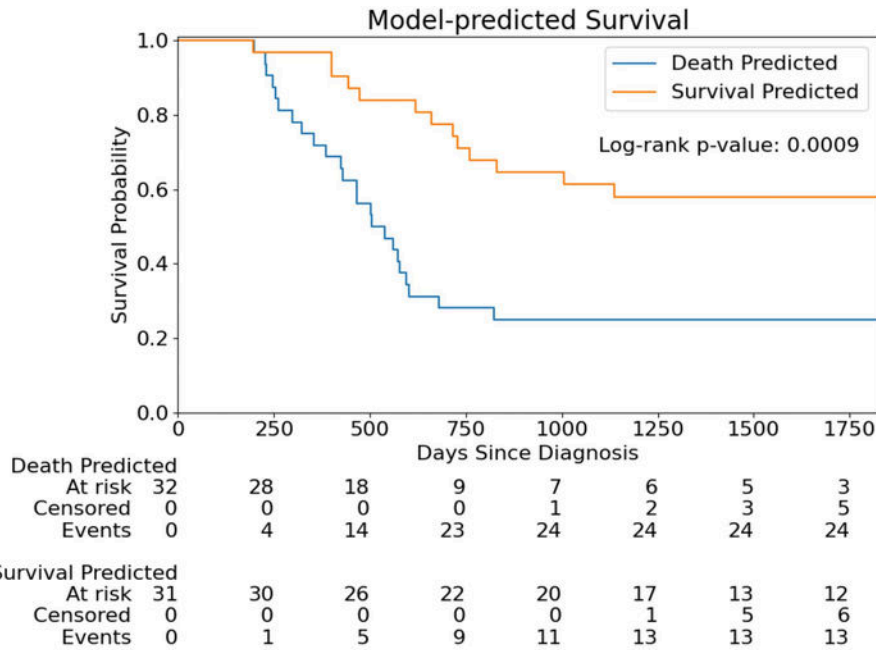
Model-predicted Survival

Figure 1: Survival prediction for patients who survived beyond 6 months, using the first 6 months of data.

Variable from EHR	# of Patients with Any Measurement (% Total)	Total # of Measurements	Summary Stats for All Observations	Summary Stats for Patients Predicted to Survive (n = 634)	Summary Stats for Patients Predicted not to Survive (n = 395)
Laboratory data	1027 (99.81%)	47672			
White blood cell count	1025 (99.61%)	24407	3.8 (0.0, 342.24)	4.0 (0.0, 342.24)	3.4 (0.0, 329.6)
Peripheral blast %	732 (71.14%)	5342	21.7 (0.0, 100.0)	19.0 (0.0, 99.0)	23.0 (0.0, 100.0)
Absolute neutrophil count	1021 (99.22%)	26665	2.0 (0.0, 3584.0)	2.43 (0.0, 3584.0)	1.4 (0.0, 1944.0)
Pathology reports	974 (94.66%)	3458			
Bone marrow cellularity	939 (91.25%)	2766	normo (hypo, hyper)	normo (hypo, hyper)	normo (hypo, hyper)
Presence of dysplasia	974 (94.66%)	3458	520 (53.39%)	224 (23.0%)	296 (30.39%)
Bone marrow blast %	808 (78.52%)	2402	4.0 (0.0, 99.0)	2.5 (0.0, 99.0)	11.0 (0.0, 97.5)
Mutation data	563 (54.71%)	867			
NPM1	556 (54.03%)	807	98 (17.63%)	68 (12.23%)	30 (5.4%)
FLT3 ITD	519 (50.44%)	698	104 (20.04%)	59 (11.37%)	45 (8.67%)
TP53	141 (13.7%)	195	20 (14.18%)	5 (3.55%)	15 (10.64%)
ASXL1	82 (7.97%)	123	16 (19.51%)	6 (7.32%)	10 (12.2%)
RUNX1	82 (7.97%)	123	16 (19.51%)	3 (3.66%)	13 (15.85%)
Karyotype data	680 (66.08%)	1789			
Loss of 17p	680 (66.08%)	1789	37 (5.44%)	11 (1.62%)	26 (3.82%)
Loss of 5q	680 (66.08%)	1789	60 (8.82%)	12 (1.76%)	48 (7.06%)
Treatment data	669 (65.01%)	8890			
HMA and/or venetoclax	669 (65.01%)	8890	213 (31.84%)	60 (8.97%)	153 (22.87%)
Chemotherapy treatment	669 (65.01%)	8890	487 (72.8%)	339 (50.67%)	148 (22.12%)

Table 1: Statistics for select variables extracted from the EHR. "Summary Stats" (last three columns) are median (min, max) for continuous and ordinal variables, and the number of patients with a positive result (% of cases with a measurement) otherwise. A unique "measurement" is a test or group of tests at a specific patient/time combination. HMA = hypomethylating agent, ITD = internal tandem duplication, cellularity is indicate by prefixes (e.g. "normo"-cellular).

Figure 1

https://doi.org/10.1182/blood-2023-190151