Artificial Intelligence, Machine Learning, and Radiomics-Based Prediction Models in Relapsed or Refractory Hodgkin Lymphoma: A Systematic Review

Javier Higuera-Ornelas, MD, MSc[1] · **Sai Harsha Yendluri, MD**[2] · Samyuktha Prabu, MD[3] · Lakshmi Chandana Velaga, MBBS[4] · Venkata Balaji Chenna, MBBS[5] · Shravani Sirikonda, MD[6] · Akhil Jain, MD[7] · Rupak Desai, MBBS[8]

1. Department of Human Genetics, Faculty of Medicine and Health Sciences, McGill University, Montreal, Quebec, Canada
2. Department of Medicine, BronxCare Health System, Bronx, New York
3. Department of Medicine, University of Central Florida College of Medicine, Orlando, Florida
4. Northwest Medical Center, Tucson, Arizona
5. Department of General Medicine, Kurnool Medical College, Kurnool, Andhra Pradesh, India
6. Department of Medicine, University of Central Florida College of Medicine, Orlando, Florida
7. Division of Hematology/Oncology, University of Iowa Hospitals and Clinics, Iowa City, Iowa
8. Independent Researcher, Outcomes Research, Atlanta, Georgia

BACKGROUND

Relapsed or refractory Hodgkin lymphoma (R/R HL) is clinically heterogeneous, and conventional prognostic tools may not fully capture the underlying imaging, molecular, and clinical complexity.

OBJECTIVE

To systematically review artificial intelligence (AI), machine learning (ML), deep learning (DL), and radiomics-based models for prediction and prognostication in R/R HL.

METHODS

- Systematic search of **PubMed and Scopus**, conducted per **PRISMA 2020**.
- Eligible: **Hodgkin lymphoma** (relapsed/refractory or relapse prediction) across frontline, salvage plus transplant, and checkpoint inhibitor settings.
- Studies had to evaluate a computational model for PFS, OS, treatment response, or risk stratification.
- Independent screening by **3 reviewers**; disagreements resolved by consensus.
- Extracted design, population, model type, inputs, outcomes and performance (AUC, C-index, HR).
- Risk of bias assessed with **PROBAST+AI**.

RESULTS

- **7 studies (894 patients)**, predominantly retrospective; **2 externally validated** multicenter models.
- Inputs: PET radiomics (3), MRI radiomics (1), spatial proteomics (1), clinical variables (1), transcriptomics (1); **3 used multimodal integration**.
- Methods ranged from SVM-based AdaBoost and Extra Trees to ridge regression, random forest, graph clustering and spatial scoring.
- **AUC 0.82–0.95** in 4 studies (internal or test set); the one externally validated AUC was **0.75**.
- Three studies showed significant Kaplan–Meier risk-group separation; **no study reported calibration** or compared against regression baselines.

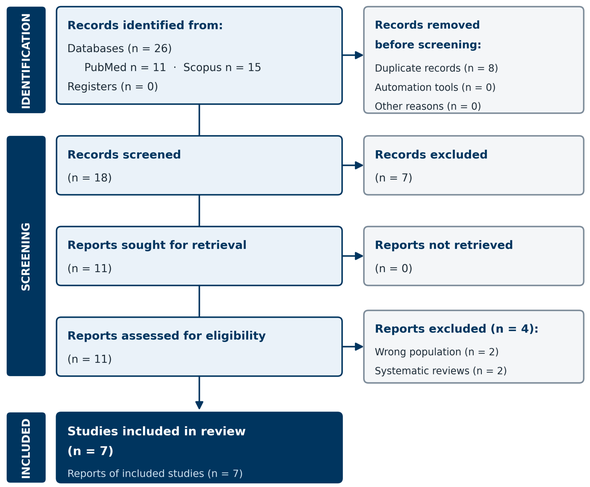
CONCLUSIONS

- AI/ML approaches show promise for risk stratification in R/R HL, particularly with **multimodal data integration**.
- Small samples, heterogeneous outcomes, and limited validation restrict clinical translation.
- Prospective, multi-institutional validation with **standardized reporting** is needed.

LIMITATIONS

- Overall risk of bias was high in 5 of 7 studies, driven chiefly by the analysis domain.
- Small samples and absent overfitting safeguards were near-universal.
- No calibration reporting and no comparison with regression baselines.
- Heterogeneous outcome definitions preclude meta-analysis.

STUDY SELECTION



CHARACTERISTICS AND PERFORMANCE OF THE 7 INCLUDED STUDIES

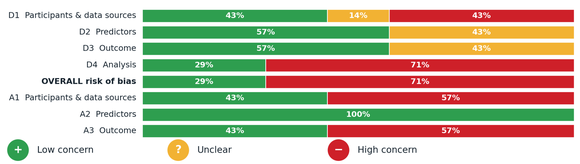
Study	Design, setting and population (n)	Model input and method	Outcome predicted	Discrimination	Risk stratification and validation
Driessen 2023	Retrospective analysis of prospective trial cohorts; multicenter, externally validated (n = 113 + 69)	^{18}F-FDG PET radiomics + clinical variables; ML backward selection with logistic regression	3-year PFS (binary landmark)	AUC 0.84 (0.74–0.93) training; 0.81 cross-validated; 0.75 (0.63–0.87) external	3-y PFS 38.1% vs 88.4% (p<0.0001); external 38.5% vs 75.0% (p=0.015); OS p=0.022
Braghini 2025	Retrospective transcriptomic analysis of classical HL (n = 130)	Multi-omic transcriptomics (single-cell + bulk deconvolution); ridge regression, hierarchical clustering	Response to treatment (binary)	AUC 0.83 (cell types); 0.87 combined; 0.81 clinical only	4 clusters: C3 lowest and C4 highest responders; no external validation
Albano 2022	Single-center cohort, newly diagnosed HL on frontline ABVD (n = 40); median follow-up 74 months	Whole-body MRI radiomics + clinical and PET-derived MTV; Extra Trees classifier	Refractory or relapsed disease (binary)	AUC 0.82 (5-fold cross-validation × 100 nested iterations)	Not reported; no external validation
Milgrom 2019	Retrospective, single institution; early-stage (I–II) classical HL (n = 251)	^{18}F-FDG PET radiomics (33 texture features); SVM-based AdaBoost classifier	Primary refractory disease (binary)	AUC 0.95 (0.87–1.00); error rate 1.8%; vs MTV 0.78, TLG 0.78, SUVmax 0.65	5 subgroups: group 5, 21% refractory and 75% died of HL vs 0% in groups 1–2; no external validation
Manson 2021	Retrospective, multicenter (14 centers); R/R HL responding to anti-PD-1, at discontinuation (n = 40; 25/15 split)	17 clinical variables; machine-learning classifier	PFS after anti-PD-1 discontinuation	C-index 0.67 (validation); error rate 32.8%	HR 1.04 per year of age (1.01–1.07), p=0.002; complete response p=0.02; no external validation
Cavinato 2021	Retrospective cohort; HL, baseline risk of therapy failure	PET radiomics (multi-view representation); recurrence-specific supervised graph clustering	Recurrence (time-to-event)	Not reported	2 clusters; median recurrence 659 days vs not reached; HR 0.16 (0.05–0.54), p=1.9×10^{-6}
Aoki 2023	Retrospective paired diagnostic–relapse biopsies with independent validation cohort (n = 71 pairs; validation n = 44)	Spatial proteomics (imaging mass cytometry) + scRNA-seq; RHL4S spatial scoring algorithm	Failure-free survival after autologous SCT	Not reported (Kaplan–Meier separation)	5-y FFS 41% vs 81% (p<0.0001); external 41% vs 83% (p=0.027); 5-y OS 88% vs 100% (p=0.046); HR 2.86 (1.63–5.02)

Discrimination values are apparent or cross-validated unless labelled external. Two studies (Driessen, Aoki) underwent external validation; no study reported calibration.

RISK OF BIAS — PROBAST+AI

Study	RISK OF BIAS					APPLICABILITY		
	D1 Participants	D2 Predictors	D3 Outcome	D4 Analysis	OVERALL risk of bias	A1 Participants	A2 Predictors	A3 Outcome
Driessen 2023	+	+	+	+	+	+	+	+
Braghini 2025	−	?	?	−	−	−	+	−
Albano 2022	−	+	+	−	−	−	+	−
Milgrom 2019	−	?	?	−	−	−	+	−
Manson 2021	+	+	+	−	−	+	+	+
Cavinato 2021	?	?	?	−	−	−	+	−
Aoki 2023	+	+	+	+	+	+	+	+

Summary across the 7 included studies



REFERENCES

1. Driessen J, et al. Blood Adv. 2023;7(21):6732-43. 2. Braghini JQ, et al. Hematol Oncol. 2025;43(5):e70135. 3. Albano D, et al. Magn Reson Imaging. 2022;86:55-60. 4. Milgrom SA, et al. Sci Rep. 2019;9(1):1322. 5. Manson G, et al. Eur J Nucl Med Mol Imaging. 2021;48(4):1144-53.
6. Cavinato L, et al. Annu Int Conf IEEE EMBC. 2021;2021:2155-8. 7. Aoki T, et al. J Clin Oncol. 2024;42(9):1077-87. 8. Page MJ, et al. BMJ. 2021;372:n71. 9. Moons KGM, et al. BMJ. 2025;388:e082505.

ACKNOWLEDGEMENTS

The authors thank the Department of Medicine, BronxCare Health System, for institutional support.

CONTACT

Sai Harsha Yendluri, MD
BronxCare Health System, Bronx, NY
dr.sai.yendluri@gmail.com