DIAGNOSTIC ACCURACY OF MACHINE-LEARNING MODELS IN NON-HODGKIN'S LYMPHOMA: A SYSTEMATIC REVIEW AND META-ANALYSIS

Beenish Saeed
Fatima Memorial Hospital College of Medicine and Dentistry

Introduction

Diagnosis of non-Hodgkin's lymphoma (NHL) is critical for staging and therapy. Machine-learning (ML) and deep-learning (DL) methods in histopathology, PET/CT radiomics and flow cytometry are aimed at improving diagnostic accuracy and workflow. Nonetheless, the diagnostic performance and external validity of ML in NHL have not been fully synthesized. This study aimed to quantify the diagnostic accuracy of ML methods for NHL and assess methodological rigor and clinical readiness.

Objectives

- Synthesize diagnostic performance (Sensitivity, Specificity, AUC) of ML models for NHL diagnosis.
- Compare performance metrics across different ML platforms (e.g., Deep Learning, Flow Cytometry, Radiomics).
- Critically appraise study quality and risk of bias using QUADAS-AI and PROBAST.

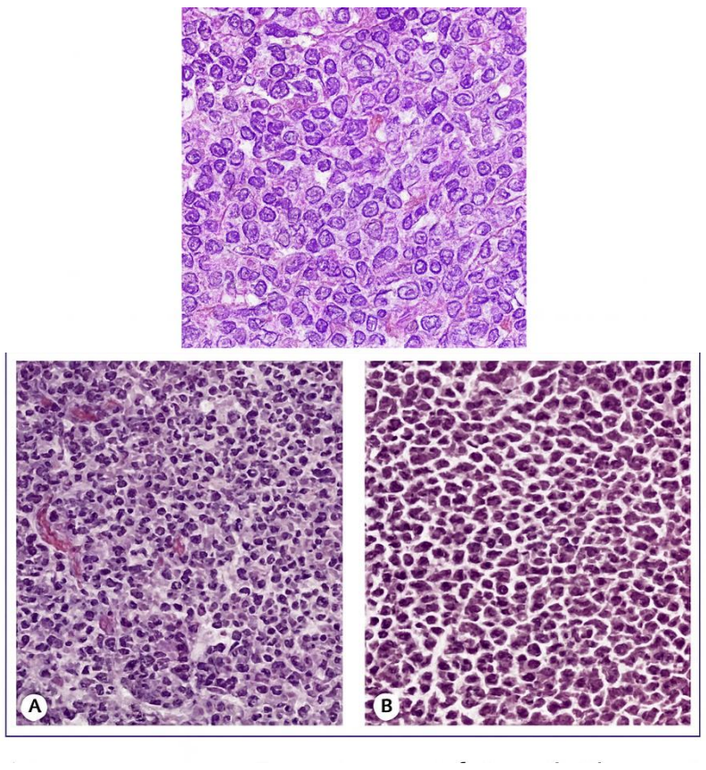

*Figure 1: Homogeneous Population of Small Cleaved Cells in Follicular Lymphoma **(B)** as opposed to Polymorphic Composition seen in Follicular Hyperplasia **(A)***

Methodology

- A systematic search was conducted in PubMed/MEDLINE, Embase, IEEE Xplore, Web of Science, and relevant preprint servers (as of *August 9, 2025*). Studies developing or validating machine-learning (ML) models for non-Hodgkin's lymphoma (NHL) diagnosis were included if they used histopathology or expert consensus as the reference standard.
- For each eligible study, key diagnostic metrics were independently extracted, including sensitivity, specificity, and area under the curve (AUC), along with details on patient population, study design, and data extraction method as shown in the Fig.1

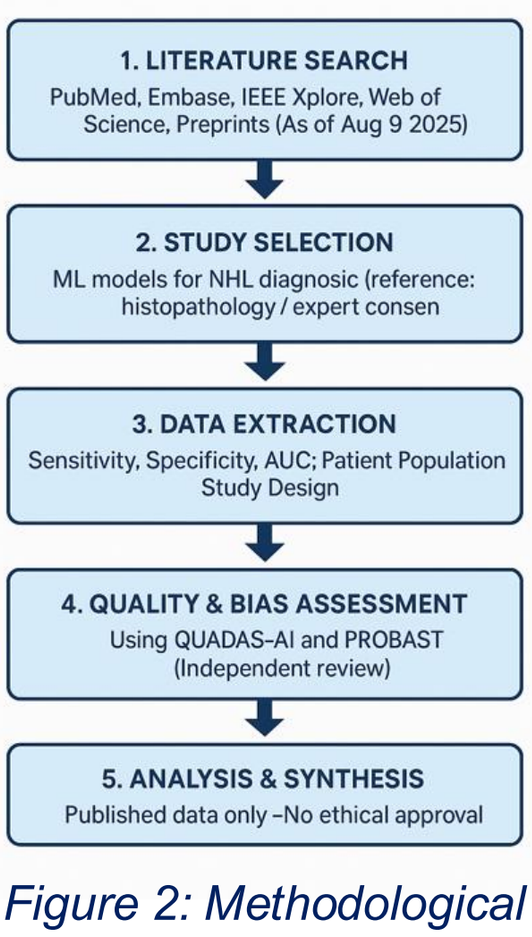

Figure 2: Methodological Framework

- Quality and risk-of-bias assessments were performed independently using QUADAS-AI and PROBAST frameworks. Because the review utilized only published data, ethical approval was not required.

Results

- A meta-analysis of 16 image-based studies showed pooled **sensitivity = 87% (95% CI: 83–91%)**, **specificity = 94% (95% CI: 92–96%)**, and **AUC = 0.97 (95% CI: 0.95–0.98),** indicating strong diagnostic accuracy.
- Deep-learning platforms for DLBCL diagnosis achieved site-specific accuracies near 100%.
- Flow cytometry ML models reached **92.7% accuracy**, with **88.5% sensitivity** and **98.8% specificity.**
- PET/CT radiomics models evaluating bone marrow involvement yielded an average **F1-score of 0.96**.

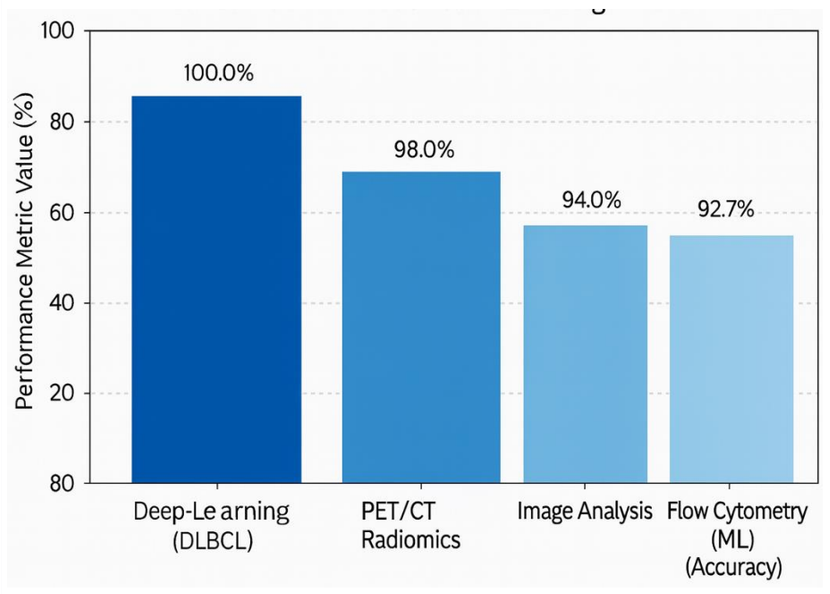

Figure 3: Performance Metrics of Machine Learning Models for NHL

- Most studies relied on internal validation, with external testing rarely performed, limiting generalizability.

Conclusion

ML approaches demonstrate high diagnostic accuracy for NHL, excelling in deep learning on histopathology and ML applied to flow cytometry and PET/CT data. However, limited external validation and methodological variability restrict clinical translation. Advancing the field requires multi-center prospective validation and standardized reporting.