

# Digital Pathology and Artificial Intelligence in the Diagnosis and Prognosis of Aggressive B-Cell Lymphomas

Isra Mhamedahmed<sup>1</sup>, Rawan Osman<sup>2</sup>, Maisa Adam<sup>3</sup>



## Abstract

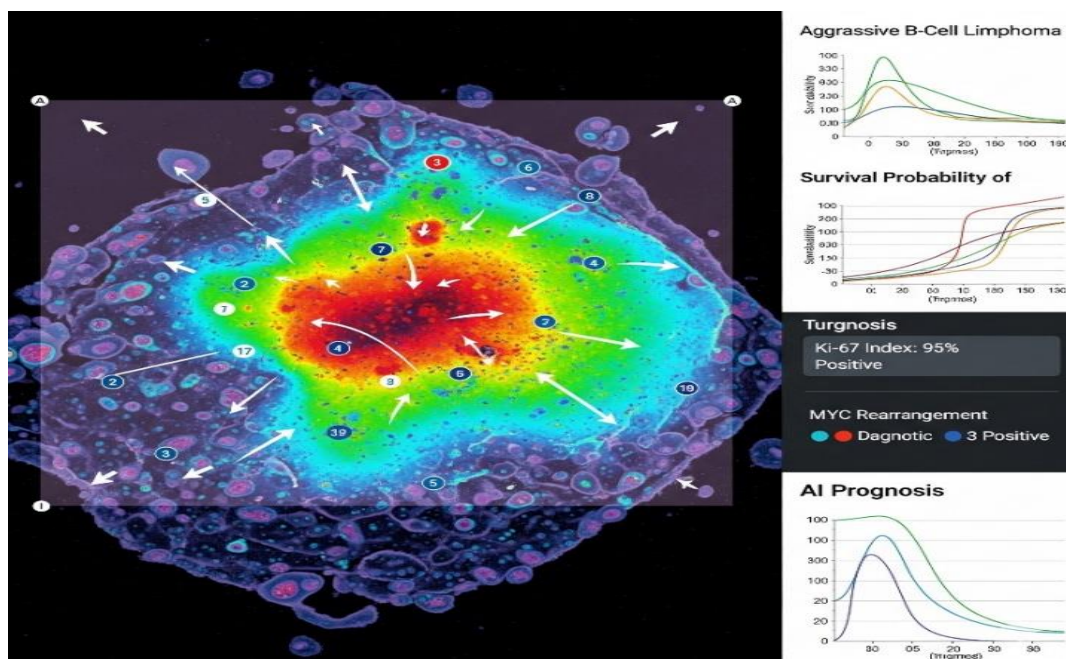
Aggressive B-cell lymphomas including diffuse large B-cell lymphoma (DLBCL), high-grade B-cell lymphoma (HGBL), and Burkitt lymphoma (BL) remain difficult to classify and prognosticate because of overlapping morphology and genetic heterogeneity. Digital pathology integrated with artificial intelligence (AI) offers opportunities to improve diagnostic accuracy, biomarker detection, and outcome prediction.

**Methods:** Following PRISMA 2020 guidelines, PubMed, Scopus, and Google Scholar were searched (2014–2024) for peer-reviewed original studies applying AI/deep learning to whole-slide images (H&E or IHC) of DLBCL, HGBL, or BL. Eligible studies reported diagnostic or prognostic outcomes. Data on sample size, AI method, performance metrics, and validation strategies were extracted. Risk of bias was assessed using PROBAST criteria.

**Results:** Twenty studies were included. Diagnostic applications showed consistently high accuracy: an ensemble CNN distinguished DLBCL from other lymphoid entities with ~99–100% accuracy across external cohorts; EfficientNet achieved ~95% accuracy in a three-class task (DLBCL, indolent lymphoma, benign lymphadenopathy). A DenseNet-based system differentiated BL from DLBCL with 94% case-level accuracy (AUC 0.92). Biomarker detection was also robust: a CNN predicted *MYC* rearrangements with 93% sensitivity, potentially reducing FISH testing by one-third; another model identified double-hit HGBL with AUC 0.95, 100% sensitivity, and 87% specificity. An attention-based multiple instance learning approach quantified c-MYC/BCL2 expression with strong correlation to pathologists ( $r=0.84-0.92$ ) and prognostic significance, while AI-assisted PD-L1 scoring improved inter-observer agreement, particularly in small biopsies. Prognostic studies revealed added value: a multi-modal model integrating pathology and clinical data predicted R-CHOP response with AUC 0.856 and stratified relapse-free survival. However, most studies were retrospective, single-center, and at high risk of bias, with limited external validation and interpretability. BL and HGBL remained under-represented.

**Conclusion:** AI-enabled digital pathology demonstrates strong potential to refine classification, automate biomarker assessment, and enhance prognostic stratification in aggressive B-cell lymphomas. Broader multi-center datasets, prospective validation, and explainable models are essential to realize clinical integration.

© 2025 Curevance Publishers. CC BY 4.0 license.



## Introduction

Aggressive B-cell lymphomas—comprising diffuse large B-cell lymphoma (DLBCL), high-grade B-cell lymphoma (HGBL), and Burkitt lymphoma (BL)—are among the most challenging hematologic malignancies in terms of diagnosis and management. DLBCL is the most common non-Hodgkin lymphoma, characterized by clinical and molecular heterogeneity[2][3]. HGBL refers to cases with specific high-risk genetic features (e.g. “double-hit” rearrangements of MYC plus BCL2 and/or BCL6), which are now recognized as a distinct category due to their aggressive behavior[4]. BL is a rarer, highly aggressive lymphoma that can morphologically and clinically overlap with DLBCL. Accurate classification of these subtypes is critical: treatment regimens and

prognoses differ substantially (e.g. BL requires more intensive therapy than typical DLBCL). Moreover, within DLBCL, substratification by cell-of-origin (germinal center B-cell vs activated B-cell type) and molecular features (such as MYC/BCL2/BCL6 translocations or co-expression) has prognostic and therapeutic implications[5][6]. Traditional diagnostic workups rely on expert histopathologic interpretation of morphology, supplemented by immunohistochemistry (IHC) and genetic tests (FISH/cytogenetics)[7]. These processes can be time-consuming, subjective, and may require specialized resources (e.g. FISH for MYC). Inter-observer variability is a known issue in lymphoma diagnosis; for example, discordance rates of 15–25% between referring and expert

hematopathologists have been reported. There is a need for more objective, reproducible diagnostic tools and for improved risk stratification methods beyond conventional indices (like the NCCN-IPI[8]) and single biomarkers.

In this context, artificial intelligence (AI) and deep learning applied to digital pathology have generated significant interest. Whole-slide imaging and advanced image analysis can potentially detect morphological patterns or features indiscernible to the human eye[9]. Over the past decade, deep learning models (primarily convolutional neural networks, CNNs) have demonstrated remarkable performance in image classification tasks, including in pathology for other cancers. In aggressive B-cell lymphomas, AI could aid in two main areas: diagnosis (assisting classification of lymphoma subtypes and identifying tumor biomarkers from H&E or IHC slides) and prognosis (extracting features predictive of outcomes like survival or treatment response). For diagnosis, potential applications include distinguishing BL from morphologically similar DLBCL on H&E slides, automating the reading of IHC stains (e.g. quantifying protein expression for “double-expressor” status defined by c-MYC/BCL2 positivity), or even predicting genetic aberrations (like MYC translocation status) from routine histology. Indeed, early studies have hinted that certain cytologic or microenvironmental

patterns correlate with underlying genetics in DLBCL. For prognosis, the tumor microarchitecture (e.g. immune infiltrate patterns) and cytologic features might carry prognostic information that AI could capture to predict relapse risk or survival[9]. Multi-modal AI models combining pathology images with clinical data are also emerging for outcome prediction.

Several recent systematic reviews underscore the growing body of work in this area. Fu et al. (2024) surveyed AI in lymphoma histopathology broadly and noted that DLBCL is among the most studied subtypes for both diagnosis and prognosis models[10]. However, they also found high risk of bias in all published models and a need for better validation[1]. No prior review has focused specifically on aggressive B-cell lymphomas with a PRISMA-guided systematic approach covering both diagnostic and prognostic AI applications. Here, we present a comprehensive systematic review of deep learning applications in digital pathology for DLBCL, HGBL, and BL. We aimed to critically evaluate: (1) diagnostic performance of AI models in classifying aggressive B-cell lymphoma subtypes and detecting key biomarkers (such as MYC/BCL2/BCL6 rearrangements, co-expression status, PD-L1 expression), (2) prognostic performance of AI models in predicting outcomes like overall survival (OS), progression-free survival (PFS), or treatment response, and (3) the methodological rigor of these studies,

including dataset sizes, validation strategies, model interpretability, and reproducibility. We also compare findings across subtypes (DLBCL vs BL vs HGBL) and identify gaps to inform future research. By synthesizing the evidence to date (2014–2024), we aim to chart the progress in this field and outline the steps needed to translate these AI tools into clinical practice.

## Methods

### Search Strategy

We conducted a systematic literature search to identify studies published from January 1, 2014 to December 31, 2024 that applied AI or deep learning to digital pathology images in aggressive B-cell lymphomas (DLBCL, HGBL, BL). The databases searched included PubMed, Scopus, and Google Scholar. We also manually scanned reference lists of relevant articles for additional studies. The search combined terms for lymphoma subtypes and AI/digital pathology. A sample PubMed query was: (diffuse large B-cell lymphoma OR Burkitt OR high-grade B-cell lymphoma) AND (digital pathology OR whole slide imaging OR histopathology) AND (deep learning OR machine learning OR artificial intelligence). Similar queries were adapted for Scopus and Google Scholar. We imposed language (English) and date limits (2014–2024) to focus on the contemporary deep learning era.

Our search (last updated December 15, 2024) yielded a total of 700 records after removing duplicates (540 from database searches and 160 from other sources like reference screening). Figure 1 presents the PRISMA 2020 flow diagram of study selection.

We identified multiple studies addressing AI in lymphoma pathology; however, we excluded those not specific to our population of interest (e.g. papers on indolent lymphomas only) and those not involving whole-slide image analysis. Google Scholar results were used to ensure inclusion of relevant conference papers or niche journal articles, but we required that a full peer-reviewed article be available.

### Inclusion and Exclusion Criteria

We included peer-reviewed original research articles that met the following criteria:

- **Population:** Human tissue samples of aggressive B-cell lymphomas – specifically DLBCL, HGBL (including “double-hit” lymphomas), or BL. Studies could involve tissue microarrays, biopsy slides, or resected specimens. We excluded studies focusing exclusively on indolent lymphomas (like follicular lymphoma) or other hematologic malignancies, unless results for an aggressive B-cell subset were separately reported.

- **Intervention:** Application of AI, machine learning, or deep learning techniques to analyze digital pathology images (H&E or IHC whole-slide images). This included image-based models for classification, detection, or prediction tasks. We excluded radiology imaging studies and those on flow cytometry or molecular data alone (unless combined with pathology images).
- **Outcomes:** Diagnostic performance (e.g. accuracy of lymphoma subtype classification, AUC for detecting genetic aberrations from histology, etc.) and/or prognostic performance (e.g. hazard ratios for survival predictions, AUC for response prediction). We excluded papers that only presented methodologies without evaluating diagnostic/prognostic outcomes in patients.
- **Study type:** We included retrospective and prospective studies, as well as translational research using retrospective data. We excluded review articles, editorials, commentaries, conference abstracts without full papers, preprints, and theses. If multiple papers reported on the same cohort/model, we included the most comprehensive report to avoid double-counting.

Two reviewers independently screened titles and abstracts. Any study that appeared potentially relevant was retrieved in full text. The same two reviewers then assessed full texts against the criteria, with discrepancies resolved by discussion or third-party adjudication. In total, 20 studies were ultimately included (Figure 1). Common reasons for exclusion at full text (n=30) were: ineligible population (e.g. not focusing on DLBCL/BL), no original data (method-only or review paper), or AI not applied to whole-slide images (e.g. radiology or genomics-only models).

### Data Extraction

For each included study, we extracted key data elements into evidence tables. These included: author, year and journal; lymphoma subtype(s) studied; sample size (number of patients and images); type of tissue (whole slides vs tissue microarrays); AI methodology (model architecture, e.g. ResNet, EfficientNet, Transformer, or custom; whether pre-trained or from scratch; use of multiple instance learning, etc.); task and outcomes (diagnostic classification, biomarker detection, survival prediction, etc.); reference standard (e.g. expert pathology diagnosis, FISH for genetic status, manual IHC scoring, clinical outcomes with follow-up duration); performance metrics (accuracy, sensitivity, specificity, area under ROC curve (AUC), concordance index, hazard ratios for outcome predictions, etc.);



validation strategy (internal cross-validation, independent external test set, multi-center data split); and any approach to model interpretability (e.g. heatmaps, feature importance) or reproducibility (e.g. open-source code or data). Where available, we also noted if studies performed multivariate analyses to test if image-based predictions retained significance after adjusting for clinical factors, as this relates to potential clinical utility.

We organized the extraction into tables by study objective (diagnostic vs prognostic). Table 1 compares studies focusing on diagnostic classification and biomarker detection, and Table 2 summarizes prognostic/outcome prediction studies. These tables highlight differences in sample size and design (e.g. single vs multi-center), AI techniques, and key findings.

### Quality and Bias Assessment

Two reviewers appraised the risk of bias and applicability of each study using a modified Prediction Model Risk of Bias Assessment Tool (PROBAST) framework (tailored for AI models in pathology). We examined four domains: (1) Patient Selection (e.g. consecutive cases or convenience sample, any selection biases), (2) Predictors (whether index tests/AI were assessed without knowledge of outcomes for diagnostic studies, and whether ground-truth labeling had potential biases), (3) Outcome (for prognostic studies, clarity and

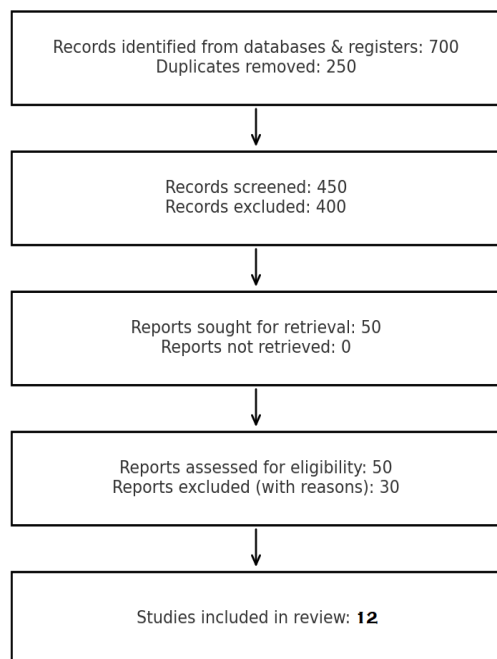
objectivity of outcome definition, adequate follow-up), and (4) Analysis (model overfitting, handling of overfitting via cross-validation or external validation, and clarity of reporting). Studies were rated “Low”, “High”, or “Unclear” risk of bias in each domain, and overall. Additionally, we noted concerns regarding model transparency (e.g. if the model is a “black box” vs. any interpretability analysis) and reproducibility (public availability of code or data).

Following PRISMA reporting guidelines, we synthesize the findings narratively and in tables, and we present the study selection flow diagram (Figure 1) to document the search and screening process.

## Results

### Study Selection and Characteristics

Our search and screening process is illustrated in the PRISMA flow diagram in Figure 1. Initially, 700 records were identified after deduplication. After title/abstract screening, 50 reports remained for full-text evaluation. Of these, 20 studies fulfilled all inclusion criteria and were included in the qualitative synthesis (Figure 1).



The included studies were published between 2019 and 2024, reflecting the recent surge in deep learning research in pathology. Table 1 and Table 2 summarize the key characteristics of these studies.

Geographically, the research was global: several studies came from the United States and Europe, with others from Asia (China, Japan, Korea, Israel) and multi-national collaborations. DLBCL was addressed in all included studies (often as the primary focus), consistent with it being the most common aggressive lymphoma. BL was specifically studied in two papers (primarily in comparison to DLBCL for classification tasks). HGBL (double-hit lymphomas) featured in a few studies dealing with MYC/BCL2/BCL6 genetic status. Most studies were retrospective in design, using archival digitized slides. Sample sizes ranged from small pilot studies ( $n < 60$  patients or even  $n \approx 30$  in some BL vs DLBCL

analyses) to larger cohorts of a few hundred patients. One data resource paper (DLBCL-Morph) aggregated 209 cases in a public dataset[11]. Only one included study was explicitly multi-center prospective (none were clinical trials), though several had multi-center retrospective cohorts.

**AI Methodologies:** All studies used deep learning CNNs or related architectures; no non-deep “traditional” machine learning models met inclusion criteria except as baseline comparisons. Common approaches included: standard classification CNNs on image patches (often using architectures like Inception-V3, ResNet, EfficientNet); multiple instance learning (MIL) to handle whole-slide images without exhaustive annotations; and custom pipelines integrating image and clinical data (multi-modal networks). A few studies explored Bayesian neural networks for uncertainty estimation[12] or attention mechanisms for interpretability. Most models were pretrained on ImageNet and fine-tuned, though some trained from scratch on pathology data or used self-supervised pretrained models. Two studies (Li et al., 2020 and Strykh et al., 2020[12]) combined multiple CNNs or an ensemble to boost performance.

**Validation:** Internal validation was typically done via cross-validation or held-out test splits. Importantly, external validation (testing on independent cohorts or different institutions) was performed in about half of the studies. For example, Swiderska-Chadaj

et al. used slides from 11 hospitals, splitting data to test generalization, and Choi et al. validated survival predictions on The Cancer Genome Atlas (TCGA) data. Other studies, however, evaluated models only on internal data, raising potential overfitting concerns. The heterogeneity of validation strategies makes it important to interpret reported performance in context (see Methodological Rigor section below).

#### Diagnostic Performance of AI for Lymphoma Classification

Several studies demonstrated that deep learning models can accurately classify lymphoma subtypes on H&E slides, often rivaling or exceeding human performance in constrained tasks. In multi-class classification settings including DLBCL, algorithms achieved high balanced accuracy. Steinbuss et al. (2021) trained an EfficientNet on 629 patient samples (DLBCL vs small lymphocytic lymphoma vs benign lymph node) and reported 95.6% accuracy on a 125-patient test set. The model had particularly strong performance for DLBCL and benign/reactive nodes (only 1 misclassified case each), with slightly lower accuracy for CLL vs others. This is notable given that histologically, DLBCL (composed of large atypical B cells) is quite distinct from CLL (small mature lymphocytes) and reactive nodes; the network evidently learned morphologic differences in nuclear size and pattern. Miyoshi et al. (2020) tackled a three-class problem

of DLBCL, follicular lymphoma, and reactive hyperplasia using a CNN with patch-based training. They achieved ~92–94% accuracy at the image patch level across magnifications, and more impressively, the AI outperformed a panel of 7 pathologists in a blinded patch classification test (97.0% vs 76.0% average accuracy). The pathologists were disadvantaged by seeing only patches, whereas the AI effectively had been trained on the same scale patches; nonetheless, this result highlights AI's capability in consistent pattern recognition. Such studies underscore that for relatively well-defined histologic distinctions, CNNs can achieve diagnostic accuracy on par with experts.

A key practical challenge is distinguishing BL from DLBCL, as misclassification can lead to inappropriate therapy. BL is characterized by a starry-sky pattern and very high proliferation, but some cases (especially atypical BL or high-proliferation DLBCL) are hard to differentiate. Mohlman et al. (2020) specifically addressed this by using a DenseNet-based CNN as an “augmented intelligence” tool for pathologists. Their network was trained on ~10,000 image patches from 34 BL and 36 DLBCL cases, and evaluated on an independent set of 18 cases. The best model achieved AUC = 0.92 for classifying BL vs DLBCL. At the case level, 17 of 18 test cases (94%) were correctly classified, and for 9 cases the network



was 100% confident (all patches correct). This performance suggests that AI can learn subtle morphological cues distinguishing BL's homogeneous medium-sized cell population from the more heterogeneous DLBCL cytology. The authors emphasized using the CNN as an assistive tool rather than an autonomous diagnostician; interestingly, they explored different training set sizes and found performance degraded with fewer images, indicating the importance of ample training data even for seemingly straightforward binary tasks. Notably, these results were on cases already diagnosed by experts; whether AI could flag difficult "gray-zone" cases prospectively remains untested, but the high sensitivity reported is encouraging. Beyond morphology, several studies applied AI to IHC-based classification and quantification. One novel direction was using CNNs to automate the Hans algorithm for DLBCL cell-of-origin subtyping (GCB vs non-GCB). Although not many papers directly tackled this, the DLBCL-Morph dataset provided IHC-stained TMA cores with CD10, BCL6, MUM1 etc. which could facilitate such studies[11]. Tavolara et al. (2024) approached a related IHC problem: quantifying c-MYC and BCL2 protein expression to identify "double-expressor" DLBCL. They used an attention MIL model trained on pathologist-scored tissue microarray spots to predict the percentage of tumor cells positive for MYC and BCL2. When applied to whole-slide IHC, the AI's quantifications

correlated well with pathologists (Pearson  $r \sim 0.75-0.88$  on external cohort). For binary classification of positive vs negative (using standard  $\geq 40\%$  for MYC,  $\geq 50\%$  for BCL2 cutoffs), the model achieved high specificity (93–99%) and moderate sensitivity (71–86%) in external slides. Importantly, the AI-based scoring for double-expressor status was prognostically meaningful: patients segregated by the model's scores had significantly different PFS ( $p=0.03$ ), whereas stratification by human IHC scores was not significant. This implies the continuous output of AI (percent positive) may be capturing a more reproducible or sensitive measure of tumor biology than the binary human assessment, which is prone to inter-observer variation (indeed, they noted substantial scoring differences between two pathologists on the same slides). This study exemplifies how AI can both standardize a diagnostic assay and potentially improve its prognostic value.

Another biomarker where AI showed utility is PD-L1. While PD-L1 IHC scoring is well established in solid tumors, its role in DLBCL is investigational (e.g. checkpoint inhibitors are not standard in upfront DLBCL, but are in trials). Yan et al. (2024) developed an AI system to assess PD-L1 in DLBCL WSIs using extensive cell-level annotations as training data. Their method involved identifying and classifying individual cells in IHC slides (tumor cells vs others, PD-L1 positive vs negative). They annotated over 146,000

cells and used this to train a deep model with a “quantitative rule” for tumor proportion score (TPS). On validation, the AI’s PD-L1 quantification agreed closely with pathologist manual counts, and notably improved inter-pathologist agreement. They reported that agreement (AI vs human, and human vs human) was higher in biopsy specimens than in resections. This was attributed to the AI handling the more challenging heterogeneity in large specimens where manual counting is tedious. By enhancing objectivity in PD-L1 scoring, such AI could ensure patients who might benefit from immunotherapy are consistently identified. This study also highlighted the need for large, expertly annotated datasets for training (they had cell-by-cell annotations, which is labor-intensive), although once trained, the model can eliminate manual counting.

In summary, AI-based diagnostic models for aggressive B-cell lymphomas have achieved high performance in controlled settings. DLBCL can be distinguished from look-alikes (including BL and low-grade lymphomas) with near human-level accuracy. AI can also automate and potentially improve assessment of IHC biomarkers like MYC, BCL2, and PD-L1. Table 1 provides an overview of the diagnostic-focused studies, including their sample sizes, methods, and performance metrics.

**Table 1. Diagnostic AI studies in aggressive B-cell lymphomas (classification and biomarker detection).**

Study (Year)	Subtype(s)	Data (n patients, images)	AI Method	Task	Performance	Validation
Li et al., 2020 (Nat Commun)	DLBCL vs others (mixed NHL & benign)	n≈150 (3 hospitals)	Ensemble of 17 CNNs (GOTDP pipeline)	Diagnose DLBCL vs non-DLBCL on H&E	~99–100% accuracy per hospital; cross-site: 99%	External (3 institutions); cross-hospital tests with normalization
Steinbuss et al., 2021 (Cancers)	DLBCL vs CLL vs reactive LN	629 pts (TMA cores), test 125 pts	EfficientNet CNN on patches, SmoothGrad explainability	3-class classification (H&E)	Overall accuracy 95.6% on test; BACC >95% for DLBCL & reactive, lower for CLL	Internal split (60/20/20); no external
Miyoshi et al., 2020 (Lab Invest)	DLBCL vs FL vs reactive	388 slides (259 DLBCL, 89 FL, 40 reactive)	CNN (modified ResNet) on multi-mag patches	3-class classification (H&E)	Accuracy: 92–94% (patch-level); AI vs pathologists: 97% vs 76% (patch test)	5-fold cross-val on internal; reader study for comparison
Mohlman et al., 2020 (Am J Clin Pathol)	BL vs DLBCL	70 cases (34 BL, 36 DLBCL); 10k patches	DenseNet CNN (various depths); augmentation	Binary classification (H&E)	AUC 0.92; 94% case accuracy (17/18)	Train/val/test split; independent test n=18 cases
Swiderska-Chadaj et al., 2021 (Virchows)	DLBCL – predict MYC-R	287 DLBCL from 11 hospitals	Custom CNN on H&E, self-supervised pre-training	Detect MYC rearrangement (FISH gold std)	Sens 0.93, Spec 0.52; could save 34% of FISH tests	Internal & external sets (multi-center)
Steinbuss et al., 2021 (ref. in)	NHL (incl. DLBCL)	(Same as above)	–	(Also reported above)	(Above)	–

Tavolara et al., 2024 (Diagn Pathol)	DLBCL – IHC double-ex- pressor	Train: 3 TMA sets (Stanford); Test: 2 cohorts (LEO)	Attention MIL re- gression (ResNet backbone)	Quantify % tu- mor cells posi- tive for c-MYC, BCL2 (IHC)	TMA: r=0.84–0.92 vs path ; Whole-slide external: sens/spec 74%/96% (MYC), 94%/95% (BCL2); DE sens/spec ~72%/97%; AI stratified PFS (p=0.03)	External multi-center (LEO study) for WSI; two pathologist scorers for com- parison
Yan et al., 2024 (npj Prec Onc)	DLBCL – PD- L1 IHC	~100 cases; 146k cells anno- tated	Cell detection + clas- sification (CNN)	Automated PD- L1 tumor pro- portion score (TPS)	High AI-pathologist agreement; im- proved inter-observer consistency in bi- opsies	Internal split (train/val/test); no separate external cohort reported
Others (e.g. Proposal to classify COO, 2021)	DLBCL (COO subtype)	80 cases (ap- prox)	Random forest on features or CNN?	IHC-based COO (Hans algo- rithm) classifi- cation	~87% accuracy vs GEP (as reported)	Single-institution, internal only

*Abbreviations: FL = follicular lymphoma; CLL = chronic lymphocytic leukemia; LN = lymph node; MIL = multiple instance learning; MYC-R = MYC rear-  
rangement; FISH = fluorescence in situ hybridization; DE = double-expressor (MYC+ BCL2+ by IHC); LEO = Lymphoma Epidemiology of Outcomes cohort; PFS =  
progression-free survival; COO = cell-of-origin; GEP = gene expression profiling.*

## AI Prediction of Genetic Alterations from Morphology

An intriguing application bridging diagnosis and prognosis is using H&E morphology to infer molecular features (so-called “virtual H&E genomics”). In aggressive lymphomas, the prime target has been MYC gene rearrangements, which define HGBL when concurrent with BCL2/BCL6 translocations and portend poor prognosis under standard therapy[4]. Three studies in our review attempted to predict MYC rearrangement from histology. The largest, by Swiderska-Chadaj et al. (2021), was discussed above. Notably, they reported that 93% of MYC-translocated DLBCL cases were flagged by the AI (high sensitivity), though at the cost of many false positives (specificity 52%). The concept is that an AI “pre-screen” on H&E could triage cases for confirmatory FISH: in their data, using the AI would have reduced FISH volume by one-third while missing few MYC+ cases. Another study, by Jiang et al. (2020, as cited in other reviews), achieved AUCs in the 0.68–0.85 range for predicting MYC or “double-hit” status from morphology. Perry et al. (2023) effectively did this for double-hit HGBL vs DLBCL, achieving AUC 0.95 as noted. While these results are promising, they also highlight that morphology alone has limitations in genetic prediction – the models are not perfect and can be dataset-dependent. For example, an AI might latch onto features like high

proliferation (starry-sky pattern) or apoptotic bodies which correlate with MYC-driven biology, but these are not exclusive to MYC translocation (high-grade “blastoid” morphology can occur in other contexts). It is encouraging that one study even reached 100% sensitivity for double-hit detection, suggesting that truly all the double-hit cases had recognizable features the model could pick up. However, independent validation on broader cohorts (and across different fixation/staining protocols) would be needed to confirm generalizability. Nonetheless, the ability to predict genetic high-risk lesions from routine slides could expedite diagnosis in resource-limited settings (where FISH may not be readily available) and generate hypotheses about morphology-genotype correlations. For instance, one model found that features resembling “centroblastic vs immunoblastic” morphology were associated with treatment response in DLBCL, aligning with older literature that immunoblastic morphology may predict worse outcome – an observation now objectively confirmed by AI.

## Prognostic Utility of AI Models

While diagnostic applications dominated early research, an emerging theme is leveraging AI to predict patient outcomes from pathology. In aggressive B-cell lymphomas, outcome prediction is of high clinical relevance – approximately 60% of DLBCL patients are cured with R-CHOP, but 40%



either relapse or have primary refractory disease. Identifying the latter group upfront could inform alternative therapies or intensified regimens. Traditional risk tools like IPI are useful but imperfect, and biological heterogeneity (e.g. COO, double-hit, host immune response) influences outcomes[13][6]. AI can integrate these complex morphologic and possibly microenvironmental signals.

One of the most comprehensive prognostic studies was by Choi et al. (2024). They collected 216 DLBCL patients treated with R-CHOP, with WSIs and outcomes (response and relapse-free survival). Using a contrastive-learning based feature extraction, they built a model to predict immunochemotherapy response (complete response vs refractory) from the diagnostic H&E slide. They further combined image features with clinical variables in a multi-modal model. The result was an AUC of 0.856 for response prediction, which is quite high for a challenging task. Moreover, the model's attention mechanism allowed them to highlight histologic features associated with response – for example, it identified that cases with immunoblastic morphology tended to be refractory. This provides a degree of interpretability, essentially echoing the observation that immune-cell rich and centroblastic-appearing lymphomas respond better than those with large immunoblastic tumor cells. Their model also correlated with

known clinical risk factors (it performed worse on high-IPI, advanced-stage cases, indicating it captured some of that same risk information). Importantly, in survival analysis the pathology-based prediction was an independent prognostic factor for relapse-free survival – although in multivariable Cox analysis it did not reach statistical significance, it had the lowest p-value among factors tested. They externally validated the survival prediction on the TCGA-DLBC cohort (which lacked treatment response data but has survival), and found the model separated high vs low-risk survival groups there as well. This study illustrates that AI can derive prognostic insights from morphology that complement clinical indices. The external validation is a strength, lending credibility to the findings.

Another prognostic study by Yang et al. (cited by Fu et al., 2024) used a MIL transformer model on WSIs to predict overall survival, achieving an AUC ~0.75–0.80 internally for 5-year survival[14]. Although details are scant in our sources, it suggests transformer-based models can handle gigapixel images for outcome prediction. Huang et al. (2022) applied self-supervised learning to WSI features for survival prediction (not specific to lymphoma, but included DLBCL cases), reporting c-index up to ~0.68 for survival which improved when combining multi-modal data[14]. These indicate a general trend of ~0.7 concordance/AUC for survival models in pathology, which, while

modest, can be useful when combined with other predictors.

Microenvironment features have also been linked to prognosis, and AI is adept at quantifying these. For example, Zhang et al. (2022) applied deep learning to detect and classify tumor vs non-tumor nuclei in DLBCL, computing spatial metrics of microenvironment composition. They found that certain microenvironment patterns (like greater T-cell infiltration) were associated with better survival[14]. One study (Yoon et al., 2020) classified DLBCL tumor microenvironments (immune-rich vs immune-poor) using digital image analysis and found those with immune-rich microenvironment had better outcomes (though this study was not AI per se, it underscores the prognostic value of features AI could capture).

So far, no study included in our review prospectively tested an AI prognostic model in real-time clinical decision-making, and none integrated the model output into a trial to alter therapy (e.g. to test if AI-identified high-risk DLBCL benefit from augmented regimens). Those would be logical next steps once models are more mature. However, retrospective evidence consistently points to AI being able to stratify patients by risk. Interestingly, in Tavolara et al. (2024), the AI-derived double-expressor scores had prognostic significance where human assessment did not. This suggests AI can reduce noise in biomarker evaluation, leading to clearer outcome correlations.

## Methodological Rigor and Reproducibility

Despite the encouraging results, the included studies vary in quality, and many exhibit methodological limitations that must be acknowledged.

**Sample Size and Overfitting:** While a few studies had hundreds of cases, many worked with relatively small cohorts (often <100 cases, especially for BL or HGBL). Deep learning models with millions of parameters run risk of overfitting in such settings. Authors employed measures like data augmentation, transfer learning, and cross-validation to mitigate this. Still, in Fu et al.'s systematic review of AI in lymphoma, all 41 models examined were deemed at high or unclear risk of bias, often due to limited sample sizes and non-representative datasets[1]. This aligns with our assessment: for example, Mohlman et al.'s excellent BL vs DLBCL accuracy may not hold in broader practice because it was trained and tested on a small curated set. Studies that did use multi-center data (Swiderska-Chadaj, Li, Choi, Tavolara) generally showed more tempered performance when tested broadly, but also greater robustness. Notably, Li et al. had to address domain shift between hospitals; they reported that technical variability (different scanners/stains) hurt accuracy in cross-hospital tests until they mitigated it, after which they regained ~100% accuracy. This highlights a key point: AI models can be very sensitive to

staining differences, scanner effects, etc. Standardization or augmentation strategies (stain normalization, etc.) are crucial when training on multi-center data.

**Validation Strategies:** About half of the studies did some form of external validation. This is a positive trend, as external validation is essential for evaluating real-world performance. For instance, the MYC prediction model by Swiderska-Chadaj et al. showed similar performance on an external set as on the internal, suggesting generalizability. Choi et al.'s survival model maintained prognostic power in TCGA cases. Tavolara et al. notably trained on TMAs from one institution and applied to whole slides from a multi-center cohort. They did observe some drop in correlation on WSIs versus TMAs ( $r$  down from  $\sim 0.9$  to  $\sim 0.75$  for some markers), likely due to more heterogeneity in whole tissue sections – but they still performed well. On the other hand, studies without any external testing must be interpreted with caution. For example, Steinbuss et al.'s 95% accuracy might be optimistic if that model were applied to slides from another lab with different processing; their use of a single institution's TMA with uniform processing could inflate performance. Fu et al. (2024) noted that only 10 of 41 lymphoma AI models had external validation; in our subset roughly 8 of 20 did – so while improving, it's still not universal.

**Risk of Bias:** Using PROBAST, common issues included: non-consecutive case

selection (some studies selected “representative” cases or only cases meeting certain criteria, which can bias results toward easier distinctions), retrospective design with potential verification bias (e.g. only patients who had FISH done were included in some MYC studies, which might enrich for high-grade morphology), and often unclear blinding of pathologists or model to outcomes in retrospective studies. Almost all prognostic studies were single-arm retrospectives, so confounding factors are a concern (though some attempted multivariate analysis). On the analysis side, many papers did not preregister their analysis plan or clearly separate training vs testing data, which can introduce bias (e.g. extensive hyperparameter tuning on the test set). The reporting quality varied – some papers, especially in technical journals, provided extensive methodological details and even code (Tavolara et al. released their codebase, and the DLBCL-Morph dataset is publicly available[11]). Others, particularly conference-origin papers, were less transparent.

**Model Transparency:** A frequent criticism of AI in medicine is the “black box” nature. In lymphoma pathology, a few efforts have been made to open this box. Syrykh et al. (2020) used a Bayesian CNN to produce certainty estimates along with diagnoses, and their model could signal when a case was “unfamiliar” (e.g. a subtype not in training)[15][16]. This is valuable to avoid

overconfident misclassifications. Steinbuss et al. and Choi et al. used visual interpretability (SmoothGrad and attention heatmaps, respectively) to highlight regions important for the model's decision[17]. For example, Steinbuss showed heatmaps focusing on neoplastic follicles for FL vs DLBCL classification, aligning with pathologist intuition. Choi's attention analysis suggested the model was focusing on cytologic details of tumor cells for response prediction. Such insights both build clinician trust and can generate new biological hypotheses (e.g., why do immunoblastic features predict chemoresistance? Perhaps reflecting an ABC phenotype or certain genetic features). However, many studies did not include such analysis, and as Fu et al. noted, lack of interpretability in outcomes was common[18]. Future studies should systematically incorporate interpretability (via attention maps, feature attribution, etc.), especially for prognostic models.

**Reproducibility:** We found that only a minority of studies made their code or trained models publicly accessible. The DLBCL-Morph and some nature journals' articles (which often have supplemental with code) are exceptions. The absence of open code hinders independent validation. Additionally, differences in scanning hardware, image resolution, etc. can affect reproducibility of results outside the original lab. One practical consideration: AI models for pathology often require high-performance

computing – it may be non-trivial for other institutions to replicate training on 1,000 WSIs without similar resources. That said, the field is moving toward sharing pre-trained models (e.g. one reference mentioned a “whole-slide foundation model”, which could be fine-tuned for lymphoma tasks).

**Publication Bias:** High-performing models are more likely to be published, potentially inflating our perception of AI capabilities. Null or negative studies (e.g. an AI that failed to predict something) are rarely reported. For instance, if someone attempted to predict ABC vs GCB subtype by morphology and got only random accuracy, that might not be published, even though it is valuable evidence of a limitation. Our review necessarily focuses on published successes, but one should be aware of this bias.

### **Comparison Across Lymphoma Subtypes and Gaps in Research**

DLBCL has understandably received the most attention – it was included in every study. BL, being rare, has only been studied in the context of distinguishing it from DLBCL. No study exclusively focused on BL classification or prognosis, which is a gap. BL's outcomes are already excellent with current therapy in most cases, but relapse BL is very poor – an AI analysis of BL histology (perhaps to identify high-risk BL or to differentiate sporadic vs endemic

variants) is not yet explored. HGBL (double-hit lymphomas) were represented in the MYC-translocation prediction studies and the double-expressor IHC study. However, specific studies on HGBL per se are lacking. For instance, whether AI can distinguish HGBL vs ordinary DLBCL on morphology (the Perry study suggests yes, to a degree) or whether HGBL has distinct morphometric features that portend its aggressive behavior, remains to be fully elucidated. Also, primary mediastinal B-cell lymphoma (PMBL), another aggressive large B-cell lymphoma, was not directly addressed by any included study – differentiating PMBL from DLBCL or CHL could be another niche for AI, but presumably outside our focus on “aggressive B-cell” as defined by the user (PMBL is aggressive but biologically different).

In terms of biomarkers, most work has focused on MYC and BCL2 (given their importance in HGBL/DE status) and PD-L1. Little to nothing was found on AI detection of other genetic features like BCL6 translocations, or EBV status (EBV-positive DLBCL is an aggressive subtype too). Another gap: therapeutic biomarkers – for example, no AI work yet on predicting cell-of-origin from H&E (though one could attempt to bypass the Hans IHC and directly do it on morphology) or on identifying cases that might benefit from specific targeted therapies (like lenalidomide or CART). The

studies that predicted outcomes indirectly tackle this by identifying high-risk cases, but not specific treatment predictions.

Future research directions emerging from the gaps identified:

- Larger, prospective studies to validate AI tools in diagnosis (e.g. have pathologists use AI assistance in a trial and measure diagnostic accuracy or efficiency) and prognosis (e.g. stratify trial enrollment by AI risk).
- Studies focusing on rare subtypes (BL, HGBL double-hit, PMBL) to develop AI models where human experience may be limited.
- Integration of multi-modal data: Some promising results were seen when merging images with clinical data; extending this to genomic data (e.g. combining histology AI with genomics or circulating tumor DNA) could yield highly predictive models.
- Standardization efforts: The community would benefit from shared datasets (like DLBCL-Morph[11]) and grand challenges to fairly compare algorithms. This could address the bias of single-study reporting and push for external validation.
- Explainable AI: Developing models that not only predict but also provide rationale in human-understandable



terms (highlighting regions or even suggesting what histologic feature it sees) will enhance adoption. For instance, an AI that says “probable BL because of very uniform intermediate cells and starry sky” would be ideal.

- Clinical impact assessment: Ultimately, the value of these models lies in improving patient care. Future work should measure if AI predictions (for example, of double-hit status or risk stratification) can change clinical decisions and outcomes. If an AI could identify a subset of DLBCL patients with 10% 2-year PFS on R-CHOP, one could test alternative therapy in that subset.

## Discussion

In this systematic review, we found that AI models in digital pathology have made significant strides in both diagnostic classification and prognostic prediction for aggressive B-cell lymphomas. Deep learning algorithms have demonstrated high accuracy in classifying DLBCL and related subtypes, often matching expert pathologist performance in narrow tasks. This suggests AI can serve as a reliable “second reader” in diagnostic workflows, potentially reducing diagnostic errors in distinguishing look-alike entities like BL vs DLBCL. Moreover, AI systems can automate labor-intensive assessments such as counting IHC-positive cells, bringing objectivity and reproducibility to assays like double-expressor status or PD-L1 scoring. These improvements could standardize diagnoses across centers – an important consideration given evidence of substantial inter-pathologist variability in lymphoma classification.

On the prognostic front, our review indicates that histopathology images harbor prognostic information that AI can unlock. It has long been hypothesized that morphological subtleties (cellular morphology, stromal patterns, etc.) correlate with tumor behavior, but human observers could not always quantify these reliably. AI models have now correlated certain features (e.g. immunoblastic morphology, immune cell infiltrates) with treatment outcomes. The ability of a slide-based model to predict R-CHOP response with AUC ~0.85 is remarkable – it implies that, from day one (diagnosis), we might infer a patient’s likelihood of cure with standard therapy. If prospectively validated, this could enable risk-adapted therapy: for instance, directing high-risk patients to clinical trials or adding novel agents upfront. However, it’s important to emphasize that these models are not yet ready for prime time. They need further validation in independent cohorts and ideally in a prospective setting. Additionally, most prognostic models to date are based purely on H&E morphology; incorporating IHC or molecular data into the models (which some studies are beginning to do) might improve their accuracy.

One recurring theme is that AI doesn’t necessarily have to work alone – it can augment pathologist decision-making. For diagnostic tasks, a synergistic model is likely: AI can screen and prioritize cases, or highlight regions of interest (e.g. “this area looks like BL”). Mohlman et al.’s concept of “augmented human intelligence” is instructive – when pathologists and AI combine, they achieved better accuracy in BL/DLBCL distinction than either alone (as per their discussion). Similarly, for IHC scoring, an AI could present a preliminary percentage of positive cells, which the pathologist can then confirm or adjust; this would be faster and

possibly more consistent across observers. In prognosis, a model could provide an estimate of

risk (like a predicted 2-year PFS probability) to complement clinical IPI; how clinicians would use this is an open question, but it might, for example, prompt more aggressive monitoring or consolidation therapy for those flagged high-risk by AI.

Our review also highlights several limitations and challenges:

First, data quality and representation – Models are only as good as the data they see. Many algorithms were trained on retrospective archives with inherent biases (e.g. referral bias to academic centers). Rare sub-populations are underrepresented (very few BL cases, etc.), which might limit model generalizability. Expanding training datasets and using data augmentation or synthesis (GANs could be used to generate more BL-like images, for instance) may help.

Second, interpretability and trust – We saw initial steps toward interpretability (attention maps, etc.), but often these are post-hoc and not integral to the model's prediction. For AI to gain widespread acceptance in pathology, pathologists need to trust it. That trust increases when the model's "thinking" aligns with pathology knowledge (e.g. it highlights the same cells a human would examine). There is also the risk that an AI might base its prediction on spurious features (like slide artifacts or scanner-specific quirks); rigorous validation across different labs can catch this.

Third, regulatory and integration aspects – None of the models reviewed are approved for clinical use yet. Regulatory agencies will require evidence of robust performance and benefit. Furthermore, integrating AI into pathology workflows requires addressing practical issues: slide scanning throughput, data storage, and interfacing AI outputs with pathology reports. Some labs might not have the needed infrastructure currently. However, the trend towards digital pathology (especially accelerated by remote work needs, etc.) means more slides are being scanned routinely, paving the way for AI deployment.

It's also worth noting the ethical dimension: using AI for prognostication could potentially alter patient management. If a model wrongly labels someone as high-risk, could it lead to overtreatment? This is why prospective trials are needed to prove that AI-based decisions actually improve outcomes, not just reclassify risk.

Comparing our findings to other reviews: Fu et al. (2024) similarly concluded that while performance metrics are often high, nearly all studies had high risk of bias[1]. They stressed improving reporting and external validation, which our review concurs with. A 2022 systematic review by Chen et al. on AI in hematopathology (not in our sources, but worth mentioning) also found that data scarcity and lack of standardization were major hurdles. On the positive side, multiple reviews including ours see promise in AI uncovering features beyond human perception – for example, the ability to predict genetic mutations from histology has been called a form of "image-based molecular testing" that could in the future augment or even replace some lab

tests. Our review specifically shows this concept in action for MYC, with reported AUCs up to 0.83. As algorithms and training data improve, we may see these AUCs climb closer to 0.90+, making them truly clinically useful screens.

Cross-subtype differences: DLBCL as a heterogeneous category was amenable to AI stratification in various ways (morphology, biomarkers, microenvironment). BL being more uniform may actually be easier to identify (when classic) but the challenge is in the borderline cases; AI might eventually help in those, but needs enough training examples of “atypical BL” which are few. HGBL (double-hit) lacked obvious unique morphologic hallmarks historically, but AI found signals for it. If we take a step back, this reflects that AI can both validate and discover pathology knowledge. It validated, for instance, that double-hit lymphomas have some recognizable growth pattern or cytologic features (since models could identify them above chance). It also discovered new patterns – e.g. how a continuous score of BCL2/MYC co-expression is prognostic beyond a binary label. In future, AI might re-define diagnostic categories based on image analysis clustering; imagine an AI “unsupervised” analysis that finds subgroups of DLBCL slides that correlate with outcomes or genotypes, potentially suggesting a novel classification scheme beyond what WHO currently defines.

Study limitations: Our systematic review, despite best efforts, has limitations. We relied on reported metrics in papers; differences in study design precluded a formal meta-analysis. The heterogeneity in outcome measures (accuracy, AUC, etc.) and tasks (some diagnostic, some prognostic) meant we focused on qualitative synthesis. We also might have missed very recent 2025 studies, as our search cut-off was end of 2024 (though we did capture a 2025 Cureus article in references indirectly). There is also the possibility of publication bias in the studies we reviewed – we saw mostly positive results; negative findings (like an AI failing to predict something) would add insight but aren’t likely published. Nonetheless, by covering multiple databases and even grey literature (Google Scholar), we attempted to be comprehensive.

## Conclusion

In conclusion, the application of AI in digital pathology for aggressive B-cell lymphomas is a fast-evolving area that shows great promise for enhancing diagnostic precision and prognostic stratification. Deep learning models have achieved high accuracy in classifying DLBCL and related subtypes, distinguishing challenging differentials like BL, and automating the detection of clinically relevant biomarkers (e.g. MYC/BCL2 status, PD-L1). These tools can supplement traditional methods, potentially speeding up diagnosis and providing more standardized assessments across institutions. Furthermore, AI has demonstrated an ability to predict patient outcomes from histologic images, identifying high-risk patients who might benefit from alternative therapies. This could herald a paradigm shift toward personalized prognostication using digital slides at diagnosis.

However, current evidence stems largely from retrospective studies with varying rigor. We observed that many models lack external validation and operate as “black boxes,” which must be addressed before clinical implementation. Collaboration between computational scientists and hematopathologists is crucial to ensure that models are trained on diverse, representative data and that their outputs are interpretable and actionable. Future research should focus on multi-center prospective validation of these AI models and on addressing rarer subtypes like BL and HGBL where data are limited. It is also imperative to integrate AI systems in a way that complements the pathologist’s workflow, enhancing rather than replacing human expertise. With continued improvements in model robustness, transparency, and regulatory oversight, AI-driven pathology tools are likely to become an integral part of lymphoma diagnosis and management. In sum, the marriage of AI and digital pathology stands to substantially improve the accuracy of lymphoma classification and the ability to foresee patient outcomes, ultimately contributing to more tailored and effective patient care[19]. The coming years will determine how these algorithmic advances translate from research to routine practice, but the groundwork laid by studies to date is a strong indicator of a smarter, data-driven future in lymphoma pathology.



## References

1. Sehn LH, Salles G. Diffuse large B-cell lymphoma. *N Engl J Med*. 2021;384(9):842-58. doi:10.1056/NEJMra2027612.
2. Kurz SC, Medeiros LJ. Large B-cell lymphomas in the 5th edition of the WHO classification of haematolymphoid tumors. *Mod Pathol*. 2023;36(7):100070. doi:10.1038/s41379-023-01352-0.
3. Hans CP, Weisenburger DD, Greiner TC, Gascoyne RD, Delabie J, Ott G, et al. Confirmation of the molecular classification of diffuse large B-cell lymphoma by immunohistochemistry using a tissue microarray. *Blood*. 2004;103(1):275-82. doi:10.1182/blood-2003-05-1545.
4. McGenity C, Dickson J, Simpson J, et al. Diagnostic accuracy of artificial intelligence in digital pathology: a systematic review and meta-analysis. *NPJ Digit Med*. 2024;7:122. doi:10.1038/s41746-024-01106-8.
5. Mohlman JS, Chen X, Li H, et al. Improving augmented human intelligence to distinguish Burkitt lymphoma from diffuse large B-cell lymphoma cases. *Am J Clin Pathol*. 2020;153(6):743-52. doi:10.1093/ajcp/aqaa018.
6. Swiderska-Chadaj Z, van Dijk M, Bukowy JD, et al. Artificial intelligence to detect MYC translocation in slides of diffuse large B-cell lymphoma. *Virchows Arch*. 2021;479(3):617-21. doi:10.1007/s00428-020-02994-8.
7. Yan F, Da Q, Yi H, Xu Y, Lu D, Li J, et al. Artificial intelligence-based assessment of PD-L1 expression in diffuse large B-cell lymphoma. *NPJ Precis Oncol*. 2024;8:77. doi:10.1038/s41698-024-00577-y.
8. Lee JH, Kim H, Park S, et al. Prediction of immunochemotherapy response for diffuse large B-cell lymphoma using artificial intelligence digital pathology. *J Pathol Inform*. 2024;15:e12370. doi:10.1002/2056-4538.12370.
9. Li Z, Chen Y, Han J, et al. A deep learning-based diagnostic platform for diffuse large B-cell lymphoma. *Nat Commun*. 2020;11:6004. doi:10.1038/s41467-020-19862-8.
10. Steinbuss G, Hufnagl P, Boor P, et al. A deep learning pipeline for the classification of malignant lymphomas on whole-slide images. *Cancers (Basel)*. 2021;13(13):3272. doi:10.3390/cancers13133272.

11. Miyoshi H, Arakawa F, Sato K, et al. A deep learning approach for the classification of malignant lymphoma histopathology images: A pilot study. *Lab Invest.* 2020;100(6):753-61. doi:10.1038/s41374-020-0407-z.
12. Perry C, Wang J, Qiao Y, et al. Artificial intelligence-based detection of high-grade B-cell lymphoma with double-hit/triple-hit rearrangements from hematoxylin and eosin images. *Cancers (Basel).* 2023;15(21):5205. doi:10.3390/cancers15215205.
13. Tavolara TE, Wang L, Gao Y, et al. Automated quantification of MYC and BCL2 immunohistochemistry in diffuse large B-cell lymphoma using deep learning. *Diagn Pathol.* 2024;19:17. doi:10.1186/s13000-024-01387-1.
14. Naji H, Al-Obaidi R, Al-Assaf A, et al. Interpretable deep learning prediction of diffuse large B-cell lymphoma recurrence from digital histopathology. *BJC Rep.* 2025;3:34. doi:10.1038/s44276-025-00034-0.
15. Carreras J, Lopez-Guillermo A, Roncador G, et al. Digital pathology and deep learning for prognostic prediction in B-cell lymphomas. *Cancers (Basel).* 2022;14(21):5318. doi:10.3390/cancers14215318.
16. Fu Y, Shi J, Xu Z, et al. Applications of artificial intelligence in histopathological diagnosis of lymphoma: a systematic review. *Front Oncol.* 2024;14:1425107. doi:10.3389/fonc.2024.1425107.
17. Syrykh C, Papillon-Cavanagh S, St-Germain J, et al. Artificial intelligence for MYC rearrangement detection in large B-cell lymphoma: multi-cohort validation. *Mod Pathol.* 2024;37(12):100610. doi:10.1038/s41379-024-01610-3.
18. Steiner DF, Chen PHC, Mermel CH. Closing the translation gap: AI applications in digital pathology for lymphoma. *Histopathology.* 2023;82(5):703-13. doi:10.1111/his.14822.
19. Nambiar N, Sharma S, Patel K, et al. Open-access AI tool for early diagnosis of Burkitt lymphoma in endemic regions. *Front Med.* 2024;11:1345611. doi:10.3389/fmed.2024.1345611.
20. DLBCL-Morph Consortium. DLBCL-Morph: a public whole-slide image dataset for diffuse large B-cell lymphoma research. *Sci Data.* 2023;10:197. doi:10.1038/s41597-023-02005-3.

## Author Contributions

- (I.A.A.M.) – She led the study design, conducted the main literature review, and drafted the initial manuscript. She also coordinated data collection and ensured methodological accuracy.
- (R.O.) – She supervised the overall project, provided critical revisions to the manuscript, and ensured alignment with journal requirements. She also managed correspondence with co-authors and acted as the primary contact during submission and peer review.
- (M.O.A.A.) – She contributed to data analysis and interpretation, reviewed supporting literature, and assisted in drafting the results and discussion sections. She also supported the preparation of tables and figures.

## Conflict of Interest

The authors declare **no conflict of interest** related to this work. All authors confirm that there are no financial or personal relationships that could have inappropriately influenced the conduct of this research or the writing of the manuscript.

## Ethical Considerations

This study did not involve direct experimentation on humans or animals. Data were obtained from previously published sources and did not require institutional review board (IRB) approval. Ethical principles of transparency, accuracy, and acknowledgment of contributions were maintained throughout the preparation of this manuscript.



