

Review

Artificial Intelligence in Lymphoma Histopathology: Systematic Review

Yao Fu¹, MMedSc; Zongyao Huang², MD; Xudong Deng³, BSc; Linna Xu², MMedSc; Yang Liu², MD; Mingxing Zhang², BMed; Jinyi Liu⁴, BMed; Bin Huang², MMedSc

¹Sichuan Tianfu New Area People's Hospital, Chengdu, China

²Department of Pathology, Sichuan Clinical Research Center for Cancer, Sichuan Cancer Hospital & Institute, Sichuan Cancer Center, University of Electronic Science and Technology of China, Chengdu, China

³Wonders Information Co., Ltd, Shanghai, China

⁴Phase I Clinical Trial Unit, Sichuan Clinical Research Center for Cancer, Sichuan Cancer Hospital & Institute, Sichuan Cancer Center, University of Electronic Science and Technology of China, Chengdu, China

Corresponding Author:

Bin Huang, MMedSc

Department of Pathology

Sichuan Clinical Research Center for Cancer, Sichuan Cancer Hospital & Institute

Sichuan Cancer Center, University of Electronic Science and Technology of China

South Renmin Road

Chengdu, 610041

China

Phone: 86 18236170185

Email: 18236170185@163.com

Abstract

Background: Artificial intelligence (AI) shows considerable promise in the areas of lymphoma diagnosis, prognosis, and gene prediction. However, a comprehensive assessment of potential biases and the clinical utility of AI models is still needed.

Objective: Our goal was to evaluate the biases of published studies using AI models for lymphoma histopathology and assess the clinical utility of comprehensive AI models for diagnosis or prognosis.

Methods: This study adhered to the Systematic Review Reporting Standards. A comprehensive literature search was conducted across PubMed, Cochrane Library, and Web of Science from their inception until August 30, 2024. The search criteria included the use of AI for prognosis involving human lymphoma tissue pathology images, diagnosis, gene mutation prediction, etc. The risk of bias was evaluated using the Prediction Model Risk of Bias Assessment Tool (PROBAST). Information for each AI model was systematically tabulated, and summary statistics were reported. The study is registered with PROSPERO (CRD42024537394) and follows the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 reporting guidelines.

Results: The search identified 3565 records, with 41 articles ultimately meeting the inclusion criteria. A total of 41 AI models were included in the analysis, comprising 17 diagnostic models, 10 prognostic models, 2 models for detecting ectopic gene expression, and 12 additional models related to diagnosis. All studies exhibited a high or unclear risk of bias, primarily due to limited analysis and incomplete reporting of participant recruitment. Most high-risk models (10/41) predominantly assigned high-risk classifications to participants. Almost all the articles presented an unclear risk of bias in at least one domain, with the most frequent being participant selection (16/41) and statistical analysis (37/41). The primary reasons for this were insufficient analysis of participant recruitment and a lack of interpretability in outcome analyses. In the diagnostic models, the most frequently studied lymphoma subtypes were diffuse large B-cell lymphoma, follicular lymphoma, chronic lymphocytic leukemia, and mantle cell lymphoma, while in the prognostic models, the most common subtypes were diffuse large B-cell lymphoma, follicular lymphoma, chronic lymphocytic leukemia, and Hodgkin lymphoma. In the internal validation results of all models, the area under the receiver operating characteristic curve (AUC) ranged from 0.75 to 0.99 and accuracy ranged from 68.3% to 100%. In models with external validation results, the AUC ranged from 0.93 to 0.99.

Conclusions: From a methodological perspective, all models exhibited biases. The enhancement of the accuracy of AI models and the acceleration of their clinical translation hinge on several critical aspects. These include the comprehensive reporting of

data sources, the diversity of datasets, the study design, the transparency and interpretability of AI models, the use of cross-validation and external validation, and adherence to regulatory guidance and standardized processes in the field of medical AI.

(*J Med Internet Res* 2025;27:e62851) doi: [10.2196/62851](https://doi.org/10.2196/62851)

KEYWORDS

lymphoma; artificial intelligence; bias; histopathology; tumor; hematological; lymphatic disease; public health; pathologists; pathology; immunohistochemistry; diagnosis; prognosis

Introduction

Lymphoma, a malignancy that originates from the lymphohematopoietic system, is recognized as one of the most prevalent hematological cancers globally. Epidemiological studies indicate that Hodgkin lymphoma (HL) and non-Hodgkin lymphoma (NHL) are prevalent malignant lymphatic disorders that pose significant public health challenges. Data from GLOBOCAN 2020 reveal that the projected global incidence and mortality for HL are 83,087 and 23,376 cases, respectively, while for NHL, these figures stand at 544,352 and 259,793, respectively [1,2]. In China, lymphoma is a considerable public health challenge. Data from the Global Burden of Diseases (GBD), Injuries, and Risk Factors Study for 2019 reveal that the age-standardized incidence rate for HL is 0.57 cases per 100,000 individuals, with an age-standardized mortality rate of 0.15 per 100,000 individuals. For NHL, the age-standardized incidence rate is significantly higher at 4.99 per 100,000 individuals, and the age-standardized mortality rate is 2.32 per 100,000 individuals [3].

Histopathology, which involves examining tissue specimens at the cellular level, is the gold standard for the diagnosis of lymphomas [4]. The conventional diagnostic process typically involves pathologists using hematoxylin-eosin (HE) staining of tissues and immunophenotyping for diagnosis. For the diagnosis of high-grade B-cell lymphomas, fluorescence in situ hybridization (FISH) is also commonly employed in conjunction [5]. However, this method has several drawbacks, including subjectivity, time-consuming procedures, and high costs [6].

Traditionally, pathologists have relied on optical microscopes to analyze pathological tissue sections. However, the advent of digital pathology has seen a shift toward the use of computers for reviewing and analyzing scanned whole slide images (WSIs). This transition is not only driven by the potential for increased efficiency but also opens new avenues for the development of automated diagnostic tools [7]. These tools have the potential to enhance the accuracy, efficiency, objectivity, and consistency of diagnoses, which is crucial in addressing the global shortage of pathologists. They can also increase diagnostic throughput and reduce the reliance on referrals and additional tests [8]. This field of research is burgeoning, and for certain types of malignant tumors, these systems are beginning to demonstrate clinical utility [9]. However, although many research models of artificial intelligence (AI) applied to lymphoma histopathology have been published, it is unclear whether there are methodological biases in these models, and the clinical utility of AI applied to lymphoma histopathology has not been summarized.

In this extensive study, we systematically reviewed the literature exploring the use of AI technologies, including traditional machine learning (ML) and deep learning (DL) methods, to assess digital pathology images for lymphoma diagnosis, prognosis, and other pertinent applications. Our review encompasses research that focuses on individual diagnostic factors, such as histological subtypes, as well as studies that perform computer-assisted tasks like tumor segmentation. We also assessed the clinical utility of these AI methods with consideration of potential biases. The objective of this review is to provide valuable insights and actionable recommendations based on the existing body of literature. Thus, this review aims to provide insights and recommendations based on published literature to improve the clinical utility of future research, including reducing the risk of bias, improving reproducibility, and increasing generalizability.

Methods

Literature Search

A thorough search was conducted across 3 prominent research databases: PubMed, Cochrane, and Web of Science. The search was restricted to peer-reviewed journals and conference proceedings to ensure the quality and credibility of the studies included. The search timeline extended from the inception of each database up to August 30, 2024. We employed MeSH terms for more precise retrieval.

Given the multitude of terms related to AI, our search strategy incorporated keywords such as “artificial intelligence,” “machine learning,” “neural network,” and “network, neural (computer),” along with “lymphoma.” We combined multiple relevant terms for each concept using the OR operator (eg, “artificial intelligence” OR “machine learning”) and then merged “lymphoma” with “artificial intelligence” using the AND operator. This approach ensured that the retrieved studies met both criteria.

Subsequently, we screened the articles based on their relevance to histopathological AI, focusing on the title and abstract. The review protocol was registered on PROSPERO (CRD42024537394) prior to the screening of search results for inclusion. Detailed search strategies and methods are provided in [Multimedia Appendix 1](#).

Literature Selection

A researcher (YF) manually removed duplicate papers with the assistance of the reference management software EndNote 20. Subsequently, another researcher (ZH) independently screened the articles for inclusion in 2 stages: the first based on titles and

abstracts, and the second based on the full text. Disagreements were discussed and arbitrated by a third researcher (JL).

The inclusion criteria required the research to evaluate the use of at least one AI approach to make diagnostic or prognostic inferences on human histopathology images from suspected or confirmed cases of lymphoma. Studies were only included if AI methods were applied directly to the digital pathology images or to features that were automatically extracted from the images. Fundamental tasks, such as segmentation and cell counting, were included as these could be used by pathologists for computer-aided diagnosis. Only conventional light microscopy images were considered, with other imaging modalities, such as fluorescence and hyperspectral imaging, excluded. Publications that did not include primary research, such as review papers, were excluded. Non-English language articles and research where a full version of the manuscript was not accessible were excluded.

In the studies included, models were deemed of interest if they adhered to the same inclusion criteria. When several models were compared against the same outcome, the model of interest was typically the newly proposed one. If this was ambiguous, the model with the best performance during the validation phase was selected. When multiple models from a single study exhibited similar modeling techniques, the one with superior validation performance was included in the assessment. Results from the same model at varying levels of precision (eg, patch level, slice level, and patient level) were not treated as distinct outcomes.

Risk of Bias Assessment

The risk of bias in the models of interest was assessed using the Prediction Model Risk of Bias Assessment Tool (PROBAST) [10]. The tool evaluates the likelihood that the reported results are distorted due to limitations in study design, conduct, and analysis. PROBAST includes 20 guiding questions categorized into 4 domains: Participants, Predictors, Outcomes, and Analysis. These questions are summarized to indicate a high risk or low risk of bias or are marked as unclear when insufficient information is available for a comprehensive assessment and no information is available to suggest a high risk of bias. It is important to note that an unclear risk of bias does not imply a methodological flaw but rather indicates incomplete reporting.

The Participants domain involves the recruitment and selection of participants to ensure the consistency and representativeness of the study population targeting the intended demographic. Relevant details include the recruitment strategy, inclusion criteria, and number of participants enrolled.

The Predictors domain addresses the consistent definition and measurement of predictive variables, which in this context often refers to the generation of digital pathology images. This encompasses methods for the fixation, staining, scanning, and digital processing of tissues prior to modeling.

The Outcomes domain involves the appropriate definition and consistent determination of ground truth labels. This includes the criteria used to ascertain diagnoses or prognoses, the

expertise of those determining these labels, and whether the labels are independent of any model outputs.

The Analysis domain encompasses statistical considerations in the evaluation of model performance to ensure valid and not overly optimistic results. It includes various factors, such as the number of participants for each outcome in the test set, the validation methods used (cross-validation, external validation, internal validation, etc), the metrics for assessing performance, and the methods to address the impact of censoring, confounding, and missing data. Some of these factors are interrelated. For example, the risk of bias due to a small dataset is somewhat mitigated by cross-validation, which increases the effective size of the test set and can be used to assess variability, reducing the optimism of the results. Additionally, the risk associated with using a small dataset depends on the type of outcome being predicted; robust analysis for a 5-class classification requires more data than a binary classification. There must also be sufficient data across all relevant patient subgroups; for instance, if multiple subtypes of lymphomas are included, it is not acceptable for 1 subtype to be represented by only a few patients. Due to these interrelated factors, there are no rigid standards for determining the appropriate size of a dataset.

Inconsistencies in methodology often lead to bias risk. For example, inconsistencies in HE staining from different research centers can lead to heterogeneity in the visual characteristics of digital pathology slides, potentially causing spurious correlations through random or systematic differences within or between subgroups in the dataset. Using a large dataset during training may enhance the model's generalizability, but this must be tightly controlled to avoid introducing systematic confounding. Inconsistencies in the determination of outcomes may mean that the results of a study are unreliable due to spurious correlations in the underlying factual labels or invalid due to misjudgment of the labels.

While PROBAST provides a framework for assessing the risk of bias, there is a degree of subjectivity in interpreting the signal questions. Therefore, each model was analyzed by 2 independent researchers (YF and ZH), with at least one computer scientist and one pathologist involved in the bias risk assessment of each model.

The Quality Assessment of Diagnostic Accuracy Studies-AI (QUADAS-AI) tool was used to evaluate the sensitivity of the included studies. QUADAS-AI is the AI-specific extension of QUADAS-2 [11] and QUADAS-C [12], and includes 4 domains for determining the risk of bias (patient selection, reference standard, index test, and flow and timing) and 3 domains for applicability issues (index test, patient selection, and reference standard) ([Multimedia Appendix 2](#)).

Data Synthesis

Data extraction was independently performed by 2 researchers (YF and MZ), using a form containing 67 fields within the categories Overview, Data, Methods, Results, and Miscellaneous. A summary of this process is provided in [Multimedia Appendix 3](#).

Information was sought from full-text articles, as well as references and supplementary materials where appropriate. Inferences were made only when both researchers were confident that this gave the correct information, with disagreements resolved through discussion. Fields that could not be confidently completed were labeled as being unclear.

All extracted data were summarized in 2 tables, 1 each for study-level and model-level characteristics. Only models of interest were included in these tables. The term model outcome refers to the model output (whether this was a clinical outcome [diagnosis or prognosis] or a diagnostically relevant outcome that could be used for computer-aided diagnosis, such as tumor segmentation). The data synthesis did not include any meta-analysis due to the diversity of the included methods and model outcomes. The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 guidelines for

reporting systematic reviews were followed, with checklists provided in [Multimedia Appendix 3](#).

Results

Results of Literature Screening

As shown in [Figure 1](#), the initial literature search identified a total of 3565 records, with 1375 being duplicates. After screening titles and abstracts, 2052 records were excluded, leaving 41 studies for inclusion in the review. All studies that met the inclusion criteria were identified through research databases, with no eligible records found in trial registries. Although the search was performed from 1949, all included studies were published since 2010, and more than 80% were published since 2020. The characteristics of these studies are summarized in [Table 1](#). The model construction parameters of the 41 included studies are detailed in [Table 2](#).

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 flowchart. AI: artificial intelligence.

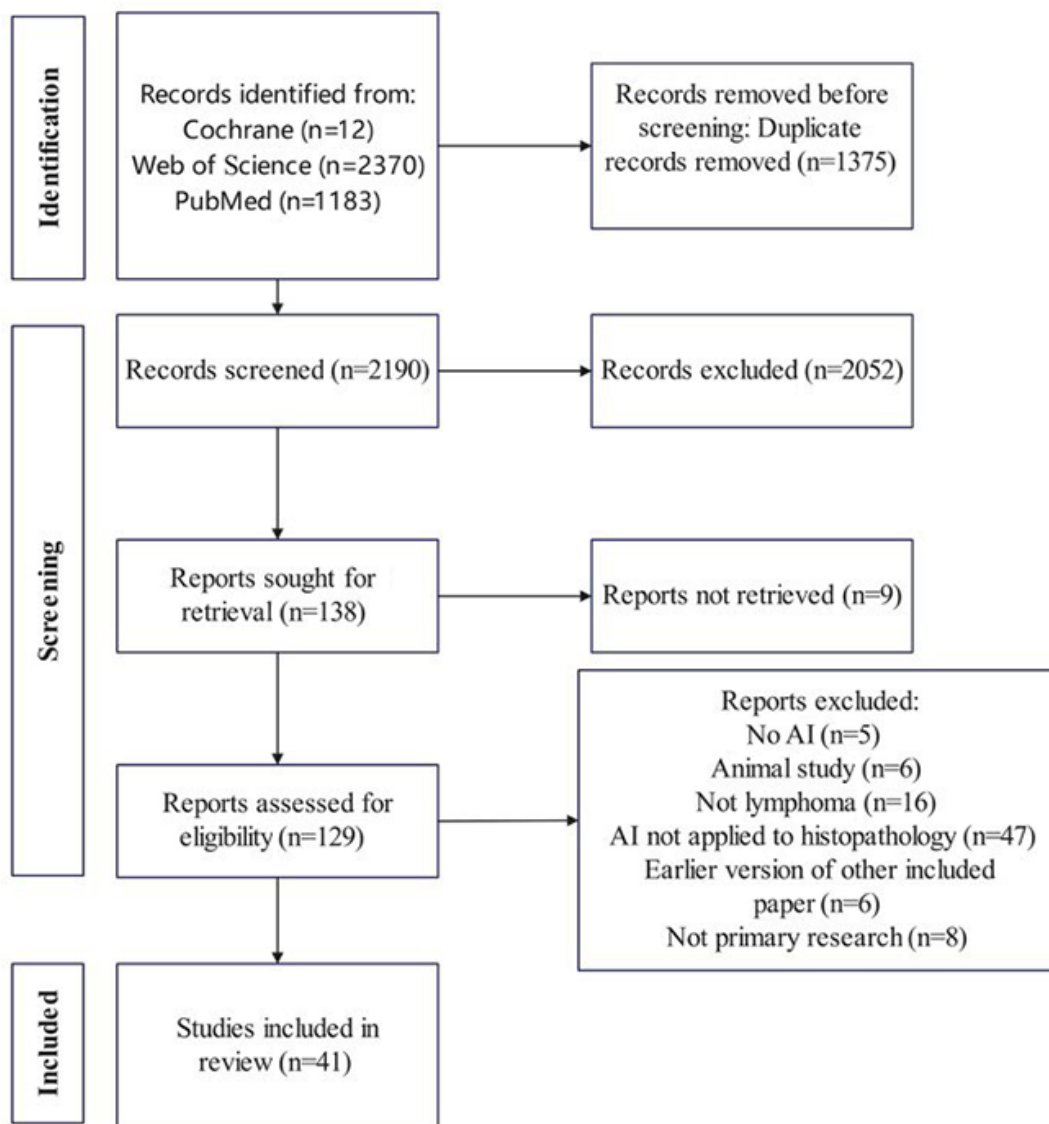


Table 1. Characteristics of the 41 studies included in this systematic review.

| Publication | Data source | Geographical distribution | Internal participants, n | Internal pathology images, n | Subtypes | Outcome type | Model task | Code available |
|--------------------------------|---------------|---------------------------|--------------------------|------------------------------|---|--------------|--|----------------|
| Achi et al [13], 2019 | Single center | America | 128 | 2560 | Burkitt lymphoma, DLBCL ^a , and small lymphocytic lymphoma | Diagnosis | Identification | None |
| Steinbuss et al [14], 2021 | Single center | Germany | 629 | Unclear | Nodal small lymphocytic lymphoma/CLL ^b and DLBCL | Diagnosis | Classification | [15] |
| Miyoshi et al [16], 2020 | Single center | Japan | 388 | Unclear | FL ^c and DLBCL | Diagnosis | Classification | None |
| Syrykh et al [17], 2020 | Single center | France | 378 | Unclear | FL | Diagnosis | Identification | [18] |
| Sereda et al [19], 2023 | Single center | Germany | 53 | Unclear | Pediatric nodular lymphocyte-predominant Hodgkin lymphoma | Prognosis | Identification, semantic segmentation | None |
| Zhang et al [20], 2020 | Single center | China | Unclear | 374 | CLL, MCL ^d , and FL | Diagnosis | Classification | None |
| Li et al [21], 2022 | Multicenter | America | 1005 | 3123 | DLBCL | Diagnosis | Classification | [22] |
| Zhang et al [23], 2021 | Single center | China | Unclear | 374 | CLL, MCL, and FL | Other | Classification | None |
| Yu et al [24], 2021 | Multicenter | China | 40 | 40 | Monomorphic epithelioid intestinal T-cell lymphoma | Other | Classification, cell nuclei segmentation | None |
| Takagi et al [25], 2023 | Single center | Japan | 842 | Unclear | DLBCL and FL | Other | Classification | None |
| Hamdi et al [26], 2023 | Single center | Saudi Arabia | Unclear | 15,000 | CLL, MCL, and FL | Diagnosis | Classification | [27] |
| Mohlman et al [28], 2020 | Single center | America | 70 | 10,818 | DLBCL and Burkitt lymphoma | Other | Classification | None |
| Karabulut et al [29], 2023 | Single center | Turkey | 11 | 60 | MF ^e | Diagnosis | Identification, classification, cell nuclei segmentation | None |
| do Nascimento et al [30], 2018 | Single center | Brazil | 347 | 15,353 | CLL, MCL, and FL | Diagnosis | Classification | None |
| Zhu et al [31], 2019 | Multicenter | China | Unclear | 374 | CLL, MCL, and FL | Other | Classification | None |
| Hashimoto et al [32], 2022 | Single center | Japan | 262 | Unclear | DLBCL, Burkitt lymphoma, and FL | Other | Classification | None |
| Michail et al [33], 2014 | Single center | Greece | Unclear | 300 | FL | Other | Identification, semantic segmentation | None |
| Kornaropoulos et al [34], 2014 | Single center | Greece | 17 | 500 | FL | Other | Classification | None |
| Chuang et al [35], 2022 | Single center | China | 103 | 309 | MCL | Prognosis | Semantic segmentation | None |

| Publication | Data source | Geographical distribution | Internal participants, n | Internal pathology images, n | Subtypes | Outcome type | Model task | Code available |
|-----------------------------------|---------------|---------------------------|--------------------------|------------------------------|---|-----------------------------|--|----------------|
| Vrabac et al [36], 2021 | Multicenter | America | 209 | Unclear | DLBCL | Prognosis | Semantic segmentation | [37] |
| Motmaen et al [38], 2023 | Single center | Germany | 83 | Unclear | HL ^f | Prognosis | Classification | None |
| Swiderska-Chadaj et al [39], 2021 | Multicenter | Holland | 287 | 354 | DLBCL | MYC translocation detection | Identification | None |
| Tagami et al [40], 2023 | Single center | Japan | 129 | 1290 | Ocular adnexal mucosa-associated lymphoid tissue lymphoma | Other | Classification | None |
| Irshaid et al [41], 2022 | Single center | America | 61 | Unclear | CLL and FL | Prognosis | Cell nuclei segmentation, classification | None |
| El Hussein et al [42], 2022 | Multicenter | America | 125 | 213 | CLL | Prognosis | Cell nuclei segmentation, classification | None |
| Chen et al [43], 2022 | Multicenter | America | 135 | 213 | CLL | Prognosis | Cell nuclei segmentation, classification | None |
| Zhang et al [44], 2024 | Multicenter | China | Unclear | Unclear | Primary central nervous system lymphoma | Diagnosis | Classification | [45] |
| Tavolara et al [46], 2024 | Single center | America | 172 | 376 | DLBCL | Other | Identification | [47] |
| Perry et al [5], 2023 | Single center | Israel | 57 | Unclear | High-grade B-cell lymphoma | Diagnosis | Classification | None |
| Yan et al [48], 2024 | Single center | China | 220 | 220 | DLBCL | Prognosis | Cell nuclei segmentation | [49] |
| Lee et al [50], 2024 | Single center | America | 216 | 251 | DLBCL | Prognosis | Cell nuclei segmentation | None |
| Duan et al [51], 2024 | Single center | China | 114 | 132 | Primary central nervous system lymphoma | Prognosis | Cell nuclei segmentation | None |
| Quan et al [52], 2024 | Single center | China | Unclear | 350 | Gastric MALT ^g lymphoma | Diagnosis | Cell nuclei segmentation | None |
| Al-Mekhlafi et al [53], 2022 | Single center | Saudi Arabia | Unclear | 1500 | CLL, FL, and MCL | Diagnosis | Classification | None |
| Somarathne et al [54], 2019 | Single center | Australia | Unclear | 374 | FL | Diagnosis | Classification | None |
| Codella et al [55], 2016 | Multicenter | America | Unclear | 374 | CLL, FL, and MCL | Diagnosis | Classification | None |
| Swiderska-Chadaj et al [56], 2020 | Multicenter | America | 91 | 157 | DLBCL | MYC translocation detection | Identification | None |
| Basu et al [57], 2022 | Unclear | India | Unclear | 700 | MF | Diagnosis | Classification | None |
| Shankar et al [58], 2023 | Single center | America | Unclear | Unclear | cHL, DLBCL, and MCL | Diagnosis | Classification | [59] |

| Publication | Data source | Geographical distribution | Internal participants, n | Internal pathology images, n | Subtypes | Outcome type | Model task | Code available |
|--------------------------|---------------|---------------------------|--------------------------|------------------------------|--|--------------|----------------|----------------|
| Soltane et al [60], 2022 | Unclear | Saudi Arabia | Unclear | 323 | cHL, nodular lymphoma predominant, Burkitt lymphoma, FL, MCL, large B-cell lymphoma, and T-cell lymphoma | Diagnosis | Classification | None |
| Tagami et al [61], 2024 | Single center | Japan | 127 | 1270 | Orbital MALT lymphoma | Diagnosis | Classification | None |

^aDLBCL: diffuse large B-cell lymphoma.

^bCLL: chronic lymphocytic leukemia.

^cFL: follicular lymphoma.

^dMCL: mantle cell lymphoma.

^eMF: mycosis fungoides.

^fHL: Hodgkin lymphoma.

^gMALT: mucosa-associated lymphoid tissue.

Table 2. Model construction parameters for the 41 included studies.

| Publication | Stain type | Original image size | Patch size | Magnification | Feature extraction | Final model | Final model prediction precision | Validation type | External validation data | Metric | Internal results | External results |
|-------------------------|------------------|---------------------|------------|---------------|--------------------|--|----------------------------------|-------------------------|--------------------------|------------------|------------------|------------------|
| Achi et al [13] | HE ^a | WSI ^b | 40×40 | 40× | Learned | CNN ^c | Patch | Internal validation | None | Accuracy | 95% | None |
| Steinbuss et al [14] | HE | WSI | 395×395 | 40× | Hand-crafted | Efficient Net | Patch | Internal validation | None | Accuracy | 95% | None |
| Miyoshi et al [16] | HE | WSI | 64×64 | 5× | Hand-crafted | CNN | Patch | 5-fold cross-validation | None | Accuracy | 94% | None |
| Srykh et al [17] | HE | WSI | 299×299 | 20× | Hand-crafted | BNN ^d | Patch | Internal validation | None | AUC ^e | 0.92-0.99 | None |
| Zhang et al [20] | HE | WSI | 224×224 | Unclear | Learned | ResNet-50 | Patch | 5-fold cross-validation | None | Accuracy | 95.4% | None |
| Li et al [21] | HE | WSI | 945×945 | 40× | Learned | 17CNN+transform | Patch | External validation | 402 | Accuracy | 100% | 100% |
| Hamdi et al [26] | HE | WSI | Unclear | Unclear | Hand-crafted | MobileNet-VGG-16, decision tree-based machine learning | WSI | Internal validation | None | AUC, accuracy | 0.99, 99.8% | None |
| Karabulut et al [29] | HE | Microscopic images | 600×600 | 200× | Learned | DL ^f | Patch | Internal validation | None | Accuracy | 94.2% | None |
| doNascimento et al [30] | HE | Microscopic images | Unclear | 1000× | Unclear | Classification using the polynomial | Unclear | Internal validation | None | Accuracy | 96%-100% | None |
| Sereda et al [19] | IHC ^g | Patch | 256×256 | Unclear | Hand-crafted | YOLOv4-tiny CNN | Patch | Internal validation | None | Accuracy | 95.43% | None |
| Chuang et al [35] | HE | WSI | 132×132 | 40× | Hand-crafted | CNN | Patch | Internal validation | None | AUC | 0.94 | None |
| Vrabac et al [36] | HE | WSI | 224×224 | 40× | Hand-crafted | Hover-Net | Patch | External validation | 179 | CI | None | 95% |
| Motmaen et al [38] | Picrosirius Red | WSI | 320×320 | 20× | Hand-crafted | YOLOv4 | Patch | Internal validation | None | AUC | 0.79 | None |
| Irshaid et al [41] | HE | WSI | 128×128 | 40× | Hand-crafted | CNN | Patch | Internal validation | None | AUC | 0.85 | None |
| El Hussein et al [42] | HE | WSI | 256×256 | 20× | Hand-crafted | Hover-Net | WSI | External validation | 28 | AUC | None | 0.93 |
| Chen et al [43] | HE | WSI | 256×256 | 20× | Hand-crafted | Hover-Net | WSI | External validation | 68 | Accuracy | None | 92.5% |

| Publication | Stain type | Original image size | Patch size | Magnification | Feature extraction | Final model | Final model prediction precision | Validation type | External validation data | Metric | Internal results | External results |
|-----------------------------|------------|---------------------|--------------------|---------------|--------------------|--------------------------------------|----------------------------------|--------------------------|--------------------------|--------------------------|------------------|------------------|
| Swiderska-Chadaj et al [39] | HE | WSI | Unclear | 20× | Learned | U-Net | WSI | External validation | 49 | Sensitivity | None | 93% |
| Zhang et al [23] | HE | WSI | Unclear | Unclear | Hand-crafted | ResNet-50 | Patch | Internal validation | None | Accuracy | 98.63% | None |
| Yu et al [24] | HE | WSI | 115×115 | 40× | Hand-crafted | Decision tree-based machine learning | Patch | Internal validation | None | AUC | 0.96 | None |
| Takagi et al [25] | HE | WSI | 224×224 | 40× | Learned | CNN | Patch | 5-fold cross-validation | None | Accuracy | 0.83 | None |
| Mohlman et al [28] | HE | WSI | 224×224 | 200× | Learned | CNN | Patch | Internal validation | None | AUC | 0.92 | None |
| Zhu et al [31] | HE | WSI | 64×64 | Unclear | Learned | VGG-16, LSTM ^h | Patch | 10-fold cross-validation | None | Overall grading accuracy | 0.98 | None |
| Hashimoto et al [32] | HE | WSI | 224×224 | 20× | Learned | CNN | Patch | 5-fold cross-validation | None | Accuracy | 68.3% | None |
| Michail et al [33] | HE | Microscopic images | Unclear | 40× | Hand-crafted | SVM ⁱ | Microscopic images | Internal validation | None | Accuracy | 97.4% | None |
| Kornaropoulos et al [34] | HE | WSI | 71×71 | Unclear | Hand-crafted | Laplacian Eigenmaps | WSI | Hold-out K-folds | None | Accuracy | 99.22% | None |
| Tagami et al [40] | HE | WSI | 208 208 | 20× | Hand-crafted | SVM | Patch | 10-fold cross-validation | None | AUC | 0.86 | None |
| Zhang et al [44] | HE | WSI | 256×256 | 40× | Learned | DL | Patch | External validation | None | AUC | 0.96 | None |
| Tavolara et al [46] | IHC | WSI | 224×224 | 40× | Hand-crafted | ResNet-50 | Patch | External validation | 108 | Sensitivity, specificity | None | 0.857, 0.991 |
| Perry et al [5] | HE | WSI | 384×384 | 40× | Hand-crafted | DL | Patch | Internal validation | None | AUC | 0.95 | None |
| Yan et al [48] | IHC | WSI | 256×256 | 40× | Hand-crafted | CNN | Patch | External validation | 61 | ICC ^j | None | 96% |
| Lee et al [50] | HE | WSI | 224×224 | 40× | Learned | ViT-S/8 | Patch | External validation | 48 | Sensitivity, specificity | None | 90.2%, 70.0% |
| Duan et al [51] | HE | WSI | 512×512 | 20× | Hand-crafted | KNN ^k | Patch | Internal validation | 46 | AUC | None | 0.92 |

| Publication | Stain type | Original image size | Patch size | Magnification | Feature extraction | Final model | Final model prediction precision | Validation type | External validation data | Metric | Internal results | External results |
|-----------------------------|------------|---------------------|----------------------|---------------|--------------------|-------------|----------------------------------|-------------------------|--------------------------|--------------------------|------------------------------------|------------------|
| Quan et al [52] | HE | WSI | 512×512 | 40× | Hand-crafted | ResNet 50 | Patch | 5-fold cross-validation | None | Sensitivity, specificity | 96.79%, SD 1.50%; 99.38%, SD 0.15% | None |
| Al-Mekhlafi et al [53] | HE | WSI | 512×512 | 40× | Hand-crafted | ResNet 50 | Patch | Internal validation | None | AUC | 0.99 | None |
| Somarathne et al [54] | HE | WSI | 227×227 | Unclear | Hand-crafted | AlexNet | Patch | External validation | 213 | AUC | None | 0.99 |
| Codella et al [55] | HE | WSI | Unclear | Unclear | Unclear | Unclear | Patch | 3-fold cross-validation | None | Accuracy | 92.3% | None |
| Swiderska-Chadaj et al [56] | HE | WSI | 512×512 | 20× | Hand-crafted | CNN | Patch | External validation | 66 | AUC | None | 0.83 |
| Basu et al [57] | HE | WSI | 224×224 | 40× | Hand-crafted | CNN | Patch | Internal validation | None | Sensitivity, specificity | 94.67%, 97.3% | None |
| Shankar et al [58] | HE | WSI | Unclear | 40× | Hand-crafted | ResNet 50 | Patch | Internal validation | None | AUC | 0.95 | None |
| Soltane et al [60] | HE | WSI | 224×224 | Unclear | Learned | ResNet 50 | Patch | 5-fold cross-validation | None | Accuracy | 91.6% | None |
| Tagami et al [61] | HE | WSI | 2048 2048 | 20× | Hand-crafted | DL | Patch | 5-fold cross-validation | None | AUC | 0.8 | None |

^aHE: hematoxylin-eosin.

^bWSI: whole slide image.

^cCNN: convolutional neural network.

^dBNN: Bayesian neural network.

^eAUC: area under the receiver operating characteristic curve.

^fDL: deep learning.

^gIHC: immunohistochemical.

^hLSTM: long short-term memory.

ⁱSVM: support vector machine.

^jICC: intraclass correlation coefficient.

^kKNN: k-nearest neighbors.

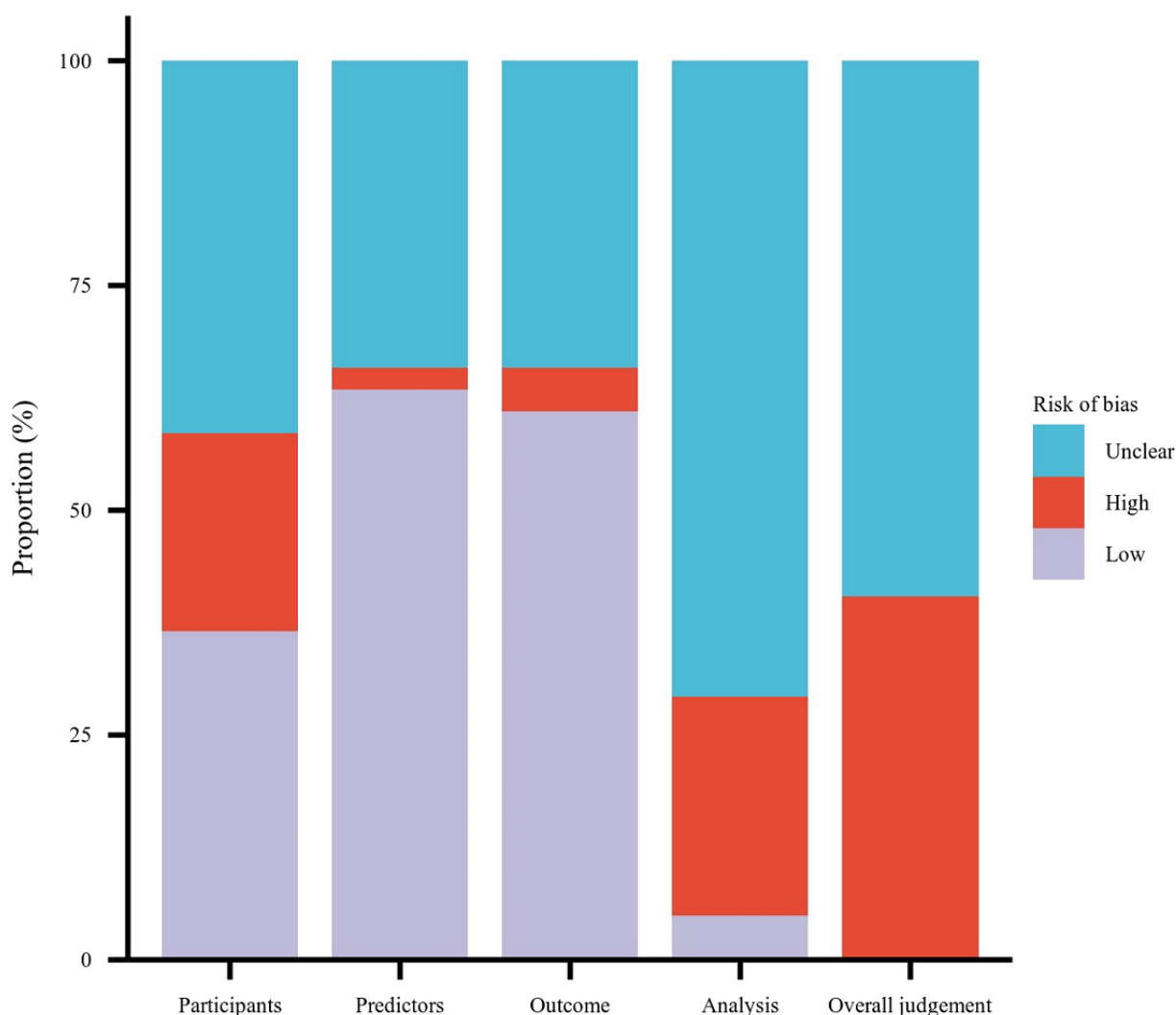
Risk of Bias Assessment

The PROBAST assessment findings are detailed in Table 3. Despite some studies encompassing multiple models of interest, each paper highlighted 1 model with superior predictive value for bias risk analysis. Most models exhibited either a high overall bias risk (13/41) or an unclear overall bias risk (28/41), with none of the models presenting a low overall bias risk (0/41).

Most high-risk models predominantly allocated their high-risk scores in the Participants domain (10/41). Conversely, most low-risk scores were concentrated in the Predictors (26/41) and Outcomes (26/41) domains. Almost all studies reported an unclear risk of bias in at least one domain, with the Participants (16/41) and Statistical Analysis (37/41) domains being the most frequently affected. Qualitative summaries are presented in Figure 2.

Table 3. Prediction Model Risk of Bias Assessment Tool (PROBAST) risk of bias assessment results for the 41 papers included in this review.

| Publication | Participants | Predictors | Outcomes | Analysis | Overall judgement |
|-----------------------------|--------------|------------|----------|----------|-------------------|
| Achi et al [13] | Unclear | Low | Unclear | Unclear | Unclear concerns |
| Steinbuss et al [14] | High | Unclear | Low | Unclear | High concerns |
| Miyoshi et al [16] | High | Low | Low | Unclear | High concerns |
| Sryrykh et al [17] | High | Low | Low | Low | High concerns |
| Sereda et al [19] | Low | Low | Low | Unclear | Unclear concerns |
| Zhang et al [20] | Unclear | Low | Unclear | Unclear | Unclear concerns |
| Li et al [21] | High | Low | Low | Unclear | High concerns |
| Zhang et al [23] | Unclear | Unclear | Low | Unclear | Unclear concerns |
| Yu et al [24] | Unclear | Low | Low | Unclear | Unclear concerns |
| Takagi et al [25] | Unclear | Low | Low | Unclear | Unclear concerns |
| Hamdi et al [26] | Unclear | Low | Unclear | Unclear | Unclear concerns |
| Mohlman et al [28] | High | Low | Low | Low | High concerns |
| Karabulut et al [29] | Unclear | Low | Low | Unclear | Unclear concerns |
| do Nascimento et al [30] | Unclear | Unclear | Unclear | Unclear | Unclear concerns |
| Zhu et al [31] | Unclear | Low | Unclear | Unclear | Unclear concerns |
| Hashimoto et al [32] | Unclear | Low | Unclear | Unclear | Unclear concerns |
| Michail et al [33] | Unclear | Unclear | Unclear | Unclear | Unclear concerns |
| Kornaropoulos et al [34] | Low | Low | Low | Unclear | Unclear concerns |
| Chuang et al [35] | Unclear | Low | Unclear | Unclear | Unclear concerns |
| Vrabac et al [36] | Low | Low | Low | Unclear | Unclear concerns |
| Motmaen et al [38] | High | Unclear | Low | Unclear | High concerns |
| Swiderska-Chadaj et al [39] | Unclear | Unclear | Unclear | Unclear | Unclear concerns |
| Tagami et al [40] | High | Unclear | Low | Unclear | High concerns |
| Irshaid et al [41] | High | Low | Low | Unclear | High concerns |
| El Hussein et al [42] | High | Low | Low | Unclear | High concerns |
| Chen et al [43] | High | Low | Low | Unclear | High concerns |
| Zhang et al [44] | Unclear | Unclear | Low | Unclear | Unclear concerns |
| Tavolara et al [46] | Low | Low | Unclear | Unclear | Unclear concerns |
| Perry et al [5] | Low | High | Low | High | High concerns |
| Yan et al [48] | Low | Low | Low | Unclear | Unclear concerns |
| Lee et al [50] | Low | Low | Low | Unclear | Unclear concerns |
| Duan et al [51] | Low | Low | High | High | High concerns |
| Quan et al [52] | Low | Unclear | Unclear | Unclear | High concerns |
| Al-Mekhlafi et al [53] | Low | Unclear | Low | Unclear | High concerns |
| Somarathne et al [54] | Low | Unclear | Low | Unclear | Unclear concerns |
| Codella et al [55] | Low | Unclear | Unclear | High | High concerns |
| Swiderska-Chadaj et al [56] | Low | Unclear | Low | Unclear | Unclear concerns |
| Basu et al [57] | Unclear | Unclear | Unclear | Unclear | Unclear concerns |
| Shankar et al [58] | Low | Low | Low | Unclear | Unclear concerns |
| Soltane et al [60] | Unclear | Low | Unclear | Unclear | Unclear concerns |
| Tagami et al [61] | Low | Low | Low | Unclear | Unclear concerns |

Figure 2. Prediction Model Risk of Bias Assessment Tool (PROBAST) risk of bias.

Data Synthesis Results

Data in the Included Literature

The number of participants across internal datasets varied significantly, with studies recruiting anywhere from 10 to 1005 patients diagnosed with lymphoma. Concurrently, the model development used a broad range of histopathological slides (count: 10-15,353). In most studies, the samples for model development were WSIs involving excised or biopsied tissues (38/41), with other samples using microscopic images (3/41). Most studies used HE-stained tissues (37/41), while others employed various immunohistochemical (IHC) staining methods (4/41). Some studies employed a multimodal analysis method that integrated pathological images with clinical information [25,35,50].

Among the included studies, most (29/41) used single-center data and few (10/41) used multicenter data. A small number of studies had unclear data sources, and the United States was the most common (12/41) source country.

Models in the Included Literature

The studies encompassed a variety of models, with the most prevalent being convolutional neural networks (CNNs),

accounting for 30 out of 41 studies. A minority of studies employed support vector machines (SVMs) and random forests (2 studies each). The CNN architectures that were explored included Mobile Net, VGG-16, Hover-Net, U-Net, and ResNet-50. These newer CNNs generally incorporated multiple standardized blocks that featured layers for convolution, normalization, activation, and pooling [23]. One study stood out by leveraging transfer learning and integrating 17 distinct DL models to create a highly accurate platform, achieving a diagnostic accuracy rate of 100% [21]. This approach significantly bolstered the model's ability to generalize. Another innovative study developed a novel architecture by applying a topological optimization method to the conventional VGG-16 model [26]. Some studies opted for a hybrid approach, combining traditional ML techniques with DL. They used CNNs for feature extraction, followed by decision tree-based methods for quantification and classification [24,26]. Notably, 1 study implemented a CNN framework grounded in Multiple Instance Learning (MIL), which autonomously concentrated on image patches from regions of interest within tumors, showcasing an advanced method for image analysis [32].

In the analysis encompassing various models, most studies predominantly used patches (33/41), with a subset operating at

the WSI level (6/41). Two distinct aggregation approaches were implemented: premodeling and postmodeling aggregation. The premodeling method necessitated the creation of slide-level features prior to modeling, whereas the postmodeling approach entailed consolidating patch-level model outputs to formulate slide-level predictions. For models that used patch images as the basis for final modeling, it was essential to segment the original images into individual patches before proceeding with modeling. The patch sizes varied from 40×40 to 2048×2048 pixels, with the most frequently employed dimensions being 224×224 pixels (9/41) and 256×256 pixels (5/41). Subsequently, a variety of feature extraction techniques were applied, encompassing both handcrafted or predefined features (27/41) and features that were automatically learned by the models (12/41).

The handcrafted features encompassed a diverse spectrum of attributes, including texture, color, cellular, and nuclear morphological characteristics. These meticulously crafted features were predominantly employed as inputs for traditional ML algorithms, such as SVM and random forest models.

In contrast, learned features were predominantly extracted through the application of CNNs, which also frequently served as the classifier of choice. Ultimately, the outputs from the patch-level models were synthesized to develop predictive models. This aggregation was achieved through various methods, such as attention-based weighted averaging, concatenation, and more sophisticated embedding techniques. These included Fisher vector encoding and k-means clustering, with the process often culminating in the selection of the maximum value to enhance the predictive accuracy of the models [36,62].

Among the papers that specified magnification levels, the most prevalent were 20× (10/41) and 40× (17/41). A handful of studies employed varying magnifications strategically to pinpoint informative tissue regions and to enhance their modeling accuracy [16,17,40].

A limited number of models integrated histopathological data with other data modalities [25,35,50]. The multimodal approaches observed in the literature included the premodeling integration of unimodal features extracted independently, as well as the amalgamation of unimodal predictions from distinct models [63]. Additionally, transformer-based methods were frequently employed for encoding the intricate relationships between different modalities [64,65]. While attention-based methods have been used in the study of other malignancies for several years [66], their application in lymphoma research is a relatively recent development. Among the studies reviewed, 1 study stood out by using a variant of the transformer architecture to encode the interplay between medical imaging data and clinical records. This study introduced a novel personalized attention mechanism (PersAM) for the classification of lymphoma subtypes, marking a significant advancement in the field [25].

Analysis in the Included Literature

Most studies relied on internal validation (30/41), while external validation using independent lymphoma datasets was seldom

conducted (11/41). In terms of internal validation, partial validation was typically executed through a 5-10-fold cross-validation approach. Some papers detailed the hyperparameter selection process using the training dataset, yet only reported evaluations on a test set derived from the same data source [24,33,67]. For external validation, models were trained on WSIs and subsequently validated on either WSIs (10/11) or tissue microarrays (TMAs) (1/11) from separate independent sources. Notably, the model of 1 study was externally validated against data from normal lymph node tissues [14]. In a particular instance, a model that achieved perfect validation accuracy (area under the receiver operating characteristic curve [AUC]=1.0) with internal validation underperformed on external cases, with an AUC ranging from 0.63 to 0.69. This discrepancy may stem from the sensitivity of ML algorithms to preprocessing steps, and neural networks, in accordance with statistical principles, necessitate a representative sample to ensure reliable inductive reasoning [12]. In another study, a comprehensive evaluation was conducted using polynomial, SVM, random forest, and decision tree classifiers to assess the efficacy of the proposed method [30].

Most models were assessed using accuracy or AUC, as well as other metrics, including sensitivity, specificity, hazard ratio, and the C-index. Even when studies reported sensitivity and specificity, CIs were not reported. In the internal validation results of all models, the AUC ranged from 0.75 to 0.99 and accuracy ranged from 68.3% to 100%. In models with external validation results, the AUC ranged from 0.93 to 0.99.

The burgeoning demand for AI methods in health care is undeniable, yet the lack of interpretability remains a significant impediment to their clinical adoption [68,69]. Enhancing the interpretability of AI models is crucial for fostering trust among medical professionals in the future AI systems they will rely upon. In a thorough analysis of the studies included, it was observed that the majority of studies (20/41) undertook efforts to analyze the interpretability of their models, with a notable minority (8/20) delving into visual interpretability analysis of the histopathological images that significantly influenced the model's prognostic assessments. Several studies meticulously characterized the spatial distribution and interrelationships of typical cells, their nuclei, and the microenvironments within the regions of interest [24,32,36,42], thereby showcasing the interpretability of their AI systems. One study presented graphical features that were correlated with clinical prognostic information [38]. A handful of studies opted for traditional ML models, such as decision trees [26], which are inherently more transparent and align closely with human reasoning processes, thus facilitating a more intuitive understanding of the decision-making process.

Clinical Utility

Among the 41 models included, 17 were diagnostic models, 10 were prognostic models, 2 were models to detect gene translocation, and 12 were other prediction- and diagnosis-related information models. The tasks of these models included identification (8/41), classification (24/41), and segmentation (9/41).

In the field of AI-based diagnostic models for lymphoma histopathology, the most common subtypes included diffuse large B-cell lymphoma (DLBCL) (5/41), follicular lymphoma (FL) (8/41), chronic lymphocytic leukemia (CLL) (5/41), and mantle cell lymphoma (MCL) (8/41). Additionally, a small number of studies developed diagnostic models for Burkitt lymphoma, central nervous system lymphoma, high-grade lymphoma, HL, and T-cell lymphoma. A study using CNNs for cell segmentation of WSIs of NHL achieved an average diagnostic accuracy of 100%, 99.73%, and 99.20% for CLL, FL, and MCL, respectively [20]. Researchers have successfully harnessed the power of both DL and traditional ML to develop a diagnostic tool with remarkable accuracy rates ranging from 95% to 100% for identifying MCL, FL, and CLL. This cutting-edge approach involves the precise segmentation of cell nuclei and the meticulous measurement of key morphological features such as area, perimeter, eccentricity, and diameter [30].

In the realm of AI-based prognostic models for lymphoma histopathology, the most common subtypes include DLBCL (3/4), CLL (2/42), HL (2/41), and FL (2/41). Sereda et al [19] used DL-based cell detection on digital slides from patients with nodular lymphocyte-predominant Hodgkin lymphoma (NLPHL) to quantitatively assess the histological patterns of lymphocyte-predominant cells. They identified 6 key features of lymphocyte-predominant cell spatial patterns and achieved a high average precision in cell detection (mean 95.24%, SD 0.17%). Furthermore, they found a strong correlation between treatment response and the density and number of lymphocyte-predominant cells ($P < .05$). Several studies have identified independent prognostic factors for DLBCL and MCL by segmenting cell nuclei and calculating the geometric characteristics of each segmented nucleus [35,36]. Several studies [41-43] addressed the clinical challenge of large cell transformation in indolent B-cell lymphomas, such as FL and CLL. They trained a CNN to predict large cell transformation based on tumor cell morphology, including the small cell proportion, chromatin pattern, presence of distinct nucleoli, and proliferation index. The machine-generated quantifications demonstrated superior reproducibility compared to estimates made by pathologists and showed a stronger correlation with the outcome data. The precise assessment and evaluation of PD-L1 biomarkers are crucial for the targeted immunotherapy triage of cancer patients. Notably, Yan et al [48] developed an AI-based image analysis method that encompasses the detection, segmentation, and classification of PD-L1+ cells for the evaluation of PD-L1 expression in patients with DLBCL. This method has produced highly correlated quantitative results compared to the subjective assessments of pathologists. However, none of the prognostic models included T-cell lymphoma subtypes in the studies.

Regarding AI-based histopathological models for detecting gene translocations in lymphoma, 2 studies focused on DLBCL as the tumor type [39,56]. Their results showed that it is possible to predict MYC translocation based on morphology alone. This would allow simple and fast prescreening, saving about 34% of genetic testing using the current algorithm.

Overall, methodologically, all studies exhibited a high or unclear risk of bias, primarily due to limited analysis and incomplete

reporting on participant recruitment. Most high-risk models (10/41) predominantly assigned high-risk classifications to participants. Almost all studies presented an unclear risk of bias in at least one domain, with the most frequent being participant selection (16/41) and statistical analysis (37/41). The primary reasons for this were insufficient analysis of participant recruitment and a lack of interpretability in outcome analyses. In the diagnostic models, the most frequently studied lymphoma subtypes were DLBCL, FL, CLL, and MCL, while in the prognostic models, the most common subtypes were DLBCL, FL, CLL, and HL. None of the prognostic models included T-cell lymphoma subtypes in the studies. In the internal validation results of all models, the AUC ranged from 0.75 to 0.99 and accuracy ranged from 68.3% to 100%. In models with external validation results, the AUC ranged from 0.93 to 0.99.

Sensitivity Analysis

To further evaluate the sensitivity of our conclusions, we conducted a sensitivity analysis by selecting diagnostic models from the literature we included and performing a QUADAS-AI diagnostic evaluation. The results are presented in [Multimedia Appendix 4](#).

Our findings revealed that out of the 17 diagnostic models considered, 15 were rated as high-risk models and 2 were deemed unclear. Most high-risk models were classified as such primarily due to their use of nonpublic datasets (13/17). A minority of studies (2/17) were rated as high risk because they failed to provide a clear description of their data sources. Therefore, we used QUADAS-AI to evaluate the diagnostic models and concluded that all the methodologies included in the diagnostic models were biased.

To further assess the sensitivity of our conclusions, we considered that the distribution of lymphoma subtypes varies by region, which could potentially bias the results. Therefore, we only included studies from the United States, as it was the most common source country (12/41). We found that in the United States, the most common subtypes in AI-based diagnostic models for lymphoma histopathology were DLBCL (5/12), FL (2/12), and MCL (2/12). Regarding AI-based prognostic models for lymphoma histopathology, the most common subtypes were DLBCL (2/12), CLL (3/12), and FL (1/12). In the models for detecting gene translocations in lymphoma histopathology, 1 study focused on DLBCL. These findings are largely consistent with our previous conclusions.

Discussion

Current Status of AI in Lymphoma Assessment

AI has significantly enhanced the precision of lymphoma diagnostics by eliminating the subjectivity often associated with human observation [16,26]. For instance, a study by Achi et al [13] demonstrated the power of CNN in accurately distinguishing between various lymphoma subtypes. Their diagnostic model, designed for 4 distinct lymphoma categories (benign lymph nodes, DLBCL, Burkitt lymphoma, and small lymphocytic lymphoma), achieved an impressive 95% accuracy rate in image prediction. Notably, in a multicenter study [21], the diagnostic accuracy of AI for DLBCL reached 100%.

Moreover, AI provides invaluable insights into the tumor microenvironment, enabling the identification and quantification of image features that surpass simple density assessments. It delves into higher-order relationships and offers a quantitative evaluation of lymphocyte aggregation patterns and the complex interplay between tumor regions. Such capabilities are pivotal for advancing cancer clinical research and the development of new therapeutics [36,70].

However, methodologically, the risk assessments of the included studies in this review were all rated as high or unclear, primarily due to incomplete reporting, absence of detailed patient source information, and inadequate explanation of the predictors used. This highlights the need to expedite the clinical translation of AI in lymphoma diagnosis, ensuring that these advanced tools are rigorously validated and seamlessly integrated into clinical practice.

Frequently omitted details include the precise origin of patient data, the total number of patients involved, the quantity of samples or images used, and the techniques employed for tissue processing and digitization. Most studies reported data from single centers, and this scarcity may stem from AI researchers not dedicating sufficient effort to comprehend these images, whether for training purposes or external validation. Information about the predictors (histopathology images and their features) was generally better reported; however, there remains an absence or inadequacy of the detailing of certain critical aspects. For example, it is often unclear whether the investigators assessed the predictors without knowing the outcomes or whether all histopathology images were processed uniformly, which could have introduced bias. Moreover, some researchers rely on a limited dataset and analyze a single test data split without implementing methods to mitigate overfitting and model optimism, such as cross-validation or external validation. These limitations are prevalent in lymphoma AI research, resulting in weak validation and an elevated risk of bias within the models.

Code sharing is crucial for enhancing the reproducibility of research findings and mitigating the effects of incomplete reporting. However, in the review of 41 papers, only 9 included codes, and the data in the other studies were either incomplete or difficult to access. To foster better reproducibility, code repositories should provide comprehensive documentation. This should include instructions for setting up the environment, an overview of the code's functionality, guidance on how to produce results, and, of course, the code itself [14,17,21,26,38,44,46,48,58].

Several studies are dedicated to enhancing the interpretability of DL tools by employing existing methods. These include post hoc techniques and supervised ML models that interpret the outcomes after DL models have generated predictions [71,72]. In the realm of AI research focused on lymphoma, most current studies offer personalized interpretability for analysis. This includes visual attention heatmaps and traditional ML highlighting the spatial locations of key feature areas and their interrelationships. However, traditional ML, often crafted in partnership with domain experts, can provide greater interpretability as it relies on manually engineered features. Despite this advantage, the process of handcrafting features is

inherently challenging and complex. It demands a substantial time commitment from pathologists or oncologists, who are responsible for developing these methods.

In recent years, there has been a notable increase in hybrid approaches that combine DL with handcrafted strategies. These methods might involve using DL algorithms for the preliminary detection of cells or elements, followed by the application of easily interpretable traditional ML techniques for making predictions. By doing so, they harness domain knowledge to ensure the biological interpretability of the approach [73].

Development of the Field

The domain of AI in lymphoma histopathology diagnosis and prognosis is experiencing rapid growth, with a notable surge in scholarly publications since 2019. Most of these studies have leveraged deep neural networks for automated feature extraction and classification. In contrast, a smaller subset of research has employed conventional ML algorithms [24,30,33,40]. Recent investigations have expanded their scope to encompass a wider array of diagnostic outcomes, such as identifying specific lymphoma subtypes [17,21], predicting prognosis [35,36], and detecting genetic translocations [39,56].

Despite advancements regarding the role of AI in lymphoma research, there has not been a noticeable trend toward larger datasets, either in the number of slides analyzed or the number of participants involved. While there is no indication that more recent studies have adopted stricter internal validation methods, there is a positive shift toward increased external validation. Prior to 2019, no studies incorporated external validation on lymphoma data; however, 10 recent publications have done this. Although these validations often involve limited data, their inclusion signifies a step forward in research methodology.

These external validations are essential for the practical application of AI models in clinical settings, as they must be robust enough to handle the visual diversity present in data from various sources. This diversity can fluctuate both between and within different data centers over time. As the field evolves, we expect to see an increase in studies that rigorously validate their models against larger, high-quality, independent datasets. This includes transparent reporting on patient recruitment and selection processes, histopathology slide preparation, and digitization techniques. Such practices will be instrumental in mitigating the biases, limited reproducibility, and restricted generalizability that currently plague much of the research in this domain.

In the realm of oncology, there has been a marked surge in the number of published multimodal research studies since 2019. Notably, this growth includes only a handful of studies focused on multimodal research in lymphoma. Histopathological examination of tissue sections continues to be the cornerstone for cancer diagnosis, yet even seasoned pathologists often seek support from biomarker assays to enhance diagnostic accuracy. Multimodal research, which amalgamates diverse data types, such as genomics, proteomics, transcriptomics, and clinical data, has been pivotal in steering the trajectory of cancer research. This approach not only consolidates information from various sources but also transforms it, offering novel insights

[74,75]. For instance, a multimodal framework can facilitate a dual-modality analysis where a pathological image is processed to yield outputs from another domain, such as genetic sequencing or different imaging formats [63]. Such a model, when adeptly trained, can be leveraged to analyze pathological images from patients without overt medical conditions, extracting valuable indicators pertinent to precision medicine, including genetic sequences. As we anticipate advancements in high-throughput technologies, alongside the fields of transcriptomics, metabolomics, and proteomics, the future holds promise for an increase in the integration of multidimensional omics data with histopathological images in multimodal analyses.

This integration is poised to significantly bolster the clinical utility of AI, enabling more precise and personalized treatment strategies in oncology. By harnessing the power of multimodal research, we can expect a future where AI plays an even more integral role in clinical decision-making, thereby enhancing patient outcomes.

Current Limitations and Future Recommendations

A considerable amount of published research lacks the necessary clinical and pathological details to evaluate potential biases effectively. As a result, it is imperative for AI researchers to meticulously document the origins of their data. This transparency is crucial for understanding the variability within the dataset and for determining whether this diversity has been adequately addressed in the research methodology. Additionally, the modeling and analytical techniques employed must be thoroughly described to ensure the reliability and reproducibility of the findings.

To further improve reproducibility, we recommend that researchers provide code and data whenever possible. Digital pathology studies on lymphoma are currently constrained by the lack of publicly available data. Furthermore, WSIs from different centers can lead to significant heterogeneity in image data due to differences in scanning equipment across various centers. This variability can introduce confounding factors that complicate the task of developing robust AI models and assessing their generalizability. Such limitations can increase the risk of bias and confusion in research findings.

To mitigate these issues, during the image preprocessing stage, it is essential to diligently address and eliminate confounding factors arising from variations in staining, the presence of bubbles, and other artifacts. This meticulous attention to detail will enhance the accuracy and reliability of AI-driven diagnostic and analytical tools.

For AI to be clinically valuable, rigorous validation is paramount, particularly considering the constraints inherent in existing datasets. We recommend that researchers employ comprehensive analytical methods, such as cross-validation and external validation, to substantiate the robustness of their findings and the capacity of their models to extend to new datasets. Moreover, it is essential to report CIs for results, with a focus on the 95% CI, especially when comparing various models. This practice aids in discerning whether observed differences in model performance are genuinely significant or merely a product of random variation. By doing so, researchers can make more informed decisions about the efficacy and reliability of different AI models in clinical settings.

Researchers are recommended to follow regulatory guidance and standardized processes in the field of medical AI, such as reporting guidelines and quality assessment tools, such as QUADAS-AI, which provide a specific framework for assessing the risk of bias and the applicability of studies of diagnostic test accuracy in AI centers.

Moreover, a lack of interpretability is a barrier to the clinical adoption of AI. Therefore, we recommend that researchers strive to demonstrate the interpretability of their models to enhance the understanding and trust of clinical and pathological professionals.

Lymphoma is a diverse type of blood cancer that includes a variety of subtypes. From the literature we have reviewed, the current application of AI in lymphoma histopathology has primarily been focused on the diagnosis and prognosis of B-cell lymphoma. There are relatively fewer AI models for HL or T-cell lymphoma and other rare subtypes. This could be attributed to the higher incidence of B-cell lymphomas compared to other subtypes [1,2]. However, given the high aggressiveness and heterogeneity of T-cell lymphoma and HL [76], we hope that in future model development, more researchers will develop AI models for T-cell lymphoma and HL from a multidimensional clinical perspective.

Conclusion

Methodologically, the diagnostic and prognostic models of AI applied to lymphoma histopathology were evaluated, and the models were found to be biased. The enhancement of the accuracy of AI models and the acceleration of their clinical translation hinge on several critical aspects. These include the comprehensive reporting of data sources, the diversity of datasets, the study design, the transparency and interpretability of AI models, the use of cross-validation and external validation, and adherence to regulatory guidance and standardized processes in the field of medical AI.

Acknowledgments

This study was supported by the National Natural Science Foundation of China (grant number: 82300216), the Fellowship of China Postdoctoral Science Foundation (grant number: 2024M760362), and the Sichuan Provincial Foundation of Science and Technology (grant number: 2025NSFSC0425).

Authors' Contributions

ZH, JL, and BH provided ideas for literature writing. YF drafted and developed literature search strategies, conducted the literature search, and wrote the manuscript. MZ, LX, and XD extracted field information and created the graphs. YL and JL discussed and arbitrated differences. ZH and BH directed the writing of the manuscript. All authors read and approved the final manuscript. Author JL is the co-corresponding author of this paper and can be reached at: Phase 1 Clinical Trial Unit, Sichuan Clinical Research Center for Cancer, Sichuan Cancer Hospital & Institute, Sichuan Cancer Center, University of Electronic Science and Technology of China, Chengdu, China; liujinyi@scszllyy.org.cn.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Literature search strategies in PubMed, Cochrane Library, and Web of Science.

[\[DOCX File, 18 KB-Multimedia Appendix 1\]](#)

Multimedia Appendix 2

Description of quality assessment based on Quality Assessment of Diagnostic Accuracy Studies-AI (QUADAS-AI) domains used to evaluate the methodological quality of the studies included.

[\[DOCX File, 16 KB-Multimedia Appendix 2\]](#)

Multimedia Appendix 3

Prediction Model Risk of Bias Assessment Tool (PROBAST) table.

[\[XLS File \(Microsoft Excel File\), 152 KB-Multimedia Appendix 3\]](#)

Multimedia Appendix 4

Quality Assessment of Diagnostic Accuracy Studies-AI (QUADAS-AI) table.

[\[XLS File \(Microsoft Excel File\), 93 KB-Multimedia Appendix 4\]](#)

Multimedia Appendix 5

PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 checklist.

[\[DOCX File, 32 KB-Multimedia Appendix 5\]](#)

References

1. Huang J, Pang WS, Lok V, Zhang L, Lucero-Prisno DE, Xu W, et al. NCD Global Health Research Group, Association of Pacific Rim Universities (APRU). Incidence, mortality, risk factors, and trends for Hodgkin lymphoma: a global data analysis. *J Hematol Oncol*. May 11, 2022;15(1):57. [\[FREE Full text\]](#) [doi: [10.1186/s13045-022-01281-9](https://doi.org/10.1186/s13045-022-01281-9)] [Medline: [35546241](https://pubmed.ncbi.nlm.nih.gov/35546241/)]
2. Chu Y, Liu Y, Fang X, Jiang Y, Ding M, Ge X, et al. The epidemiological patterns of non-Hodgkin lymphoma: global estimates of disease burden, risk factors, and temporal trends. *Front Oncol*. 2023;13:1059914. [\[FREE Full text\]](#) [doi: [10.3389/fonc.2023.1059914](https://doi.org/10.3389/fonc.2023.1059914)] [Medline: [37333805](https://pubmed.ncbi.nlm.nih.gov/37333805/)]
3. GBD 2021 Diseases and Injuries Collaborators. Global incidence, prevalence, years lived with disability (YLDs), disability-adjusted life-years (DALYs), and healthy life expectancy (HALE) for 371 diseases and injuries in 204 countries and territories and 811 subnational locations, 1990-2021: a systematic analysis for the Global Burden of Disease Study 2021. *Lancet*. May 18, 2024;403(10440):2133-2161. [\[FREE Full text\]](#) [doi: [10.1016/S0140-6736\(24\)00757-8](https://doi.org/10.1016/S0140-6736(24)00757-8)] [Medline: [38642570](https://pubmed.ncbi.nlm.nih.gov/38642570/)]
4. Pathological Technology Expert Group of Medical Technician Committee of Chinese Medical Doctor Association, Standardization Department of Pathological Equipment Branch of China Medical Equipment Association, Pathological Technology Group, Pathology Professional Committee. [Chinese expert consensus on detection techniques and interpretation of lymphoma gene rearrangement (2023 version)]. *Zhonghua Bing Li Xue Za Zhi*. Jun 08, 2023;52(6):558-565. [doi: [10.3760/cma.j.cn112151-20230106-00010](https://doi.org/10.3760/cma.j.cn112151-20230106-00010)] [Medline: [37263919](https://pubmed.ncbi.nlm.nih.gov/37263919/)]
5. Perry C, Greenberg O, Haberman S, Herskovitz N, Gazy I, Avinoam A, et al. Image-based deep learning detection of high-grade b-cell lymphomas directly from hematoxylin and eosin images. *Cancers (Basel)*. Oct 29, 2023;15(21):5205. [\[FREE Full text\]](#) [doi: [10.3390/cancers15215205](https://doi.org/10.3390/cancers15215205)] [Medline: [37958379](https://pubmed.ncbi.nlm.nih.gov/37958379/)]
6. Ferry JA. Scientific advances and the evolution of diagnosis, subclassification and treatment of lymphoma. *Arch Med Res*. Nov 2020;51(8):749-764. [doi: [10.1016/j.arcmed.2020.05.022](https://doi.org/10.1016/j.arcmed.2020.05.022)] [Medline: [32553461](https://pubmed.ncbi.nlm.nih.gov/32553461/)]

7. Baidoshvili A, Bucur A, van Leeuwen J, van der Laak J, Kluin P, van Diest PJ. Evaluating the benefits of digital pathology implementation: time savings in laboratory logistics. *Histopathology*. Nov 2018;73(5):784-794. [doi: [10.1111/his.13691](https://doi.org/10.1111/his.13691)] [Medline: [29924891](https://pubmed.ncbi.nlm.nih.gov/29924891/)]
8. Stenzinger A, Alber M, Allgäuer M, Jurmeister P, Bockmayr M, Budczies J, et al. Artificial intelligence and pathology: From principles to practice and future applications in histomorphology and molecular profiling. *Semin Cancer Biol*. Sep 2022;84:129-143. [doi: [10.1016/j.semcancer.2021.02.011](https://doi.org/10.1016/j.semcancer.2021.02.011)] [Medline: [33631297](https://pubmed.ncbi.nlm.nih.gov/33631297/)]
9. Raciti P, Sue J, Retamero JA, Ceballos R, Godrich R, Kunz JD, et al. Clinical validation of artificial intelligence-augmented pathology diagnosis demonstrates significant gains in diagnostic accuracy in prostate cancer detection. *Arch Pathol Lab Med*. Oct 01, 2023;147(10):1178-1185. [FREE Full text] [doi: [10.5858/arpa.2022-0066-OA](https://doi.org/10.5858/arpa.2022-0066-OA)] [Medline: [36538386](https://pubmed.ncbi.nlm.nih.gov/36538386/)]
10. Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST Group. PROBAST: A tool to assess the risk of bias and applicability of prediction model studies. *Ann Intern Med*. Jan 01, 2019;170(1):51-58. [FREE Full text] [doi: [10.7326/M18-1376](https://doi.org/10.7326/M18-1376)] [Medline: [30596875](https://pubmed.ncbi.nlm.nih.gov/30596875/)]
11. Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. Oct 18, 2011;155(8):529-536. [FREE Full text] [doi: [10.7326/0003-4819-155-8-201110180-00009](https://doi.org/10.7326/0003-4819-155-8-201110180-00009)] [Medline: [22007046](https://pubmed.ncbi.nlm.nih.gov/22007046/)]
12. Yang B, Mallett S, Takwoingi Y, Davenport CF, Hyde CJ, Whiting PF, QUADAS-C Group, et al. QUADAS-C: A tool for assessing risk of bias in comparative diagnostic accuracy studies. *Ann Intern Med*. Nov 2021;174(11):1592-1599. [FREE Full text] [doi: [10.7326/M21-2234](https://doi.org/10.7326/M21-2234)] [Medline: [34698503](https://pubmed.ncbi.nlm.nih.gov/34698503/)]
13. Achi HE, Belousova T, Chen L, Wahed A, Wang I, Hu Z, et al. Automated diagnosis of lymphoma with digital pathology images using deep learning. *Ann Clin Lab Sci*. Mar 2019;49(2):153-160. [Medline: [31028058](https://pubmed.ncbi.nlm.nih.gov/31028058/)]
14. Steinbuss G, Kriegsmann M, Zgorzelski C, Brobeil A, Goepfert B, Dietrich S, et al. Deep learning for the classification of non-Hodgkin lymphoma on histopathological images. *Cancers (Basel)*. May 17, 2021;13(10):2419. [FREE Full text] [doi: [10.3390/cancers13102419](https://doi.org/10.3390/cancers13102419)] [Medline: [34067726](https://pubmed.ncbi.nlm.nih.gov/34067726/)]
15. Deep Learning for the Classification of Non-Hodgkin Lymphoma on Histopathological Images. GitHub. URL: <https://github.com/AG-Computational-Diagnostic/pacltune-pup> [accessed 2025-01-24]
16. Miyoshi H, Sato K, Kabeya Y, Yonezawa S, Nakano H, Takeuchi Y, et al. Deep learning shows the capability of high-level computer-aided diagnosis in malignant lymphoma. *Lab Invest*. Oct 29, 2020;100(10):1300-1310. [FREE Full text] [doi: [10.1038/s41374-020-0442-3](https://doi.org/10.1038/s41374-020-0442-3)] [Medline: [32472096](https://pubmed.ncbi.nlm.nih.gov/32472096/)]
17. Srykh C, Abreu A, Amara N, Siegfried A, Maisongrosse V, Frenois FX, et al. Accurate diagnosis of lymphoma on whole-slide histopathology images using deep learning. *NPJ Digit Med*. 2020;3:63. [FREE Full text] [doi: [10.1038/s41746-020-0272-0](https://doi.org/10.1038/s41746-020-0272-0)] [Medline: [32377574](https://pubmed.ncbi.nlm.nih.gov/32377574/)]
18. Accurate diagnosis of lymphoma on whole-slide histopathology images using deep learning. GitHub. URL: <https://github.com/ArnaudAbreu/DiagFLFH> [accessed 2025-01-24]
19. Sereda S, Shankar A, Weber L, Ramsay AD, Hall GW, Hayward J, et al. Digital pathology in pediatric nodular lymphocyte-predominant Hodgkin lymphoma: correlation with treatment response. *Blood Adv*. Oct 24, 2023;7(20):6285-6289. [FREE Full text] [doi: [10.1182/bloodadvances.2023010652](https://doi.org/10.1182/bloodadvances.2023010652)] [Medline: [37611165](https://pubmed.ncbi.nlm.nih.gov/37611165/)]
20. Zhang J, Cui W, Guo X, Wang B, Wang Z. Classification of digital pathological images of non-Hodgkin's lymphoma subtypes based on the fusion of transfer learning and principal component analysis. *Med Phys*. Sep 2020;47(9):4241-4253. [doi: [10.1002/mp.14357](https://doi.org/10.1002/mp.14357)] [Medline: [32593219](https://pubmed.ncbi.nlm.nih.gov/32593219/)]
21. Li D, Bledsoe JR, Zeng Y, Liu W, Hu Y, Bi K, et al. A deep learning diagnostic platform for diffuse large B-cell lymphoma with high accuracy across multiple hospitals. *Nat Commun*. Nov 26, 2020;11(1):6004. [FREE Full text] [doi: [10.1038/s41467-020-19817-3](https://doi.org/10.1038/s41467-020-19817-3)] [Medline: [33244018](https://pubmed.ncbi.nlm.nih.gov/33244018/)]
22. Li D, Bledsoe J, Zeng Y, Liu W, Hu Y, Bi K, et al. A deep learning diagnostic platform for diffuse large B-cell lymphoma with high accuracy across multiple hospitals. UMass Chan Medical School. URL: <https://fts.umassmed.edu/> [accessed 2025-01-24]
23. Zhang X, Zhang K, Jiang M, Yang L. Research on the classification of lymphoma pathological images based on deep residual neural network. *Technol Health Care*. 2021;29(S1):335-344. [FREE Full text] [doi: [10.3233/THC-218031](https://doi.org/10.3233/THC-218031)] [Medline: [33682770](https://pubmed.ncbi.nlm.nih.gov/33682770/)]
24. Yu W, Li C, Wang R, Yeh C, Chuang S. Machine learning based on morphological features enables classification of primary intestinal t-cell lymphomas. *Cancers (Basel)*. Oct 30, 2021;13(21):5463. [FREE Full text] [doi: [10.3390/cancers13215463](https://doi.org/10.3390/cancers13215463)] [Medline: [34771625](https://pubmed.ncbi.nlm.nih.gov/34771625/)]
25. Takagi Y, Hashimoto N, Masuda H, Miyoshi H, Ohshima K, Hontani H, et al. Transformer-based personalized attention mechanism for medical images with clinical records. *J Pathol Inform*. 2023;14:100185. [FREE Full text] [doi: [10.1016/j.jpi.2022.100185](https://doi.org/10.1016/j.jpi.2022.100185)] [Medline: [36691660](https://pubmed.ncbi.nlm.nih.gov/36691660/)]
26. Hamdi M, Senan EM, Jadhav ME, Olayah F, Awaji B, Alalayah KM. Hybrid models based on fusion features of a CNN and handcrafted features for accurate histopathological image analysis for diagnosing malignant lymphomas. *Diagnostics (Basel)*. Jul 04, 2023;13(13):2258. [FREE Full text] [doi: [10.3390/diagnostics13132258](https://doi.org/10.3390/diagnostics13132258)] [Medline: [37443652](https://pubmed.ncbi.nlm.nih.gov/37443652/)]

27. Hamdi M, Senan EM, Jadhav ME, Olayah F, Awaji B, Alalayah KM. Hybrid models based on fusion features of a CNN and handcrafted features for accurate histopathological image analysis for diagnosing malignant lymphomas (Dataset). kaggle. URL: <https://www.kaggle.com/datasets/obulisainaren/multi-cancer> [accessed 2025-01-24]
28. Mohlman JS, Leventhal SD, Hansen T, Kohan J, Pascucci V, Salama ME. Improving augmented human intelligence to distinguish Burkitt lymphoma from diffuse large B-cell lymphoma cases. *Am J Clin Pathol*. May 05, 2020;153(6):743-759. [doi: [10.1093/ajcp/aqaa001](https://doi.org/10.1093/ajcp/aqaa001)] [Medline: [32067039](https://pubmed.ncbi.nlm.nih.gov/32067039/)]
29. Karabulut YY, Dinç U, Köse E, Türsen Ü. Deep learning as a new tool in the diagnosis of mycosis fungoides. *Arch Dermatol Res*. Jul 2023;315(5):1315-1322. [doi: [10.1007/s00403-022-02521-1](https://doi.org/10.1007/s00403-022-02521-1)] [Medline: [36571610](https://pubmed.ncbi.nlm.nih.gov/36571610/)]
30. do Nascimento MZ, Martins AS, Azevedo Tosta TA, Neves LA. Lymphoma images analysis using morphological and non-morphological descriptors for classification. *Comput Methods Programs Biomed*. Sep 2018;163:65-77. [doi: [10.1016/j.cmpb.2018.05.035](https://doi.org/10.1016/j.cmpb.2018.05.035)] [Medline: [30119858](https://pubmed.ncbi.nlm.nih.gov/30119858/)]
31. Zhu H, Jiang H, Li S, Li H, Pei Y. A novel multispace image reconstruction method for pathological image classification based on structural information. *Biomed Res Int*. 2019;2019:3530903. [FREE Full text] [doi: [10.1155/2019/3530903](https://doi.org/10.1155/2019/3530903)] [Medline: [31111048](https://pubmed.ncbi.nlm.nih.gov/31111048/)]
32. Hashimoto N, Ko K, Yokota T, Kohno K, Nakaguro M, Nakamura S, et al. Subtype classification of malignant lymphoma using immunohistochemical staining pattern. *Int J Comput Assist Radiol Surg*. Jul 2022;17(7):1379-1389. [FREE Full text] [doi: [10.1007/s11548-021-02549-0](https://doi.org/10.1007/s11548-021-02549-0)] [Medline: [35147848](https://pubmed.ncbi.nlm.nih.gov/35147848/)]
33. Michail E, Dimitropoulos K, Koletsis T, Kostopoulos I, Grammalidis N. Morphological and textural analysis of centroblasts in low-thickness sliced tissue biopsies of follicular lymphoma. *Annu Int Conf IEEE Eng Med Biol Soc*. 2014;2014:3374-3377. [doi: [10.1109/EMBC.2014.6944346](https://doi.org/10.1109/EMBC.2014.6944346)] [Medline: [25570714](https://pubmed.ncbi.nlm.nih.gov/25570714/)]
34. Kornaropoulos EN, Niazi MKK, Lozanski G, Gurcan MN. Histopathological image analysis for centroblasts classification through dimensionality reduction approaches. *Cytometry A*. Mar 2014;85(3):242-255. [FREE Full text] [doi: [10.1002/cyto.a.22432](https://doi.org/10.1002/cyto.a.22432)] [Medline: [24376080](https://pubmed.ncbi.nlm.nih.gov/24376080/)]
35. Chuang W, Yu W, Lee Y, Zhang Q, Chang H, Shih L, et al. Deep learning-based nuclear morphometry reveals an independent prognostic factor in mantle cell lymphoma. *Am J Pathol*. Dec 2022;192(12):1763-1778. [FREE Full text] [doi: [10.1016/j.ajpath.2022.08.006](https://doi.org/10.1016/j.ajpath.2022.08.006)] [Medline: [36150505](https://pubmed.ncbi.nlm.nih.gov/36150505/)]
36. Vrabac D, Smit A, Rojansky R, Natkunam Y, Advani RH, Ng AY, et al. DLBCL-Morph: Morphological features computed using deep learning for an annotated digital DLBCL image set. *Sci Data*. May 20, 2021;8(1):135. [FREE Full text] [doi: [10.1038/s41597-021-00915-w](https://doi.org/10.1038/s41597-021-00915-w)] [Medline: [34017010](https://pubmed.ncbi.nlm.nih.gov/34017010/)]
37. DLBCL-Morph: Morphological features computed using deep learning for an annotated digital DLBCL image set. GitHub. URL: <https://github.com/stanfordmlgroup/DLBCL-Morph> [accessed 2025-01-24]
38. Motmaen I, Sereda S, Brobeil A, Shankar A, Braeuninger A, Hasenclever D, et al. Deep-learning based classification of a tumor marker for prognosis on Hodgkin's disease. *Eur J Haematol*. Nov 2023;111(5):722-728. [doi: [10.1111/ejh.14066](https://doi.org/10.1111/ejh.14066)] [Medline: [37549921](https://pubmed.ncbi.nlm.nih.gov/37549921/)]
39. Swiderska-Chadaj Z, Hebeda KM, van den Brand M, Litjens G. Artificial intelligence to detect MYC translocation in slides of diffuse large B-cell lymphoma. *Virchows Arch*. Sep 2021;479(3):617-621. [FREE Full text] [doi: [10.1007/s00428-020-02931-4](https://doi.org/10.1007/s00428-020-02931-4)] [Medline: [32979109](https://pubmed.ncbi.nlm.nih.gov/32979109/)]
40. Tagami M, Nishio M, Katsuyama-Yoshikawa A, Misawa N, Sakai A, Haruna Y, et al. Machine learning model with texture analysis for automatic classification of histopathological images of ocular adnexal mucosa-associated lymphoid tissue lymphoma of two different origins. *Curr Eye Res*. Dec 2023;48(12):1195-1202. [doi: [10.1080/02713683.2023.2246696](https://doi.org/10.1080/02713683.2023.2246696)] [Medline: [37566457](https://pubmed.ncbi.nlm.nih.gov/37566457/)]
41. Irshaid L, Bleiberg J, Weinberger E, Garritano J, Shallis RM, Patsenker J, et al. Histopathologic and machine deep learning criteria to predict lymphoma transformation in bone marrow biopsies. *Arch Pathol Lab Med*. Jan 02, 2022;146(2):182-193. [FREE Full text] [doi: [10.5858/arpa.2020-0510-OA](https://doi.org/10.5858/arpa.2020-0510-OA)] [Medline: [34086849](https://pubmed.ncbi.nlm.nih.gov/34086849/)]
42. El Hussein S, Chen P, Medeiros LJ, Wistuba II, Jaffray D, Wu J, et al. Artificial intelligence strategy integrating morphologic and architectural biomarkers provides robust diagnostic accuracy for disease progression in chronic lymphocytic leukemia. *J Pathol*. Jan 2022;256(1):4-14. [FREE Full text] [doi: [10.1002/path.5795](https://doi.org/10.1002/path.5795)] [Medline: [34505705](https://pubmed.ncbi.nlm.nih.gov/34505705/)]
43. Chen P, El Hussein S, Xing F, Aminu M, Kannapiran A, Hazle JD, et al. Chronic lymphocytic leukemia progression diagnosis with intrinsic cellular patterns via unsupervised clustering. *Cancers (Basel)*. May 13, 2022;14(10):2398. [FREE Full text] [doi: [10.3390/cancers14102398](https://doi.org/10.3390/cancers14102398)] [Medline: [35626003](https://pubmed.ncbi.nlm.nih.gov/35626003/)]
44. Zhang X, Zhao Z, Wang R, Chen H, Zheng X, Liu L, et al. A multicenter proof-of-concept study on deep learning-based intraoperative discrimination of primary central nervous system lymphoma. *Nat Commun*. May 04, 2024;15(1):3768. [FREE Full text] [doi: [10.1038/s41467-024-48171-x](https://doi.org/10.1038/s41467-024-48171-x)] [Medline: [38704409](https://pubmed.ncbi.nlm.nih.gov/38704409/)]
45. A multicenter proof-of-concept study on deep learning-based intraoperative discrimination of primary central nervous system lymphoma. GitHub. URL: <https://github.com/Kepler1647b/LGNet/tree/main> [accessed 2025-01-24]
46. Tavolara TE, Niazi MKK, Feldman AL, Jaye DL, Flowers C, Cooper LAD, et al. Translating prognostic quantification of c-MYC and BCL2 from tissue microarrays to whole slide images in diffuse large B-cell lymphoma using deep learning. *Diagn Pathol*. Jan 19, 2024;19(1):17. [FREE Full text] [doi: [10.1186/s13000-023-01425-6](https://doi.org/10.1186/s13000-023-01425-6)] [Medline: [38243330](https://pubmed.ncbi.nlm.nih.gov/38243330/)]

47. Translating prognostic quantification of c-MYC and BCL2 from tissue microarrays to whole slide images in diffuse large B-cell lymphoma using deep learning. GitHub. URL: https://github.com/cialab/tma_to_wsi [accessed 2025-01-24]
48. Yan F, Da Q, Yi H, Deng S, Zhu L, Zhou M, et al. Artificial intelligence-based assessment of PD-L1 expression in diffuse large B cell lymphoma. *NPJ Precis Oncol*. Mar 27, 2024;8(1):76. [FREE Full text] [doi: [10.1038/s41698-024-00577-y](https://doi.org/10.1038/s41698-024-00577-y)] [Medline: [38538739](https://pubmed.ncbi.nlm.nih.gov/38538739/)]
49. Artificial intelligence-based assessment of PD-L1 expression in diffuse large B cell lymphoma. GitHub. URL: <https://github.com/yanfang-research/Digital-PD-L1-Scoring> [accessed 2025-01-24]
50. Lee JH, Song G, Lee J, Kang S, Moon KM, Choi Y, et al. Prediction of immunochemotherapy response for diffuse large B-cell lymphoma using artificial intelligence digital pathology. *J Pathol Clin Res*. May 2024;10(3):e12370. [FREE Full text] [doi: [10.1002/2056-4538.12370](https://doi.org/10.1002/2056-4538.12370)] [Medline: [38584594](https://pubmed.ncbi.nlm.nih.gov/38584594/)]
51. Duan L, He Y, Guo W, Du Y, Yin S, Yang S, et al. Machine learning-based pathomics signature of histology slides as a novel prognostic indicator in primary central nervous system lymphoma. *J Neurooncol*. Jun 2024;168(2):283-298. [FREE Full text] [doi: [10.1007/s11060-024-04665-8](https://doi.org/10.1007/s11060-024-04665-8)] [Medline: [38557926](https://pubmed.ncbi.nlm.nih.gov/38557926/)]
52. Quan J, Ye J, Lan J, Wang J, Hu Z, Guo Z, et al. A deep learning model fusion algorithm for the diagnosis of gastric Mucosa-associated lymphoid tissue lymphoma. *Biomedical Signal Processing and Control*. Jun 2024;92:106064. [doi: [10.1016/j.bspc.2024.106064](https://doi.org/10.1016/j.bspc.2024.106064)]
53. Al-Mekhlafi ZG, Senan EM, Mohammed BA, Alazmi M, Alayba AM, Alreshidi A, et al. Diagnosis of histopathological images to distinguish types of malignant lymphomas using hybrid techniques based on fusion features. *Electronics*. Sep 10, 2022;11(18):2865. [doi: [10.3390/electronics11182865](https://doi.org/10.3390/electronics11182865)]
54. Somaratne U, Wong K, Parry J, Sohel F, Wang X, Laga H. Improving Follicular Lymphoma Identification using the Class of Interest for Transfer Learning. 2019. Presented at: 2019 Digital Image Computing: Techniques and Applications (DICTA); December 02-04, 2019; Perth, WA, Australia. [doi: [10.1109/DICTA47822.2019.8946075](https://doi.org/10.1109/DICTA47822.2019.8946075)]
55. Codella N, Moradi M, Matasar M, Sveda-Mahmood T, Smith J. Lymphoma diagnosis in histopathology using a multi-stage visual learning approach. In: *Proceedings Volume 9791, Medical Imaging 2016: Digital Pathology*. 2016. Presented at: SPIE Medical Imaging; February 27-March 3, 2016; San Diego, CA. [doi: [10.1117/12.2217158](https://doi.org/10.1117/12.2217158)]
56. Swiderska-Chadaj Z, Hebeda K, van den Brand M, Litjens G. Predicting MYC translocation in HE specimens of diffuse large B-cell lymphoma through deep learning. In: *Proceedings Volume 11320, Medical Imaging 2020: Digital Pathology*. 2020. Presented at: SPIE Medical Imaging; February 15-20, 2020; Houston, TX. [doi: [10.1117/12.2549650](https://doi.org/10.1117/12.2549650)]
57. Basu S, Agarwal R, Srivastava V. Deep discriminative learning model with calibrated attention map for the automated diagnosis of diffuse large B-cell lymphoma. *Biomedical Signal Processing and Control*. Jul 2022;76:103728. [doi: [10.1016/j.bspc.2022.103728](https://doi.org/10.1016/j.bspc.2022.103728)]
58. Shankar V, Yang X, Krishna V, Tan B, Silva O, Rojansky R, et al. LymphoML: An interpretable artificial intelligence-based method identifies morphologic features that correlate with lymphoma subtype. *Proceedings of Machine Learning Research*. 2023;225:528-558. [FREE Full text]
59. LymphoML: An interpretable artificial intelligence-based method identifies morphologic features that correlate with lymphoma subtype. GitHub. URL: <https://github.com/rajpurkarlab/LymphoML> [accessed 2025-01-24]
60. Soltane S, Alshreef S, Eldin SM. Classification and diagnosis of lymphoma's histopathological images using transfer learning. *Computer Systems Science and Engineering*. 2022;40(2):629-644. [doi: [10.32604/csse.2022.019333](https://doi.org/10.32604/csse.2022.019333)]
61. Tagami M, Nishio M, Yoshikawa A, Misawa N, Sakai A, Haruna Y, et al. Artificial intelligence-based differential diagnosis of orbital MALT lymphoma and IgG4 related ophthalmic disease using hematoxylin-eosin images. *Graefes Arch Clin Exp Ophthalmol*. Oct 03, 2024;262(10):3355-3366. [doi: [10.1007/s00417-024-06501-1](https://doi.org/10.1007/s00417-024-06501-1)] [Medline: [38700592](https://pubmed.ncbi.nlm.nih.gov/38700592/)]
62. Samsi S, Krishnamurthy AK, Gurcan MN. An efficient computational framework for the analysis of whole slide images: application to follicular lymphoma immunohistochemistry. *J Comput Sci*. Sep 01, 2012;3(5):269-279. [FREE Full text] [doi: [10.1016/j.jocs.2012.01.009](https://doi.org/10.1016/j.jocs.2012.01.009)] [Medline: [22962572](https://pubmed.ncbi.nlm.nih.gov/22962572/)]
63. Qiao Y, Zhao L, Luo C, Luo Y, Wu Y, Li S, et al. Multi-modality artificial intelligence in digital pathology. *Brief Bioinform*. Nov 19, 2022;23(6):bbac367. [FREE Full text] [doi: [10.1093/bib/bbac367](https://doi.org/10.1093/bib/bbac367)] [Medline: [36124675](https://pubmed.ncbi.nlm.nih.gov/36124675/)]
64. Ruchti A, Neuwirth A, Lowman AK, Duenweg SR, LaViolette PS, Bukowy JD. Homologous point transformer for multi-modality prostate image registration. *PeerJ Comput Sci*. 2022;8:e1155. [FREE Full text] [doi: [10.7717/peerj-cs.1155](https://doi.org/10.7717/peerj-cs.1155)] [Medline: [36532813](https://pubmed.ncbi.nlm.nih.gov/36532813/)]
65. Dong W, Yang Q, Wang J, Xu L, Li X, Luo G, et al. Multi-modality attribute learning-based method for drug-protein interaction prediction based on deep neural network. *Brief Bioinform*. May 19, 2023;24(3):bbad161. [doi: [10.1093/bib/bbad161](https://doi.org/10.1093/bib/bbad161)] [Medline: [37114624](https://pubmed.ncbi.nlm.nih.gov/37114624/)]
66. Lu MY, Williamson DFK, Chen TY, Chen RJ, Barbieri M, Mahmood F. Data-efficient and weakly supervised computational pathology on whole-slide images. *Nat Biomed Eng*. Jun 01, 2021;5(6):555-570. [FREE Full text] [doi: [10.1038/s41551-020-00682-w](https://doi.org/10.1038/s41551-020-00682-w)] [Medline: [33649564](https://pubmed.ncbi.nlm.nih.gov/33649564/)]
67. Frankenstein Z, Uraoka N, Aypar U, Aryeequaye R, Rao M, Hameed M, et al. Automated 3D scoring of fluorescence in situ hybridization (FISH) using a confocal whole slide imaging scanner. *Appl Microsc*. Apr 09, 2021;51(1):4. [FREE Full text] [doi: [10.1186/s42649-021-00053-y](https://doi.org/10.1186/s42649-021-00053-y)] [Medline: [33835321](https://pubmed.ncbi.nlm.nih.gov/33835321/)]

68. Madabhushi A, Lee G. Image analysis and machine learning in digital pathology: Challenges and opportunities. *Med Image Anal.* Oct 2016;33:170-175. [[FREE Full text](#)] [doi: [10.1016/j.media.2016.06.037](https://doi.org/10.1016/j.media.2016.06.037)] [Medline: [27423409](#)]
69. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat Mach Intell.* May 2019;1(5):206-215. [[FREE Full text](#)] [doi: [10.1038/s42256-019-0048-x](https://doi.org/10.1038/s42256-019-0048-x)] [Medline: [35603010](#)]
70. Xu Z, Li Y, Wang Y, Zhang S, Huang Y, Yao S, et al. A deep learning quantified stroma-immune score to predict survival of patients with stage II-III colorectal cancer. *Cancer Cell Int.* Oct 30, 2021;21(1):585. [[FREE Full text](#)] [doi: [10.1186/s12935-021-02297-w](https://doi.org/10.1186/s12935-021-02297-w)] [Medline: [34717647](#)]
71. Steiner DF, MacDonald R, Liu Y, Truszkowski P, Hipp JD, Gammage C, et al. Impact of deep learning assistance on the histopathologic review of lymph nodes for metastatic breast cancer. *Am J Surg Pathol.* Dec 2018;42(12):1636-1646. [[FREE Full text](#)] [doi: [10.1097/PAS.0000000000001151](https://doi.org/10.1097/PAS.0000000000001151)] [Medline: [30312179](#)]
72. Liu Y, Kohlberger T, Norouzi M, Dahl GE, Smith JL, Mohtashamian A, et al. Artificial intelligence-based breast cancer nodal metastasis detection: insights into the black box for pathologists. *Arch Pathol Lab Med.* Jul 2019;143(7):859-868. [[FREE Full text](#)] [doi: [10.5858/arpa.2018-0147-OA](https://doi.org/10.5858/arpa.2018-0147-OA)] [Medline: [30295070](#)]
73. Wang X, Janowczyk A, Zhou Y, Thawani R, Fu P, Schalper K, et al. Prediction of recurrence in early stage non-small cell lung cancer using computer extracted nuclear features from digital H-E images. *Sci Rep.* Oct 19, 2017;7(1):13543. [[FREE Full text](#)] [doi: [10.1038/s41598-017-13773-7](https://doi.org/10.1038/s41598-017-13773-7)] [Medline: [29051570](#)]
74. Zhao L, Dong Q, Luo C, Wu Y, Bu D, Qi X, et al. DeepOmix: A scalable and interpretable multi-omics deep learning framework and application in cancer survival analysis. *Comput Struct Biotechnol J.* 2021;19:2719-2725. [[FREE Full text](#)] [doi: [10.1016/j.csbj.2021.04.067](https://doi.org/10.1016/j.csbj.2021.04.067)] [Medline: [34093987](#)]
75. d'Este SH, Nielsen MB, Hansen AE. Visualizing glioma infiltration by the combination of multimodality imaging and artificial intelligence, a systematic review of the literature. *Diagnostics (Basel).* Mar 25, 2021;11(4):592. [[FREE Full text](#)] [doi: [10.3390/diagnostics11040592](https://doi.org/10.3390/diagnostics11040592)] [Medline: [33806195](#)]
76. Horwitz SM, Ansell S, Ai WZ, Barnes J, Barta SK, Brammer J, et al. T-Cell lymphomas, version 2.2022, NCCN Clinical Practice Guidelines in Oncology. *J Natl Compr Canc Netw.* Mar 2022;20(3):285-308. [doi: [10.6004/jnccn.2022.0015](https://doi.org/10.6004/jnccn.2022.0015)] [Medline: [35276674](#)]

Abbreviations

- AI:** artificial intelligence
- AUC:** area under the receiver operating characteristic curve
- CLL:** chronic lymphocytic leukemia
- CNN:** convolutional neural network
- DL:** deep learning
- DLBCL:** diffuse large B-cell lymphoma
- FL:** follicular lymphoma
- HE:** hematoxylin-eosin
- HL:** Hodgkin lymphoma
- IHC:** immunohistochemical
- MCL:** mantle cell lymphoma
- ML:** machine learning
- NHL:** non-Hodgkin lymphoma
- PROBAST:** Prediction Model Risk of Bias Assessment Tool
- QUADAS-AI:** Quality Assessment of Diagnostic Accuracy Studies-AI
- SVM:** support vector machine
- WSI:** whole slide image

Edited by A Schwartz; submitted 03.06.24; peer-reviewed by M Zhang, S Marletta; comments to author 11.09.24; revised version received 03.11.24; accepted 07.01.25; published 14.02.25

Please cite as:

Fu Y, Huang Z, Deng X, Xu L, Liu Y, Zhang M, Liu J, Huang B
Artificial Intelligence in Lymphoma Histopathology: Systematic Review
J Med Internet Res 2025;27:e62851
URL: <https://www.jmir.org/2025/1/e62851>
doi: [10.2196/62851](https://doi.org/10.2196/62851)
PMID:

©Yao Fu, Zongyao Huang, Xudong Deng, Linna Xu, Yang Liu, Mingxing Zhang, Jinyi Liu, Bin Huang. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 14.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org/>, as well as this copyright and license information must be included.