Artificial intelligence in lymphoma histopathology: a systematic review

Yao Fu, Mingxing Zhang, Xudong Deng, Linna Xu, Yang Liu, Zongyao Huang

Submitted to: Journal of Medical Internet Research
on: June 03, 2024

Table of Contents

Artificial intelligence in lymphoma histopathology: a systematic review

Yao Fu[1]; Mingxing Zhang[2]; Xudong Deng[3]; Linna Xu[2]; Yang Liu[2]; Zongyao Huang[2]

[1]Tianfu New District People's Hospital, Sichuan, China Chengdu CN
[2]Department of Pathology, Sichuan Clinical Research Center for Cancer, Sichuan Cancer Hospital & Institute, Sichuan Cancer Center, Affiliated Cancer Hospital of University of Electronic Science and Technology of China, Chengdu, China. Chengdu CN
[3]Wonders Information Co., Ltd. Chengdu CN

Corresponding Author:
Zongyao Huang
Department of Pathology, Sichuan Clinical Research Center for Cancer, Sichuan Cancer Hospital & Institute, Sichuan Cancer Center, Affiliated Cancer Hospital of University of Electronic Science and Technology of China, Chengdu, China.
Department of Pathology, Sichuan Clinical Research Center for Cancer, Sichuan Cancer Hospital & Institute, Sichuan Cancer Center, Affiliated Cancer Hospital of University of Electronic Science and Technology of China, Chengdu, China.
Chengdu
CN

Abstract

Background: Lymphoma is a malignant tumor originating from the lymphoid hematopoietic system and is one of the most common hematological tumors worldwide. According to epidemiological data, Hodgkin's lymphoma (HL) and non-Hodgkin's lymphoma (NHL) are common malignant lymphatic diseases that threaten public health. GLOBOCAN 2020 statistics show that the estimated number of HL cases and deaths worldwide is 83087 and 23376, respectively, while that of NHL is 54 352 and 25 993, respectively.
Histopathology, the examination of tissue specimens at the cellular level, is the gold standard for lymphoma diagnosis. The conventional diagnostic process is that pathologists usually use hematoxylin-eosin (H&E) stained tissue to make a diagnosis, but the diagnosis has the disadvantages of subjectivity and time-consuming. In diagnosing difficult cases, the general pathologist may seek the help of a subspecialty lymphoma pathologist, and/or use ancillary tests such as immunohistochemistry (IHC). Referral and ancillary testing are critical to the accuracy of the diagnostic process, but at the cost of making its diagnostic cycle longer and more expensive. It is becoming increasingly common for pathologists to use computers to review whole slide images (WSI) of analytical scans. These tools can often improve diagnostic accuracy, efficiency, objectivity, and consistency. These tools can help alleviate the global pathologist workforce shortage, improve diagnostic throughput, and reduce the need for referrals and ancillary testing

Objective: Artificial intelligence (AI) has great potential in the diagnosis, prognosis, or gene prediction of lymphoma. We aimed to summarize the performance of AI models for diagnostic or prognostic purposes based on published studies on histopathological images used in lymphoma.

Methods: This study followed the Systematic Review Reporting Program guidelines. A literature search of PubMed, Cochrane Library, and Web of Science was conducted from inception until February 28, 2024. Included in the standard requirement will artificial intelligence for the prognosis of human lymphoma tissue pathology images or diagnosis, gene mutation, etc. The risk of bias was assessed using PROBAST. Information for each model was tabulated, and summary statistics were reported. The study was registered with PROSPERO (CRD42024537394) and followed PRISMA 2020 reporting guidelines.

Results: The search identified 3414 records, of which 31 articles were eligible for inclusion. With a total of 42 models, these studies included 16 diagnostic models, 8 prognostic models, 1 model to detect whether the gene was ectopic, and 17 other diagnostic related models. Common tasks include identification task (6/42), classification task (31/42), segmentation tasks (5/42). Models were developed using 10 to 84139 histopathology slides from 10 to 1005 lymphoma patients. In all studies have found higher risk bias or not clear. In the high-risk model, high-risk scores appeared in the participant, outcome, and analysis sections (3/31). Most of the low-risk domains appeared in the predictors (21/31) and outcomes (18/31) sections. Almost all papers had an unclear risk of bias in at least one domain, the most common being the domains of participants (15/31) and statistical analysis (29/31). In the overall practicality evaluation, most of the models were unclear (16/31), high-risk models (12/31), and low-risk models (2/31). the most common reason is the analysis of the participant recruitment limited and no interpretability of the results analysis.

Conclusions: applying artificial intelligence to lymphoma tissue pathology diagnosis or prognosis of the purpose of image limited study, found no one model can prove to be prepared for the implementation of the real world. Key aspects of accelerating the clinical translation of AI include comprehensive reporting of data sources and modeling approaches, interpretability of AI models, and improved quantitative assessments using cross validation and external validation. Clinical Trial: The study was registered with PROSPERO (CRD42024537394)

(JMIR Preprints 03/06/2024:62851)
DOI: https://doi.org/10.2196/preprints.62851

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Artificial intelligence in lymphoma histopathology: a systematic review

Yao Fu[1] Mingxing Zhang[2] Linna Xu[2] Xudong Deng[3] Yang Liu[2] Zongyao Huang[2]*

Affiliations
[1]Tianfu New District People's Hospital, Sichuan, China
[2]Department of Pathology, Sichuan Clinical Research Center for Cancer, Sichuan Cancer Hospital & Institute, Sichuan Cancer Center, Affiliated Cancer Hospital of University of Electronic Science and Technology of China, Chengdu, China.
[3]Wonders Information Co., Ltd.

[*]**Corresponding author**

E-mail address
Yao Fu: yaofu77@outlook.com
Mingxing Zhang: zhangmingxing@scszlyy.org.cn
Linna Xu: xulinna2921@163.com
Xudong Deng: xdengxd@gmail.com
Yang Liu: liuyang@scszlyy.org.cn
Zongyao Huang: huangzongyao@scszlyy.org.cn

Author for correspondence:
Zongyao Huang:(Ph.D., huangzongyao@scszlyy.org.cn, Tel:(86)18582861120)

Abstract

Objective: Artificial intelligence (AI) has great potential in the diagnosis, prognosis, or gene prediction of lymphoma. We aimed to summarize the performance of AI models for diagnostic or prognostic purposes based on published studies on histopathological images used in lymphoma.
Methods: This study followed the Systematic Review Reporting Program guidelines. A literature search of PubMed, Cochrane Library, and Web of Science was conducted from inception until February 28, 2024. Included in the standard requirement will artificial intelligence for the prognosis of human lymphoma tissue pathology images or diagnosis, gene mutation, etc. The risk of bias was assessed using PROBAST. Information for each model was tabulated, and summary statistics were reported. The study was registered with PROSPERO (CRD42024537394) and followed PRISMA 2020 reporting guidelines.
Results: The search identified 3414 records, of which 31 articles were eligible for inclusion. With a total of 42 models, these studies included 16 diagnostic models, 8 prognostic models, 1 model to detect whether the gene was ectopic, and 17 other diagnostic related models. Common tasks include identification task (6/42), classification task (31/42), segmentation tasks (5/42). Models were developed using 10 to 84139 histopathology slides from 10 to 1005 lymphoma patients. In all studies have found higher risk bias or not clear. In the high-risk

model, high-risk scores appeared in the participant, outcome, and analysis sections (3/31). Most of the low-risk domains appeared in the predictors (21/31) and outcomes (18/31) sections. Almost all papers had an unclear risk of bias in at least one domain, the most common being the domains of participants (15/31) and statistical analysis (29/31). In the overall practicality evaluation, most of the models were unclear (16/31), high-risk models (12/31), and low-risk models (2/31). the most common reason is the analysis of the participant recruitment limited and no interpretability of the results analysis.

Conclusion: applying artificial intelligence to lymphoma tissue pathology diagnosis or prognosis of the purpose of image limited study, found no one model can prove to be prepared for the implementation of the real world. Key aspects of accelerating the clinical translation of AI include comprehensive reporting of data sources and modeling approaches, interpretability of AI models, and improved quantitative assessments using cross validation and external validation.

Keywords lymphoma, Artificial intelligence, Bias

INTRODUCTION

Lymphoma is a malignant tumor originating from the lymphohematopoietic system and is one of the most common hematological malignancies worldwide. According to epidemiological data, Hodgkin's lymphoma (HL) and non-Hodgkin's lymphoma (NHL) are common malignant lymphatic system diseases that pose a threat to public health. The statistical data from GLOBOCAN 2020 show that the estimated number of new cases and deaths worldwide for HL are 83,087 and 23,376, respectively, while for NHL, they are 544,352 and 259,793, respectively(1, 2)。In China, lymphoma also constitutes a significant public health issue. According to data from the Global Burden of Diseases(GBD), Injuries, and Risk Factors for the year 2019, China accounts for approximately 10.8% of the global incidence and 9.8% of the mortality cases for HL, and 20.1% of the incidence and 17.4% of the mortality for NHL(3).

Histopathology, which is the examination of tissue specimens at the cellular level, is the gold standard for the diagnosis of lymphoma (4).The conventional diagnostic process involves pathologists using hematoxylin and eosin (H&E) stained tissue for diagnosis. however, this method has drawbacks such as subjectivity and time consumption. In cases of difficult diagnosis, general pathologists may seek the assistance of specialized lymphoma pathologists and/or use auxiliary tests, such as immunohistochemistry (IHC). Referrals and auxiliary tests are crucial for the accuracy of the diagnostic process but at the cost of a longer diagnostic cycle and higher expenses(5).

Traditionally, pathologists analyze pathological tissue sections using optical microscopes. However, it is becoming increasingly common for pathologists to use computers to review and analyze scanned whole slide images (WSI). Although the use

of digital pathology may be driven by efficiency benefits, it indeed creates opportunities for pathologists to develop automated tools(6).These tools can typically improve the accuracy, efficiency, objectivity, and consistency of diagnoses. They can help alleviate the global shortage of pathologists, increase diagnostic throughput, and reduce the need for referrals and auxiliary tests(7). This is an increasingly active field of research, and for some malignant tumors, these systems are beginning to achieve clinical utility (8).

In this study, we systematically reviewed all literature in which artificial intelligence (AI) technology, including traditional machine learning (ML) and deep learning methods, was applied to digital pathology images for the diagnosis, prognosis, or other applications in lymphoma. This includes studies focusing on single diagnostic factors such as histological subtypes, as well as studies performing computer-assisted diagnostic tasks such as tumor segmentation. The review describes the state of the field, outlines which diagnostic and prognostic tasks have been addressed, and assesses factors related to the clinical utility of these methods, such as the risk of bias. Although lymphoma is a particularly difficult disease to detect and diagnose, and there is a shortage of available pathologists, AI models have not yet been implemented in the clinical practice for this disease. 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, enhancing reproducibility, and improving generalizability.

RESULTS

As depicted in Figure 1, the literature search initially identified a total of 3414 records, of which 819 were duplicates. During the screening of titles and abstracts, 2519records were excluded, including six records for which the full articles were not accessible. The remaining 31 studies were included in the review. All accepted studies were originally identified through the search of research databases, with no records from trial registries meeting the inclusion criteria. Although the search retrieved literature dating back to 1949, all studies that met the inclusion criteria were published since 2010, and over 80% of the included literature has been published since 2020. The characteristics of the studies are presented in Table 1. The 31 accepted articles encompass 42 models of interest, with detailed information provided in Table 2

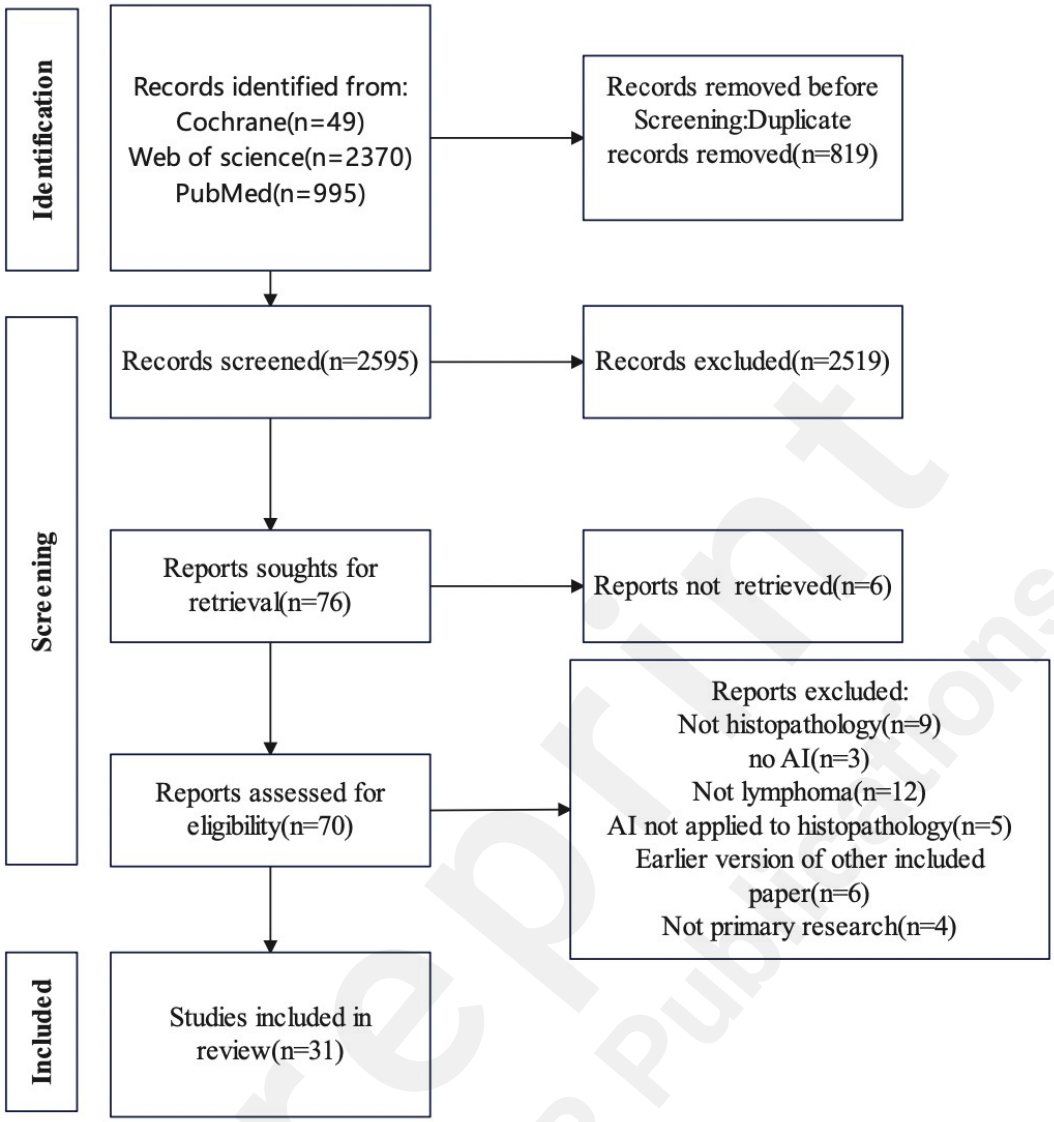


Fig. 1 PRISMA 2020 flowchart. PRISMA 2020 flowchart of the study identification and selection process for the systematic review. Records were screened on titles and abstracts alone, and reports were assessed based on the full-text content. CENTRAL Central Register of Controlled Trials. WHO-ICTRP World Health Organisation International Clinical Trial Registry Platform.

Table 1. Characteristics of the 31 studies included in this systematic review.

PMID	Publication	Data source	Models of interest	Subtypes	Outcome type	Model ta
31028058(9)	Hanadi El Achi 2019	Micro Brightfield Bioscience (Williston, VT USA)	1	Burkitt lymphoma, diffuse large B-cell lymphoma, Burkitt lymphoma, small lymphocytic lymphoma	Diagnosis	Identification
34067726(10)	Georg Steinbuss 2021	Institute of Clinical Pathology, University of Heidelberg	1	nodal small lymphocytic lymphoma/chronic lymphocytic leukemia, diffuse large B-cell lymphoma.	Diagnosis	classification
32472096(11)	Hiroaki Miyoshi 2020	unclear	3	follicular lymphoma □ diffuse large B-cell lymphoma. follicular lymphoma	Diagnosis	classification
32377574(12)	Charlotte Syrykh 2020	Toulouse University Cancer Institute and Dijon University Hospital, France	1		Diagnosis	Identification
37611165(13)	Sergej Sereda2023	unclear	1	pediatric nodular lymphocyte-predominant Hodgkin lymphoma	Prognosis	Identification, segmentation

32593219(14)	Jianfei Zhang 2020	unclear	1	chronic lymphocytic leukemia mantle cell lymphoma follicular lymphoma	Diagnosis	classification
33244018(15)	Dongguang Li 2020	unclear	1	diffuse large B-cell lymphoma	Diagnosis	classification
33682770(16)	Xiaoli Zhang 2021	unclear	3	chronic lymphocytic leukemia mantle cell lymphoma follicular lymphoma	other	classification
34771625(17)	Wei-Hsiang Yu 2021	17 hospitals in Taiwan	3	monomorphic epitheliotropic intestinal T-cell lymphoma	other	Classification, c segmentation
33835321(18)	Ziv Frankenstein2021	unclear	1	diffuse large B-cell lymphoma follicular lymphoma	other	cell nuclei seg gene signals detec
36691660(19)	Yusuke Takagi 2023	unclear	1	diffuse large B-cell lymphoma follicular lymphoma	other	classification
37167882(20)	Guanghui Yang 2023	Chinese WeChat public websites	1	diffuse large B-cell lymphoma follicular lymphoma	Diagnosis	classification
37443652(21)	Mohammed Hamdi 2023	unclear	2	chronic lymphocytic leukemia mantle cell lymphoma follicular lymphoma	Diagnosis	classification
32067039(22)	Jeffrey S Mohlman 2020	Department of Pathology, Scientific Computing and Imaging Institute, University of Utah, Salt Lake City.	1	Diffuse large B-cell lymphoma Burkitt lymphoma	other	classification
36571610(23)	Yasemin Yuyucu Karabulut 2023	unclear	1	Mycosis Fungoides	Diagnosis	Identification, classification, c segmentation
30119858(24)	Marcelo Zanchetta do Nascimento2018	unclear	4	chronic lymphocytic leukemia mantle cell lymphoma follicular lymphoma	Diagnosis	classification
32854491(25)	Yosep Chong 2020	Yeouido, and Seoul St. Mary's Hospital, College of Medicine, The Catholic University of Korea	1	chronic lymphocytic leukemia mantle cell lymphoma follicular lymphoma Diffuse large B-cell lymphoma	Diagnosis	classification
31111048(26)	Honglin Zhu 2019	unclear	1	chronic lymphocytic leukemia mantle cell lymphoma follicular lymphoma	other	classification
22962572(27)	Siddharth Samsi 2012	unclear	1	follicular lymphoma	Diagnosis	Identification
35147848(28)	Noriaki Hashimoto 2022	unclear	1	Diffuse large B-cell lymphoma Burkitt lymphoma follicular lymphoma	other	classification
25570714(29)	Emmanouil Michail 2014	unclear	1	follicular lymphoma	other	Identification, segmentation
24376080(30)	Evgenios N Kornaropoulos 2014	The Ohio State University with Institutional Review Board (IRB)	2	follicular lymphoma	other	classification
36150505(31)	Wen-Yu Chuang 2022	he archives of Departments of Pathology of two medical centers in Taiwan	1	mantle cell lymphoma	prognosis	semantic segment
34017010(32)	Damir Vrabac 2021	Stanford Hospital	1	Diffuse large B-cell lymphoma	prognosis	semantic segment
37549921(33)	Ila Motmaen 2023	unclear	1	Hodgkin lymphoma	prognosis	classification
36382626(34)	Antonio Santisteban-Espejo 2022	the Pathology Department of the Puerta del Mar University Hospital	1	Hodgkin lymphoma	prognosis	classification
32979109(35)	Zaneta Swiderska-Chadaj 2021	11 hospitals in the Netherlands	1	Diffuse large B-cell lymphoma	MYC translocation detection	Identification
37566457(36)	Mizuki Tagami 2023	unclear	3	Ocular Adnexal Mucosa-associated Lymphoid Tissue Lymphoma	other	classification
34086849(37)	Lina Irshaid 2022	unclear	1	chronic lymphocytic leukemia follicular lymphoma	prognosis	cell segmentation cla
34505705(38)	Siba El Hussein 2022	UTMDACC	1	chronic lymphocytic leukemia	prognosis	cell segmentation cla

35626003(39)	Pingjun Chen 2022	MDACC	1		chronic lymphocytic leukemia	prognosis	cell segmentation□cla				

Table 2. Characteristics of the 42 models of interest from the 31 papers included in this systematic review, grouped by model outcome.

Outcome type	PMID	internal participants	internal pathology image	stain type	original image size	patch size(pixels)	Magnification(s)	Feature extraction	Final model	Final model Prediction precision	cl
Diagnosis	31028058(9)	128	2560	H&E	WSI	40X40	40X	learned	CNN	Patch	unclear
	34067726(10)	629	84139	H&E	WSI	395×395	40X	Hand- crafted	Efficient Net	Patch	Tumor, s
	32472096(11)	388	unclear	H&E	WSI	64X64	5X	Hand- crafted	CNN	Patch	unclear
	32472096(11)	388	unclear	H&E	WSI	64X64	20X	Hand- crafted	CNN	Patch	unclear
	32472096(11)	388	unclear	H&E	WSI	64X64	40X	Hand- crafted	CNN	Patch	unclear
	32377574(12)	378	>378	H&E	WSI	299×299	4X□20X	Hand- crafted	BNN	Patch	tumor-tu
	32593219(14)	unclear	374	H&E	WSI	224×224	unclear	learned	ResNet-50	Patch	cellular morphol ar area
	33244018(15)	1005	3123	H&E	WSI	945×945	40X	learned	17CNN+transform	Patch	Cancer,
	37167882(20)	unclear	12015	H&E	microscopic images	unclear	unclear	learned	GAN	microscopic images	cellular morphol ar area
	37443652(21)	unclear	15000	H&E	WSI	unclear	unclear	Hand- crafted	MobileNet-VGG16□decision-tree-based machine learning	WSI	Cancer,
	37443652(21)	unclear	15000	H&E	WSI	unclear	unclear	Hand- crafted	MobileNet-VGG16□XGBoost	WSI	Cancer,
	36571610(23)	11	60	H&E	microscopic images	600X600	200X	learned	DL	Patch	cellular morphol ar area
	30119858(24)	347	15353	H&E	microscopic images	unclear	1000X	unclear	classification using the polynomial	unclear	cellular morphol ar area
	30119858(24)	347	15353	H&E	microscopic images	unclear	1000X	unclear	random forest	unclear	cellular morphol ar area
	30119858(24)	347	15353	H&E	microscopic images	unclear	1000X	unclear	decision tree	unclear	cellular morphol ar area
	30119858(24)	347	15353	H&E	microscopic images	unclear	1000X	unclear	SVM	unclear	cellular morphol ar area
	32854491(25)	602	>602	IHC	WSI	unclear	unclear		decision-tree-based machine learning	WSI	tumor-tu
	22962572(27)	unclear	12	IHC	WSI	512X512	40X	Hand- crafted	DL	WSI	cellular morphol
Prognosi	37611165(13)	53	>53	IHC	Patch	256×256	unclear	Hand- crafted	YOLOv4-tiny CNN	Patch	nuclear neighbor proportio
	36150505	103	309	H&E	WSI	132X132	40X	Hand- crafted	CNN	Patch	cellular

	05(31)										morphol ar area
	340170 10(32)	209		H&E	WSI	224X224	40X	Hand- crafted	Hover-Net	Patch	cellular morphol ar area
	375499 21(33)	83	>83	Picrosir ius Red	WSI	320X320	20X	Hand- crafted	YOLOv4	Patch	MMP9 negative
s	363826 26(34)	16	80	H&E &IHC	WSI	unclear	40X	Hand- crafted	CNN	WSI	cellular morphol ar area
	340868 49(37)	61	unclear	H&E	WSI	128X128	40X	Hand- crafted	CNN	Patch	cellular morphol
	345057 05(38)	125	213	H&E	WSI	256×256	20X	Hand- crafted	Hover-Net	WSI	cellular morphol ar area
	356260 03(39)	135	213	H&E	WSI	256×256	20X	Hand- crafted	Hover-Net	WSI	cellular morphol ar area
Gene translocation detection	329791 09(35)	287	354	H&E	WSI	unclear	20X	learned	U-Net	WSI	*MYC*+□
	336827 70(16)	uncle ar	374	H&E	WSI	unclear	unclear	Hand- crafted	GA-BP	Patch	cellular morphol ar area
	336827 70(16)	uncle ar	374	H&E	WSI	unclear	unclear	Hand- crafted	BP	Patch	cellular morphol ar area
	336827 70(16)	uncle ar	374	H&E	WSI	unclear	unclear	Hand- crafted	ResNet-50	Patch	cellular morphol ar area
	347716 25(17)	40	40	H&E	WSI	115x115	40X	Hand- crafted	decision-tree-based machine learning	Patch	cellular morphol ar area
	347716 25(17)	40	40	H&E	WSI	115x115	40X	Hand- crafted	decision-tree-based machine learning	Patch	cellular morphol ar area
	347716 25(17)	40	40	H&E	WSI	115x115	40X	Hand- crafted	CNN	Patch	cellular morphol ar area
	338353 21(18)	10	>10	H&E &IHC	WSI	unclear	20X	Hand- crafted	SHIMARIS PAFQ	WSI	cellular morphol ar area
Others	366916 60(19)	842	>842	H&E	WSI	224×224	40X	learned	CNN	Patch	cellular morphol
	320670 39(22)	70	10818	H&E	WSI	224×224	200X	learned	CNN	Patch	cellular morphol ar area
	311110 48(26)	uncle ar	374	H&E	WSI	64X64	unclear	learned	VGG-16□LSTM	Patch	unclear
	351478 48(28)	262	unclear	H&E	WSI	224X224	20X	learned	CNN	Patch	cellular morphol ar area
	255707 14(29)	uncle ar	300	H&E	microscopic images	unclear	400X	Hand- crafted	SVM	microscopic images	cellular morphol ar area
	243760 80(30)	17	500	H&E	WSI	71x71	unclear	Hand- crafted	Classification using orthogonal bases	WSI	cellular morphol ar area
	243760 80(30)	17	500	H&E	WSI	71x71	unclear	Hand- crafted	Laplacian Eigenmaps	WSI	cellular morphol ar area
	375664 57(36)	129	1290	H&E	WSI	2048X2048	4X	Hand- crafted	SVM	Patch	cellular morphol
	375664 57(36)	129	1290	H&E	WSI	2048X2048	20X□	Hand- crafted	SVM	Patch	cellular morphol

37566457(36)	129	1290	H&E	WSI	2048X2048	40X	Hand- crafted	SVM	Patch	cellular morphol

Risk of bias assessment

The results of the PROBAST assessment are presented in Table 3. Although some studies included multiple models of interest, each paper has identified one model with a higher predictive value for bias risk analysis. All models showed either a high overall bias risk (3/31) or an unclear overall bias risk (28/31). None of the models had a low overall bias risk (0/31). The high-risk models had high-risk scores in the Participants, Outcomes, and Analysis sections (3/31). Most of the low-risk scores were found in the Predictors (21/31) and Outcomes (18/31) sections. Almost all papers had an unclear risk of bias in at least one domain, with the most common being the Participants (15/31) and Statistical Analysis (29/31) domains. In the overall utility assessment, the majority were classified as unclear models (16/31), followed by high-risk models (12/31), and low-risk models (2/31). A qualitative summary can be found in Figure 3.

Table3.PROBAST risk of bias assessment results for the 31 papers included in this review.

PMID	Participants	Predictors	Analysis	Outcome
31028058(9)	Unclear	Low	Unclear	Unclear
34067726(10)	High	Unclear	Unclear	Low
32472096(11)	High	Low	Unclear	Low
32377574(12)	High	Low	Low	Low
37611165(13)	Low	Low	Unclear	Low
32593219(14)	Unclear	Low	Unclear	Unclear
33244018(15)	High	Low	Unclear	Low
33682770(16)	Unclear	Unclear	Unclear	Low
34771625(17)	Unclear	Low	Unclear	Low
33835321(18)	High	Low	Unclear	Low
36691660(19)	Unclear	Low	Unclear	Low
37167882(20)	Unclear	Low	Unclear	Unclear
37443652(21)	Unclear	Low	Unclear	Unclear

32067039(22)	High	Low	Low	Low
36571610(23)	Unclear	Low	Unclear	Low
30119858(24)	Unclear	Unclear	Unclear	Unclear
32854491(25)	Low	Unclear	High	high
31111048(26)	Unclear	Low	Unclear	Unclear
22962572(27)	Unclear	Unclear	Unclear	Unclear
35147848(28)	Unclear	Low	Unclear	Unclear
25570714(29)	Unclear	Unclear	Unclear	Unclear
24376080(30)	Low	Low	Unclear	Low
36150505(31)	Unclear	Low	Unclear	Unclear
34017010(32)	Low	Low	Unclear	Low
37549921(33)	High	Unclear	Unclear	Low
36382626(34)	Unclear	Low	Unclear	Unclear
32979109(35)	Unclear	Unclear	Unclear	High
37566457(36)	High	Unclear	Unclear	Low
34086849(37)	High	Low	Unclear	Low
34505705(38)	High	Low	Unclear	Low
35626003(39)	High	Low	Unclear	Low

Data synthesis results

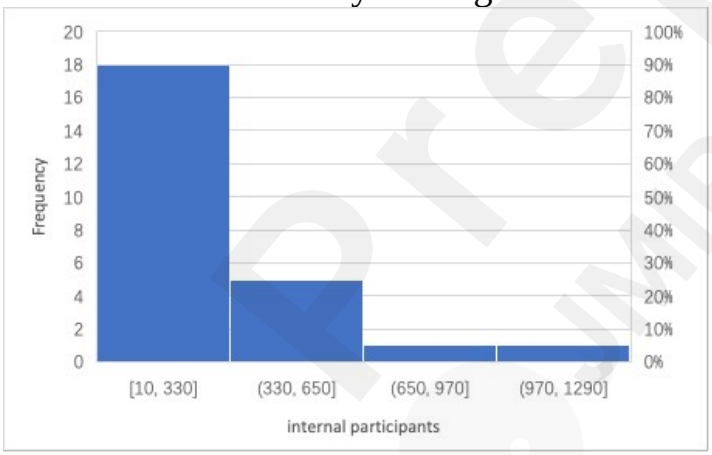
Data in included literature

The number of participants in the internal datasets varied by an order of magnitude, with each study including 10 to 1,005 lymphoma patients. Most studies utilized data from NHL patients (28/31), with B-cell lymphoma being the most prevalent (26/31), Among B-cell lymphomas, follicular lymphoma subtype

and diffuse large B-cell lymphoma subtype accounted for relatively high proportion, which were 16/26 and 12/26, respectively.and fewer studies focused on T-cell lymphoma (2/31). Studies involving HLwere less common (3/31). Only one study explicitly included prospective data collection, and in this study, the internal training set consisted of 209 patients with an overall low risk of applicability, making it suitable for a subset of external validation.

As depicted in Figure 2, the number of pathological slides used typically far exceeds the number of patient participants. In most studies, the samples for model development were Whole Slide Images (WSIs) containing excised or biopsied tissue (26/31), with other samples using single Tissue Microarray (TMA) core images (3/31) or pre-trimmed digital pathology images (2/31). Most studies utilized Hematoxylin and Eosin (H&E) stained tissues (25/31), while others employed various Immunohistochemical (IHC) staining methods (4/31), and two papers used both H&E and IHC staining (2/31). One study employed a multimodal analysis method that integrated pathological images with clinical information [19].

The origin of the data was unclear in most studies (16/31). Among the studies with identifiable data sources, some used data from a single center (11/31), while a smaller number of studies utilized data from multiple centers (4/31). Countries from which the data originated included China, Taiwan, the United States, France, South Korea, Spain, and the Netherlands. The United States was the most common country of origin for the data.



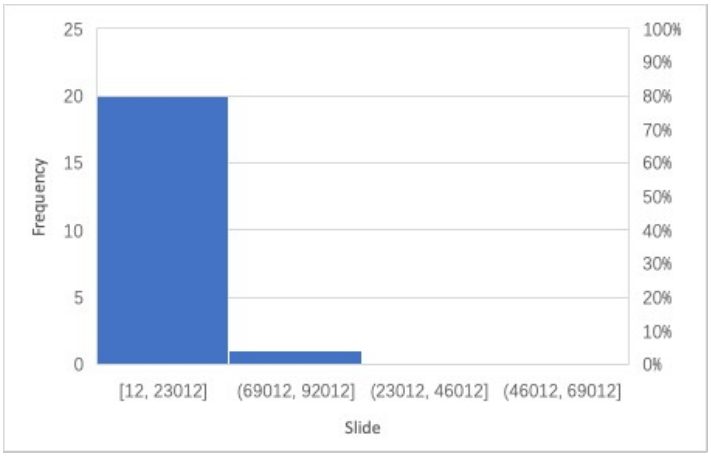


Fig. 2 Number of patients and slides per model. Histograms showing the number of lymphoma patients and histopathology slides used in model development. Many of these values are uncertain due to incomplete reporting, as reflected in Table 2.

Methods in included literature

Among the 31 papers reviewed, there were a total of 42 models of interest, with each paper containing between 1 to 4 such models. These included 16 diagnostic models, 8 prognostic models, 1 to detect whether the gene was translocation model and 17 models for other predictive diagnostic-related information. The tasks of these models encompassed identification tasks (6/42), classification tasks (31/42), and segmentation tasks (5/42).

Various models were utilized in the included studies. In the final analysis of the models, the most common types were Convolutional Neural Networks (CNN) (18/31), Support Vector Machines (SVMs) (2/31), and Random Forests (2/31), with one study employing an Adversarial Neural Network (1/31). CNN architectures included MobileNet-VGG16, HoVer-Net, U-Net, and ResNet-50. New CNNs typically employed multiple standardized blocks involving convolutional, normalization, activation, and/or pooling layers(16).One study utilized transfer learning (15),and integrated 17 deep learning models to establish a high-precision deep learning platform, enhancing the model's generalization capability and achieving a diagnostic accuracy rate of 100%. Another study generated their novel architecture using a topological optimization method on the standard VGG16(21). Some studies combined traditional machine learning with deep learning, using CNNs to extract features followed by decision tree-based methods for quantification and classification (17, 21). One study employed a CNN based on Multiple Instance Learning (MIL), which can automatically focus on image patches from the tumor region of interest (28).

In the models included in the analysis, most studies commonly used patches (19/31), with some models operating at the Whole Slide Image (WSI) level

(8/31). Two different aggregation methods were used: pre-modeling aggregation and post-modeling aggregation. The former required the generation of slide-level features before modeling, while the latter involved aggregating patch-level model outputs for slide-level predictions. For models using patch images as the final modeling images, it was necessary to segment the original images into patches for individual processing before modeling, with patch sizes ranging from 40×40 to 2048×2048 pixels, with the most common sizes being 224×224 pixels (4/31) and 256×256 pixels (3/31). A range of feature extraction techniques were then employed, including handcrafted/predefined features (19/31) and features automatically learned by models (10/31). Handcrafted features included a wide array of texture, color, cellular, and nuclear morphological features. These handcrafted features were typically used as inputs for classical machine learning methods, such as SVM and Random Forest models. Learned features were often extracted using CNNs, which were also commonly used for classification. Finally, patch-level model outputs were aggregated to generate predictive models using methods such as attention-based weighted averaging, concatenation, and more complex embedding techniques like Fisher Vector encoding or k-means clustering, by taking the maximum value(27, 32)

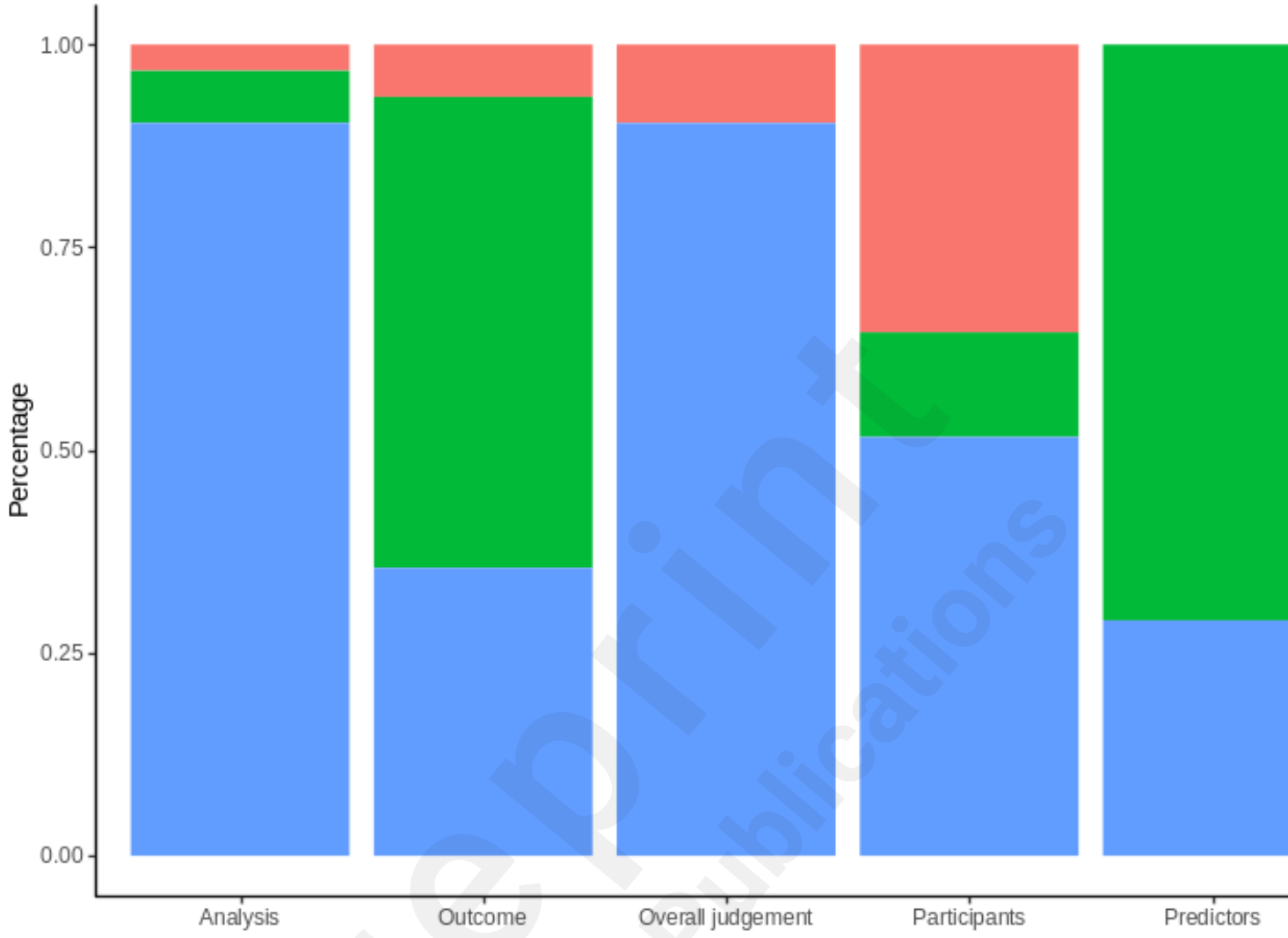


Fig. 3 PROBAST risk of bias results.

The machine learning models in the included studies commonly undertook tasks including identification (6/42), classification (31/42), and segmentation (5/42), with some models encompassing identification, segmentation, and classification tasks(17, 23, 29) One study also employed segmentation tasks to detect gene signals in WSI images identified by FISH (18). Several studies also used segmentation to determine the relationship between various features of tissue, cell, and nuclear count and spatial patterns (13). In the papers where the magnification could be determined, the most common modeling magnifications were ×20 (6/31) and ×40 (12/31). A few studies used different magnifications to select informative tissue areas and for modeling (11, 12, 36).

Very few models integrated histopathological data with other modalities (1/31). Multimodal approaches included the pre-modeling connection of separately extracted unimodal features and the merging of unimodal predictions from different models(40), while transformer-based methods were also commonly used for encoding the relationships between modalities (41, 42). Attention-based

methods have been applied to other malignancies for several years (43), but have only recently emerged in lymphoma research. In the included studies, one study employed a variant of the Transformer architecture to encode the relationship between medical images and clinical records, proposing a personalized attention mechanism (PersAM) method for the classification of lymphoma subtypes(19).

Data synthesis results

In internal validation, less than half of the model outcomes were assessed using cross-validation (7/31), and external validation using independent lymphoma data was rarely performed (7/31). The cross-validation methods included k-fold (4/31) with 5-10 folds. Some papers described the selection of hyperparameters using the training set but only evaluated using a test set from the same data source(17, 18, 29). The models for external validation were trained on WSIs)and validated on WSIs (6/7) or TMAs (1/7) from independent data sources. One study's model was externally validated using data from normal lymph node tissue(10). In one study, a model that demonstrated perfect validation accuracy (AUC = 1.0) with internal validation failed to perform effectively on external cases, with an AUC ranging between 0.63 and 0.69. This may be due to the sensitivity of machine learning techniques to preprocessing steps, as neural networks, adhering to statistical laws, require a representative sample to achieve coherent inductive reasoning [12]. Another study employed polynomial, support vector machine, random forest, and decision tree classifiers to evaluate the performance of the proposed method(24).

Most classification models were assessed using accuracy, balanced accuracy, and/or the area under the receiver operating characteristic curve (AUC). One study reported only a p-value, demonstrating that WSIs can be incorporated into the histopathological assessment of Hodgkin's lymphoma as an appropriate method for quantifying different cellular populations(34).Performance of survival models is typically reported using AUC, with other metrics including p-values, accuracy, hazard ratios, and the C-index, which is similar to AUC.

The demand for artificial intelligence methods is growing, and the lack of interpretability is a barrier to their clinical use (44, 45). The interpretability of AI can help to enhance the trust of medical professionals in future AI systems. Upon analyzing the included studies, it was found that the majority have conducted interpretability analysis of the models (18/31), with a smaller subset conducting visual interpretability analysis of the histopathological images that influenced the model's prognostic judgment (7/18). Several studies characterized the spatial location and relationships of typical cells and their surrounding cells, nuclei, and microenvironments of the regions of interest(17, 28, 32, 38), demonstrating interpretability. One study presented graphical features associated with clinical prognostic information (33). A few studies employed traditional machine learning modeling, such as decision trees (21, 25), which are inherently more understandable as they simulate human thought processes to make decisions.

DISCUSSION

In this review, which includes studies on artificial intelligence for the purpose of lymphoma histopathology diagnosis or prognosis, in general, there were more research models for B-cell lymphoma than for T-cell lymphoma, the overall risk assessment is either high risk or unclear, with no studies classified as low risk. Due to incomplete reporting, studies frequently present an unclear risk of bias in the Participants and Analysis domains of PROBAST. Information that is often missing includes the specific source of patients, the number of patients included, the number of samples/images used, whether any patients/images were excluded, and the methods of processing and digitizing the tissue. Additionally, most studies reporting on data sources are single-center studies, with only one study utilizing images from an accessible public dataset(20), which may be attributed to AI researchers not spending sufficient time to understand these images, whether as a training set or for external validation. Overall, only four papers were found to have a low risk of bias regarding participants, which included a clear and reasonable patient recruitment strategy and selection criteria(13, 25, 30, 32). Information about predictors (histopathological images and their features) is generally well described and interpreted, but there is still a lack of description or execution of some key details, such as whether the predictors were assessed without knowledge of the outcomes, or it is unclear whether all tissue samples underwent similar processing, which could lead to a risk of bias due to visual heterogeneity. Furthermore, some researchers use a limited amount of data and analyze a single test data split without employing any methods to address overfitting and model optimism (cross-validation, external validation). These limitations are common in lymphoma AI research, leading to weak validation and a high risk of bias in the models.

Sharing code can help mitigate the impact of incomplete reporting and significantly improve reproducibility, but out of 31 papers, only five shared codes, some of which appeared to be incomplete or inaccessible. Better code repositories include detailed documentation to aid reproducibility, including information on environment setup, functionality overview, result production, and the code itself(10, 12, 15, 21, 33)

Some studies aim to provide interpretability for DL tools using current methods, including post hoc approaches or supervised ML models to interpret the outputs after DL models have made predictions(46, 47). In the field of AI research for lymphoma, most studies currently provide personalized interpretability for analysis, including visual attention heatmaps, traditional ML showcasing the spatial location of feature areas, and relationships. However, traditional ML, which is often developed in collaboration with domain experts, can offer more interpretability as it is based on handcrafted features. Nevertheless, handcrafting features is a challenging and non-trivial task due to the significant time investment required by pathologists or oncologists in developing such methods.

In recent years, there has been a rise in hybrid approaches that integrate DL and handcrafted strategies, which may involve using DL algorithms for initial detection of cells or elements, followed by relying on easily interpretable traditional ML methods for predictions, thereby leveraging domain knowledge to ensure the biological interpretability of the approach(48).

Development of the field

The field of AI in lymphoma histopathology diagnosis and prognosis is rapidly evolving, with a significant increase in published research since the beginning of 2019. Most of this research involves the use of deep neural networks for automatic feature extraction and classification, while a minority employs traditional machine learning models(17, 25, 29, 36) Recent studies have investigated a broader range of diagnostic outcomes, including the diagnosis of lymphoma subtypes (12, 15) prognosis (31, 32) and the detection of genetic translocations (35)

Despite the progress of AI in lymphoma research, there has been no clear trend in the size of the datasets used over time, in terms of either the number of slides or the number of participants. Similarly, there is no evidence to suggest that more recent studies have included stricter internal validation, but the frequency of external validation has been increasing. Before 2020, none of the studies included any external validation on lymphoma data, but seven recently published studies have(15, 20, 25, 32, 35, 38, 39). Although these external validations are often limited to a small amount of data, with only two of them featuring larger datasets (15, 25), the inclusion of any external validation indicates progress over previous research. These validations are crucial for the clinical utility of these models, as real-world implementation requires robustness to visual heterogeneity from different sources, which can vary between and within data centers over time. As the field continues to mature, we anticipate seeing more studies that validate their models using larger, high-quality independent datasets, including explicit reporting on patient recruitment and selection protocols, histopathology slide preparation, and digitization. This will help to reduce the bias, limited reproducibility, and limited generalizability that are prevalent in most current research in the field.

Regarding multimodal research in oncology, there has been a rapid increase in published studies since the beginning of 2019, with only one study on multimodal research in lymphoma published in 2023(19). The assessment of tissue sections remains the gold standard for cancer diagnosis, and even trained pathologists may require assistance from certain biomarker tests. Multimodal research, which integrates multi-dimensional information such as genomics, proteomics, transcriptomics, and clinical information, has contributed to the

direction of cancer research(49, 50). Multimodality can enable a fusion approach of two modalities, where a pathologic image can be input and another modality (sequencing, other formats of images) can be output(40).This method does not simply summarize information from pathologic images but transforms it. Thus, a well-trained multimodal model can be applied to the pathologic images of patients without medical conditions to obtain indicators related to precision medicine, such as genetic sequences. We hope that with the development of high-throughput technologies, transcriptomics, metabolomics, proteomics, and other technologies, there will be more multimodal analyses that integrate multidimensional omics with histopathological images in the future. This will help to promote the application of artificial intelligence in clinical scenarios

Current limitations and future recommendations

A significant portion of published work does not provide sufficient clinical and pathological information to assess the risk of bias. Consequently, AI researchers must thoroughly report the provenance of data to understand the degree of heterogeneity within the dataset and to ascertain whether this has been appropriately considered in the study design. The modeling and analytical methods must also be comprehensively reported to enhance reliability and reproducibility.

To further improve reproducibility, we recommend that researchers provide code and data whenever possible. Digital pathology studies in lymphoma are currently constrained by the lack of publicly available data. Additionally, datasets often lack multicenter contributions, and the heterogeneity of images from different scanners at various centers can introduce confounding factors, making it challenging for researchers to train robust models or assess generalizability. This poses risks of bias and confusion. Therefore, we recommend that datasets should, as much as possible, include multicenter sources to enhance the generalizability and clinical applicability of models. During the image preprocessing phase, efforts should be made to eliminate confounding factors caused by staining variations, bubbles, and other artifacts.

For AI to achieve clinical utility, more robust validation is essential, especially considering the limitations of available datasets. We suggest that researchers consistently conduct thorough analyses using cross-validation, external validation, and other techniques to ensure the robustness of the results and the ability to generalize to new data. It is also crucial to report confidence intervals for the results (typically 95% confidence intervals), particularly when comparing multiple models. This helps to discern whether one model is genuinely superior to another or if the differences are due to chance.

The 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

METHODS

Literature search

A comprehensive search was conducted across three major research databases: PubMed, Cochrane, and Web of Science. The databases were limited to peer-reviewed journals and conference proceedings to ensure the integrity of the included studies. The search spanned from the inception of each database up to February 28, 2024. The search strategy was composed of three distinct aspects: artificial intelligence, lymphoma, and histopathology. The search query was formulated as follows: (("deep learning" OR "Deep learning" OR "Whole slide image" OR "artificial intelligence" OR "machine learning" OR "neural network" OR "support vector" OR "digital pathology") AND ("lymphoid neoplasm" OR lymphoma OR lymphoproliferative")). For each aspect, multiple relevant terms were combined using the OR operator (e.g., "artificial intelligence" OR "machine learning"), and then these terms were merged using the AND operator to ensure that the retrieved studies met all three criteria. Each search engine utilized the broadest possible set of search fields, except for Cochrane, which imposed restrictions to avoid searching within the reference lists of articles, a feature not available in other search engines. The terms "ML" (Machine Learning) and "AI" (Artificial Intelligence) were limited to specific domains due to the diversity of their potential meanings. To ensure the most rigorous literature search, no restrictions were placed on publication dates or article types during the search process.

Many artificial intelligence methods are based on statistical models, such as logistic regression, which can blur the lines between disciplines. The search method adopted in previous reports involved searching for typical artificial intelligence methods by name (e.g., neural networks) and for other methods by whether the authors described their work as artificial intelligence. The review protocol was registered on PROSPERO (CRD42024537394) before the search results were screened for inclusion."

Literature selection

A researcher (Y.F.) manually removed duplicate papers with the assistance of the reference management software EndNote 20. Subsequently, another researcher (ZY.H.) independently screened the articles for inclusion in two 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 (Y.L.).

The inclusion criteria required that studies evaluate the application of at least one AI method for the diagnosis, prognostic inference, or genetic mutation prediction of suspected or confirmed lymphoma cases using human histopathology images. Studies were included only if the AI method was directly applied to digital pathology images or if features were automatically extracted

from the images. Basic tasks such as segmentation and cell counting were included as these can be utilized by pathologists for computer-assisted diagnosis. Conventional optical microscopy images, including whole slide images scanned by scanners for fluorescence in situ hybridization (FISH), were considered, while other imaging modalities such as hyperspectral imaging were excluded. Publications that were not primary research, such as review papers, were excluded. Non-English articles and studies without access to the full manuscript were also excluded.

If the models in the included studies met the same inclusion criteria, they were models of interest. If multiple model outcomes were evaluated within the same study, one model of interest was selected for that study, regardless of whether the modeling methods were similar or not. The same model outcomes at different levels of accuracy (e.g., patch level, slide level, patient level) were not considered to be different models of interest. Models did not need to be entirely independent; for example, the output of one model of interest could serve as the input for another model of interest, provided that the model performance of each model was assessed separately

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)(51),The tool evaluates the likelihood that reported results are distorted due to limitations in study design, conduct, and analysis. PROBAST includes 20 guiding questions categorized into four 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 the 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/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 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 five-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 lymphoma are included, it is not acceptable for one 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 H&E 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 two independent researchers (Y.F., ZY.H.), with at least one computer scientist and one pathologist involved in the bias risk assessment of each model.

Data synthesis

Data extraction was independently conducted by two researchers (Y.F and MX.Z.), utilizing a structured table that encompasses 67 fields, including categories such as literature titles, data, methods, results, and more. A summary of this process is provided in Supplementary Table 1.

When translating the full text, references, and appropriate supplementary materials are consulted to ensure the information is accurate. Inferences are only made when both researchers are confident that the correct information has been identified, and any discrepancies are resolved through discussion. Fields that could not be conclusively determined are marked as unclear.

All extracted data are consolidated into two tables, one for study-level characteristics and the other for model-level features. These tables include only the models of interest. Model outcomes refer to the outputs of the models, which could be clinical

outcomes (diagnosis/prognosis) or diagnostic-related results that can be used for computer-assisted diagnosis, such as the classification of subtypes. Due to the diversity in inclusion methods and model outcomes, a meta-analysis was not included in the data synthesis. We adhered to the PRISMA 2020 guidelines for reporting systematic reviews, with the checklist provided in the Supplementary Table.

Abbreviations

AI, artificial intelligence
HL, Hodgkin's lymphoma
NHL, non-Hodgkin's lymphoma
H&E, hematoxylin and eosin
IHC, immunohistochemistry
WSI, whole slide images
ML, machine learning
DL, deep learning
CNN,Convolutional Neural Networks
SVM, support vector machine
TMA, Tissue Microarray
AUC, receiver operating characteristic curve

Declarations

Ethics approval and consent to participate
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author Contributions
ZY.H. Provided ideas for literature writing, YF drafted and developed literature search strategies, conducted literature search, wrote the manuscript, and MX.Z., LN.X., XD,D.drew graphs. Y.L. discusses and arbitrates differences.ZY.H. directed the writing of the manuscript. The authors read and approved the final manuscript.

Funding
No funding.

Availability of data and materials
Not applicable.

REFERENCES

1. Huang J, Pang WS, Lok V, Zhang L, Lucero-Prisno DE, 3rd, Xu W, et al. Incidence, mortality, risk factors, and trends for Hodgkin lymphoma: a global data analysis. J Hematol Oncol. 2022;15(1):57.
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.
3. 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. 2024.
4. [Chinese expert consensus on detection techniques and interpretation of lymphoma gene rearrangement (2023 version)]. Zhonghua Bing Li Xue Za Zhi. 2023;52(6):558-65.
5. Ferry JA. Scientific Advances and the Evolution of Diagnosis, Subclassification and Treatment of Lymphoma. Arch Med Res. 2020;51(8):749-64.
6. 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. 2018;73(5):784-94.
7. 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. 2022;84:129-43.
8. 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. 2023;147(10):1178-85.
9. 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. 2019;49(2):153-60.
10. Steinbuss G, Kriegsmann M, Zgorzelski C, Brobeil A, Goeppert B, Dietrich S, et al. Deep Learning for the Classification of Non-Hodgkin Lymphoma on Histopathological Images. Cancers (Basel). 2021;13(10).
11. 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. 2020;100(10):1300-10.
12. Syrykh 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.
13. 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. 2023;7(20):6285-9.
14. 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. 2020;47(9):4241-53.
15. 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. 2020;11(1):6004.
16. 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):335-44.
17. Yu WH, Li CH, Wang RC, Yeh CY, Chuang SS. Machine Learning Based on Morphological Features Enables Classification of Primary Intestinal T-Cell Lymphomas. Cancers (Basel). 2021;13(21).
18. 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. 2021;51(1):4.
19. 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.
20. Yang G, Qin Z, Mu J, Mao H, Mao H, Han M. Efficient diagnosis of hematologic malignancies using bone marrow microscopic images: A method based on MultiPathGAN and MobileViTv2. Comput Methods Programs Biomed. 2023;237:107583.
21. 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). 2023;13(13).
22. 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. 2020;153(6):743-59.
23. Karabulut YY, Dinç U, Köse E, Türsen Ü. Deep learning as a new tool in the diagnosis of mycosis fungoides. Arch Dermatol Res. 2023;315(5):1315-22.

24. 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. 2018;163:65-77.
25. Chong Y, Lee JY, Kim Y, Choi J, Yu H, Park G, et al. A machine-learning expert-supporting system for diagnosis prediction of lymphoid neoplasms using a probabilistic decision-tree algorithm and immunohistochemistry profile database. J Pathol Transl Med. 2020;54(6):462-70.
26. 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.
27. 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. 2012;3(5):269-79.
28. 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. 2022;17(7):1379-89.
29. Michail E, Dimitropoulos K, Koletsa 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;2014:3374-7.
30. Kornaropoulos EN, Niazi MK, Lozanski G, Gurcan MN. Histopathological image analysis for centroblasts classification through dimensionality reduction approaches. Cytometry A. 2014;85(3):242-55.
31. Chuang WY, Yu WH, Lee YC, Zhang QY, Chang H, Shih LY, et al. Deep Learning-Based Nuclear Morphometry Reveals an Independent Prognostic Factor in Mantle Cell Lymphoma. Am J Pathol. 2022;192(12):1763-78.
32. 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. 2021;8(1):135.
33. 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. 2023;111(5):722-8.
34. Santisteban-Espejo A, Bernal-Florindo I, Perez-Requena J, Atienza-Cuevas L, Catalina-Fernandez I, Fernandez-Valle MDC, et al. Identification of prognostic factors in classic Hodgkin lymphoma by integrating whole slide imaging and next generation sequencing. Mol Omics. 2022;18(10):1015-28.
35. 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. 2021;479(3):617-21.
36. 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. 2023;48(12):1195-202.
37. 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. 2022;146(2):182-93.
38. 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. 2022;256(1):4-14.
39. 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). 2022;14(10).
40. Qiao Y, Zhao L, Luo C, Luo Y, Wu Y, Li S, et al. Multi-modality artificial intelligence in digital pathology. Brief Bioinform. 2022;23(6).
41. 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.
42. 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. 2023;24(3).
43. 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. 2021;5(6):555-70.
44. Madabhushi A, Lee G. Image analysis and machine learning in digital pathology: Challenges and opportunities. Med Image Anal. 2016;33:170-5.
45. Rudin C. Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead. Nat Mach Intell. 2019;1(5):206-15.
46. 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. 2018;42(12):1636-46.
47. 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. 2019;143(7):859-68.
48. 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. 2017;7(1):13543.
49. 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;19:2719-25.
50. 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). 2021;11(4).
51. Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. Ann Intern Med. 2019;170(1):51-8.

Supplementary Files