

SYSTEMATIC REVIEW

Open Access



# Automated scoring methods for quantitative interpretation of Tumour infiltrating lymphocytes (TILs) in breast cancer: a systematic review

Nurkhairul Bariyah Baharun<sup>1,2\*</sup>, Afzan Adam<sup>3</sup>, Mohamed Afiq Hidayat Zailani<sup>1,5</sup>, Nasir M. Rajpoot<sup>4</sup>, Qiaoyi XU<sup>3</sup> and Reena Rahayu Md Zin<sup>1</sup>

## Abstract

Tumour microenvironment (TME) of breast cancer mainly comprises malignant, stromal, immune, and tumour infiltrating lymphocyte (TILs). Assessment of TILs is crucial for determining the disease's prognosis. Manual TIL assessments are hampered by multiple limitations, including low precision, poor inter-observer reproducibility, and time consumption. In response to these challenges, automated scoring emerges as a promising approach. The aim of this systematic review is to assess the evidence on the approaches and performance of automated scoring methods for TILs assessment in breast cancer. This review presents a comprehensive compilation of studies related to automated scoring of TILs, sourced from four databases (Web of Science, Scopus, Science Direct, and PubMed), employing three primary keywords (artificial intelligence, breast cancer, and tumor-infiltrating lymphocytes). The PICOS framework was employed for study eligibility, and reporting adhered to the PRISMA guidelines. The initial search yielded a total of 1910 articles. Following screening and examination, 27 studies met the inclusion criteria and data were extracted for the review. The findings indicate a concentration of studies on automated TILs assessment in developed countries, specifically the United States and the United Kingdom. From the analysis, a combination of semantic segmentation and object detection ( $n = 10$ , 37%) and convolutional neural network (CNN) ( $n = 11$ , 41%), become the most frequent automated task and ML approaches applied for model development respectively. All models developed their own ground truth datasets for training and validation, and 59% of the studies assessed the prognostic value of TILs. In conclusion, this analysis contends that automated scoring methods for TILs assessment of breast cancer show significant promise for commodification and application within clinical settings.

**Keywords** Automated scoring, Breast cancer, Tumor-infiltrating lymphocytes, Digital pathology

\*Correspondence:

Nurkhairul Bariyah Baharun  
p110546@siswa.ukm.edu.my

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

## Introduction

Tumour microenvironment (TME) of breast cancer mainly comprises malignant, stromal and immune cells. TME plays a pivotal role in the tumorigenesis, progression, and metastatic spread of many cancers including breast cancer. Tumour infiltrating lymphocytes (TILs) are part of TME and evaluating it is crucial for determining the disease's prognosis. TILs include various type of lymphocytes that have migrated into the TME and play an important role to fight against cancerous cells particularly in highly proliferative breast cancers such as triple negative breast cancer (TNBC) and human epidermal growth factor receptor 2 (HER-2) positive breast cancer subtypes. TILs are a diverse group of immune cell types including cytotoxic CD8+T-cells, natural killer (NK) cells, macrophages, T-helper cells, immune suppressing B-cells and regulatory CD4+T-cells [1]. TILs can be found in the tumor-associated stroma or embedded in the tumor area. TILs in direct cell-to-cell contact with tumor cells, with no stroma between them known as intratumoral lymphocytes (iTILs), whereas stromal tumor-infiltrating lymphocytes (sTILs) are scattered or grouped TILs between carcinoma cells, and do not interact directly with tumor cells. TILs assessment has also been shown to provide important prognostic information for various types of solid tumors, including breast cancer [2]. Vicent et al. (2022) revealed that node negative TNBC with high sTILs ( $\geq 75\%$ ) have an excellent prognosis [3]. They provided evidence supporting the potential prognostic function of stromal TILs in TNBC patient. Therefore, TILs have been proposed as potential biomarkers for routine histopathological examinations and have been suggested for evaluating residual disease after neoadjuvant chemotherapy (NACT) [4]. TILs scoring also has been regarded as a vital part of the TNBC prognosis workflow and HER-2 positive breast cancer [5]. A recent study also suggested that immunotherapy potential can be evaluated or predicted using WSI-based assessments [6]. In addition to the immunological aspects of breast cancer management, effective treatment often involves multimodal strategies, including surgery, chemotherapy, targeted therapy, and radiation therapy. Radiation therapy, particularly post-mastectomy radiation therapy (PMRT), is a crucial component for patients at high risk of local recurrence. PMRT has been shown to significantly improve survival rates by targeting residual microscopic disease that may persist post-surgery [7].

Various studies have shown that TILs and the spatial characterization of WSIs in histopathological sections provide diagnostic and prognostic values for TNBC and HER-2 positive breast cancers [8–10]. Therefore, accurate detection and quantification of TILs are important tasks for researchers to develop a standard and reproducible

method for the clinical validity of TILs scoring, which can preferably be validated in several independent populations and thus provide biomarkers with strong prognostic and predictive power for cancer progression and therapeutic efficacy.

Incorporating TILs into routine clinical practice for TNBC is supported by international clinical and pathology standard (St. Gallen 2019, WHO 2019, and ESMO 2019) [11–13]. However, visual TILs assessment (VTA) or manual TILs assessment is susceptible to high inter-observer variability and ambiguity due to the lack of adequate standardization and training [9, 13, 14]. Subsequently, the TIL Working Group (TIL-WG), produced several published guidelines to standardize VTA in solid tumors and to enhance reproducibility and clinical adoption [2, 15]. The use of a standardized methodology for TILs assessment as a reference standard will help to resolve many issues associated with TILs scoring in future studies. Nevertheless, visual TILs assessments have some limitations, including inter-reader variability, time constraints in routine practice, and subjectivity, which may introduce bias [9, 16]. To overcome these issues, automated image analysis methods are required to reduce labour costs and provide consistent and accurate TILs evaluation. Research into automated TILs scoring has gained more notable surge and showing increasing trend from 2017. This is because automated scoring models for TILs assessment offering more accurate, efficient, and reproducible method of assessing TILs compared to manual scoring.

Automated TILs quantification refers to the method that utilizes computational algorithm and image analysis techniques to quantify TILs, while deep learning (DL) methods is a subset of automated methods that utilize artificial neural networks to learn from the data. DL approaches can learn complex pattern and hierarchical presentation leading to high-performance TILs quantification providing highly accurate and robust TILs detection. Automated TILs assessments have great potential to address the fundamental limitations of visual TILs assessments. There is evidence that computational algorithms have been successfully commercialized and used in medicine, such as pap smear cytology analysers [17], blood analysers [18] and automated immunohistochemistry (IHC) procedures for ER, PR, HER-2, and Ki67 [19–23]. Despite that, the automated TILs assessments have been utilized in various clinical application. The use of this approach remains limited in breast cancer due to the lack of standardized protocols and guidelines for automated TIL quantification, causes inconsistency between studies and makes it challenging to compare outcomes. TILs assessment in breast cancer plays a significant roles as prognostic biomarker and predictive value for treatment

response. Recent advances in the field of computational pathology suggest that automated TILs scoring methods carry a significant potential for commercialization and deployment in a clinical setting. This systematic review was conducted to access the evidence on the approaches and performance of automated TILs scoring method in breast cancer.

## Materials and methods

### Study design

This systematic review was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guideline [24]. The protocol for this systematic review was registered in The International Prospective Register of Systematic Reviews (PROSPERO) database (Registration number CRD42023418519).

### Systematic search strategy

To ensure comprehensive review for the relevant studies reported on the approaches and performance of the automated scoring techniques for TILs assessments in breast cancer, the systematic search was conducted across four major databases: Scopus, Web of Science, PubMed, and Science Direct. Two main search techniques were used: an advanced search technique and a manual search of the four main databases, as mentioned above. The main keywords used in the searches were “artificial intelligence,” “tumor-infiltrating lymphocytes,” and “breast cancer.” Enriched keywords for artificial intelligence, include automated scoring, digital pathology, automated quantification, digitalization, computational, and standardized methodology. For tumor-infiltrating lymphocytes, the enriched keywords are lymphocytes, immune microenvironment, tumor-derived activated cells and for breast cancer, the enriched keywords include breast neoplasm, Triple negative breast cancer ER-Negative PR-Negative HER-2 negative breast cancer, ER-negative PR-negative HER-2 negative breast neoplasms, and lymphocyte-predominant breast cancer. The Boolean operators “OR” and/or “AND” were used to combine the keywords for the advanced search process. Manual search, handpicking, backward tracking, and forward tracking were used to identify the relevant articles. The articles were restricted to those published between January 2017 and November 2023. The search was enhanced using an asterisk (\*) as a wildcard sign to include different word endings. A summary of the search strategies is presented in Table 1.

### Article screening and quality assessment

We employed Rayyan AI, an advanced tool specifically developed for systematic reviews, to manage and

screen the papers effectively. The Population, Intervention, Comparator, Outcome, Study (PICOS) criteria was applied to define search strategies, and guide articles selection (Table 2). This review identified cross sectional diagnostic accuracy study that focused on TILs assessment in breast cancer that were compared with manual ground truth dataset and the performance of the developed model. Selected articles identified through the literature search were screened for eligibility. The inclusion criteria were as follows: (i) the article meets all requirements in the P.I.C.O.S criteria (ii) the articles were published between January 2017 and November 2023, and (iii) the articles used the English language. The publication years were chosen to ensure study maturity, as proposed by Kraus, (2020) [25]. Eligible studies were restricted to the English language to ensure clarity and accessibility of the extracted data, and to avoid linguistics confusion for the investigators. Linares (2018) recommended using English as an inclusion criteria for investigator because reading articles in languages other than English could result in confusion, higher time-consumption and cost [26]. Studies were excluded if: (i) the articles did not fulfil the PICOS criteria, (ii) the articles were not original research, including letters, newsletters, editorials, book chapters, or case studies; (iii) duplication of studies; or (iv) studies reported using other than English language.

The selection process involved three phases: First, the title and the abstract of the articles were screened by one investigator, and those that failed to meet the inclusion criteria were excluded. Then, the remaining articles were carefully examined and screened by two investigators based on the abstract and in cases of uncertainty, the full paper was downloaded. Lastly, the quality and risk of bias in each study were assessed by two investigators using the QUADAS-2 tool to characterize each study in four domains (Patient election; Index test; Reference standard; Flow and timing) [27]. Any disagreements regarding article selection and categorization were resolved through discussion.

### Data extraction and analysis

The data extraction process involved manual retrieval of information from each study, including the first author's name, publication year, study location or country, sample size, sample type, source of sample, staining type, scanner type, method for detecting TILs, approach to developing ground truth, the measured location of TILs, clinical outcome, and correlation between automated approach and pathologist consensus on TILs assessment. Subsequently, the collected data was organized and compiled into tables.

**Table 1** Summary of search strategy

Main keywords	Enriched keywords	Search string
Artificial intelligence Tumour infiltrating lymphocyte Breast cancer & Triple negative breast cancer	Artificial intelligence = Automated scoring, digital pathology, automated quantification, digitalization, computational, standardized methodology Tumour infiltrating lymphocyte = lymphocytes, immune microenvironment, Tumor-Derived Activated Cells Breast cancer = Breast neoplasm, Triple negative breast cancer = Triple-Negative Breast Neoplasms, ER-Negative PR-Negative HER-2 Negative Breast Cancer, ER-Negative PR-Negative HER-2 Negative Breast Neoplasms, lymphocyte-predominant breast cancer	<b>Scopus</b> TITLE-ABS-KEY (("artificial intelligence" OR "automated scoring" OR "digital pathology" OR "automated quantification" OR "digitalization" OR "computational") AND ("tumour infiltrating lymphocyte" OR "lymphocyte" OR "immune microenvironment" OR "tumor microenvironment") AND ("breast cancer" OR "breast neoplasm" OR "triple negative breast cancer" OR "triple negative breast neoplasm" OR "ER-negative PR-negative HER-2 negative breast cancer"))
		<b>Web of Science (WOS)</b> ALL=((("Artificial Intelligence" OR "Automated scoring" OR "digital pathology" OR "automated quantification" OR "digitalization" OR "computational") AND ("Tumour infiltrating lymphocyte" OR "lymphocyte" OR "immune microenvironment" OR "Tumor-Derived Activated Cell" OR "tumour associated leukocyte" OR "tumor microenvironment") AND ("Breast cancer" OR "Breast neoplasm" OR "Triple negative breast cancer" OR "Triple-Negative Breast Neoplasm" OR "ER-Negative PR-Negative HER-2 Negative Breast Cancer" OR "ER-Negative PR-Negative HER-2 Negative Breast Neoplasm")))
		<b>Science Direct</b> (("Artificial Intelligence" OR "Automated scoring" OR "digital pathology" OR "automated quantification") AND ("Tumour infiltrating lymphocyte" OR "immune microenvironment") AND ("Breast cancer" OR "Triple negative breast cancer" OR "ER-Negative PR-Negative HER-2 Negative Breast Cancer"))
		<b>PubMed</b> (("Artificial Intelligence" OR "Automated scoring" OR "digital pathology" OR "automated quantification" OR "digitalization" OR "computational") AND ("Tumour infiltrating lymphocyte" OR "lymphocyte" OR "immune microenvironment" OR "Tumor-Derived Activated Cell" OR "tumour associated leukocyte" OR "tumor microenvironment") AND ("Breast cancer" OR "Breast neoplasm" OR "Triple negative breast cancer" OR "Triple-Negative Breast Neoplasm" OR "ER-Negative PR-Negative HER-2 Negative Breast Cancer" OR "ER-Negative PR-Negative HER-2 Negative Breast Neoplasm"))

**Table 2** PICOS criteria for inclusion eligibility

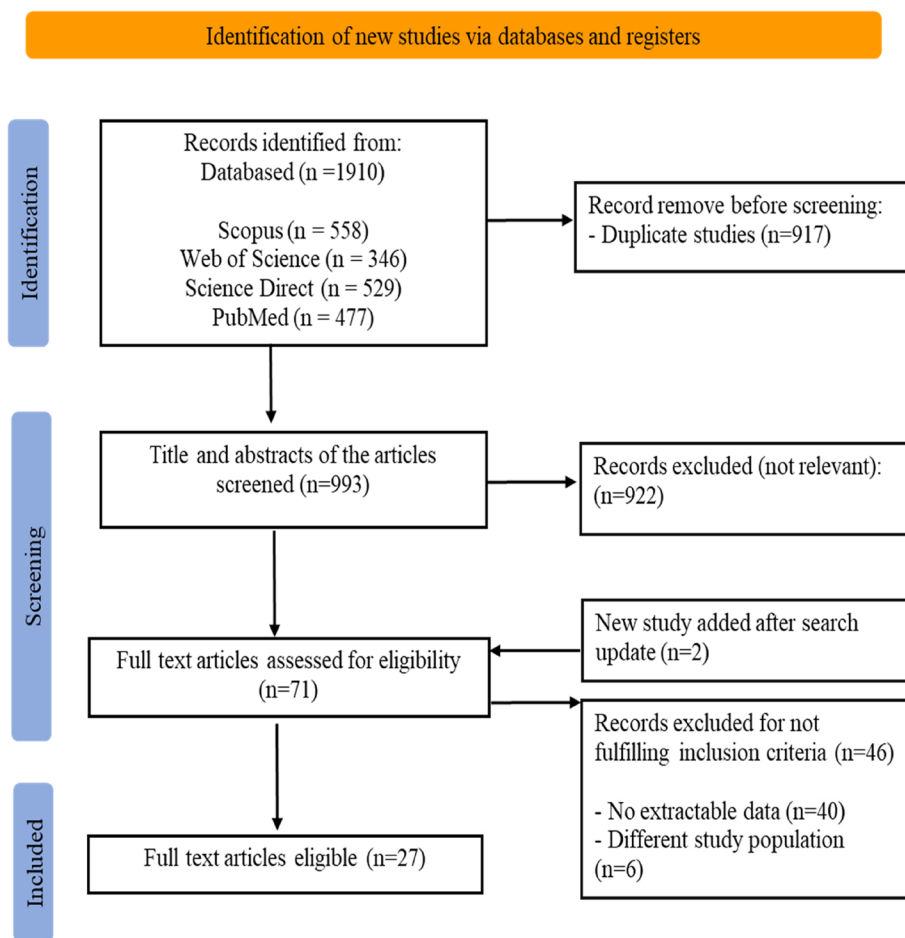
Parameter	Inclusion criteria	Data extraction
Population (P)	Breast cancer cases.	Location, sample size, source of sample, type of staining, type of scanner.
Intervention (I)	Computational or machine learning approach for TILs assessment.	Approaches for computational or machine learning for TILs assessment.
Comparator (C)	Manual ground truth dataset.	Approaches for ground truth development.
Outcome (O)	Performance of the model developed.	Clinical outcome, correlation between automated approach and pathologist consensus on TILs assessment.
Study (S)	Cross sectional diagnostic accuracy study.	Study design.

**Results**

**Article selection**

The article selection process for this study was summarized using the PRISMA flow chart (Fig. 1). A total of 1910 relevant articles were identified through the electronic databases search, Scopus ( $n=558$ ), Web of Science ( $n=346$ ), Science Direct ( $n=529$ ) and PubMed ( $n=477$ ). All identified articles were then imported to the Rayyan AI tool. The tool detected and marked 917

duplicate articles from the various databases. The algorithm developed by Rayyan AI accurately identified these duplicate items, which were then eliminated from further analysis. Title and abstract screening were performed for the remaining articles ( $n=993$ ). Following the titles and abstracts screening, 922 articles were excluded for further analysis. A total of 73 (including 2 new study added) full-text articles were then assessed in the subsequent stage, in which 46 articles were excluded due to irrelevant



**Fig. 1** Flow chart for the article's selection process

data and different study populations. Finally, 27 articles were included in this systematic review following a discussion among the investigators, and full data extraction was performed for these articles.

**Risk of bias**

The quality assessment found that the overall risks of bias for the selected studies were low to moderate. These studies applied appropriate approaches to the research questions and the reported findings were consistent in their data sources, data collection, and analysis. The analysis risk of bias as shown in Fig. 2.

**Study characteristics**

The geographical distribution of studies on automated TIL scoring methods in breast cancer, as shown in Fig. 3, including eight countries: USA ( $n=11$ , 41%), UK ( $n=4$ , 15%), Netherlands ( $n=4$ , 15%), Lithuania ( $n=3$ , 11%), China ( $n=2$ , 7%), Republic of Korea ( $n=1$ , 4%), Denmark ( $n=1$ , 4%), and Pakistan ( $n=1$ , 4%). The sample size of these studies was varied, ranging from 5 to 3760 subjects. A total of 16 studies (59%) employed private dataset that predominantly originated from the main author country. For studies that provide information about the type of scanner utilized to produce whole slide images (WSI), Aperio scanner become the most frequent scanner used in the studies ( $n=6$ , 22%) to produce WSIs. For staining, the specimen slides were predominantly stained with Haematoxylin and Eosin (H&E) stain ( $n=16$ , 59%), while some studies opted for immunohistochemistry (IHC) stain ( $n=8$ , 30%), transmembrane glycoproteins, particularly CD8 and CD3 antigens, became the most

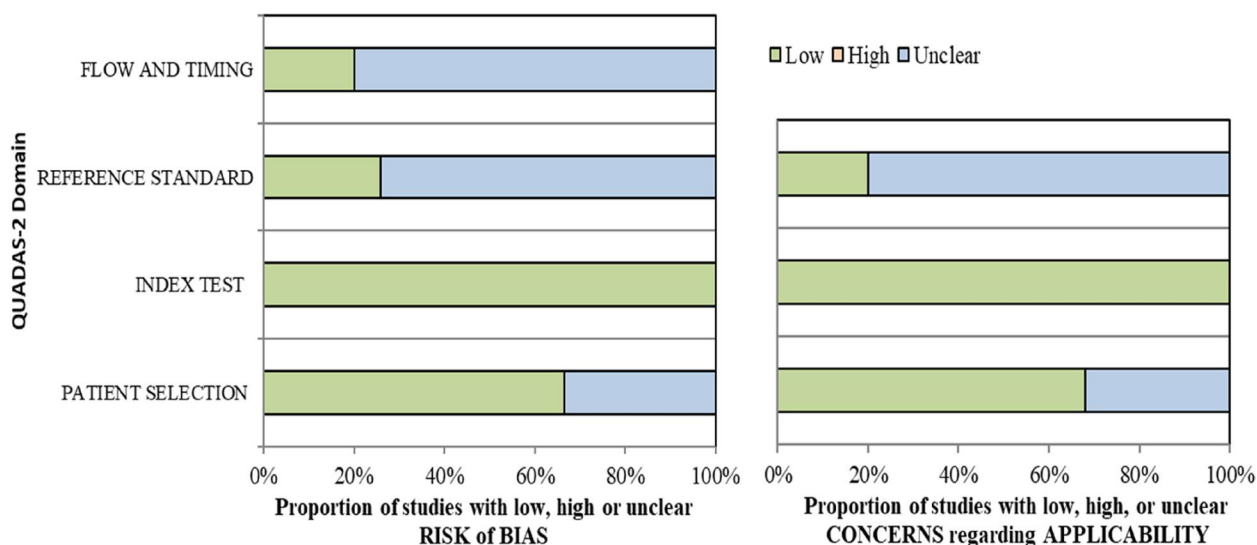
common marker used to highlight TILs in the IHC staining method. The summary of the study characteristics is shown in Tables 3 and 4.

**Features detection approach**

Multiple approaches like object detection, semantic segmentation, patch classification, and a combination of these methods are used for TILs automated detection method and approach. A combination of semantic segmentation and object detection was the most common automated task performed for the computational TILs scoring model development ( $n=10$ , 37%), followed by patch classification ( $n=4$ , 15%) and a combination of image segmentation & object detection ( $n=3$ , 11%). List of these different approaches were summarized in Fig. 4.

**Machine learning (ML) approach**

From the analysis, most of the studies ( $n=22$ , 82%) implementing ML approach for their TILs assessment model development. Of these, four studies utilized multiple ML approach for their model development [28–31]. There were 6 different type of ML approaches used, and the convolutional neural network (CNN) was found to be the most common DL approach applied ( $n=11$ , 41%), followed by the fully convolutional neural network (FCNN) (Fig. 3). CNNs are commonly used for tasks like image classification, object detection, and feature extraction. While fully convolutional neural network (FCN) is a type of convolutional neural network (CNN) commonly used in spatial tasks such as semantic segmentation, object detection, and image reconstruction.



**Fig. 2** Risk of bias evaluation according to QUADAS-2 tool. Template adapted from the <https://www.bristol.ac.uk/population-healthsciences/projects/quadas/resources>

**Table 3** Summary of selected studies synthesized in the systematic review pertaining to the sample population, approaches and prognostic value selected studies synthesized in the systematic

No	First author/ Year/ Reference	Country	Sample size	Type of sample	Source of sample	Staining type	Type of scanner
1	Heindl et al. (2018) [32]	UK	963	ER+ breast cancer.	Private dataset (UK)	H&E	N/A
2	Swiderska-Chadaj et al. (2018) [28]	Netherlands	58	Breast, colon, prostate	Multiple private dataset (Netherlands)	IHC (CD3 & CD8)	Pannoramic 250 Flash II at 40x magnification
3	Amgad et al. (2019) [33]	USA	151	TNBC	TCGA dataset.	H&E	N/A
4	McIntire et al. (2019) [34]	USA	66	TNBC	Private dataset (New York, USA)	IHC (CD8)	Aperio AT2 scanner at 20x magnification
5	Amgad et al. (2019) [35]	USA	120	Breast cancer	Private dataset (USA)	H&E	Aperio scanner at 20x magnification
6	Swiderska-Chadaj et al. (2019) [29]	Netherlands	83	Breast, prostate and colon cancer	Multiple private dataset (Netherlands)	IHC (CD3 and CD8)	Pannoramic 250 Flash II at 20x magnification
7	Le et al. (2020) [36]	USA	976	Breast cancer	SEER and TCGA dataset.	H&E	N/A
8	Lu et al. (2020) [37]	China	1000	Breast cancer	TCGA dataset.	H&E	N/A
9	Mi et al. (2020) [38]	USA	5	TNBC	Private dataset (USA)	IHC (CD3, CD4, CD8, CD20, and FOXP3)	Aperio Scanscope AT at 20x magnification
10	Entenberg et al. (2020) [39]	USA	3760	Breast cancer	Private dataset (USA)	IHC (CD3)	Pannoramic 250 Flash II at 40x magnification
11	Rasmusson et al. (2020) [40]	Lithuania	102	Breast cancer and colorectal cancer	Private dataset (Lithuania)	IHC (CD8)	ScanScope XT Slide Scanner at 20x objective
12	Zilenaite et al. (2020) [41]	Lithuania	101	Breast cancer	Private dataset (Lithuania)	IHC (CD8)	ScanScope XT Slide Scanner at 20x objective
13	Budginaita et al. (2020) [42]	Lithuania	104	Breast and colorectal cancer	TCGA and private dataset (Lithuania)	H&E	Aperio ScanScope XT Slide Scanner at 20x magnification
14	Sun et al. (2021) [43]	China	265	TNBC	TCGA and Private dataset (China)	H&E	Aperio, ScanScope AT2 at 40x magnification
15	Balkenhol et al. (2021) [44]	Netherlands	94	TNBC	Multiple private dataset (Netherlands)	H&E and IHC (FOXP3, CD8, CD3)	Pannoramic 250 Flash II and Vectra spectral imaging system
16	Thagaard et al. (2021) [45]	Denmark	204	TNBC	Private (Denmark) & TCGA dataset	H&E & IHC (CD3 and CD79a)	N/A
17	Krijgsman et al. (2021) [46]	Netherlands	200	ER+ breast cancer.	Private dataset (Netherlands)	H&E and IHC (CD8)	N/A
18	Zormpas-Petridis et al. (2021) [47]	UK	223	TNBC, melanoma and neuroblastoma	TCGA and private dataset (Serbia)	H&E	Aperio ImageScope and Hamamatsu Nano Zoomer
19	Zafar et al. (2022) [48]	Pakistan	43	Breast, prostate, and colon	Public dataset	IHC (CD3 and CD8)	N/A
20	Danielle et al. (2022) [49]	USA	2072	Breast cancer	Private (USA) & TCGA dataset	H&E	N/A
21	Albusayli et al. (2023) [31]	UK	429	TNBC	TCGA and private dataset (Australia)	H&E	Hamamatsu Nano Zoomer scanner at 40x magnification
22	Yosofvand et al. (2023) [30]	USA	63	Breast cancer	Private dataset (USA)	H&E	N/A
23	Rong et al. (2023) [50]	USA	1061	Breast and lung cancer	Public dataset (NuCLS portal)	H&E	N/A
24	Choi et al. (2023) [51]	Korea	402	Breast cancer	Private dataset from 2 institution (USA and Korea)	H&E	Pannoramic 1000, 3DHISTECH at 40x magnification

**Table 3** (continued)

No	First author/ Year/ Reference	Country	Sample size	Type of sample	Source of sample	Staining type	Type of scanner
25	Makhlouf et al. (2023) [52]	UK	2231	Luminal breast cancer	Private dataset (UK)	H&E	Philips IntelliSite Ultra-Fast Scanner and Panoramic 250 Flash III at 40x magnification
26	Bhattarai et al. (2024) [53]	USA	76	TNBC	Private dataset (USA)	H&E	Hamamatsu Nano-Zoomer 2.0-HT C9600-13 at 40x magnification
27	Fisher et al. (2024) [54]	USA	164	TNBC	Private dataset (USA and Ireland)	H&E	Hamamatsu Nano-Zoomer 2.0-HT C9600-13 at 40x magnification

### Ground Truth and validation

For the development of ground truth dataset, different approaches for TILs assessment have different ground truth requirements. Ground truth dataset is essential for automated model training and validation particularly for tasks like image segmentation. It consists of input data with manually annotated images by the pathologists that serve as the reference standard for evaluating the performance of the automated model. This systematic review revealed that traced region boundaries was the most used approach among the studies for the ground truth dataset development ( $n=15$ , 56%), followed by labelled patches (i.e. labelling the images as with Yes/No TILs) ( $n=4$ , 15%). Moreover, 70% of the model developed in these studies were aligned with the visual TILs assessment guidelines established by the TIL-WG (Table 4). The correlation between pathologist consensus and the automated TILs assessment model was found in 12 studies (58%), where majority of them ( $n=8$ ) demonstrated a moderate to strong correlation ( $R$ -value 0.6–0.98), one study demonstrated moderate agreement ( $\kappa=0.57$ ) and another study demonstrated fair to moderate agreement (ICC value 0.40–0.70).

### Prognostic predictive value of TILs

The prognostic value of TILs was assessed in 16 (59%) eligible studies for inclusion in the analysis as illustrated in Tables 3 and 4. From the 16 studies, 10 studies (63%) analysed prognostic value of sTILs, 5 studies (31%) analysed prognostic value of both sTILs and iTILs and only one study (6%) analysed prognostic value of only iTILs with different variable as shown in Tables 3 and 4. The strengths and limitations of each study included in this review as shown in Fig. 5.

### Discussion

#### Sample population

The choice of the sample population is a key factor in the development of automated models. Opting for a representative and diverse sample population is essential for developing an accurate, effective, and robust automated model for TILs assessment. A large and diverse training dataset could provide a more accurate and robust model. The largest sample size observed in this review was from a private source, comprising 3760 breast cancer samples [39]. Although the use of private datasets requires careful handling to ensure privacy, 56% of the studies in this review employed private datasets either from their own institutions or external sources for model training and validation. Public datasets from the cancer genome atlas program (TCGA), the Surveillance, Epidemiology, and End Results (SEER), and nucleus classification, localization, and segmentation (NuCLS) were used in most studies included in this review. Available public datasets may not always be representative of the target population, prompting an increasing number of studies to opt for private datasets for automated model training and validation.

The type of stain and imaging modality also had a significant impact on automated model development. According to published guidelines, TILs assessment in invasive breast cancer requires a pathologist to select the tumor region and delineate stromal areas to assess the percentage of sTILs within the boundaries of the entire tumor [55]. H&E staining was used to stain histology slides for TIL assessment because its practicality, widely available and provided a clear presentation of tissue architecture [2, 4, 15, 56]. Due to TILs heterogeneity distribution, an in-situ approach, such as IHC staining, is another technique that can improve image analysis by identifying the spatial patterns of TILs distribution. It



**Table 4** Summary of selected studies synthesized in the systematic review pertaining to the sample population, approaches and prognostic value of automated TILs scoring methods on breast cancer from the published literature

No	TILs detection method		Ground truth development.			Location of TILs measured	Prognostic value of TILs for breast cancer	Risk of bias
	Automated task	Automated approach	Approach	Follow TIL-WG guidelines (Yes/No)	Number of pathologists			
1	Object detection + inferred TIL location	SVM classifier using morphology TILs detection method.	Labelled nuclei	No	1	Stromal	TILs have a significant prognostic value for late recurrence (5–10 years after diagnosis) in ER + breast cancer patients treated with endocrine therapy.	++
2	Patch classification, semantic segmentation, bounding box	Multiple ML pipeline	Labelled nuclei within labelled region	No	N/A	Stromal	N/A	++
3	Semantic segmentation	FCNN	Traced region boundaries	No	25	Stromal	N/A	++
4	Semantic segmentation + Object detection	Non-ML pipeline	Labelled nuclei within labelled region	No	6	Stromal	N/A	++
5	Semantic segmentation + Object detection	FCNN	Traced region boundaries	Yes	2	Stromal	Poor survival outcomes for Low-sTILs	+
6	Patch classification + Semantic segmentation + Object detection	Multiple ML pipeline	Traced region boundaries	No	4	Stromal	N/A	++
7	Patch classification	CNN	Labelled patches (Yes/No)	No	3	Stromal	Higher levels of TIL infiltration were associated with better OS in patients with breast cancer.	++
8	Patch classification	CNN	Labelled patches (Yes/No)	No	N/A	Stromal	TIL spatial features have potential impact on patient outcomes in breast cancer.	++
9	Image segmentation + Object detection	Non-ML pipeline	Traced region boundaries	No	N/A	Stromal	N/A	++
10	Image segmentation + Object detection	Non-ML pipeline	Traced region boundaries	No	5	Stromal	N/A	++
11	Systematic grid-based subsampling	Non-ML pipeline	Traced region boundaries.	No	1	Stromal	High sTIL better OS for breast cancer	++
12	Image segmentation + Object detection	Non-ML pipeline	Labelled patches (Low/medium/high)	No	N/A	Stromal and intratumoural	Higher CD8 (both sTILs and iTILs) associated with higher OS probabilities	++
13	Nuclei segmentation + Object detection	CNN	Labelled patches (Yes/No)	No	2	Stromal	N/A	+

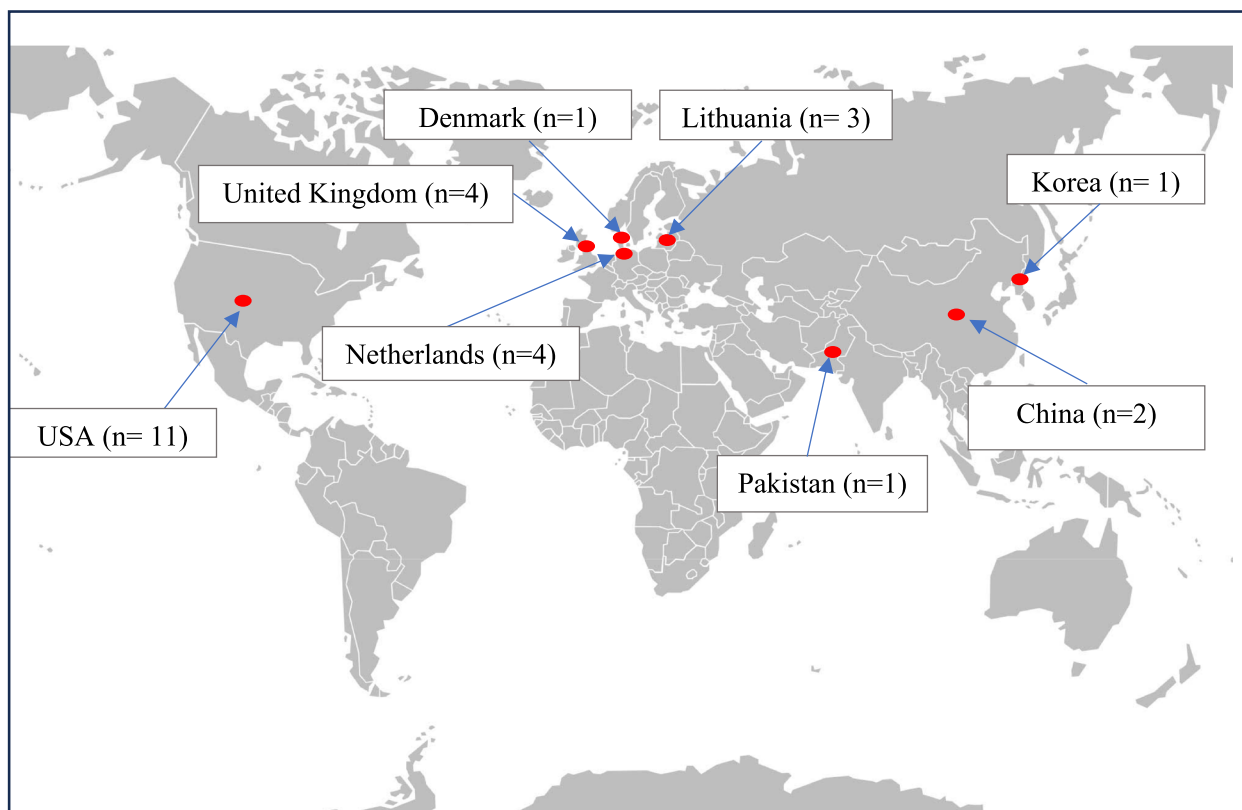
**Table 4** (continued)

No	TILs detection method		Ground truth development.			Location of TILs measured	Prognostic value of TILs for breast cancer	Risk of bias
	Automated task	Automated approach	Approach	Follow TIL-WG guidelines (Yes/No)	Number of pathologists			
14	Patch classification	CNN	Traced region boundaries	Yes	3	Stromal	Significant and independent factors for DFS	++
15	Object detection	CNN	Traced region boundaries	No	N/A	Stromal and intratumoural	The presence of CD3, CD8 and FOXP3 all correlated with an improved survival. High iTILs (CD3 and CD8) count correlated with improved survival.	++
16	Semantic segmentation + Object detection	CNN	Traced region boundaries	Yes	3	Stromal	Higher sTILs scores associated with significantly prolonged OS for both automated and manual TILs assessment but not for RFS.	+
17	Semantic segmentation + Object detection	FCNN	Traced region boundaries	No	N/A	Stromal	High CD8 shows statistically significant correlations with OS but not significant for DFS.	++
18	Semantic segmentation + Object detection	CNN	Qualitative density score	No	1	Stromal	N/A	++
19	Semantic segmentation + Object detection	CNN	Labelled patches (Yes/No)	No	1	Stromal	N/A	++
20	Patch classification	CNN	Labelled patches (Low/medium/high)	No	3	Intratumoural	High iTILs associated with increase survival, prolonged progression-free interval.	++
21	Semantic segmentation + Object detection	Multiple ML pipeline	Traced region boundaries	Yes	2	Stromal and intratumoural	N/A	+
22	Patch classification + Object detection	Multiple ML pipeline	Labelled patches (Yes/No)	No	32	Stromal	Increase sTILs reduce the hazard ratio and lead to better survival outcomes.	++
23	Semantic segmentation + Object detection	R-CNN	Traced region boundaries.	Yes	4	Stromal	N/A	++

**Table 4** (continued)

No	TILs detection method		Ground truth development.		Location of TILs measured	Prognostic value of TILs for breast cancer	Risk of bias
	Automated task	Automated approach	Approach	Follow TIL-WG guidelines (Yes/No)			
24	Semantic segmentation + Object detection	CNN	Traced region boundaries.	Yes	Stromal	A high Digi-sTILs score was significantly associated with higher breast cancer disease-specific survival.	++
25	Semantic segmentation + Object detection	CNN	Traced region boundaries.	Yes	Stromal and intratumoural	High sTILs and tTILs had significantly shorter survival.	+
26	Image segmentation + Object detection	Mask R-CNN	Labelled nuclei	Yes	Stromal and intratumoural	High counts of tTILs are generally associated with better response to NAC. High counts of sTILs are associated with better prognosis and improved survival outcomes in TNBC patients.	+
27	Patch classification + segmentation + object detection	SVM classifier	Traced region boundaries	No	Stromal and intratumoural	High number of sTILs and tTILs is strongly associated with pathological complete response (pCR) in TNBC patients.	+

\*Abbreviation: tTILs/ITILs Intratumoural Tumour Infiltrating lymphocytes, sTILs Stromal Tumour Infiltrating Lymphocytes, CNN Convolutional neural network, FCNN Fully Convolutional Neural Network, R-CNN Region-based Convolutional Neural Network, N/A (Not available), NAC Neoadjuvant chemotherapy, NL Machine Learning



**Fig. 3** Geographical distribution of studies on automated TIL scoring methods in breast cancer based on published literature

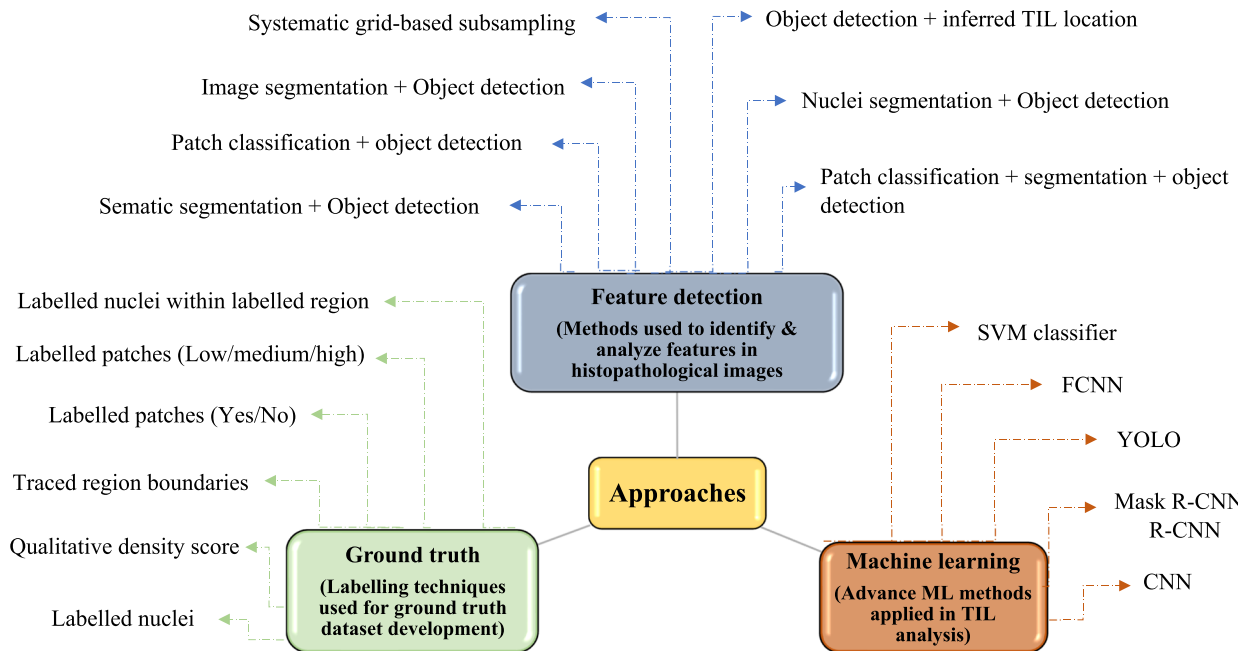
also allows for the discrimination between many relevant subtypes of TILs that have different roles in the TME [46]. Through the application of IHC staining, TILs will be specifically highlighted, which will help improve algorithm specificity; thus, the misclassification of subpopulations of TILs can also be reduced [57].

A previous study that utilized IHC as a staining method for assessing TIL in colon cancer found that the assay was reproducible, objective, and robust [58]. Furthermore, IHC assays have been validated for the assessment of TIL biomarkers such as CD3+, CD8+, and FOXP3+, and have been found to produce reliable results [57]. Another study found that CD4+ lymphocytes were the most common subtype in the tumor stroma and at tumor edges, whereas CD8+ lymphocytes were the most common in tumor nests and FOXP3+ lymphocytes were the least common in all compartments [1]. From this review, 59% of the selected studies utilized H&E to stain their slides for, and the remaining studies utilized either IHC alone or a combination of both H&E and IHC. Notably, CD8+ was the most employed marker for highlighting TILs alongside CD3+, CD4+, CD20+, and FOXP3+. Because of its higher accuracy and reliability in identifying spatial patterns of TILs distribution, an increasing

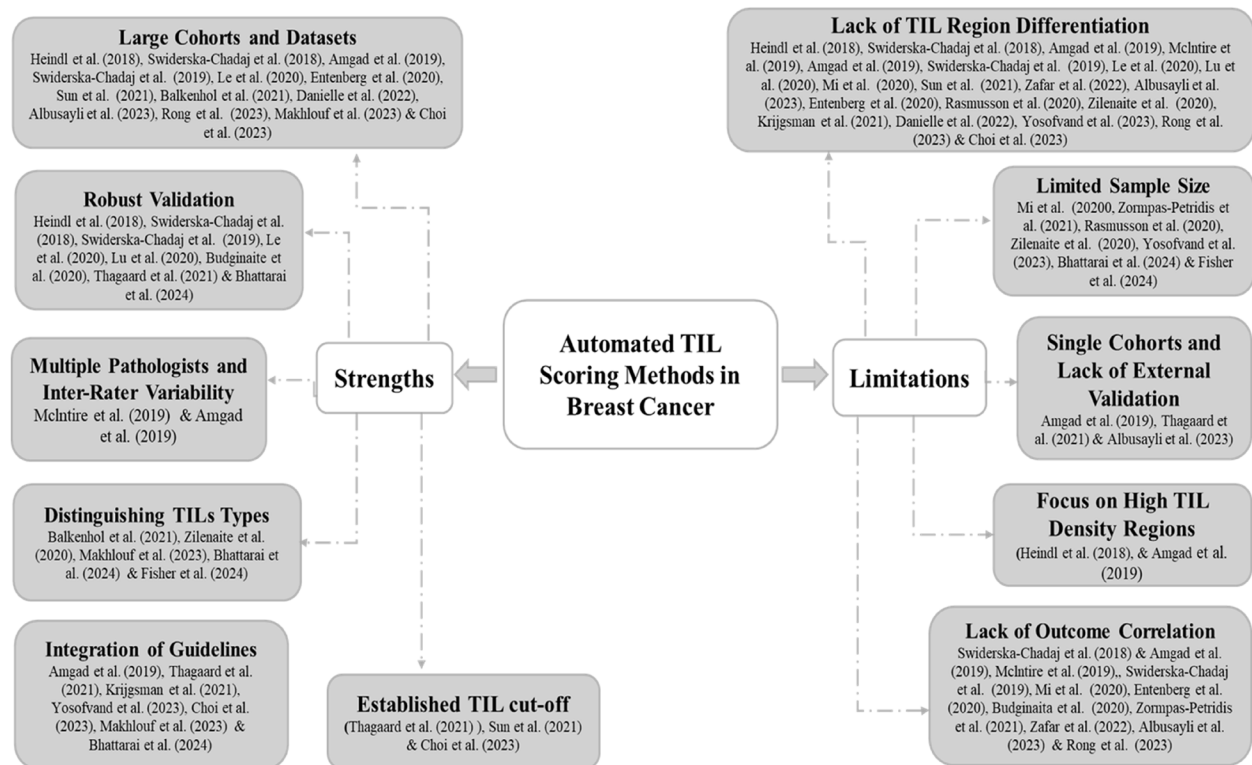
number of studies have used IHC for slide staining for ground truth development to generate more objective ground truths for their automated models.

#### TILs assessment approaches

The findings from this study show that a range of approaches have been proposed for the segmentation and detection of tumor-infiltrating lymphocytes (TILs) in breast cancer, with most studies applying automatic scoring based on a deep-learning approach to develop a model for TILs assessment. Five of the selected studies applied only an algorithm-based approach for model development [34, 38–41], where the semantic segmentation and object detection approaches were frequently used to develop their model. This technique has been demonstrated to improve the accuracy and efficiency of TILs segmentation [56]. CNN is the most common deep-learning approach applied to develop an automated TILs assessment model. It is due to its unique capabilities in learning hierarchical features, preserving spatial context and achieving robust performance. TILs analysis requires the accurate identification and segmentation of intra-tumoural stromal areas in patches or individual TILs. The findings of this study demonstrate the potential of



**Fig. 4** Summary of different approaches for automated model and ground truth dataset development from the published literature. Abbreviation: Convolutional neural network (CNN), Fully Convolutional Network (FCN), Region-based Convolutional Neural Network (R-CNN), You Only Look Once (YOLO), Support Vector Machine (SVM), Log-Structured Merge-tree (LSM)



**Fig. 5** Summary of strengths and limitations for automated TILs scoring methods on breast cancer from the published literature

machine learning and advanced algorithms for the precise segmentation and identification of TILs, which are crucial for cancer prognosis and treatment.

Apart from that, 74% of the studies did not adhere to the TIL-WG guidelines for visual assessment of TIL. The reason for this deviation was due to their failure to properly segment crucial confounding cells or regions (such as artifacts, central necrosis, regressive hyalinization, DCIS, or fibrosis) within their model. Andreas et al. (2018) [32] revealed a weak correlation between automated and manual TILs assessments. This was because of the automated scores, which included regions of the tumor that were excluded from the pathological evaluation. It is necessary to specifically segment regions for exclusion from the analysis when segmenting the regions in which TILs will be analysed. In addition, some of the developed models did not recognize stromal regions or identify individual TILs, and they were only focusing on hotspots for TILs assessment, as demonstrated in studies by Andreas et al. (2018) [32] and Amgad et al. (2019) [33]. To fully adhere to the TIL-WG guidelines, a high-quality segmentation is required to specifically calculate TILs in the intratumoural stroma region and to exclude key cofounder regions from the analysis of the entire slide. This is important to ensure robust and reliable interpretation of TILs density. Thus, segmentation of the region of interest is important during model development. These regions are major sources of variation in visual TILs assessments. Therefore, the development of computational algorithms that can perform high-quality segmentation tasks can increase reproducibility and consistency in the TILs assessment for breast cancers [9, 14].

Validation and training of the model developed are very important. Two models are available to evaluate this type of model creation workflow. In the conventional open assessment approach, the algorithm can be tested on an independent held-out testing set after training it on a set of manually annotated datasets. Alternatively, a closed-loop approach may be used, in which pathologists can use the algorithm's output to reconsider their initial judgments on the held-out set after being exposed to the results of the algorithm. Most studies that translate manual guidelines for TILs quantification into an automated approach rely on traditional open-assessment frameworks [57, 58]. Thus, the development of the ground-truth dataset established by pathologist annotations of TILs scoring can be used to facilitate external validation of other algorithms to make the model more reliable and robust to a variety of biological, staining, and scanning settings. Recently, researchers have attempted to develop and provide a sTILdensity annotated dataset in H&E-stained invasive breast cancer specimens for automated model validation that relies on guidelines established by

the international TILs-WG. The validation dataset was established by pathologist annotations serving as a tool for evaluating the accuracy of algorithms to quantify the density of stroma TILs [34].

Validation and training of the developed automated model are important tasks to be performed by researchers. These include pre-analytical validation (Pre-AV), analytical validation (AV), clinical validation (CV), and clinical utility [13, 59, 60]. Pre-AV, which focuses on actions taken before the application of algorithms, includes processes such as specimen preparation, slide quality, WSI scanning requirements for magnification and resolution, and the image format. AV refers to accuracy and reproducibility, CV for grouping patients into clinically significant subgroups, and clinical utility for total benefit in the clinical setting; while considering practices and methodologies [5]. According to the College of American Pathologist (CAP) guidelines [61], it is necessary for researchers to perform in-house validation of the developed automated model, and Pre-AV and AV are the most suitable for in-house validation. followed by an external validation. In addition, AV depends on the availability of quality "ground truth" annotations; thus, the development of open-access and large-scale datasets is important to facilitate this matter.

#### Ground truth development

Pathology's concept of "ground truth" can be vague and frequently subjective, particularly when dealing with H&E. The findings of this systematic review revealed that most developed automated models relied on the pathologist's manual annotation to train and validate their model where traced region boundaries are commonly used to develop their ground-truth dataset for automated model training and validation. The ground-truth dataset is important as reference annotations for automated model training and validation. According to Amgad et al. (2020), TIL scoring needs to capture the concepts of stromal and intratumoural TILs as well as confounding morphologies specific to particular tumour locations, subtypes, and histologic patterns. Thus, the development of ground truth dataset is crucial to include iTILs during pathologist assessment of TILs. TILs-WG has suggested that the percentage of intratumoural stroma occupied by TILs should be calculated, and that the algorithm developed for automated TILs should follow published visual guidelines. However, the segmentation of intra-tumoural stroma requires exhaustive boundary annotations that are tedious and prone to high annotation errors [42, 62].

This systematic review showed six studies measured both stromal and intratumoural TILs [31, 41, 44, 52–54], only one study measured intratumoural TILs only [21], and the remaining studies measured only stromal TILs

in their assessment in accordance with the recommendations from the TIL-WG [55]. Two of the studies that measured both iTILs and sTILs emphasized the significance of examining spatial patterns of TILs that can indicate immune functional phenotypes and disease prognosis, in addition to TILs numbers [39, 42]. Another study also showed the importance of both sTILs and iTILs for pathologic complete response in advanced breast cancer, in which it was concluded that iTILs can be used to determine neoadjuvant chemotherapy (NACT) in patients with early-stage breast cancer [63]. Khoury et al. (2018) highlighted the difficulties in detecting and assessing iTILs due to their heterogeneity which may have influenced the superiority of sTILs and several specific circumstances, including when they are low in numbers or embedded in the tumor [64]. Therefore, further methodological research is required to characterize this variable with greater accuracy. Nevertheless, in the case of H&E-stained sections of invasive breast carcinoma, observing iTILs is more challenging owing to increased heterogeneity, making their identification more difficult without supplementary staining.

Reporting inter-observer agreement in the manual ground truth dataset development is essential to ensure the reliability and consistency of the TILs assessments. A recent study evaluated their model using multiple expert annotation and observed moderate interobserver agreement [65]. This finding showed the difficulty of defining a clear and objective ground truth for model training and validation. Three studies reported an excellent agreement [38, 43, 45] in TILs assessment, one study reported moderate to substantial agreement between pathologists from manual TILs assessment [49] and one study reported a strong correlation between pathologists from manual TILs assessment [39]. Interobserver agreement in the manual assessment of tumor-infiltrating lymphocytes (TILs) has consistently been emphasized in previous studies [66–68]. They describe that interobserver agreement is important to show reliability of TILs assessment, standardization and enhances the clinical relevance of TILs as biomarker for treatment decision and patient outcomes in breast cancer.

The correlation between pathologist consensus and an automated model in the context of histological image analysis is an important aspect to consider. This analysis offers valuable perspectives on the clinical relevance of the model and potential for assisting pathologists in their diagnostic tasks. Comparison of the automated immune scores with a pathologist's score following recommendations for TIL evaluation in breast cancer in a subset of samples showed an overall strong correlation in most of the selected studies with only three studies showed moderate agreement between pathologist

consensus and automated TILs assessment [29, 43, 52]. However, two studies showed weak correlation [32, 46]. Factors such as different scope of evaluation, discrepancy in the regions evaluated, spatial heterogeneity of TILs infiltration, potential sampling bias, technical variability, and biological variability may have contributed to the weak correlation between the pathologist's TIL score and the automated immune scores observed in their studies [32, 46].

#### Prognostic value of TILs

In the era of immunotherapy, the relevance of the TILs in predicting patient outcomes, as well as the prospective influence of chemotherapy and hormone treatment for breast cancer, has been proven [69–71]. As an example, a pooled analysis of 3771 breast cancer patients treated with NACT showed that patients diagnosed with TNBC and HER-2 positive breast cancer reported longer disease-free survival with 10% increase in TILs [69]. Moreover, the cell abundance and spatial patterns of TILs can be indicative of immune functional phenotypes and disease prognosis, and therefore, should be identified for improvement of clinical management and health outcomes [69–71]. Research conducted by Bernardo et al. (2022) demonstrated a significant prognostic value of the spatial distributions of CD3+ and CD8+ TILs among early breast cancer patients. Therefore, automated TILs assessments should be able to provide information regarding the distribution of TILs in relation to stromal and tumour cells.

The prognostic values of TILs for breast cancer were assessed in 16 studies that were synthesized in this systematic review (Tables 3 and 4). The long-term prognosis indicators that were used in most studies were overall survival (OS). All studies have reported the presence of TILs as a favourable prognostic factor for treatment outcomes in breast cancer except for one study by Makhoul et al. (2023) that revealed high sTILs and iTILs, corresponding to significantly shorter survival for luminal breast cancer [52]. A general observation of our data for the patterns of hazard ratio (HR) found that most studies reported HR values of less than 1, indicating a higher level of TILs was associated with a reduced risk of disease recurrence or death [35, 36, 40, 41, 43–46, 49]. Notably, only three studies [31, 32, 52] deviated from this trend with HR values more than 1 indicating that an elevated presence of TILs is associated with an increased risk of disease recurrence or death for patients with breast cancer.

The observed primary issue pertaining to the TILs scoring and its prognostic value was that no official clinically relevant TILs cut-off points have been suggested. The cut-off points for TILs may differ depending on the

cancer type and clinical context, due to biological variability, clinical heterogeneity, outcome relevance and they can be established by conducting statistical analysis and validating them in separate cohorts. An ideal TILs cut-off point is necessary for clinical decision-making or risk management. This systematic review found three studies that addressed the TILs' cut-off points [43, 45, 51]. Thaggaard et al. (2021) revealed that by using > 10% cutoff point for the manual sTIL assessment, helps stratify patients into distinct prognostic groups, while Sun et al. (2021) using various cutoff points, to stratify patients into TILs-High and TILs-Low groups and another study used a cut-off point of 10% for low TILs, 1–49% for intermediate TILs, and 50% or higher for high TILs [51]. These studies proved that the TILs cut-off values for patient stratification varied between different ethnicities and thus, further research and testing on an independent cohort should be conducted to determine the ideal TILs cut-off points [43, 45]. By categorizing patients based on this cut-off point, researchers can identify differences in outcomes and prognosis, allowing for a more personalized approach for treatment and management of breast cancer patients.

### Strength and limitations

This systematic review provides a comprehensive compilation of recent studies related to the types of approaches and performance of automated TILs scoring model implementing P.I.C.O.S criteria and adhering to the PRISMA guideline. This study also employed Rayyan AI to effectively manage and refine the process of screening and selecting of articles. Rayyan AI facilitates the process of identifying duplicates and applying inclusion and exclusion criteria, resulting in a precise and efficient review process. Previous studies have shown that Rayyan AI is effective in reducing the screening time while retaining a high level of accuracy [72, 73]. The finding of this study provides insights toward the evaluation of methodological quality, also helps to identify the strengths and limitations of different automated TILs assessment approaches which can guide future research directions. However, the potential limitation of this review could be in the heterogeneity of the study designs, sample characteristics, and approaches for model development which caused challenges in the data pooling process, synthesizing the results and to determine firm conclusions.

### Recommendation

Developing standardized guidelines for the assessment of algorithms including training, pre-AV, AV, and CV that closely capture visual guidelines and standards are factors that need to be considered during automated model development. Conducting further research to validate the use of computational TILs assessment in clinical practice,

as well as developing tools and resources to facilitate the adoption of computational TILs assessment in clinical practice is very important to achieve the full potential of computational TILs assessment in improving patient outcomes, precision medicine, and enhancing the quality of patients care. It also important to ensure standardization and consistency for regulatory approval and guidelines development. In addition, pathologist judgement is crucial to make manual adjustment for the assessments that did not adequately represent certain tumour subtypes or variations, and for tumour that exhibit heterogeneity in TILs distribution. Thus, combining an automated model with a pathology specialist enables a more comprehensive and accurate evaluation of TILs in breast cancer.

### Conclusion

In conclusion, this analysis contends that automated scoring methodologies for TILs assessment in the context of breast cancer show significant promise for commodification and application within clinical settings. Future direction of standardizing algorithms and validating clinical utility are important aspect to consider for integrating computational TILs assessment into routine clinical practice. Collaborative efforts between computer scientist for the automated model development and pathologists offer promising avenue for comprehensive and accurate TILs assessment. Ultimately, breast cancer management will benefit from precision medicine and improved patient care.

### Abbreviations

AV	Analytical validation
CAP	College of American pathologist
CD3	Cluster of differentiation 3
CD4	Cluster of differentiation 4
CD8	Cluster of differentiation 8
CD20	Cluster of differentiation 20
CNN	Convolutional Neural Network
CV	Clinical validation
DCIS	Ductal carcinoma in-situ
DIA	Digital image analysis
ER	Estrogen receptor
FCN	Fully Convolutional <i>network</i>
FOXP3	Forkhead box P3 (FOXP3) protein
H&E	Haematoxylin and eosin
HER-2	Human epidermal growth factor receptor 2
IHC	Immunohistochemistry
iTILs	Intra tumoural tumor infiltrating lymphocytes
NACT	Neoadjuvant chemotherapy
PR	Progesterone receptor
PRISMA	Preferred Reporting Items for Systematic Review and Meta-Analysis
RCNN	Region-based Convolutional Neural Network
SOTA	State-of-the-art
sTILs	Stroma tumor infiltrating lymphocytes
TCGA	The cancer genome atlas program
TDC-LC	Two-Phase Deep CNN based Lymphocyte Counter
TME	Tumour microenvironment
TILs	Tumour infiltrating lymphocytes
TIL-WG	Tumour infiltrating lymphocytes-working group
TNBC	Triple-negative breast cancer
VTA	Visual TILs assessment



WHO World Health Organization  
 WOS Web of science  
 WSI Whole slide image

### Acknowledgements

The National University of Malaysia (UKM).

### Authors' contributions

N.K.B.B. and M.A.H.Z performed the literature search, reviewed the literature, and wrote the manuscript; R.R.M.Z was involved in the literature search, reviewed the literature and the final manuscript; A. A. reviewed the literature and the final manuscript; N.M.R. and Q. X. was involved in the conception and design of the manuscript and revised the manuscript. All authors have read and approved the final manuscript.

### Funding

The authors would like to thank the Universiti Kebangsaan Malaysia (UKM) for the grant (FF-2021-444).

### Availability of data and materials

Upon request from the corresponding author on reasonable request.

### Declarations

#### Ethics approval and consent to participate

Ethic approval by the Faculty of Medicine, The National University of Malaysia (UKM) (Ethic approval number: JEP-2021-724).

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare no competing interests.

#### Author details

<sup>1</sup>Department of Pathology, Faculty of Medicine, The National University of Malaysia, Jalan Yaacob Latif, Bandar Tun Razak, 56000 Cheras, Kuala Lumpur, Wilayah Persekutuan 56000, Malaysia. <sup>2</sup>Department of Medical Diagnostic, Faculty of Health Sciences, Universiti Selangor, Jalan Zirkon A7/7, Seksyen 7, Shah Alam, Selangor 40000, Malaysia. <sup>3</sup>Centre for Artificial Intelligence Technology (CAIT), Faculty of Information Science & Technology, The National University of Malaysia, Bangi, Selangor 43600, Malaysia. <sup>4</sup>Department of Computer Science, University of Warwick, 6 Lord Bhattacharyya Way, Coventry CV4 7EZ, UK. <sup>5</sup>Department of Pathology and Forensic Pathology, Faculty of Medicine, MAHSA University, Bandar Saujana Putra, Malaysia.

Received: 30 April 2024 Accepted: 18 September 2024

Published online: 30 September 2024

### References

- Verma R, Hanby AM, Horgan K, Verghese ET, Volpato M, Carter CR, et al. Levels of different subtypes of tumour-infiltrating lymphocytes correlate with each other, with matched circulating lymphocytes, and with survival in breast cancer. *Breast Cancer Res Treat.* 2020;183(1):49–59. <https://doi.org/10.1007/s10549-020-05757-5>.
- Hendry S, Salgado R, Gevaert T, Russell PA, John T, Thapa B, et al. Assessing tumor-infiltrating lymphocytes in solid tumors. *Adv Anat Pathol.* 2017;24:235–51.
- De Jong VMT, Wang Y, Ter Hoeve ND, Opdam M, Stathonikos N, Józwiak K, et al. Prognostic value of stromal tumor-infiltrating lymphocytes in young, node-negative, triple-negative breast Cancer patients who did not receive (neo)adjuvant systemic therapy. *J Clin Oncol.* 2022;40(21):2361–74.
- Dieci MV, Radosevic-Robin N, Fineberg S, van den Eynden G, Ternes N, Penault-Llorca F, et al. Update on tumor-infiltrating lymphocytes (TILs) in breast cancer, including recommendations to assess TILs in residual disease after neoadjuvant therapy and in carcinoma in situ: a report of the International Immuno-Oncology Biomarker Working Group on breast cancer. *Semin Cancer Biol.* 2018;52:16–25 Academic.
- Amgad M, Stovgaard ES, Balslev E, Thagaard J, Chen W, Dudgeon S, et al. Report on computational assessment of tumor infiltrating lymphocytes from the International Immuno-Oncology Biomarker Working Group. *NPJ Breast Cancer.* 2020;6(1):16.
- Inge LJ, Dennis E. Development and applications of computer image analysis algorithms for scoring of PD-L1 immunohistochemistry. *Immuno-Oncol Technol.* 2020;6:2–8 Elsevier Inc.
- Demir H, Gul OV, Aksu T. Investigation of skin dose of post-mastectomy radiation therapy for the halcyon and tomotherapy treatment machine: comparison of calculation and in vivo measurements. *Radiat Meas.* 2024;173:107112.
- von Minckwitz G, Procter M, de Azambuja E, Zardavas D, Benyunes M, Viale G, et al. Adjuvant pertuzumab and trastuzumab in early HER2-Positive breast Cancer. *N Engl J Med.* 2017;377(2):122–31.
- Denkert C, Wienert S, Poterie A, Loibl S, Budczies J, Badve S, et al. Standardized evaluation of tumor-infiltrating lymphocytes in breast cancer: results of the ring studies of the international immuno-oncology biomarker working group. *Mod Pathol.* 2016;29(10):1155–64.
- Piccart-Gebhart M, Holmes E, Baselga J, De Azambuja E, Dueck AC, Viale G, et al. Adjuvant lapatinib and trastuzumab for early human epidermal growth factor receptor 2-positive breast cancer: results from the randomized phase III adjuvant lapatinib and/or trastuzumab treatment optimization trial. *J Clin Oncol.* 2016;34(10):1034–42.
- Cardoso F, Kyriakides S, Ohno S, Penault-Llorca F, Poortmans P, Rubio IT, et al. Early breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol.* 2019;30(8):1194–220. <https://doi.org/10.1093/annonc/mdz173>.
- Morigi C. Highlights of the 16th St Gallen international breast Cancer Conference, Vienna, Austria, 20–23 March 2019: personalised treatments for patients with early breast cancer. *Ecancermedicalscience.* 2019;13(March):20–3.
- Balic M, Thomssen C, Würstlein R, Gnant M, Harbeck N. St. Gallen/Vienna 2019: a brief summary of the consensus discussion on the optimal primary breast cancer treatment. *Breast Care.* 2019;14:103–10.
- Wein L, Savas P, Luen SJ, Virassamy B, Salgado R, Loi S. Clinical validity and utility of Tumor-infiltrating lymphocytes in routine clinical practice for breast cancer patients: current and future directions. *Front Oncol.* 2017;7:156 Frontiers Media S.A.
- Salgado R, Denkert C, Demaria S, Sirtaine N, Klauschen F, Pruner G, et al. The evaluation of tumor-infiltrating lymphocytes (TILs) in breast cancer: recommendations by an International TILS Working Group 2014. *Ann Oncol.* 2015;26:259–71 Oxford University Press.
- Brunyé TT, Mercan E, Weaver DL, Elmore JG. Accuracy is in the eyes of the pathologist: the visual interpretive process and diagnostic accuracy with digital whole slide images. *J Biomed Inf.* 2017;66:171–9.
- Stoler MH. Glandular Lesions of the Uterine Cervix. The United States and Canadian Academy of Pathology. 2000;3(3):261.
- Vis JY, Huisman A. Verification and quality control of routine hematology analyzers. *Int J Lab Hematol.* 2016;38:100–9 Blackwell Publishing Ltd.
- Perkel JM. Immunohistochemistry for the 21st century. *Sci.* 2016;351:1098–100.
- Lloyd MC, Allam-Nandyala P, Purohit CN, Burke N, Coppola D, Bui MM. Using image analysis as a tool for assessment of prognostic and predictive biomarkers for breast cancer: how reliable is it? *J Pathol Inf.* 2010;1(1):29.
- Holten-Rossing H, Møller Talman ML, Kristensson M, Vainer B. Optimizing HER2 assessment in breast cancer: application of automated image analysis. *Breast Cancer Res Treat.* 2015;152(2):367–75.
- Gavrielides MA, Lenz P, Badano A, Hewitt SM. Silver Spring, MD 20993 (marios.gavrielides@fda.hhs.gov). *Arch Pathol Lab Med.* 2011;62(2):233–42.
- Hamilton PW, Bankhead P, Wang Y, Hutchinson R, Kieran D, McArt DG, et al. Digital pathology and image analysis in tissue biomarker research. *Methods.* 2014;70(1):59–73.
- Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ.* 2021;372:n71.

25. Kraus S, Breier M, Dasi-Rodríguez S. The art of crafting a systematic literature review in entrepreneurship research. *Int Entrep Manag J*. 2020;16(3):1023–42.
26. Linares-Espinós E, Hernández V, Domínguez-Escrig JL, Fernández-Pello S, Hevia V, Mayor J, Padilla-Fernández B, Ribal MJ. Metodología de una revisión sistemática. *Methodology of a systematic review*. *Actas Urol Esp (Engl Ed)*. 2018;42(8):499–506.
27. Whiting PF, Reitsma JB, Leeflang MMG, Sterne JAC, Bossuyt PMM, Rutjes AWSS, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155(4):529–36 Available from.
28. Swiderska-Chadaj Z, Pinckaers H, Van Rijthoven M, Balkenhol M, Melnikova M, Geessink O, et al. Convolutional neural networks for lymphocyte detection in immunohistochemically stained whole-slide images. 1st Conference on the Medical Imaging with Deep Learning. 2018.
29. Swiderska-Chadaj Z, Pinckaers H, van Rijthoven M, Balkenhol M, Melnikova M, Geessink O, et al. Learning to detect lymphocytes in immunohistochemistry with deep learning. *Med Image Anal*. 2019;58:101547. <https://doi.org/10.1016/j.media.2019.101547>.
30. Yosofvand M, Khan SY, Dhakal R, Nejat A, Moustaid-Moussa N, Rahman RL, et al. Automated detection and scoring of Tumor-infiltrating lymphocytes in breast Cancer histopathology slides. *Cancers (Basel)*. 2023;15(14):3635.
31. Albusayli R, Graham JD, Pathmanathan N, Shaban M, Raza SEA, Minhas F, et al. Artificial intelligence-based digital scores of stromal tumour-infiltrating lymphocytes and tumour-associated stroma predict disease-specific survival in triple-negative breast cancer. *J Pathol*. 2023;260(1):32–42.
32. Heindl A, Sestak I, Naidoo K, Cuzick J, Dowsett M, Yuan Y. Relevance of spatial heterogeneity of immune infiltration for predicting risk of recurrence after endocrine therapy of ER+ breast Cancer. *J Natl Cancer Inst*. 2018;110(2):166–75.
33. Amgad M, Elfandy H, Hussein H, Atteya LA, Elsebaie MAT, Abo Elnasr LS, et al. Structured crowdsourcing enables convolutional segmentation of histology images. *Bioinformatics*. 2019;35(18):3461–7.
34. McIntire PJ, Zhong E, Patel A, Khani F, D'Alfonso TM, Chen Z, et al. Hotspot enumeration of CD8+ tumor-infiltrating lymphocytes using digital image analysis in triple-negative breast cancer yields consistent results. *Hum Pathol*. 2019;85:27–32.
35. Amgad M, Sarkar A, Srinivas C, Redman R, Ratra S, Bechert CJ, et al. Joint region and nucleus segmentation for characterization of tumor infiltrating lymphocytes in breast cancer. *SPIE-the International Society for Optical Engineering*. 2019;20:10956, 109560M. <https://doi.org/10.1117/12.2512892>.
36. Le H, Gupta R, Hou L, Abousamra S, Fassler D, Torre-Healy L, et al. Utilizing automated breast cancer detection to identify spatial distributions of tumor-infiltrating lymphocytes in invasive breast cancer. *Am J Pathol*. 2020;190(7):1491–504.
37. Lu Z, Xu S, Shao W, Wu Y, Zhang J, Han Z, et al. Deep-learning-based characterization of Tumor-infiltrating lymphocytes in breast cancers from histopathology images and Multiomics Data. *JCO Clin Cancer Inf*. 2020;4:480–90.
38. Mi H, Gong C, Sulam J, Fertig EJ, Szalay AS, Jaffee EM, et al. Digital pathology analysis quantifies spatial heterogeneity of CD3, CD4, CD8, CD20, and FoxP3 immune markers in triple-negative breast Cancer. *Front Physiol*. 2020;11:11.
39. Entenberg D, Oktay MH, D'Alfonso T, Ginter PS, Robinson BD, Xue X et al. Validation of an automated quantitative digital pathology approach for scoring tmem, a prognostic biomarker for metastasis. *Cancers (Basel)*. 2020;12(4).
40. Rasmusson A, Zilenaite D, Nestarenkaite A, Augulis R, Laurinaviciene A, Ostapenko V, et al. Immunogradient indicators for antitumor response assessment by automated tumor-stroma interface zone detection. *Am J Pathol*. 2020;190(6):1309–22. <https://doi.org/10.1016/j.ajpath.2020.01.018>.
41. Zilenaite D, Rasmusson A, Augulis R, Besusparis J, Laurinaviciene A, Plan-coulaine B, et al. Independent prognostic value of intratumoral heterogeneity and immune response features by automated digital immunohistochemistry analysis in early hormone receptor-positive breast carcinoma. *Front Oncol*. 2020;10(June):1–13.
42. Budginaita E, Morknas M, Laurinavicius A, Treigys P. Deep learning model for cell nuclei segmentation and lymphocyte identification in whole slide histology images. *Inform*. 2021;32(1):23–40.
43. Sun P, He J, Chao X, Chen K, Xu Y, Huang Q, et al. A computational tumor-infiltrating lymphocyte Assessment Method comparable with visual reporting guidelines for triple-negative breast Cancer. *EBioMedicine*. 2021;70:70.
44. Balkenhol MC, Ciompi F, Świderska-Chadaj Ż, van de Loo R, Intezar M, Otte-Höller I, et al. Optimized tumour infiltrating lymphocyte assessment for triple negative breast cancer prognostics. *Breast*. 2021;56:78–87.
45. Thagaard J, Stovgaard ES, Vognsen LG, Hauberg S, Dahl A, Ebstrup T, et al. Automated quantification of stiel density with h&e-based digital image analysis has prognostic potential in triple-negative breast cancers. *Cancers*. 2021;13(12):3050. <https://doi.org/10.3390/cancers13123050>.
46. Krijgsman D, Van Leeuwen MB, Van Der Ven J, Almeida V, Vlutters R, Halter D, et al. Quantitative whole Slide Assessment of Tumor-infiltrating CD8-Positive lymphocytes in ER-Positive breast Cancer in relation to clinical outcome. *IEEE J Biomed Heal Inf*. 2021;25(2):381–92.
47. Zormpas-Petridis K, Noguera R, Ivanovic DK, Roxanis I, Jamin Y, Yuan Y. SuperHistopath: a deep learning pipeline for mapping tumor heterogeneity on low-resolution whole-slide digital histopathology images. *Front Oncol*. 2021;10(January):1–13.
48. Muhammad Mohsin Zafar. Detection of tumour infiltrating lymphocytes in CD3 and CD8 stained histopathological images using a two-phase deep CNN. *Photodiagnosis Photodyn Ther*. 2022;37:102676.
49. Danielle J. Fassler 1, Luke A. Torre-Healy 1, Rajarsi Gupta 1, Alina M. Hamilton 2 SK 1, Sarah C. Van Alsten 2, Yuwei Zhang 1, Tahsin Kurc 1, Richard A. Moffitt 1 MAT 2, Saltz KAH 3 and J. Spatial Characterization of Tumor-Infiltrating Lymphocytes and Breast Cancer Progression. *Cancers*. 2022;14(9):2148. <https://doi.org/10.3390/cancers14092148>.
50. Rong R, Sheng H, Jin KW, Wu F, Luo D, Wen Z, et al. A deep learning approach for histology-based nucleus segmentation and tumor microenvironment characterization. *Mod Pathol*. 2023;36(8):100196. <https://doi.org/10.1016/j.modpat.2023.100196>.
51. Choi S, Cho SI, Jung W, Lee T, Choi SJ, Song S, et al. Deep learning model improves tumor-infiltrating lymphocyte evaluation and therapeutic response prediction in breast cancer. *npj Breast Cancer*. 2023;9(1):1–13.
52. Makhlof S, Wahab N, Toss M, Ibrahim A, Lashen AG, Atallah NM, et al. Evaluation of tumour infiltrating lymphocytes in luminal breast cancer using artificial intelligence. *Br J Cancer*. 2023;129:1747–58.
53. Bhattarai S, Saini G, Li H, Seth G, Fisher TB, Janssen EAM, et al. Predicting Neoadjuvant Treatment Response in Triple-negative breast Cancer using machine learning. *Diagnostics*. 2024;14(1):1–13.
54. Fisher TB, Saini G, Rekha TS, Krishnamurthy J, Bhattarai S, Callagy G, et al. Digital image analysis and machine learning-assisted prediction of neoadjuvant chemotherapy response in triple-negative breast cancer. *Breast Cancer Res*. 2024;26(1):1–13. <https://doi.org/10.1186/s13058-023-01752-y>.
55. Denkert C, Salgado R, Demaria S. Standardized evaluation of tumor-infiltrating lymphocytes (TIL) in breast cancer for daily clinical and research practice or clinical trial setting a tutorial prepared by the International Working Group for TIL in breast cancer. 2014.
56. Hendry S, Salgado R, Gevaert T, Russell PA, John T, Thapa B, et al. Assessing tumor-infiltrating lymphocytes in solid tumors: a practical review for pathologists and proposal for a standardized method from the International Immunooncology Biomarkers Working Group: part 1: assessing the host immune response, TILs in invasive breast carcinoma and ductal carcinoma in situ, metastatic tumor deposits and areas for further research. *Adv Anat Pathol*. 2017;24:235–51 Lippincott Williams and Wilkins.
57. Singh U, Cui Y, Dimaano N, Mehta S, Pruitt SK, Yearley J, et al. Analytical validation of quantitative immunohistochemical assays of tumor infiltrating lymphocyte biomarkers. *Biotech Histochem*. 2018;93(6):411–23.
58. Pagès F, Mlecnik B, Marliot F, Bindea G, Ou FS, Bifulco C, et al. International validation of the consensus immunoscore for the classification of colon cancer: a prognostic and accuracy study. *Lancet*. 2018;391(10135):2128–39.
59. Zhang X, Liu K, Zhang K, Li X, Sun Z, Wei B. SAMS-Net: Fusion of attention mechanism and multi-scale features network for tumor infiltrating lymphocytes segmentation. *Math Biosci Eng*. 2023;20(2):2964–79.
60. Gurcan MN. Histopathological image analysis: path to Acceptance through evaluation. *Microsc Microanal*. 2016;22:1004–5.
61. Fauzi MFA, Pennell M, Sahiner B, Chen W, Shana'Ah A, Hemminger J, et al. Classification of follicular lymphoma: the effect of computer aid on pathologists grading clinical decision-making, knowledge support systems, and theory. *BMC Med Inf Decis Mak*. 2015;15(1):115.

62. Garcia V, Elfer K, Peeters DJE, Ehinger A, Werness B, Ly A, et al. Development of training materials for pathologists to provide machine learning validation data of tumor-infiltrating lymphocytes in breast cancer. *Cancers (Basel)*. 2022;14(10).
63. Sirinukunwattana K, Raza SEA, Tsang YW, Snead DRJ, Cree IA, Rajpoot NM. Locality sensitive deep learning for detection and classification of nuclei in routine Colon cancer histology images. *IEEE Trans Med Imaging*. 2016;35(5):1196–206.
64. Khoury T, Nagrale V, Opyrchal M, Peng X, Wang D, Yao S. Prognostic significance of Stromal Versus Intratumoral infiltrating lymphocytes in different subtypes of breast Cancer treated with cytotoxic neoadjuvant chemotherapy.
65. Verghese G, Li M, Liu F, Lohan A, Kurian NC, Meena S, et al. Multiscale deep learning framework captures systemic immune features in lymph nodes predictive of triple negative breast cancer outcome in large-scale studies. *J Pathol*. 2023;260:376–89.
66. Van Bockstal MR, François A, Altinay S, Arnould L, Balkenhol M, Broeckx G, et al. Interobserver variability in the assessment of stromal tumor-infiltrating lymphocytes (sTILs) in triple-negative invasive breast carcinoma influences the association with pathological complete response: the IVITA study. *Mod Pathol*. 2021;34(12):2130–40.
67. Swisher SK, Wu Y, Castaneda CA, Lyons GR, Yang F, Tapia C, et al. Inter-observer Agreement between pathologists assessing tumor-infiltrating lymphocytes (TILs) in breast Cancer using methodology proposed by the International TILs Working Group. *Ann Surg Oncol*. 2016;23(7):2242–8.
68. Tramm T, Di Caterino T, Jylling AMB, Lelkaitis G, Lænkholm AV, Ragó P, et al. Standardized assessment of tumor-infiltrating lymphocytes in breast cancer: an evaluation of inter-observer agreement between pathologists. *Acta Oncol (Madr)*. 2018;57(1):90–4. <https://doi.org/10.1080/0284186X.2017.1403040>.
69. Loi S, Drubay D, Adams S, Pruner G, Francis PA, Lacroix-Triki M, et al. Tumor-infiltrating lymphocytes and prognosis: a pooled individual patient analysis of early-stage triple-negative breast cancers. *J Clin Oncol*. 2019;37:559–69.
70. Luen SJ, Griguolo G, Nuciforo P, Campbell C, Fasani R, Cortes J, Untch M, Lin SJ, Savas P, Fox SB, Serena Di Cosimo, Antonio Llombart Cussac, Evandro de Azambuja, Piccart-Gebhart MJ, and SL. On-treatment changes in tumor-infiltrating lymphocytes (TIL) during neoadjuvant HER2 therapy (NAT) and clinical outcome. *J Clin Oncol*. 2019;15.
71. Schmid P, Salgado R, Park YH, Muñoz-Couselo E, Kim SB, Sohn J, et al. Pembrolizumab plus chemotherapy as neoadjuvant treatment of high-risk, early-stage triple-negative breast cancer: results from the phase 1b open-label, multicohort KEYNOTE-173 study. *Ann Oncol*. 2020;31(5):569–81. <https://doi.org/10.1016/j.annonc.2020.01.072>.
72. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. *Syst Rev*. 2016;5(1):210.
73. Valizadeh A, Moassefi M, Nakhostin-Ansari A, Hosseini Asl SH, Saghab Torbati M, Aghajani R, et al. Abstract screening using the automated tool Rayyan: results of effectiveness in three diagnostic test accuracy systematic reviews. *BMC Med Res Methodol*. 2022;22(1):1–15. <https://doi.org/10.1186/s12874-022-01631-8>.

## Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.