



Systematic Review

Advanced Deep Learning Approaches in Detection Technologies for Comprehensive Breast Cancer Assessment Based on WSIs: A Systematic Literature Review

Qiaoyi Xu ^{1,2,*} , Afzan Adam ¹ , Azizi Abdullah ¹ and Nurkhairul Bariyah ³

¹ Center for Artificial Intelligence Technology, Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, Bangi 43600, Selangor, Malaysia; afzan@ukm.edu.my (A.A.); azizia@ukm.edu.my (A.A.)

² College of Intelligent Manufacturing, Zibo Vocational Institute, Zibo 255314, China

³ Department of Pathology, Faculty of Medicine, Universiti Kebangsaan Malaysia, Jalan Yaacob Latif, Bandar Tun Razak, Cheras 56000, Wilayah Persekutuan Kuala Lumpur, Malaysia; p110546@siswa.ukm.edu.my

* Correspondence: p114345@siswa.ukm.edu.my

Abstract: Background: Breast cancer is one of the leading causes of death among women worldwide. Accurate early detection of lymphocytes and molecular biomarkers is essential for improving diagnostic precision and patient prognosis. Whole slide images (WSIs) are central to digital pathology workflows in breast cancer assessment. However, applying deep learning techniques to WSIs presents persistent challenges, including variability in image quality, limited availability of high-quality annotations, poor model interpretability, high computational demands, and suboptimal processing efficiency. **Methods:** This systematic review, guided by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA), examines deep learning-based detection methods for breast cancer published between 2020 and 2024. The analysis includes 39 peer-reviewed studies and 20 widely used WSI datasets. **Results:** To enhance clinical relevance and guide model development, this study introduces a five-dimensional evaluation framework covering accuracy and performance, robustness and generalization, interpretability, computational efficiency, and annotation quality. The framework facilitates a balanced and clinically aligned assessment of both established methods and recent innovations. **Conclusions:** This review offers a comprehensive analysis and proposes a practical roadmap for addressing core challenges in WSI-based breast cancer detection. It fills a critical gap in the literature and provides actionable guidance for researchers, clinicians, and developers seeking to optimize and translate WSI-based technologies into clinical workflows for comprehensive breast cancer assessment.



Academic Editor: Daniele Giansanti

Received: 18 March 2025

Revised: 24 April 2025

Accepted: 28 April 2025

Published: 30 April 2025

Citation: Xu, Q.; Adam, A.; Abdullah, A.; Bariyah, N. Advanced Deep

Learning Approaches in Detection

Technologies for Comprehensive

Breast Cancer Assessment Based on

WSIs: A Systematic Literature Review.

Diagnostics **2025**, *15*, 1150. [https://](https://doi.org/10.3390/diagnostics15091150)

doi.org/10.3390/diagnostics15091150

Copyright: © 2025 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article

distributed under the terms and

conditions of the Creative Commons

Attribution (CC BY) license

([https://creativecommons.org/](https://creativecommons.org/licenses/by/4.0/)

[https://creativecommons.org/](https://creativecommons.org/licenses/by/4.0/)

[https://creativecommons.org/](https://creativecommons.org/licenses/by/4.0/)

[https://creativecommons.org/](https://creativecommons.org/licenses/by/4.0/)

Keywords: breast cancer; detection algorithm; whole slide images; deep learning; systematic literature review

1. Introduction

Breast cancer remains one of the most prevalent and deadly malignancies affecting women worldwide, posing a significant threat to both health outcomes and quality of life [1–3]. Despite remarkable advances in medical diagnostics and therapeutic strategies, the molecular heterogeneity and diverse clinical manifestations of breast cancer continue to complicate its detection and management [4]. Biomarker and lymphocyte detection are critical components in breast cancer assessment, offering essential guidance for diagnosis, classification, and prognosis [5].

Among these, the estrogen receptor (ER) serves as a key molecular marker for subtype classification and supports the development of personalized treatment regimens, directly influencing patient outcomes [6]. The proliferation index Ki-67 Antigen (Ki67) provides independent predictive value for treatment response, while the joint evaluation of ER and progesterone receptor (PR) further refines hormone therapy decisions [7]. Overexpression of human epidermal growth factor receptor 2 (HER2) correlates with increased tumor aggressiveness and recurrence risk, making HER2-positive cases suitable for targeted therapy. In addition, tumor-infiltrating lymphocytes (TILs) are considered reliable prognostic indicators, especially for triple-negative breast cancer (TNBC) patients, where TIL density correlates with both disease-free and overall survival [8]. Accurate identification of these biomarkers and immune features forms the foundation for effective clinical decision-making.

Recent advancements in whole-slide imaging (WSI) have enabled digital pathology to visualize entire tissue sections at ultra-high resolution, capturing both the spatial distribution of biomarkers and the microenvironmental context of lymphocytes [9,10]. These large-format digitized slides—typically prepared using Hematoxylin and Eosin (H&E) staining for structural visualization and Immunohistochemistry (IHC) for highlighting specific molecular targets—provide pathologists with richer information but also introduce significant technical challenges. The sheer volume of image data—often exceeding one billion pixels per slide—and the complex spatial patterns of diagnostic features render manual analysis time-consuming, subjective, and difficult to scale [11,12]. Traditional workflows depend heavily on expert interpretation, leading to potential variability and limited reproducibility in large-scale clinical studies.

To address these limitations, deep learning has emerged as a promising approach for automating WSI analysis. While segmentation tasks aim to label each pixel according to tissue type or structure [13], detection-based techniques are particularly well-suited for clinical applications. They generate interpretable, quantitative outputs such as cell counts, biomarker localization, and lymphocyte spatial distributions, which align more directly with diagnostic workflows [13,14]. This review focuses on detection-oriented methods in WSI analysis and their applications across four critical clinical tasks: diagnosis, classification, grading, and prognosis.

Nonetheless, the integration of deep learning into WSI-based breast cancer assessment remains nontrivial. The ultra-high resolution of WSIs demands substantial computational resources and presents design challenges for conventional model architectures [15,16]. Furthermore, WSIs exhibit a multi-scale nature—ranging from microscopic cell morphology to macroscopic tissue organization—necessitating models capable of learning across spatial hierarchies [17]. The analysis is further complicated by sparse diagnostic features, inter-sample heterogeneity, and staining variations. Compounding these issues is the scarcity of large, well-annotated datasets, which limits the scalability of supervised learning approaches [18]. Interpretability also remains a major concern, as the opaque decision-making processes of deep learning models reduce clinical trust [19]. Effective detection algorithms must therefore balance sensitivity to subtle morphological cues with robustness against noise and artifacts—especially when identifying key features such as biomarker expression and lymphocyte infiltration.

Despite a growing body of literature in this domain, existing reviews often remain narrowly focused. Many concentrate on segmentation techniques or algorithmic performance metrics, without sufficiently addressing the clinical integration of detection methods or their practical utility [20–22]. Additionally, critical issues such as dataset bias, computational burden, and feedback mechanisms from clinical deployment are frequently

overlooked. This gap highlights the need for a more holistic and clinically grounded synthesis of detection-focused research in breast cancer WSIs.

To this end, this review aims to address three key research questions:

1. What types of datasets are used for comprehensive breast cancer assessment using WSIs?
2. What are the main challenges associated with comprehensive breast cancer assessment using WSIs?
3. How do WSIs impact the accuracy and reliability of advanced deep learning approaches for comprehensive breast cancer assessment?

Figure 1 summarizes the task types, datasets, AI models, and evaluation criteria that frame this review of deep learning-based breast cancer assessment using WSIs. This study systematically reviews the landscape of relevant approaches published between 2020 and 2024, following PRISMA guidelines [23]. It synthesizes recent methodological advancements, identifies unresolved challenges, and outlines future research directions. In particular, we emphasize the clinical significance of detection tasks, the scalability of AI-based diagnostic tools, and the potential for deep learning to transform biomarker discovery and personalized treatment planning in digital pathology.

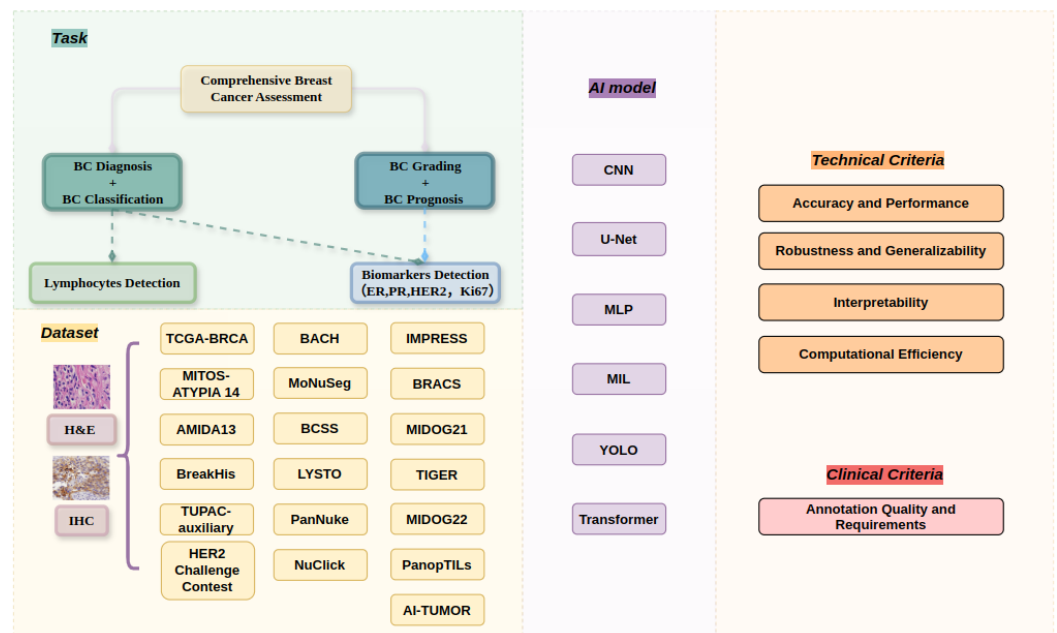


Figure 1. Overview of detection tasks, datasets, models, and evaluation criteria in WSI-based breast cancer assessment.

2. Methods

The study followed the PRISMA guideline [24] in this systematic literature review. PRISMA provides a consistent, repeatable process for locating, assessing, and selecting pertinent research. It also provides guidance on how to choose, recognize, and evaluate studies [25,26].

Figure 2 shows the PRISMA procedure used for this systematic literature review. The next subsection provides details on the review process.

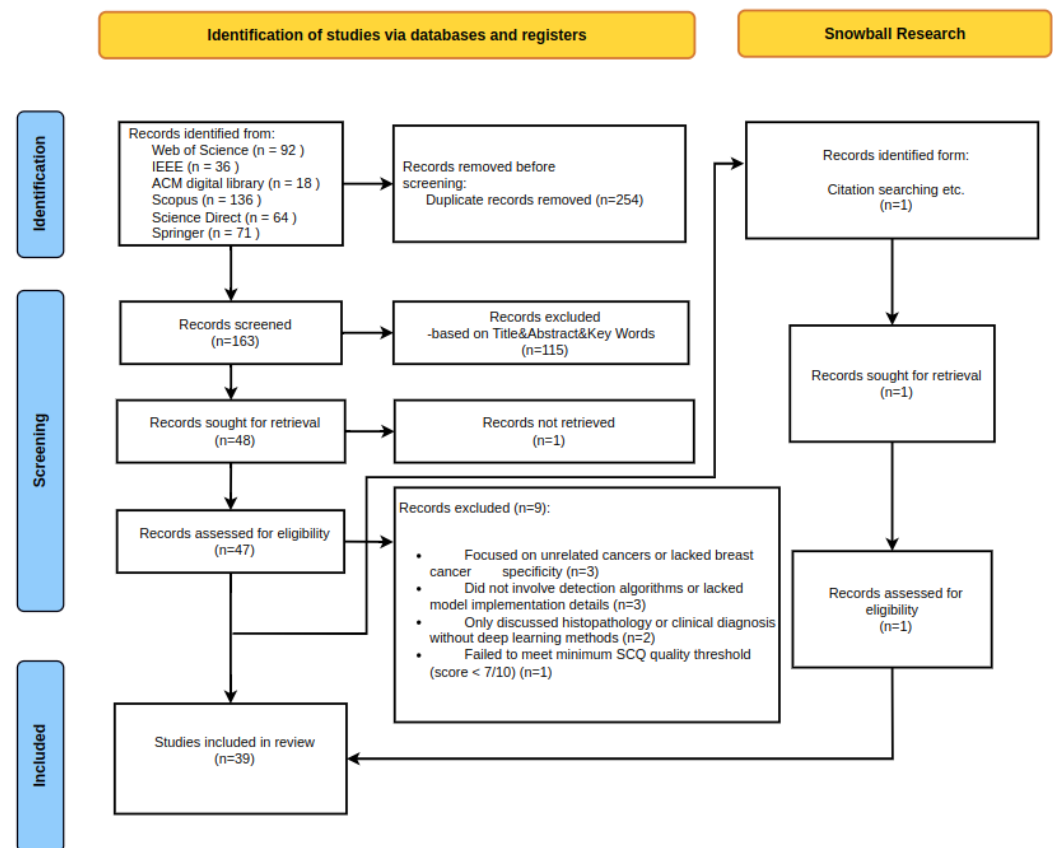


Figure 2. PRISMA flow diagram of the SLR.

2.1. Data Sources and Search Strategy

To identify relevant academic publications, 8 major bibliographic databases were used, including Scopus, IEEE Xplore Library, Web of Science, SpringerLink, ACM Digital Library, and ScienceDirect. Consequently, the systematic literature review focused on articles published between 2020 and 2024 to ensure the inclusion of the most recent and relevant research findings. The search strategy employed combinations of key terms like the following:

- “breast cancer detection” AND “deep learning”
- “breast cancer diagnosis” AND “deep learning”
- “convolutional neural networks” AND “breast cancer”
- “lymphocytes detection” OR “biomarkers detection” AND “breast cancer”
- “H&E stained images” AND “deep learning” AND “breast cancer”
- “immunohistochemistry” AND “deep learning” AND “breast cancer”
- “automated breast cancer diagnosis” OR “AI in breast cancer screening”

These search parameters were created to encompass the depth of knowledge regarding deep learning applications in the detection and analysis of breast cancer.

2.2. Selection Criteria

An exhaustive search strategy combining both automated and manual methods initially retrieved 417 academic publications. Throughout the entire screening process, the inclusion and exclusion criteria outlined in Table 1 were applied consistently. After eliminating 254 duplicate records, 163 unique publications were screened based on their titles, abstracts, and keywords. At this stage, 115 papers were excluded due to a lack of domain relevance or methodological inadequacy.

Table 1. Inclusion and exclusion criteria.

Inclusion Criteria	Exclusion Criteria
Articles related to advanced deep learning-based techniques for comprehensive breast cancer assessment, focusing on diagnosis, classification, grading, and prognosis	Duplicates
Emphasis on peer-reviewed journal articles and conference papers	Articles not addressing breast cancer detection, classification, grading, or prognosis strategies
Studies employing WSI techniques and related fields, particularly those focusing on breast cancer diagnosis and prognosis	Studies not related to WSI techniques and their application in breast cancer assessment
Articles that meet rigorous quality-assessment standards and have a clear focus on the two major tasks of breast cancer assessment	Articles not meeting quality-assessment standards
Articles addressing breast cancer-detection techniques through WSIs, focusing on key points like biomarkers and lymphocytes	Studies not utilizing WSIs or not focusing on the specified key points for breast cancer assessment

Following this, 48 full-text articles were retrieved for in-depth evaluation. One article was excluded due to inaccessibility, and the remaining 47 underwent full eligibility assessment. Nine additional studies were excluded for the following reasons: three did not focus specifically on breast cancer or its primary clinical tasks (e.g., detection, classification, prognosis); three lacked concrete details regarding deep learning algorithms or implementation strategies; two relied solely on conventional pathology without computational approaches; and one failed to meet the minimum quality threshold on the Standard Quality Checklist (SCQ).

This systematic and criteria-driven selection process resulted in 39 high-quality studies being included in the final review. The characteristics of these selected publications are summarized in Table 2, providing a foundation for subsequent quality assessment and synthesis.

Table 2. Summary of included studies with evaluation and data scale (LD = lymphocyte detection; BD = biomarker detection; - = not specified in the article).

Source	Year	Task	Model	Result	Dataset and Scale
1 [27]	2020	BD	AlexNet	F1 score = 0.73	Private Dataset 50 Ki67 WSIs 50 H&E WSIs
2 [28]	2020	LD	NucTraL + BCF	Accuracy = 0.9691	BreakHis Benign: 637 WSIs Malignant: 1390 WSIs
3 [29]	2020	LD	Deep-CNN + FSRM	Accuracy = 98.8%, F1 score = 0.967	BreakHis Benign: 2480 WSIs Malignant: 5429 WSIs

Table 2. Cont.

Source	Year	Task	Model	Result	Dataset and Scale
4 [30]	2020	LD	Faster R-CNN	F1 score: Scattered lymphocytes: 0.9615 Agglomerated lymphocytes: 0.8645 Artifact area: 0.8197	LYSTO 1228 WSIs
5 [31]	2021	BD	U-Net-piNET	F1 score: Ki67-: 0.868; Ki67+: 0.804	Private Dataset 1142 WSIs
6 [32]	2021	BD	U-Net	Pearson correlation coefficient of $r = 0.95$	Private Dataset 200 WSIs
7 [33]	2021	BD	DLS	AUC = 0.60	TCGA-BRCA 3274 WSIs
8 [34]	2021	LD	UBCNN	-	MoNuSeg 120 WSIs
9 [35]	2021	LD	Micro-Net + MLP	F1 score = 0.87	TCGA-BRCA 4281 nuclei images
10 [36]	2021	LD	UNMaSk	F1 score: H&E: 0.856 IHC: 0.9037	Private Dataset 178 WSIs
11 [37]	2021	LD	CNN	AUROC = 0.864	TCGA-BRCA 7075 WSIs
12 [38]	2021	LD & BD	PathoNet	F1 score = 0.7928	SHIDC-BC-Ki-67 2357 WSIs
13 [39]	2021	LD	SegNet + U-Net	F1 score = 0.91	BCa 100 ROIs
14 [40]	2021	LD	WeakSTIL	-	TCGA-BRCA 286 WSIs
15 [41]	2021	LD	CNN	-	TCGA-BRCA 2358 WSIs
16 [42]	2022	LD	TDC-LC	F1 score = 0.892	LYSTO 19663 patches
17 [43]	2022	LD	DDTNet	F1 score = 0.907	TCGA-BRCA 865 WSIs
18 [44]	2023	BD	GAN+CNN+U-Net	-	Private Dataset 321 samples
19 [45]	2023	LD	CB-HVT Net	F1 score = 0.88	LYSTO & NuClick 21312 images
20 [46]	2023	LD	AC-Former	F1 score = 0.796	TCGA-BRCA 452 images
21 [47]	2023	LD	-	-	LYSTO 83 WSIs

Table 2. Cont.

Source	Year	Task	Model	Result	Dataset and Scale
22 [48]	2023	LD	CBAM-Residual U-Net	F1 score = 0.9	TNBC 81 WSIs
23 [49]	2023	LD	U-Net +Mask R-CNN	F1 score = 0.941	Private Dataset 63 WSIs
24 [50]	2023	LD	MobileNetV2 +U-Net	F1 score = 0.927	Private Dataset 30 WSIs
25 [51]	2023	LD	DC-Lym-AF	F1 score = 0.84	LYSTO & NuClick 871 images
26 [52]	2023	LD	FMDet	F1 score = 0.773	MIDOG21, MITOS-ATYPIA 14, AMIDA13, TUPAC-auxiliary 403 WSIs
27 [53]	2023	LD	DeepLabV3	F1 score = 71.23	OCELOT, TIGER 17041patches
28 [54]	2023	LD	BCF-Lym-Detector	F1 score: LYSTO: 0.93 NuClick: 0.84	LYSTO & NuClick 16326 images
29 [55]	2023	LD	CNN	F1 score: Immune cell: 0.82 Tumour cell: 0.92 Stromal cell: 0.81	Private Dataset 2549 images
30 [56]	2023	LD	PROACTING	AUC = 0.88	IMPRESS 1053 images
31 [57]	2024	BD	Based on CNN	AUC = 0.72	TCGA-BRCA 14435 images
32 [58]	2024	BD	ViT	F1 score = 0.9269	Her 2 challenge contest 172 WSIs
33 [59]	2024	LD	Cell-Tissue-ViT	F1 score = 0.7243	OCELOT 400 Patches
34 [60]	2024	LD	Tissue Context + Cell Refinement	F1-score = 0.7473	OCELOT 400 Patches
35 [61]	2024	LD	PathoNet	F1 score: SHIDC-B-Ki-67: 0.842 LSOC-Ki-67: 0.909	SHIDC-BC-Ki-67, LSOC-Ki-67 2447 WSIs
36 [62]	2024	LD	DeepMitosisNet	F1 score = 0.93	MITOS-ATYPIA 14 2127 images
37 [63]	2024	LD	CellViT	F1 score = 0.81	PanNuke 7959 images
38 [64]	2024	LD	YOLOv5	F1 score = 0.7438	TIGER 1879 ROIs
39 [65]	2024	LD	MuTILs	AUROC = 93.0	PanopTILs 1317 ROIs

2.3. Quality Assessment

Assessing the quality of evidence is crucial in a systematic literature review (SLR), as methodological biases may influence outcomes and lead to misinterpretation. To ensure the reliability and rigor of the included studies, this review adopted the Standard SCQ proposed by [66], which comprises ten evaluation items. Following the approach of [57], only studies that provided a “yes” response to at least seven SCQ items were included.

Among the 39 studies selected after quality filtering, the SCQ score distribution was as follows: 10 studies received a full score (10/10), 15 studies scored 9/10, 11 studies scored 8/10, and 3 studies met the minimum threshold with 7/10. Studies scoring below this threshold were excluded to maintain methodological rigor. This distribution reflects the overall quality consistency of the selected literature and ensures that only robust and dependable studies were included in the final synthesis. The SCQ evaluation and data extraction processes were closely integrated to enhance the validity and significance of the review outcomes. Table 3 outlines the SCQ criteria used in this study.

Table 3. Checklist for quality assessment.

Checklist for Quality Assessment	SCQ No
Is the report coherent and easy to read?	SCQ1
Is the research’s purpose well-defined?	SCQ2
Is the procedure for gathering data clearly laid out?	SCQ3
Have the settings of diversity been thoroughly examined?	SCQ4
Can the study’s conclusions be trusted?	SCQ5
Is there a connection between the information, analysis, and conclusion?	SCQ6
Is the process of experimentation and approach transparent?	SCQ7
Are the research methods sufficiently documented?	SCQ8
If they are credible, are they important?	SCQ9
Is it possible to duplicate the research findings?	SCQ10

2.4. Data Extraction and Synthesis

The study noted the pertinent information from each study, such as the publisher, authors, and year of publication, and gathered data for our SLR on the deep learning techniques used, reported accuracy, and assessment criteria. To answer the study’s research questions, the authors particularly examined the results during the data synthesis phase. The authors used a variety of visualization approaches and tools, including tables and diagrams, to make this analysis easier.

3. Results and Meta-Analysis

The meta-analysis of the search results from our systematic literature review is shown in this section. It starts with a summary of the chosen articles before going into each of the study questions that were developed and stated in the introduction.

3.1. Overview of Selected Studies

Figure 3 provides a chronological overview of the articles selected for this SLR, detailing the number of publications related to advanced deep learning methods for breast cancer cell detection using digital pathology images from 2020 to 2024. The diagram indicates a growing trend in this research area over the past few years, particularly from 2023 to 2024, when the number of published articles significantly increased. Most of the articles considered for this study were published after 2023. Specifically, the highest number of articles published in a single year was 13 in 2023 and 9 in 2024, followed by 10 papers in 2021 and 4 papers in 2020. In contrast, only 2 papers were reviewed in 2022, which some literature attributes to the impact of the COVID-19 pandemic on research output.

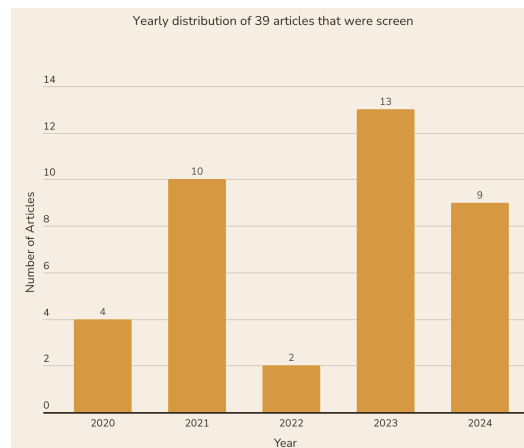


Figure 3. Diagram showing the number of articles published per year from 2020 to 2024.

3.2. Research Question 1: What Types of Datasets Are Employed for Comprehensive Breast Cancer Assessment Using WSIs?

This research evaluates 20 key WSIs datasets pivotal to advancing deep learning applications in breast cancer detection, diagnosis, classification, grading, and prognosis. These datasets, each contributing uniquely to comprehensive breast cancer assessment, reflect the rapid evolution and diversification of digital pathology resources.

The following attributes define the state of the breast cancer WSI datasets:

1. Scale and Origin:

- The basic resources are large-scale public repositories such as TCGA-BRCA (3111 WSIs) [67].
- Focused, high-resolution data are available from specialized datasets like PanNuke (200,000 nuclei across 19 tissue types) and BACH (400 patches, 30 WSIs) [68,69].
- For particular study goals, derivative datasets (MoNuSeg, BCSS, TIGER) expand on primary sources [70].

2. Research Focus and Annotation Granularity:

- Datasets cover a wide range of anatomical structures, such as nuclei in MoNuSeg and tumor-infiltrating lymphocytes in PanopTILs, to specific cellular structures.
- Increasing model sophistication is reflected in the variation in annotation detail from whole-slide to pixel-level [71].

3. Multi-modal Integration:

- Datasets that integrate clinical and genetic information with histopathological images, such as TCGA-BRCA, provide opportunities for more thorough and comprehensive analysis [72,73].

4. Ethical Considerations and Diversity:

- More recent datasets, such as AI-TUMOR, highlight the diversity of patient demographics and the use of ethical data collection techniques [74].

Together, these datasets enable a wide range of deep learning uses, ranging from simple cell identification to intricate tumor categorization and prognostic modeling. In addition to driving innovation in deep learning architectures suited to the particular difficulties of breast cancer analysis, they act as benchmarks for model development and validation [75,76].

These datasets development follows the field's move toward more extensive, fully annotated resources on a wider scale. This pattern suggests a developing sector ready

to use cutting-edge computational techniques to address the intricate, multidimensional character of breast cancer research, along with the integration of multimodal data and a growing emphasis on ethical issues.

Table 4 lists the various datasets and highlights their unique properties that were employed for WSI-based comprehensive breast cancer assessment. Each dataset plays a crucial role in advancing deep learning methodologies by providing diverse and detailed data for model training and validation.

Table 4. Key WSI datasets relevant to comprehensive breast cancer assessment.

Name	Year	Key Features	Contribution	Link
TCGA-BRCA [77]	-	3111 H&E-stained WSIs from 1086 female and 12 male breast cancer patients. Includes matched gene expression data and clinical information.	Widely used to create derivative datasets like MoNuSeg, BCSS, LYSTO, and TIGER.	https://portal.gdc.cancer.gov/projects/TCGA-BRCA (accessed on 18 July 2024)
MITOS-ATYPIA 14	2013	2400 high-power field (HPF) images from 11 breast cancer patients, scanned by two devices: 1200 at 1539×1376 pixels and 1200 at 1663×1485 pixels, all at $40\times$ magnification.	Provides a comprehensive external test set for evaluating the robustness of mitosis-detection models across different imaging conditions.	https://mitos-atypia-14.grand-challenge.org/Description/ (accessed on 18 July 2024)
AMIDA13 [78]	2013	606 HPF images (311 training, 295 test) from 23 subjects, each 2000×2000 pixels (0.25 mm^2)	Established a benchmark for mitosis-detection algorithms, later integrated into larger datasets for broader impact.	-
BreakHis [79] (accessed on 18 July 2024)	2016	9109 breast tissue images from 82 patients at various magnifications ($40\times$ – $400\times$). Includes 2480 benign and 5429 malignant samples, all 700×460 pixel RGB images. Samples collected via partial mastectomy.	Provides a diverse, well-annotated dataset for developing and evaluating breast cancer classification algorithms. Enables research on automated diagnosis across different magnifications and tumor types, potentially improving clinical diagnostic accuracy.	https://web.inf.ufpr.br/vri/databases/breast-cancer-histopathological-database-breakhis/ (accessed on 18 July 2024)
TUPAC-auxiliary [78]	2016	73 breast cancer WSIs from 3 pathology centers, scanned by 2 types of scanners at $40\times$ magnification.	Integrates multi-center data, enhancing the scope of mitosis-detection research in breast cancer.	https://tupac.grand-challenge.org/ (accessed on 18 July 2024)
HER2 Challenge Contest [80]	2016	High-resolution breast cancer histology dataset. Comprises 100 gigapixel WSIs. Annotated with expert pathologist HER2 scores and percentage assessments.	Pioneering benchmark for automated HER2 quantification in digital pathology. Facilitates the development of AI tools to enhance diagnostic consistency and efficiency in breast cancer assessment.	https://warwick.ac.uk/fac/cross_fac/tia/data/her2contest/ (accessed on 18 July 2024)
BACH [81]	2018	Includes 400 H&E-stained patches (2048×1536 resolution) and 30 WSIs with pixel-level annotations	Useful for training models with pixel-level cancer type annotations	https://www.kaggle.com/datasets/truthisneverlinear/bach-breast-cancer-histology-images (accessed on 18 July 2024)
MoNuSeg [82]	2018	WSIs from 30 organs, creating 1000×1000 pixel sub-images with nuclear annotations.	Ensures variation in nuclear appearance, enhancing model training on diverse tissue samples.	https://monuseg.grand-challenge.org/Data/ (accessed on 18 July 2024)
BCSS [83]	2019	151 WSIs with representative regions of interest (ROI) selected, contributing to the TIGER dataset.	Helps in understanding tumor-infiltrating lymphocytes in HER2+ and triple-negative breast cancers.	https://bcsegmentation.grand-challenge.org/ (accessed on 18 July 2024)
LYSTO [47]	2019	LYSTO data set comprises 20,000 images from 43 patients with breast, colon, and prostate cancers. It's structured with a patient-level split: 19 for training, 9 for validation, and 6 for testing, ensuring diverse representation across cancer types and stages.	Enables development of cross-cancer lymphocyte-assessment algorithms. Patient-level data division supports robust AI model validation, advancing generalized lymphocyte-detection tools.	https://lysto.grand-challenge.org/ (accessed on 18 July 2024)

Table 4. Cont.

Name	Year	Key Features	Contribution	Link
PanNuke [84]	2020	Contains 200,000 nuclei categorized into five clinically significant classes, with high-resolution patches scanned at 20× or 40× magnification	Supports nuanced classification of different tissue types in breast cancer research	https://sites.google.com/view/panoptils (accessed on 18 July 2024)
NuClick [85]	2020	NuClick dataset comprises 871 images derived from 440 WSIs, strategically partitioned into 471 training, 99 validation, and 300 testing images. The dataset encompasses various cancer types, with careful consideration given to maintaining separation between patient samples across different sets.	Supports AI model development for cancer diagnostics with real-world applicability. Thoughtful data division prevents patient-level leakage, enhancing model reliability and clinical relevance.	https://github.com/navidstuv/NuClick (accessed on 18 July 2024)
IMPRESS [86]	2020	Large-scale linguistic dataset comprising over 25,000 sentence pairs. Utilizes a rich vocabulary of 3000+ lexical items with grammatical annotations. Generated using advanced natural language processing techniques, following established Natural language inference dataset formats.	Advances pragmatic inference research by providing a comprehensive benchmark for evaluating NLI models' understanding of presuppositions and implicatures. Facilitates the development of more nuanced language understanding systems capable of grasping subtle linguistic phenomena.	https://github.com/facebookresearch/Impres?tab=readme-ov-file (accessed on 18 July 2024)
BRACS [87]	2021	4539 ROIs from 547 H&E WSIs, meticulously categorized into different lesion types.	Facilitates detailed subtyping in breast cancer research.	https://www.bracs.icar.cnr.it/ (accessed on 18 July 2024)
SHIDC-BC-Ki-67 [38]	2021	2357 tru-cut biopsy images of invasive ductal carcinoma, collected from 2017 to 2020. Comprises 1656 training and 701 test samples, all expertly annotated for Ki-67 markers.	Addresses the scarcity of comprehensive Ki-67 marked datasets in breast cancer research. Facilitates development of deep learning models for accurate Ki-67 assessment, potentially enhancing diagnostic precision and treatment planning for invasive ductal carcinoma.	https://shiraz-hidc.com/ki-67-dataset/ (accessed on 18 July 2024)
MIDOG21 [87]	2021	280 breast cancer WSIs at 8000 × 8000 pixels, scanned by 4 different devices.	Addresses scanner variability in mitosis detection, promoting robust algorithm development across diverse imaging equipment.	https://midog2021.grand-challenge.org/ (accessed on 18 July 2024)
TIGER [88]	2022	Includes H&E-stained WSIs of Her2 positive and Triple Negative breast cancer. TIGER comprises cases from clinical routine and a phase 3 clinical trial. Includes annotations for lymphocytes, plasma cells, invasive tumors, and stroma.	First challenge for fully automated assessment of TILs in breast cancer. Aims to validate AI-based TIL scores for clinical use.	https://tiger.grand-challenge.org/ (accessed on 18 July 2024)
MIDOG22 [89]	2022	50 canine mast cell tumor cases, scanned by a single device type.	Expands mitosis detection to veterinary pathology, offering cross-species validation for detection algorithms.	https://midog.deepmicroscopy.org/download-dataset/ (accessed on 18 July 2024)
PanopTILs [65]	2023	Includes annotations for 814,886 nuclei from 151 patients, with WSIs scanned at 20× magnification.	Enhances understanding of the role of TILs in breast cancer detection.	https://sites.google.com/view/panoptils (accessed on 18 July 2024)
AI-TUMOR [73]	2024	2500 WSIs with pixel-level annotations for tumor regions, normal tissue, and surrounding stroma.	Focuses on reducing biases in AI models, ensuring better generalizability.	https://github.com/nickryd/ATLAS/blob/main/README.md (accessed on 18 July 2024)

3.3. Research Question 2: What Are the Main Challenges Associated with Comprehensive Breast Cancer Assessment Using WSIs?

WSI has emerged as a pivotal component in the realm of digital pathology, particularly for the application of deep learning methodologies in the detection of breast cancer [90]. Nonetheless, numerous substantial obstacles hinder its optimal implementation. These

obstacles encompass a spectrum of technical challenges associated with image processing and analytical procedures, as well as overarching issues pertaining to data integrity, and annotation practices. Table 5 summarizes the main issues raised in recent research papers, together with their descriptions, implications for breast cancer detection, and relevant research references.

Table 5. Main challenges in WSIs for comprehensive breast cancer assessment.

Challenge	Description	Impact on Breast Cancer Cell Detection	Source
Size and Complexity of WSIs	WSIs can be several gigabytes in size, containing millions of pixels that need to be processed.	Requires substantial computational resources, high-performance GPUs, and large memory capacities; storage and management of large datasets are logistically challenging.	[28–30]
Variability in Image Quality and Resolution	Inconsistencies in image quality and resolution across different scanners and datasets.	Leads to variations in training data quality, reducing model generalization and accuracy across different scanners.	[27,28,31,39,42,45,51,54,63,64]
Annotation Challenges	High-quality annotations are labor-intensive, time-consuming, and prone to variability.	Inconsistent annotations can introduce biases, affecting model robustness and generalizability.	[29,40,47,57,60,61]
Feature Extraction Difficulties	Complex and subtle patterns in breast cancer tissues are challenging to capture accurately.	Traditional methods often fall short, necessitating advanced deep learning techniques for accurate detection.	[29,30,32–38,40,41,43–60,62,63,65]

3.4. Research Question 3: How Do WSIs Affect the Accuracy and Reliability of Advanced Deep Learning Approaches for Comprehensive Breast Cancer Assessment?

WSI plays a critical role in enhancing the accuracy and reliability of deep learning techniques for comprehensive assessments of breast cancer [91]. Deep learning algorithms can examine complicated tissue structures and cellular configurations with more ease thanks to the high-resolution and finely detailed features of WSIs. This is crucial for distinguishing between cancerous and benign cells. However, as Table 6 summarizes, several critical variables related to WSIs significantly impact the accuracy and reliability of deep learning models when it comes to breast cancer diagnosis.

As Table 6 shows, these factors are interrelated and impact model performance. Reliability in various circumstances is contingent upon constant image quality and standardization, whereas high resolution improves detection accuracy but raises processing demands. Strong annotations and a wide range of training data are essential for the resilience and generalizability of the model. WSI resolution makes it easier to obtain the necessary detail for accurate analytical evaluations, but it also creates problems with data processing and standardization methods. It is crucial to guarantee that WSIs are of superior quality, coherent, and comprehensively annotated, while concurrently tackling the intricacies linked to data heterogeneity and standardization since it is essential to the creation of reliable models that may be successfully applied in clinical settings.

In summary, WSIs have a major influence on the precision and dependability of deep learning techniques used in the detection and assessment of breast cancer. Developing deep learning models that are therapeutically useful requires addressing these variables simultaneously. In order to further improve the accuracy and dependability of deep learning techniques in breast cancer evaluation utilizing WSIs, future research should concentrate on maximizing these factors.

Table 6. The impact of WSIs on the accuracy and reliability of deep learning methods for comprehensive breast cancer assessment.

Factor	Description	Impact on Accuracy and Reliability
Resolution of WSIs	High-resolution images provide detailed information, crucial for detecting subtle differences between normal and cancerous tissues, especially in early-stage cancers.	Enhances detection accuracy but increases computational requirements.
Quality and Consistency	Variations in staining, lighting, and scanner types can introduce noise and artifacts, affecting model predictions.	Inconsistent quality can lead to inaccuracies; standardization techniques are vital.
Availability of Annotated WSIs	High-quality annotations by expert pathologists are essential for training accurate models. Obtaining such annotations is resource-intensive.	Poor annotations reduce model accuracy; improving annotation quality is key.
Diversity and Generalizability of Training	Models trained on diverse WSIs, covering a range of tissue types and patient demographics, are more likely to generalize well across different clinical settings.	Enhances generalization and reliability, but lack of standardization can hinder performance.

4. Criteria for Comprehensive Breast Cancer-Assessment WSI Algorithms

A number of important factors must be considered when evaluating detection criteria for a thorough assessment of breast cancer based on WSI algorithms to guarantee efficacy, consistency, and clinical applicability. These criteria can be divided into two factors: technological and clinical. Each of these factors is vital in establishing the technology's overall value and suitability.

Technical Criteria:

1. **Accuracy and Performance Metrics:** According to [91], sensitivity, specificity, accuracy, recall, AUC, and F1 score are essential for accurately identifying cancer cells while reducing false positives.
2. **Robustness and Generalizability:** Algorithms must manage common problems like noise and artifacts while operating consistently over a range of datasets, scanners, and staining processes [92].
3. **Interpretability and Explainability:** Clinical trust depends on model openness and error analysis capabilities [93].
4. **Computational Efficiency:** Algorithms should be appropriate for a range of computational contexts, with respectable processing speeds and efficient resource consumption [94].

Clinical Criteria:

Annotation Quality and Requirements: Preference for algorithms that reduce reliance on resource-intensive annotations by performing well with minimum or semi-supervised learning [94].

Together, these criteria ensure that the algorithms chosen are not only sound from a methodological standpoint but also practical, significant, and useful in clinical settings. Through a methodical approach to these components, scientists and medical professionals can evaluate and choose algorithms that are suitable for the complex problems involved in WSIs-based breast cancer cell detection, improving diagnostic accuracy and patient outcomes. To drive progress in breast cancer-detection techniques, researchers must prioritize

optimizing baseline models while also carrying out thorough assessments of those models. These models serve as essential baselines to assess the effectiveness of WSI-based detection techniques in clinical settings. Targeted optimization techniques can significantly enhance important characteristics like clinical relevance, accuracy, and robustness, addressing the many problems that come with assessing breast cancer. This systematic approach guarantees that the chosen algorithms enhance patient outcomes and boost diagnostic accuracy.

4.1. Baseline Models for Detection Technologies Applied in Comprehensive Breast Cancer Assessment Based on WSI Algorithms

To better visualize the methodological progression in WSI-based breast cancer detection over recent years, Figure 4 presents a Sankey diagram capturing the dynamic interplay between publication year, detection task, and model architecture from 2020 to 2024. Each stream represents a flow of research attention, where the thickness of the connection reflects the frequency of model usage for specific tasks. As shown in the figure, two major trends emerge: first, a shift in detection focus—from early emphasis on biomarkers to increasing attention on lymphocyte detection, and more recently, to frameworks addressing both targets simultaneously; and second, a transition in model design—from dominant use of Convolutional Neural Network (CNN) and U-Net toward more sophisticated or hybrid approaches involving Transformers and Generative Adversarial Network (GAN), reflecting growing demands for richer spatial modeling and generalization across WSI domains.

These evolving trends are closely reflected in the selection and adaptation of baseline model architectures for different detection tasks. Over the five-year span, CNN and U-Net remained dominant. CNNs were predominantly applied to biomarkers detection, where classification or sparse detection was needed to identify ER/PR/HER2-positive cells. In contrast, U-Net architectures were widely adopted for lymphocyte detection, due to their pixel-level precision and strong segmentation capabilities—critical for detecting densely distributed immune cells with indistinct boundaries. Notably, in tasks combining biomarkers and lymphocyte detection, U-Net-based frameworks were still preferred for their ability to support multi-task outputs, such as simultaneous localization and segmentation.

From 2022 onward, architectural diversification accelerated. Transformer-based models gained traction, especially in biomarker detection, by leveraging self-attention mechanisms to capture long-range contextual dependencies in high-resolution WSI data. Hybrid approaches, such as CNN+Transformer and GAN+CNN+U-Net, emerged around 2023, integrating the spatial locality of CNNs, the generative robustness of GANs, and the global modeling power of Transformers—enabling more adaptive and domain-generalizable detection systems. Meanwhile, lighter or exploratory models like Multilayer Perceptron (MLP), Multiple Instance Learning (MIL), and You Only Look Once (YOLO) appeared after 2021, though their use remained limited due to challenges in dense detection and precise localization on WSIs.

Evaluation metrics across these models varied with task and output granularity. Classification and sparse detection tasks typically employed AUC, accuracy, and F1-score, while segmentation-oriented models were assessed using Dice coefficient, IoU, and boundary-aware metrics. For multi-output detection models, task-specific metrics were reported independently, reflecting the complexity of comprehensive breast cancer assessment.

Overall, segmentation-centric models have remained the backbone of WSI-based detection. Their ability to handle high-resolution images through patch-based processing, preserve pixel-level detail, and support dense prediction tasks makes them especially suited for WSI applications, where targets are often overlapping, small, and structurally complex. The increasing use of hybrid and Transformer-based models reflects a broader trend toward unifying global and local representation learning for improved clinical utility.

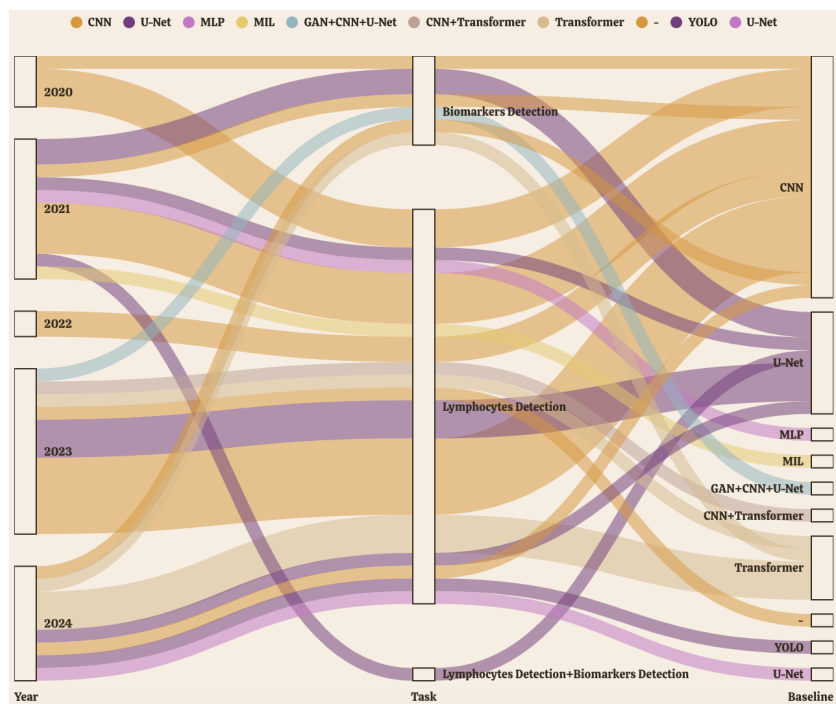


Figure 4. Evolution of baseline-detection models for breast cancer assessment in WSI from 2020 to 2024. (Flow width indicates the relative usage frequency of each model type, with thicker streams showing more widely adopted models. Different colors represent distinct model architectures tracked across the study period).

4.2. Optimizing and Improving Existing Baselines Based on Evaluation Criteria

A multidisciplinary approach is required to optimize current baseline models for breast cancer cell detection using WSI, considering technological and clinical factors. Several important categories can be used to group this optimization process:

4.2.1. Enhancing Model Performance

In the realm of deep learning-based detection techniques for computer-aided diagnosis (CAD) of breast cancer, strategies to enhance model performance have become increasingly diverse and sophisticated. Primarily, ensemble learning frameworks effectively improve detection accuracy and generalization by integrating various deep learning architectures such as U-Net, GANs, and CNNs [36,42–46,48–51,53–55,58,59,62,65]. Additionally, the application of multi-task learning paradigms, which allow models to jointly learn multiple related tasks such as lesion segmentation, classification, and malignancy grading within a unified architecture, has gained attention for its ability to leverage shared representations, reduce overfitting, and improve generalization across heterogeneous lesions. This approach not only provides more comprehensive pathological information but also enhances the model's robustness and diagnostic performance [43,53,56,59,61,63]. Furthermore, multimodal data fusion strategies integrate genomic data with WSIs [27,41], or combine multi-level data from cellular and tissue levels [37,57,60], significantly boosting diagnostic precision and interpretability by capturing complementary biological features. Training on cross-institutional, multi-center datasets enhances the model's domain adaptability, mitigating the adverse effects of data distribution shifts on performance [34].

Regarding model architectures, increasing network depth equips models with stronger feature extraction capabilities, enabling the capture of more complex pathomorphological features [30]. The introduction of diverse convolutional modules, such as residual convolutional blocks [38], parallel convolutional blocks [35], dilated convolutional blocks [51], and color deconvolution [50], further enhances the model's feature representation ability

and multi-scale information capture. Recently, the integration of multiple attention mechanisms, including combinations of spatial attention, channel attention, and self-attention, has gained widespread application in medical image analysis, improving detection precision and efficiency by enhancing the model's focus on key regions and features.

Lastly, post-processing optimization techniques, including morphological opening operations, watershed algorithms [32,38], and advanced methods like HoVer-Net [63], further elevate the precision and consistency of detection outcomes, particularly in cell instance segmentation and separation of adhered structures. The synergistic effect of these methods is driving the gradual implementation and application of breast cancer CAD systems in clinical practice.

While recent advances have significantly improved detection accuracy, they often come at the expense of interpretability and computational efficiency. In deep learning-based CAD systems, increasing model complexity to boost performance tends to reduce transparency and raises deployment barriers. For example, models like CB-HVT Net, which integrate PVT, ResNet variants, and attention mechanisms, feature high parameter counts and computational demands that hinder real-time deployment [45,54]. These challenges highlight the need for balanced approaches that maintain strong accuracy while enhancing explainability and deployability.

4.2.2. Improving Robustness and Generalizability

In the realm of deep learning-based breast cancer-detection research, robustness enhancement strategies predominantly converge on two pivotal methodologies. Cross-disease data integration and validation, referring to the inclusion of datasets from multiple cancer types to expose the model to a broader range of pathological variations, substantially augments model robustness through the incorporation of multi-disease datasets and the execution of pan-cancer experiments [34,37,46,48,50,54]. This approach facilitates the model's acquisition of heterogeneous pathological features, thereby enhancing its discriminative capacity across diverse cancer types and elevating its adaptability within complex clinical milieus. Concurrently, the implementation of semi-supervised learning paradigms efficaciously addresses the constraints imposed by the paucity of annotated data [29]. By synergistically leveraging limited labeled datasets in conjunction with voluminous unlabeled data, this technique not only mitigates the reliance on extensive manual annotation processes but also markedly amplifies the model's generalization prowess in data-constrained environments.

However, while cross-disease integration can enhance the model's adaptability to diverse histopathological morphologies, it may also introduce label noise and inconsistencies in annotation standards. This, in turn, can lead to domain shift and reduce detection specificity for certain cancer types, such as breast cancer. For example, when independent models were constructed for five cancer types in TCGA and pan-cancer training was applied, the performance of pan-cancer models declined in certain tasks (e.g., PD-L1 prediction in STAD), suggesting that disease-specific signals may be diluted in a multi-cancer setting [37]. Moreover, many semi-supervised frameworks rely on heuristic thresholds or consistency regularization strategies, which often require task-specific tuning, thereby limiting their generalizability and scalability in real-world clinical applications.

4.2.3. Increasing Interpretability and Explainability

Interpretability techniques in deep learning have emerged as essential components for fostering trust, transparency, and clinical acceptability in breast cancer assessment. The Human-Interpretable Features (HIF) paradigm aims to bridge the semantic gap between model outputs and clinical understanding by aligning predictions with visually

and diagnostically meaningful image features [37], whereas saliency-based visualization methods offer intuitive heatmaps that localize regions contributing most to the model's decision-making process, thereby enhancing interpretability for end-users [40]. While the integration of such methods has demonstrably improved the transparency and perceived reliability of AI-assisted diagnostic systems, several limitations persist. HIF-based strategies typically rely on predefined, handcrafted feature sets, which may insufficiently capture the complex and abstract representations encoded by deep neural networks, thus constraining their explanatory power and generalizability across datasets or imaging modalities. In parallel, saliency visualizations are often susceptible to input perturbations and architectural variations, producing unstable and sometimes misleading attributions. Moreover, the post hoc nature of most interpretability tools, coupled with a lack of standardized validation protocols, raises concerns regarding their clinical robustness and reproducibility. These limitations underscore an urgent need for the development of principled, rigorously evaluated interpretability frameworks that can yield consistent, meaningful, and clinically actionable explanations.

4.2.4. Optimizing Computational Efficiency

In research applying deep learning techniques to breast cancer assessment, various strategies have been employed to enhance model efficiency. The primary approach involves precise localization of ROI through methods such as Gaussian kernel annotation [32] and micro-block selection techniques [31,64], enabling models to focus on key pathological features. This not only improves accuracy but also reduces computational costs. Additionally, researchers have utilized pre-training strategies to optimize model architecture [29,31], accelerating model convergence, improving task initialization, and reducing parameter count and architectural complexity, thereby lowering computational demands while maintaining high performance. However, these efficiency-oriented strategies are not without limitations. ROI localization methods often rely on heuristic rules or expert-defined annotations, which may introduce bias and reduce scalability across datasets with varying staining characteristics or image resolutions. Moreover, such narrowly focused techniques risk omitting relevant contextual cues essential for accurate diagnosis. Similarly, the effectiveness of pre-training heavily depends on the relevance and quality of the source domain; mismatched pre-training can result in suboptimal initialization and diminished downstream performance. Furthermore, existing studies rarely provide systematic evaluations of the trade-offs between architectural simplification and diagnostic accuracy, leaving the optimal balance between efficiency and performance largely unexplored.

4.2.5. Addressing Data Quality and Annotation Challenges

In deep learning-based breast cancer-detection research, data quality and annotation challenges are primarily mitigated through two methodologies: weak supervision learning and segmentation map generation. Weak supervision enables models to extract features from limited or imprecise annotations, reducing dependence on fully annotated datasets [29,31,41]. Segmentation map generation techniques create synthetic annotations, providing finer-grained information on regions of interest, thus compensating for incomplete or noisy labels [39]. These approaches synergistically enhance model robustness and performance in the face of data quality and annotation challenges in breast cancer-detection tasks. However, both methods present notable limitations. Weakly supervised models are vulnerable to label noise and may overfit to coarse annotations, limiting their generalizability. Meanwhile, segmentation map generation often relies on heuristic or rule-based pseudo-labels that may introduce bias, particularly in complex tumor microenvironments. The lack of standardized

validation procedures for these synthetic annotations also raises concerns regarding their clinical reliability and reproducibility.

Table 7 concisely summarizes the baseline models and their optimization strategies, aiding in the understanding of how these models can be improved to enhance the efficiency of comprehensive breast cancer assessment based on detection technologies.

Table 7. Baseline models for WSI detection technologies and the strategies for optimizing and improving these models based on specific criteria.

Criteria	Baseline Models	Optimization and Improvement Strategies
Accuracy and Performance Metrics	CNN, GAN, U-Net, Transformer, MIL, MLP	<ul style="list-style-type: none"> • Ensemble learning frameworks • Multi-task learning paradigms • Multimodal data fusion strategies • Cross-institutional training • Advanced convolutional modules • Multiple attention mechanisms
Robustness and Generalizability	CNN, U-Net, Transformer	<ul style="list-style-type: none"> • Cross-disease data integration and validation • Semi-supervised learning paradigms
Interpretability and Explainability	CNN, MIL	<ul style="list-style-type: none"> • HIF approach • Saliency visualization techniques
Computational Efficiency	CNN, U-Net, YOLO	<ul style="list-style-type: none"> • Precise localization of ROI • Pre-training strategies
Annotation Quality and Requirements	CNN, U-Net	<ul style="list-style-type: none"> • Weak supervision learning • Segmentation map generation techniques

5. Discussion and Potential Solutions for Improving WSIs for Breast Cancer Cell Detection

The application of advanced deep learning methodologies to WSIs for comprehensive breast cancer assessment represents a significant advancement in digital pathology. A systematic analysis of WSI datasets, associated challenges, and their impact on deep learning models has yielded five critical evaluation metrics: model performance, data integration and preprocessing, architectural optimization, robustness and generalizability, and interpretability and clinical application. This analytical framework provides a comprehensive evaluation of current technologies while illuminating future research trajectories.

The diversity of WSIs datasets, ranging from large-scale public repositories such as TCGA-BRCA to specialized high-resolution datasets like PanNuke and BACH, significantly influences model performance and data integration strategies. The multimodal nature of datasets such as TCGA-BRCA offers a robust foundation for developing comprehensive assessment models. However, dataset heterogeneity presents substantial challenges, particularly in terms of model generalizability. Innovative approaches have emerged to address these challenges. Notable among these is the integration of U-Net, GANs, and CNNs [43], which demonstrates exceptional performance in handling diverse data. Additionally, the multi-task learning paradigm described in [53] enhances model adaptability to varied WSIs data types through simultaneous segmentation, classification, and grading tasks.

The primary challenges in WSIs applications—including image size, quality variability, and annotation difficulties—directly impact data preprocessing strategies and model architecture design. These challenges have catalyzed innovative solutions. The panoramic segmentation method proposed by Liu et al. effectively addresses large-sized WSIs processing, markedly improving computational efficiency. In data integration and preprocessing, [27] showcases a multimodal fusion strategy combining genomic data with WSIs, enhancing diagnostic accuracy while partially mitigating data heterogeneity issues. Nev-

ertheless, the efficient processing of large-scale WSIs data remains a significant challenge, necessitating further research into advanced preprocessing techniques.

The impact of WSIs on deep learning model accuracy and reliability reveals a dichotomy: high-resolution data enhances detection precision while simultaneously presenting computational and standardization challenges. This contradiction has driven innovations in model architectures. The introduction of residual convolutional blocks [38] significantly enhances feature extraction capabilities, while the attention mechanism described in [63] improves the model's capacity to identify key regions. These optimizations directly address the computational challenges posed by high-resolution WSIs while enhancing the model's ability to capture complex pathological features.

Regarding robustness and generalizability, cross-disease dataset integration [34] and semi-supervised learning methods [29] demonstrate significant advantages in managing diverse WSIs datasets and limited annotated data. These approaches enhance model adaptability across varied clinical settings and effectively address the challenge of WSIs data annotation difficulties.

Interpretability and clinical application remain significant hurdles for WSI-based deep learning models. HIF methods [37] and saliency visualization techniques [40] have advanced the interpretability of model outputs, crucial for fostering clinical trust and adoption. However, the seamless integration of these technologies into clinical workflows requires further investigation. The identified evaluation metrics not only provide a comprehensive assessment of existing models but also delineate key directions for future research. Significant opportunities persist in addressing WSIs dataset diversity, processing large-scale high-resolution data, and improving model robustness and interpretability.

6. Conclusions, Implication, and Recommendations for Future Research

This review has systematically examined deep learning-based detection methods for breast cancer using WSIs, highlighting substantial progress in model accuracy, robustness, and data efficiency. Nonetheless, significant challenges remain in computational scalability, interpretability, annotation quality, and clinical applicability. To address these issues, future research should be guided by both technical feasibility and clinical relevance. In the short term, efforts should focus on developing lightweight, interpretable architectures optimized for WSI-scale processing to support real-time, resource-aware deployment. Enhancing weakly supervised and semi-supervised learning frameworks through uncertainty modeling and confidence-guided label refinement represents a technically viable strategy for improving annotation robustness. Medium-term priorities include designing domain-adaptive and resolution-consistent models to address data heterogeneity across institutions and staining variations. In the long term, the establishment of clinically validated interpretability protocols and the construction of large-scale, standardized WSI datasets should be pursued to support reproducibility, benchmarking, and translational impact. Prioritizing these directions will facilitate a more effective alignment between algorithmic innovation and real-world clinical integration, advancing the role of AI in precision breast cancer diagnostics.

Author Contributions: Writing—original draft preparation and editing, Q.X.; Supervision, review, and editing, A.A. (Afzan Adam) and A.A. (Azizi Abdullah); Medical terminology and case validation, N.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Ministry of Higher Education Malaysia through the Fundamental Research Grant Scheme (FRGS), grant code: FRGS/1/2024/ICT02/UKM/02/5.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

PRISM	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
WSI	Whole Slide Image
WSIs	Whole Slide Images
ER	Estrogen Receptor
PR	progesterone Receptor
HER2	Human Epidermal Growth Factor Receptor 2
Ki-67	Ki-67 Antigen
TIL	Tumor-Infiltrating Lymphocyte
H&E	Hematoxylin and Eosin
IHC	Immunohistochemistry
TNBC	Triple-Negative Breast Cancer
LD	Lymphocyte Detection
BD	Biomarker Detection
SCQ	Systematic Quality Criteria
SLR	Systematic Literature Review
HPF	High-Power Field
CNN	Convolutional Neural Network
GAN	Generative Adversarial Network
MLP	Multilayer Perceptron
MIL	Multiple Instance Learning
YOLO	You Only Look Once
CAD	Computer-Aided Diagnosis
ROI	Region of Interest
HIF	Human-Interpretable Features

References

- Faroughi, F.; Fathnezhad-Kazemi, A.; Sarbakhsh, P. Factors Affecting Quality of Life in Women With Breast Cancer: A Path Analysis. *BMC Women's Health* **2023**, *23*, 578.
- Khairi, S.S.M.; Bakar, M.A.A.; Alias, M.A.; Bakar, S.A.; Liong, C.Y.; Rosli, N.; Farid, M. Deep learning on histopathology images for breast cancer classification: A bibliometric analysis. *Healthcare* **2021**, *10*, 10. [[CrossRef](#)] [[PubMed](#)]
- Saifullah, S.; Dżewski, R. Enhancing breast cancer diagnosis: A CNN-based approach for medical image segmentation and classification. In *International Conference on Computational Science*; Springer: Berlin/Heidelberg, Germany, 2024; pp. 155–162.
- Mahmoud, R.; Ordóñez-Morán, P.; Allegrucci, C. Challenges for triple negative breast cancer treatment: Defeating heterogeneity and cancer stemness. *Cancers* **2022**, *14*, 4280. [[CrossRef](#)]
- El Bairy, K.; Haynes, H.R.; Blackley, E.; Fineberg, S.; Shear, J.; Turner, S.; De Freitas, J.R.; Sur, D.; Amendola, L.C.; Gharib, M.; et al. The tale of TILs in breast cancer: A report from the international immuno-oncology biomarker working group. *NPJ Breast Cancer* **2021**, *7*, 150. [[PubMed](#)]
- Pauzi, S.H.M.; Saari, H.N.; Roslan, M.R.; Azman, S.N.S.S.K.; Tauan, I.S.; Rusli, F.A.M.; Aizuddin, A.N. A comparison study of HER2 protein overexpression and its gene status in breast cancer. *Malays. J. Pathol.* **2019**, *41*, 133–138.
- Finkelman, B.S.; Zhang, H.; Hicks, D.G.; Turner, B.M. The evolution of Ki-67 and breast carcinoma: Past observations, present directions, and future considerations. *Cancers* **2023**, *15*, 808. [[CrossRef](#)]
- Zahari, S.; Syafruddin, S.E.; Mohtar, M.A. Impact of the Cancer Cell Secretome in driving breast Cancer progression. *Cancers* **2023**, *15*, 2653. [[CrossRef](#)]
- Gupta, R.; Le, H.; Van Arnem, J.; Belinsky, D.; Hasan, M.; Samaras, D.; Kurc, T.; Saltz, J.H. Characterizing immune responses in whole slide images of cancer with digital pathology and pathomics. *Curr. Pathobiol. Rep.* **2020**, *8*, 133–148.
- Roostee, S.; Ehinger, D.; Jönsson, M.; Phung, B.; Jönsson, G.; Sjö Dahl, G.; Staaf, J.; Aine, M. Tumour immune characterisation of primary triple-negative breast cancer using automated image quantification of immunohistochemistry-stained immune cells. *Sci. Rep.* **2024**, *14*, 21417. [[CrossRef](#)]

11. Liu, Y.; Han, D.; Parwani, A.V.; Li, Z. Applications of artificial intelligence in breast pathology. *Arch. Pathol. Lab. Med.* **2023**, *147*, 1003–1013.
12. Shakhawat, H.; Hossain, S.; Kabir, A.; Mahmud, S.H.; Islam, M.M.; Tariq, F. Review of artifact detection methods for automated analysis and diagnosis in digital pathology. In *Artificial Intelligence for Disease Diagnosis and Prognosis in Smart Healthcare*; CRC Press: Boca Raton, FL, USA, 2023; pp. 177–202.
13. Li, X.; Li, C.; Rahaman, M.M.; Sun, H.; Li, X.; Wu, J.; Yao, Y.; Grzegorzec, M. A comprehensive review of computer-aided whole-slide image analysis: From datasets to feature extraction, segmentation, classification and detection approaches. *Artif. Intell. Rev.* **2022**, *55*, 4809–4878. [[CrossRef](#)]
14. Wu, Y.; Cheng, M.; Huang, S.; Pei, Z.; Zuo, Y.; Liu, J.; Yang, K.; Zhu, Q.; Zhang, J.; Hong, H.; et al. Recent advances of deep learning for computational histopathology: Principles and applications. *Cancers* **2022**, *14*, 1199. [[CrossRef](#)] [[PubMed](#)]
15. Carriero, A.; Groenhoff, L.; Vologina, E.; Basile, P.; Albera, M. Deep Learning in Breast Cancer Imaging: State of the Art and Recent Advancements in Early 2024. *Diagnostics* **2024**, *14*, 848. [[CrossRef](#)] [[PubMed](#)]
16. Duggento, A.; Conti, A.; Mauriello, A.; Guerrisi, M.; Toschi, N. Deep computational pathology in breast cancer. *Semin. Cancer Biol.* **2021**, *72*, 226–237. [[CrossRef](#)]
17. Luo, L.; Wang, X.; Lin, Y.; Ma, X.; Tan, A.; Chan, R.; Vardhanabhuti, V.; Chu, W.C.W.; Cheng, K.T.; Chen, H. Deep learning in breast cancer imaging: A decade of progress and future directions. *IEEE Rev. Biomed. Eng.* **2024**, *18*, 130–151. [[CrossRef](#)]
18. Al-Thelaya, K.; Gilal, N.U.; Alzubaidi, M.; Majeed, F.; Agus, M.; Schneider, J.; Househ, M. Applications of discriminative and deep learning feature extraction methods for whole slide image analysis: A survey. *J. Pathol. Inform.* **2023**, *14*, 100335. [[CrossRef](#)]
19. Hulsens, T. Explainable Artificial Intelligence (XAI): Concepts and Challenges in Healthcare. *AI* **2023**, *4*, 652–666. [[CrossRef](#)]
20. Mahmood, T.; Li, J.; Pei, Y.; Akhtar, F.; Imran, A.; Rehman, K.U. A brief survey on breast cancer diagnostic with deep learning schemes using multi-image modalities. *IEEE Access* **2020**, *8*, 165779–165809. [[CrossRef](#)]
21. Mridha, M.F.; Hamid, M.A.; Monowar, M.M.; Keya, A.J.; Ohi, A.Q.; Islam, M.R.; Kim, J.M. A comprehensive survey on deep-learning-based breast cancer diagnosis. *Cancers* **2021**, *13*, 6116. [[CrossRef](#)]
22. Wen, Z.; Wang, S.; Yang, D.M.; Xie, Y.; Chen, M.; Bishop, J.; Xiao, G. Deep learning in digital pathology for personalized treatment plans of cancer patients. *Semin. Diagn. Pathol.* **2023**, *40*, 109–119. [[CrossRef](#)]
23. Rabilloud, N.; Allaupe, P.; Acosta, O.; De Crevoisier, R.; Bourgade, R.; Loussouarn, D.; Rioux-Leclercq, N.; Khene, Z.E.; Mathieu, R.; Bensalah, K.; et al. Deep learning methodologies applied to digital pathology in prostate cancer: A systematic review. *Diagnostics* **2023**, *13*, 2676. [[CrossRef](#)] [[PubMed](#)]
24. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* **2021**, *372*, n71. [[CrossRef](#)] [[PubMed](#)]
25. Knobloch, K.; Yoon, U.; Vogt, P.M. Preferred reporting items for systematic reviews and meta-analyses (PRISMA) statement and publication bias. *J. Cranio-Maxillofac. Surg.* **2011**, *39*, 91–92. [[CrossRef](#)] [[PubMed](#)]
26. Tricco, A.C.; Lillie, E.; Zarin, W.; O'Brien, K.K.; Colquhoun, H.; Levac, D.; Moher, D.; Peters, M.D.; Horsley, T.; Weeks, L.; et al. PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. *Ann. Intern. Med.* **2018**, *169*, 467–473. [[CrossRef](#)]
27. Swiderska-Chadaj, Z.; Gallego, J.; Gonzalez-Lopez, L.; Bueno, G. Detection of Ki67 hot-spots of invasive breast cancer based on convolutional neural networks applied to mutual information of H&E and Ki67 whole slide images. *Appl. Sci.* **2020**, *10*, 7761. [[CrossRef](#)]
28. George, K.; Faziludeen, S.; Sankaran, P.; Joseph, P.K. Breast cancer detection from biopsy images using nucleus guided transfer learning and belief based fusion. *Comput. Biol. Med.* **2020**, *124*, 103954. [[CrossRef](#)]
29. Krithiga, R.; Geetha, P. Deep learning based breast cancer detection and classification using fuzzy merging techniques. *Mach. Vis. Appl.* **2020**, *31*, 63. [[CrossRef](#)]
30. Evangeline, I.K.; Precious, J.G.; Pazhanivel, N.; Kirubha, S.P.A. Automatic Detection and Counting of Lymphocytes from Immunohistochemistry Cancer Images Using Deep Learning. *J. Med Biol. Eng.* **2020**, *40*, 735–747. [[CrossRef](#)]
31. Geread, R.S.; Sivanandarajah, A.; Brouwer, E.R.; Wood, G.A.; Androustos, D.; Faragalla, H.; Khademi, A. piNET: An automated proliferation index calculator framework for Ki67 breast cancer images. *Cancers* **2021**, *13*, 11. [[CrossRef](#)]
32. Krijgsman, D.; van Leeuwen, M.B.; van der Ven, J.; Almeida, V.; Vlutters, R.; Halter, D.; Kuppen, P.J.K.; van de Velde, C.J.H.; Wimberger-Friedl, R. Quantitative whole slide assessment of tumor-infiltrating CD8-positive lymphocytes in ER-positive breast cancer in relation to clinical outcome. *IEEE J. Biomed. Health Inform.* **2021**, *25*, 381–392. [[CrossRef](#)]
33. Gamble, P.; Jaroensri, R.; Wang, H.; Tan, F.; Moran, M.; Brown, T.; Flament-Auvigne, I.; Rakha, E.A.; Toss, M.; Dabbs, D.J.; et al. Determining breast cancer biomarker status and associated morphological features using deep learning. *Commun. Med.* **2021**, *1*, 14. [[CrossRef](#)] [[PubMed](#)]
34. Lu, X.; You, Z.; Sun, M.; Wu, J.; Zhang, Z. Breast cancer mitotic cell detection using cascade convolutional neural network with U-Net. *Math. Biosci. Eng.* **2021**, *18*, 673–695. [[CrossRef](#)] [[PubMed](#)]

35. Budginaite, E.; Morkunas, M.; Laurinavicius, A.; Treigys, P. Deep Learning Model for Cell Nuclei Segmentation and Lymphocyte Identification in Whole Slide Histology Images. *Informatica* **2021**, *32*, 23–40. [[CrossRef](#)]
36. Narayanan, P.L.; Raza, S.E.A.; Hall, A.H.; Marks, J.R.; King, L.; West, R.B.; Hernandez, L.; Guppy, N.; Dowsett, M.; Gusterson, B.; et al. Unmasking the immune microecology of ductal carcinoma in situ with deep learning. *NPJ Breast Cancer* **2021**, *7*, 19. [[CrossRef](#)]
37. Diao, J.A.; Wang, J.K.; Chui, W.F.; Mountain, V.; Gullapally, S.C.; Srinivasan, R.; Mitchell, R.N.; Glass, B.; Hoffman, S.; Rao, S.K.; et al. Human-interpretable image features derived from densely mapped cancer pathology slides predict diverse molecular phenotypes. *Nat. Commun.* **2021**, *12*, 1613. [[CrossRef](#)]
38. Negahbani, F.; Sabzi, R.; Jahromi, B.P.; Firouzabadi, D.; Movahedi, F.; Shirazi, M.K.; Majidi, S.; Dehghanian, A. PathoNet introduced as a deep neural network backend for evaluation of Ki-67 and tumor-infiltrating lymphocytes in breast cancer. *Sci. Rep.* **2021**, *11*, 8489. [[CrossRef](#)]
39. Diaz Guerrero, R.E.; Oliveira, J.L. Improvements in lymphocytes detection using deep learning with a preprocessing stage. In Proceedings of the 2021 IEEE 34th International Symposium on Computer-Based Medical Systems (CBMS), Aveiro, Portugal, 7–9 June 2021; pp. 178–182. [[CrossRef](#)]
40. Schirris, Y.; Engelaer, M.; Panteli, A.; Horlings, H.M.; Gavves, E.; Teuwen, J. WeakSTIL: Weak whole-slide image level stromal tumor infiltrating lymphocyte scores are all you need. *Proc. SPIE Med. Imaging: Digit. Comput. Pathol.* **2022**, *12039*, 120390B. [[CrossRef](#)]
41. Ektefaie, Y.; Yuan, W.; Dillon, D.A.; Lin, N.U.; Golden, J.A.; Kohane, I.S.; Yu, K.H. Integrative multiomics-histopathology analysis for breast cancer classification. *NPJ Breast Cancer* **2021**, *7*, 147. [[CrossRef](#)]
42. Zafar, M.M.; Rauf, Z.; Sohail, A.; Khan, A.R.; Obaidullah, M.; Khan, S.H.; Lee, Y.S.; Khan, A. Detection of tumour infiltrating lymphocytes in CD3 and CD8 stained histopathological images using a two-phase deep CNN. *Photodiagnosis Photodyn. Ther.* **2022**, *37*, 102676. [[CrossRef](#)]
43. Zhang, X.; Zhu, X.; Tang, K.; Zhao, Y.; Lu, Z.; Feng, Q. DDTNet: A dense dual-task network for tumor-infiltrating lymphocyte detection and segmentation in histopathological images of breast cancer. *Med. Image Anal.* **2022**, *78*, 102415. [[CrossRef](#)]
44. Chen, Y.; Li, H.; Janowczyk, A.; Toro, P.; Corredor, G.; Whitney, J.; Lu, C.; Koyuncu, C.F.; Mokhtari, M.; Buzzy, C.; et al. Computational pathology improves risk stratification of a multi-gene assay for early stage ER+ breast cancer. *NPJ Breast Cancer* **2023**, *9*, 40. [[CrossRef](#)] [[PubMed](#)]
45. Ali, M.L.; Rauf, Z.; Khan, A.; Sohail, A.; Ullah, R.; Gwak, J. CB-HVT Net: A Channel-Boosted Hybrid Vision Transformer Network for Lymphocyte Detection in Histopathological Images. *IEEE Access* **2023**, *11*, 115740–115750. [[CrossRef](#)]
46. Huang, J.; Li, H.; Wan, X.; Li, G. Affine-Consistent Transformer for Multi-Class Cell Nuclei Detection. In Proceedings of the 2023 International Conference on Computer Vision (ICCV 2023), Paris, France, 2–3 October 2023; pp. 21327–21336. [[CrossRef](#)]
47. Jiao, Y.; van der Laak, J.; Albarqouni, S.; Li, Z.; Tan, T.; Bhalerao, A.; Cheng, S.; Ma, J.; Pocock, J.; Pluim, J.P.W.; et al. LYSTO: The Lymphocyte Assessment Hackathon and Benchmark Dataset. *IEEE J. Biomed. Health Inform.* **2024**, *28*, 1161–1172. [[CrossRef](#)] [[PubMed](#)]
48. Shah, H.A.; Kang, J.M. An Optimized Multi-Organ Cancer Cells Segmentation for Histopathological Images Based on CBAM-Residual U-Net. *IEEE Access* **2023**, *11*, 111608–111621. [[CrossRef](#)]
49. Yosofvand, M.; Khan, S.Y.; Dhakal, R.; Nejat, A.; Moustaid-Moussa, N.; Rahman, R.L.; Moussa, H. Automated Detection and Scoring of Tumor-Infiltrating Lymphocytes in Breast Cancer Histopathology Slides. *Cancers* **2023**, *15*, 3635. [[CrossRef](#)]
50. Huang, P.W.; Ouyang, H.; Hsu, B.Y.; Chang, Y.R.; Lin, Y.C.; Chen, Y.A.; Hsieh, Y.H.; Fu, C.C.; Li, C.F.; Lin, C.H.; et al. Deep-learning based breast cancer detection for cross-staining histopathology images. *Heliyon* **2023**, *9*, e13171. [[CrossRef](#)]
51. Rauf, Z.; Sohail, A.; Khan, S.H.; Khan, A.; Gwak, J.; Maqbool, M. Attention-guided multi-scale deep object detection framework for lymphocyte analysis in IHC histological images. *Microscopy* **2023**, *72*, 27–42. [[CrossRef](#)]
52. Wang, X.; Zhang, J.; Yang, S.; Xiang, J.; Luo, F.; Wang, M.; Zhang, J.; Yang, W.; Huang, J.; Han, X. A generalizable and robust deep learning algorithm for mitosis detection in multicenter breast histopathological images. *Med. Image Anal.* **2023**, *84*, 102703. [[CrossRef](#)]
53. Ryu, J.; Puche, A.V.; Shin, J.; Park, S.; Brattoli, B.; Lee, J.; Jung, W.; Cho, S.I.; Paeng, K.; Ock, C.Y.; et al. OCELOT: Overlapped Cell on Tissue Dataset for Histopathology. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Paris, France, 2–3 October 2023; pp. 23902–23912. [[CrossRef](#)]
54. Rauf, Z.; Khan, A.R.; Sohail, A.; Alquhayz, H.; Gwak, J.; Khan, A. Lymphocyte detection for cancer analysis using a novel fusion block based channel boosted CNN. *Sci. Rep.* **2023**, *13*, 14047. [[CrossRef](#)]
55. Makhlof, S.; Wahab, N.; Toss, M.; Ibrahim, A.; Lashen, A.G.; Atallah, N.M.; Ghannam, S.; Jahanifar, M.; Lu, W.; Graham, S.; et al. Evaluation of Tumour Infiltrating Lymphocytes in Luminal Breast Cancer Using Artificial Intelligence. *Br. J. Cancer* **2023**, *129*, 1747–1758. [[CrossRef](#)]

56. Aswolinskiy, W.; Munari, E.; Horlings, H.M.; Mulder, L.; Bogina, G.; Sanders, J.; Liu, Y.H.; van den Belt-dusebout, A.W.; Tessier, L.; Balkenhol, M.; et al. PROACTING: Predicting Pathological Complete Response to Neoadjuvant Chemotherapy in Breast Cancer from Routine Diagnostic Histopathology Biopsies with Deep Learning. *Breast Cancer Res.* **2023**, *25*, 142. [CrossRef] [PubMed]
57. Genc-Nayebi, N.; Abran, A. A systematic literature review: Opinion mining studies from mobile app store user reviews. *J. Syst. Softw.* **2017**, *125*, 207–219. [CrossRef]
58. Kabir, S.; Vranic, S.; Al Saady, R.M.; Khan, M.S.; Sarmun, R.; Alqahtani, A.; Abbas, T.O.; Chowdhury, M.E.H. The utility of a deep learning-based approach in Her-2/neu assessment in breast cancer. *Expert Syst. Appl.* **2024**, *238*, 122051. [CrossRef]
59. Li, Z.; Li, W.; Mai, H.; Zhang, T.; Xiong, Z. Enhancing Cell Detection in Histopathology Images: A ViT-Based U-Net Approach. In *Graphs in Biomedical Image Analysis, and Overlapped Cell on Tissue Dataset for Histopathology, 5th MICCAI Workshop, Proceedings of the 5th International Workshop on Graphs in Biomedical Image Analysis (GRAIL) / Workshop on Overlapped Cell on Tissue—Cell Detection from Cell-Tissue Interaction Challenge (OCELOT), MICCAI 2023, Vancouver, BC, Canada, 23 September–4 October 2023*; Ahmadi, S.A., Pereira, S., Eds.; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2024; Volume 14373, pp. 150–160. [CrossRef]
60. Millward, J.; He, Z.; Nibali, A. Dense Prediction of Cell Centroids Using Tissue Context and Cell Refinement. In *Graphs in Biomedical Image Analysis, and Overlapped Cell on Tissue Dataset for Histopathology, 5th MICCAI Workshop, Proceedings of the 5th International Workshop on Graphs in Biomedical Image Analysis (GRAIL) and the Workshop on Overlapped Cell on Tissue (OCELOT), MICCAI 2023, Vancouver, BC, Canada, 23 September–4 October 2023*; Ahmadi, S.A., Pereira, S., Eds.; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2024; Volume 14373, pp. 138–149. [CrossRef]
61. Karol, M.; Tabakov, M.; Markowska-Kaczmar, U.; Fulawka, L. Deep Learning for Cancer Cell Detection: Do We Need Dedicated Models? *Artif. Intell. Rev.* **2024**, *57*, 53. [CrossRef]
62. Lakshmanan, B.; Anand, S.; Raja, P.S.V.; Selvakumar, B. Improved DeepMitosisNet Framework for Detection of Mitosis in Histopathology Images. *Multimed. Tools Appl.* **2023**, *83*, 43303–43324. [CrossRef]
63. Hoerst, F.; Rempe, M.; Heine, L.; Seibold, C.; Keyl, J.; Baldini, G.; Ugurel, S.; Siveke, J.; Gruenwald, B.; Egger, J.; et al. CellViT: Vision Transformers for Precise Cell Segmentation and Classification. *Med. Image Anal.* **2024**, *94*, 103143. [CrossRef]
64. Marzouki, A.; Guo, Z.; Zeng, Q.; Kurtz, C.; Loménie, N. Optimizing Lymphocyte Detection in Breast Cancer Whole Slide Imaging through Data-Centric Strategies. *arXiv* **2024**, arXiv:2405.13710.
65. Liu, S.; Amgad, M.; More, D.; Rathore, M.A.; Salgado, R.; Cooper, L.A.D. A panoptic segmentation dataset and deep-learning approach for explainable scoring of tumor-infiltrating lymphocytes. *NPJ Breast Cancer* **2024**, *10*, 52.
66. Keele, S. *Guidelines for Performing Systematic Literature Reviews in Software Engineering*; Technical Report, Version 2.3, EBSE-2007-01; Software Engineering Group, School of Computer Science and Mathematics; Keele University: Keele, UK; Department of Computer Science, University of Durham: Durham, UK, 2007.
67. Tafavvoghi, M.; Bongo, L.A.; Shvetsov, N.; Busund, L.T.R.; Møllersen, K. Publicly available datasets of breast histopathology H&E whole-slide images: A scoping review. *J. Pathol. Inform.* **2024**, *15*, 100363.
68. Abdel-Nabi, H.; Ali, M.; Awajan, A.; Daoud, M.; Alazrai, R.; Suganthan, P.N.; Ali, T. A Comprehensive Review of the Deep Learning-Based Tumor Analysis Approaches in Histopathological Images: Segmentation, Classification and Multi-Learning Tasks. *Clust. Comput.* **2023**, *26*, 3145–3185. [CrossRef]
69. Basu, A.; Senapati, P.; Deb, M.; Rai, R.; Dhal, K.G. A survey on recent trends in deep learning for nucleus segmentation from histopathology images. *Evol. Syst.* **2024**, *15*, 203–248.
70. Wang, L.; Pan, L.; Wang, H.; Liu, M.; Feng, Z.; Rong, P.; Chen, Z.; Peng, S. DHU-net: Dual-branch hierarchical global–local fusion network for whole slide image segmentation. *Biomed. Signal Process. Control* **2023**, *85*, 104976.
71. Liu, S.; Amgad, M.; Rathore, M.A.; Salgado, R.; Cooper, L.A.D. A panoptic segmentation approach for tumor-infiltrating lymphocyte assessment: Development of the MuTILs model and PanopTILs dataset. *MedRxiv* **2022**. [CrossRef]
72. Mondello, A.; Dal Bo, M.; Toffoli, G.; Polano, M. Machine learning in onco-pharmacogenomics: A path to precision medicine with many challenges. *Front. Pharmacol.* **2024**, *14*, 1260276.
73. Rydzewski, N.R.; Shi, Y.; Li, C.; Chrostek, M.R.; Bakhtiar, H.; Helzer, K.T.; Bootsma, M.L.; Berg, T.J.; Harari, P.M.; Floberg, J.M.; et al. A platform-independent AI tumor lineage and site (ATLAS) classifier. *Commun. Biol.* **2024**, *7*, 314.
74. Liu, Y.; He, X.; Yang, Y. Tumor immune microenvironment-based clusters in predicting prognosis and guiding immunotherapy in breast cancer. *J. Biosci.* **2024**, *49*, 19.
75. Ahmed, F.; Abdel-Salam, R.; Hamnett, L.; Adewunmi, M.; Ayano, T. Improved Breast Cancer Diagnosis through Transfer Learning on Hematoxylin and Eosin Stained Histology Images. *arXiv* **2023**, arXiv:2309.08745.
76. Nasir, E.S.; Parvaiz, A.; Fraz, M.M. Nuclei and glands instance segmentation in histology images: A narrative review. *Artif. Intell. Rev.* **2023**, *56*, 7909–7964.
77. National Cancer Institute. TCGA-BRCA Project. Available online: <https://portal.gdc.cancer.gov/projects/TCGA-BRCA> (accessed on 18 July 2024).

78. Veta, M.; Van Diest, P.J.; Willems, S.M.; Wang, H.; Madabhushi, A.; Cruz-Roa, A.; Gonzalez, F.; Larsen, A.B.L.; Vestergaard, J.S.; Dahl, A.B.; et al. Assessment of algorithms for mitosis detection in breast cancer histopathology images. *Med. Image Anal.* **2015**, *20*, 237–248.
79. Spanhol, F.A.; Oliveira, L.S.; Petitjean, C.; Heutte, L. A dataset for breast cancer histopathological image classification. *IEEE Trans. Biomed. Eng.* **2015**, *63*, 1455–1462. [[CrossRef](#)]
80. Qaiser, T.; Mukherjee, A.; Reddy Pb, C.; Munugoti, S.D.; Tallam, V.; Pitkäaho, T.; Lehtimäki, T.; Naughton, T.; Berseth, M.; Pedraza, A.; et al. Her 2 challenge contest: A detailed assessment of automated HER 2 scoring algorithms in whole slide images of breast cancer tissues. *Histopathology* **2018**, *72*, 227–238. [[CrossRef](#)] [[PubMed](#)]
81. Aresta, G.; Araújo, T.; Kwok, S.; Chennamsetty, S.S.; Safwan, M.; Alex, V.; Marami, B.; Prastawa, M.; Chan, M.; Donovan, M.; et al. Bach: Grand challenge on breast cancer histology images. *Med. Image Anal.* **2019**, *56*, 122–139. [[CrossRef](#)] [[PubMed](#)]
82. Kumar, N.; Verma, R.; Anand, D.; Zhou, Y.; Onder, O.F.; Tsougenis, E.; Chen, H.; Heng, P.A.; Li, J.; Hu, Z.; et al. A Multi-Organ Nucleus Segmentation Challenge. *IEEE Trans. Med. Imaging* **2020**, *39*, 1380–1391. [[CrossRef](#)] [[PubMed](#)]
83. Ortega-Ruiz, M.A.; Roman-Rangel, E.; Reyes-Aldasoro, C.C. Multiclass Semantic Segmentation of Immunostained Breast Cancer Tissue with a Deep-Learning Approach. *medRxiv* **2022**. [[CrossRef](#)]
84. Gamper, J.; Koohbanani, N.A.; Benes, K.; Graham, S.; Jahanifar, M.; Khurram, S.A.; Azam, A.; Hewitt, K.; Rajpoot, N. Pannuke dataset extension, insights and baselines. *arXiv* **2020**, arXiv:2003.10778.
85. Koohbanani, N.A.; Jahanifar, M.; Tajadin, N.Z.; Rajpoot, N. NuClick: A deep learning framework for interactive segmentation of microscopic images. *Med. Image Anal.* **2020**, *65*, 101771. [[CrossRef](#)]
86. Jeretic, P.; Warstadt, A.; Bhooshan, S.; Williams, A. Are natural language inference models IMPPPRESSive? Learning IMPLICature and PRESupposition. *arXiv* **2020**, arXiv:2004.03066.
87. Brancati, N.; Anniciello, A.M.; Pati, P.; Riccio, D.; Scognamiglio, G.; Jaume, G.; De Pietro, G.; Di Bonito, M.; Foncubierta, A.; Botti, G.; et al. BRACS: A Dataset for BReAst Carcinoma Subtyping in H&E Histology Images. *arXiv* **2021**, arXiv:2111.04740.
88. Grand Challenge. Tiger Challenge. 2022. Available online: <https://tiger.grand-challenge.org/> (accessed on 18 July 2024).
89. Aubreville, M.; Stathonikos, N.; Donovan, T.A.; Klopffleisch, R.; Ammeling, J.; Ganz, J.; Wilm, F.; Veta, M.; Jabari, S.; Eckstein, M.; et al. Domain generalization across tumor types, laboratories, and species—Insights from the 2022 edition of the Mitosis Domain Generalization Challenge. *Med. Image Anal.* **2024**, *94*, 103155. [[CrossRef](#)]
90. Madoori, P.K.; Sannidhi, S.; Anandam, G. Artificial Intelligence and Computational Pathology: A comprehensive review of advancements and applications. *Perspectives* **2023**, *11*, 30. [[CrossRef](#)]
91. Kumar, N.; Gupta, R.; Gupta, S. Whole slide imaging (WSI) in pathology: Current perspectives and future directions. *J. Digit. Imaging* **2020**, *33*, 1034–1040. [[CrossRef](#)] [[PubMed](#)]
92. Wagner, S.J.; Matek, C.; Boushehri, S.S.; Boxberg, M.; Lamm, L.; Sadafi, A.; Winter, D.J.E.; Marr, C.; Peng, T. Built to last? Reproducibility and reusability of deep learning algorithms in computational pathology. *Mod. Pathol.* **2024**, *37*, 100350. [[CrossRef](#)] [[PubMed](#)]
93. Nasarian, E.; Alizadehsani, R.; Acharya, U.R.; Tsui, K.L. Designing interpretable ML system to enhance trust in healthcare: A systematic review to proposed responsible clinician-AI-collaboration framework. *Inf. Fusion* **2024**, *108*, 102412. [[CrossRef](#)]
94. Hossain, M.S.; Shahriar, G.M.; Syeed, M.M.M.; Uddin, M.F.; Hasan, M.; Hossain, M.S.; Bari, R. Tissue artifact segmentation and severity assessment for automatic analysis using WSI. *IEEE Access* **2023**, *11*, 21977–21991. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.