Image and Vision Computing 162 (2025) 105708

Contents lists available at ScienceDirect

Image and Vision Computing

journal homepage: www.elsevier.com/locate/imavis

A review of breast cancer histopathology image analysis with deep learning: Challenges, innovations, and clinical integration

Inayatul Haq [a,b,c], Zheng Gong [a,c], Haomin Liang [c], Wei Zhang [c], Rashid Khan [c], Lei Gu [d], Roland Eils [e], Yan Kang [f], Bingding Huang [a,c,*]

[a] College of Applied Sciences, Shenzhen University, Shenzhen 518060, China
[b] Guangdong Key Laboratory for Biomedical Measurements and Ultrasound Imaging, National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, School of Biomedical Engineering, Shenzhen University Medical School, Shenzhen 518060, China
[c] College of Big Data and Internet, Shenzhen Technology University, Shenzhen 518118, China
[d] Epigenetics Laboratory, Max Planck Institute for Heart and Lung Research, Bad Nauheim 61231, Germany
[e] Center for Digital Health, Berlin Institute of Health at Charité -Universitätsmedizin Berlin, Berlin 10178, Germany
[f] College of Health Science and Environmental Engineering, Shenzhen Technology University, Shenzhen 518118, China

ARTICLE INFO

Keywords:
Histopathology
Breast cancer
Deep learning
Detection
Diagnosis
Image analysis

ABSTRACT

Breast cancer (BC) is the most frequently diagnosed cancer among women and a leading cause of cancer-related mortality globally. Accurate and timely diagnosis is essential for improving patient outcomes. However, traditional histopathological assessments are labor-intensive and subjective, leading to inter-observer variability and diagnostic inconsistencies, especially in resource-limited settings. Furthermore, variability in tissue staining, limited availability of standardized annotated datasets, and subtle morphological patterns complicate the consistent characterization of tumors. Deep learning (DL) has recently emerged as a transformative technology in breast cancer pathology, providing automated and objective solutions for cancer detection, classification, and segmentation from histopathological images. This review systematically evaluates advanced deep learning (DL) architectures, including convolutional neural networks (CNNs), generative adversarial networks (GANs), autoencoders, deep belief networks (DBNs), extreme learning machines (ELMs), and transformer-based models such as Vision Transformers (ViTs) as well as transfer learning, attention-based explainable AI techniques, and multimodal integration to address these diagnostic challenges. Analyzing 199 references, including 182 peer-reviewed studies published between 2014 and 2025 and 17 reputable online sources (websites, databases, etc.), we identify key innovations, limitations, and opportunities for future research. Furthermore, we explore the critical roles of synthetic data augmentation, explainable AI (XAI), and multimodal integration to enhance clinical trust, model interpretability, and diagnostic precision, ultimately facilitating personalized and efficient patient care.

1. Introduction

1.1. Background

Breast cancer (BC) is the most commonly diagnosed cancer among women globally and remains a significant public health challenge. In 2022, approximately 2.3 million women were diagnosed with breast cancer globally, leading to 670,000 deaths [1]. In the United States (U. S.), breast cancer is expected to account for 30% of all new cancer diagnoses in women in 2025, with an estimated 316,950 new cases and 42,250 deaths [2,3]. The lifetime risk for a woman developing breast cancer in the U.S. is approximately 13% [4]. The disease burden is even more pronounced in middle- and low-income countries, where higher mortality rates result from limited healthcare infrastructure, lack of awareness, and late-stage diagnoses [5]. For instance, India accounts for 13.5% (178,361) of all cancer cases and 10.6% (90,408) of all cancer-related deaths, with similar patterns observed in Egypt and sub-Saharan Africa, where inadequate screening and poor access to specialized healthcare further worsening patient outcomes [6]. Fig. 1 illustrates regional disparities in breast cancer incidence worldwide based on 2022 data, comparing the age-standardized rate (ASR), crude rate per 100,000 population, and total number of cases across six global

---

* Corresponding author.
E-mail address: huangbingding@sztu.edu.cn (B. Huang).

https://doi.org/10.1016/j.imavis.2025.105708
Received 30 December 2024; Received in revised form 6 July 2025; Accepted 19 August 2025

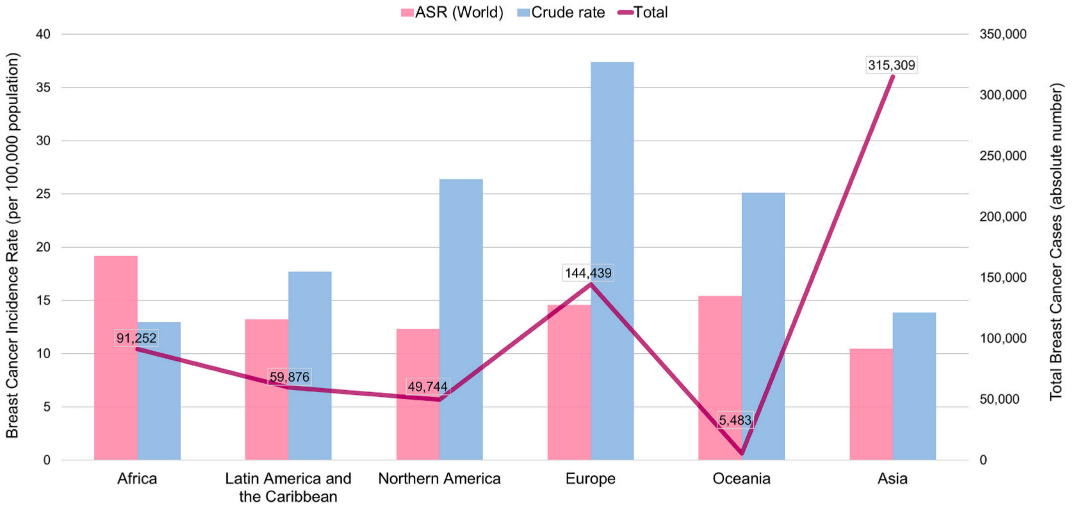

Fig. 1. Breast cancer incidence across global regions in 2022, showing age-standardized rates (ASR), crude rates per 100,000 population, and total case counts. Bar plots display ASR and crude rates, while the magenta line represents the total number of cases. Europe and Northern America have the highest crude rates, whereas Asia reports the highest case numbers due to its large population [7].

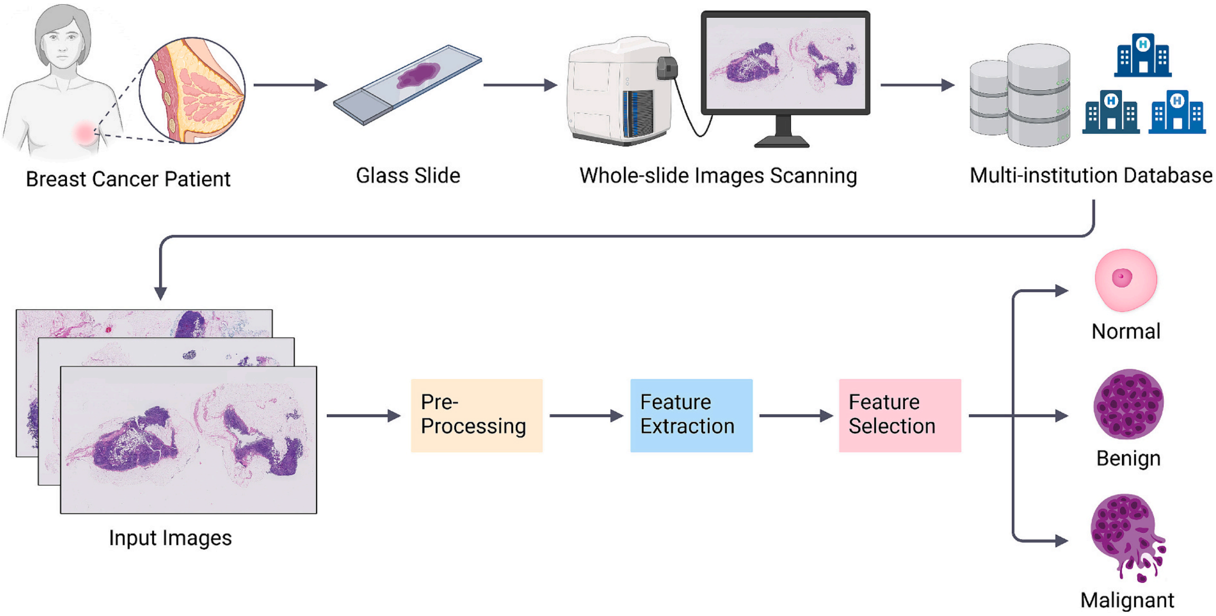

Fig. 2. Workflow of a computer-aided diagnosis (CAD) system for breast cancer classification. Whole-slide images (WSIs) undergo preprocessing, feature extraction, and selection, ultimately leading to classification as normal, benign, or malignant [13]. WSI patches shown in this figure are retrieved from [14].

regions. Although Europe and Northern America exhibit higher incidence rates, Asia reports the largest number of absolute cases, highlighting the impact of population size on the cancer burden. This visualization underscores the geographic variation in disease prevalence and the need for region-specific screening and diagnostic strategies [7].

Given the increasing incidence and mortality rates, early and accurate detection of breast cancer (BC) is crucial for improving patient survival. Early-stage diagnosis can improve survival rates from 30% to over 50%, underscoring the importance of robust screening and diagnostic strategies [8]. Currently, histopathology, mammography, ultrasound, and MRI remain the most reliable imaging modalities for breast cancer diagnosis [9]. However, while indispensable, traditional histopathological diagnosis methods have inherent limitations that affect clinical outcomes. Manual assessments by pathologists are subjective, time-consuming, and often result in considerable inter-observer variability. Diagnostic accuracy heavily depends on individual expertise, which can create discrepancies, particularly in resource-limited settings where expert pathologists are scarce. Furthermore, variability in staining protocols, tissue preparation techniques, and imaging devices further exacerbates diagnostic inconsistencies across institutions. These limitations underscore the necessity of advanced solutions, such as deep learning (DL), which automates image analysis tasks, reduces subjectivity and inter-observer variability, and enhances diagnostic accuracy, consistency, and efficiency in breast cancer pathology [10].

The evolution of breast cancer detection methodologies has progressed from manual examinations to advanced technologies, including digital mammography (DM), breast MRI (BMRI), breast ultrasound (BUS), and histopathological analysis using whole slide imaging (WSI) and AI-driven models [11]. Over the past decade, computer-aided diagnosis (CAD) systems have been integrated into clinical workflows, assisting radiologists and pathologists by improving detection accuracy, sensitivity, and specificity [12]. Fig. 2 presents the end-to-end workflow

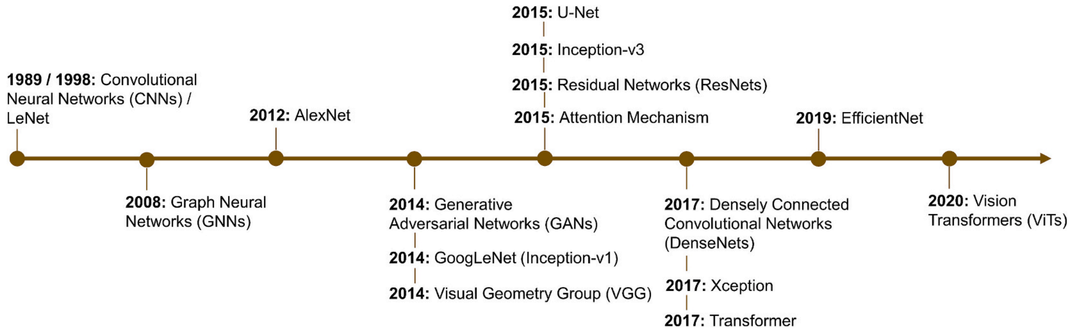

Fig. 3. Timeline of key Deep learning advancements, from CNNs (1989) to ViTs (2020), highlighting significant breakthroughs in Neural Network architectures [19].

of a CAD system used for classifying breast cancer histopathology images. The pipeline starts with digitized whole-slide images, which undergo preprocessing to standardize input quality. Key features are then extracted and selected to accurately categorize tissue samples into normal, benign, or malignant classes. This structured approach supports automated, scalable, and objective diagnosis in clinical pathology.

Recent advancements in artificial intelligence (AI) and deep learning (DL) have reformed breast cancer diagnostics by using vast amounts of medical imaging data to enhance detection accuracy. Convolutional Neural Networks (CNNs), Generative Adversarial Networks (GANs), and Vision Transformers (ViTs) have demonstrated superior performance in automated cancer classification, segmentation, and detection, often exceeding human-level accuracy in histopathology image analysis [15]. In particular, CNN-based models such as ResNet, VGG16, and U-Net have been widely adopted for detecting, classifying, and segmenting breast tumors in histopathology images [16,17]. GANs have also gained traction for synthetic data generation, data augmentation, and domain adaptation, addressing the challenges posed by small datasets in medical imaging [18]. Fig. 3 illustrates a timeline of key advancements in deep learning architectures, emphasizing significant developments from CNNs (1989) to Transformer-based models (2017) and Vision Transformers (ViTs) (2020). These milestones reflect the evolution of neural network architectures for computer vision and related tasks [19].

Deep learning (DL)-based breast cancer (BC) segmentation approaches have also evolved significantly, progressing from manual segmentation to semi-automated and now fully automated approaches that incorporate artificial intelligence (AI) [20]. In the early 2000s, segmentation methods such as threshold-based techniques, region-growing approaches, and clustering algorithms were commonly employed. Over time, these methods evolved into more sophisticated architectures, such as U-Net, Mask R-CNN, and PSP-Net, demonstrating high accuracy in tumor segmentation for breast cancer (BC) pathology [21]. These DL models deliver consistent and reproducible results, enhancing diagnostic efficiency and reducing the workload for pathologists. Despite advances, deep learning still faces challenges in breast cancer histopathology, including data variability, limited annotated datasets, and poor interpretability, which hinder its clinical adoption [22]. Explainable AI (XAI), multimodal learning, and hybrid deep learning (DL) architectures have been explored to address these limitations. Researchers integrate XAI techniques, such as Grad-CAM, SHAP, and attention mechanisms, to enhance interpretability in deep learning (DL) models. Furthermore, multimodal approaches combining histopathology images, radiology scans, and genomic data aim to enhance diagnostic accuracy and enable personalized treatment strategies [23].

1.2. Scope of this study

This review critically examines the advancements in deep learning (DL) for breast cancer (BC) histopathology, analyzing 182 peer-reviewed studies published between 2014 and 2024. The paper explores:

- CNN-based, GAN-based, and Transformer-based models for breast cancer detection.
- Transfer learning and synthetic data augmentation to improve model generalization.
- Explainable AI and multimodal integration for enhanced clinical interpretability.
- Challenges in AI-driven breast cancer histopathology and potential solutions for clinical implementation.

This review addresses current gaps in breast cancer (BC) histopathology image analysis. It aims to provide a comprehensive synthesis of research findings while proposing future directions for integrating deep learning (DL) in breast cancer (BC) diagnostics.

1.3. Motivation of the study

Breast cancer remains the leading cause of cancer-related mortality among women worldwide, making timely and accurate pathological diagnosis vital for effective patient management. Traditional histopathological diagnostic approaches, while essential, have significant limitations, including subjective interpretation, substantial inter-observer variability, dependence on individual pathologist expertise, and considerable time demands. These issues lead to inconsistent diagnoses, delayed treatment decisions, and compromised clinical outcomes, particularly affecting healthcare quality in resource-limited settings. Deep learning (DL) technologies have emerged as powerful tools to overcome these limitations by providing automated, objective, and reproducible analysis of histopathological images. DL approaches have demonstrated substantial potential in improving diagnostic accuracy by reducing inter-observer variability, significantly accelerating diagnostic workflows, and enhancing the standardization of histopathology processes. Furthermore, DL facilitates the extraction of subtle histopathological features that are often difficult to interpret through manual observation alone, thus offering improved diagnostic precision and efficiency. The integration of DL technologies holds promise for augmenting pathologist expertise, improving diagnostic reliability, and supporting personalized patient care, making it an indispensable innovation in modern breast cancer (BC) pathology.

1.4. Contribution of the study

- This paper comprehensively reviews state-of-the-art deep learning (DL) architectures, including convolutional neural networks (CNNs), generative adversarial networks (GANs), and Transfer Learning, to explore their applications in breast cancer detection, classification, and segmentation.
- Identifying research gaps focuses on challenges like data variability, the scarcity of annotated datasets, and the limited interpretability of models in clinical applications.

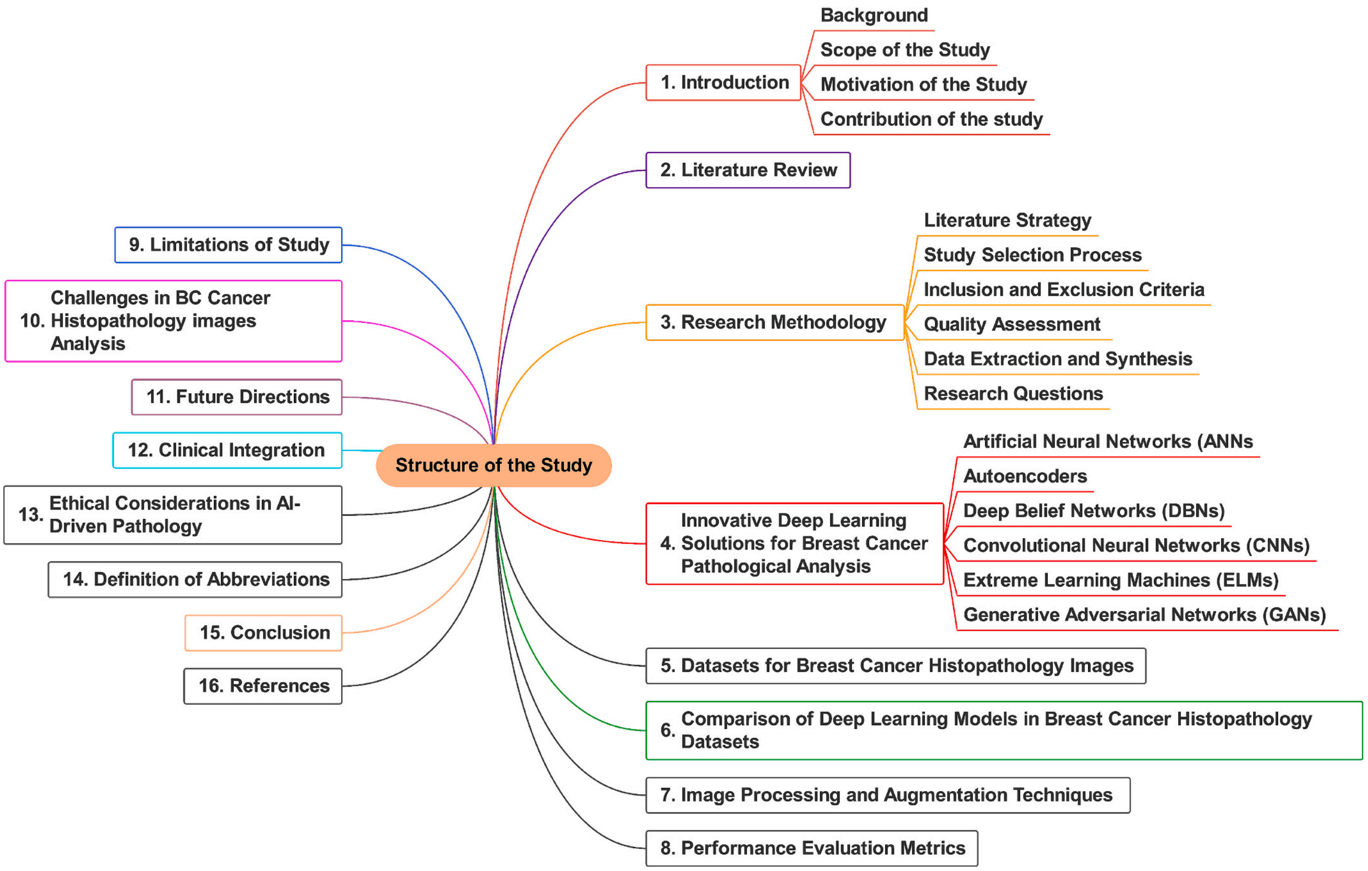

Fig. 4. Overview of the paper's structure, illustrating the main sections and their interconnections, including methodology, deep learning approaches, datasets, evaluation, and supporting topics.

- Evaluating emerging trends involves examining solutions such as explainable AI, synthetic data generation, and multimodal data integration to enhance diagnostic accuracy.
- Focusing on practical integration involves developing strategies to incorporate AI tools into clinical workflows while addressing ethical, bias, and privacy concerns.
- Future research directions propose strategies to enhance model robustness, scalability, and applicability across diverse clinical contexts.

1.5. *Structure of the study*

Fig. 4 illustrates the structured organization of this study, detailing the systematic approach to exploring deep learning (DL) in breast cancer (BC) histopathology. The framework encompasses key components, including the introduction, literature review, and research methodology, followed by an in-depth analysis of DL techniques, datasets, image processing, and evaluation metrics. Furthermore, it addresses challenges in BC image analysis, ethical considerations, clinical integration, and future research directions. This structured workflow provides a comprehensive and logical progression for understanding advancements, limitations, and potential applications of DL in BC histopathology.

2. Literature review

Deep learning (DL) has significantly advanced breast cancer (BC) histopathology, enhancing diagnostic accuracy and efficiency. Below is a categorized overview of studies employing various DL methodologies in BC histopathology:

2.1. *Convolutional neural networks (CNN-based studies)*

Deep learning (DL), particularly Convolutional Neural Networks (CNNs), has significantly advanced breast cancer (BC) classification using histology images. One approach compares handcrafted feature extraction methods (Bag of Words and Locality Constrained Linear Coding) with a CNN-based model trained on the BreakHis dataset. The CNN architecture consists of five convolutional layers and two fully connected layers, utilizing ReLU activation, max pooling, dropout, and L2 regularization for optimal training. The findings reveal that CNNs outperform traditional feature-based classifiers [24]. To further refine CNN-based classification, Discrimination of Tumor Heterogeneity Using Deep CNNs employed Google's Inception architectures to classify histopathology images from breast cancer tissues. Given the extensive tumor heterogeneity observed in breast cancer (BC), DL ensemble methods were employed to improve accuracy. The study achieved high precision in distinguishing various cancer subtypes, demonstrating the robustness of CNN-based models in handling complex histopathological variations [25].

In addition to tumor classification, CNNs have been widely adopted for the analysis of histological images. The study on the Classification of breast cancer histology images using CNNs proposed an approach for analyzing hematoxylin and eosin (H&E) stained breast biopsy images. The method achieved an accuracy of 77.8% in four-class classification and 83.3% in carcinoma/non-carcinoma classification, confirming the efficacy of CNNs in differentiating malignant and benign histopathological structures [26]. CNNs have also demonstrated their utility in breast biopsy analysis. A study [27] utilized CNNs on breast biopsies and successfully differentiated benign from malignant lesions, reinforcing the clinical utility of CNN-based DL in assisting radiologists and pathologists in decision-making [27]. Residual learning-based CNNs improve breast cancer histopathology classification by enhancing feature extraction and model generalization. The ResHist model (152-layer residual CNN), trained on BreakHis, achieved 92.52% accuracy using skip connections to alleviate vanishing gradients, proving effective in BC diagnostics [28].

2.2. Generative adversarial networks (GAN-based studies)

Generative Adversarial Networks (GANs) have reformed breast cancer (BC) histopathology by addressing key challenges such as data scarcity, synthetic image generation, and domain adaptation. These models have been widely applied in histopathology, radiology, and genomic analysis, providing more robust and generalizable deep learning (DL) frameworks for cancer detection and classification.

A key breakthrough in GAN-based applications is Accurate Reconstruction of Pan-Cancer Gene Expression Profiles Using GANs, where researchers applied a GAN-driven approach to predict actionable genomic alterations in BC, including PIK3CA mutation and homologous recombination deficiency (HRD) status. The model effectively reconstructed pan-cancer gene expression profiles, showing the potential of GANs in genomic data analysis and precision oncology [29]. Using labeled and unlabeled data, NAS-SGAN, a semi-supervised GAN framework, enhances nuclear atypia scoring in breast cancer grading. Trained on the MITOS-ATYPIA14 dataset, it achieved 98.22%–98.44% accuracy, improving feature extraction for atypical nuclei detection. This approach reduces annotation dependency and demonstrates the potential of adversarial learning in histopathology [30]. Expanding the application of adversarial learning in histopathology, the study on BC Classification Using ACGAN proposed an ACGAN-based approach for synthetic histopathological image generation. This model, combined with ResNet-50 feature extraction and XGBoost classification, significantly improved BC diagnosis by enhancing dataset diversity and balance [31].

Beyond imaging modalities, GANs have been employed in histopathological image synthesis to expand training datasets for DL models. A study in [32] synthesis of diagnostic quality histopathology images using GANs focused on training GAN architectures to generate high-quality histopathology images across multiple BC subtypes. These synthetic images were validated to be diagnostically relevant, making them highly useful for augmenting datasets and training AI-based histopathology systems [32]. The PTGAN model enhances domain adaptation in histopathology image classification by integrating GANs with prototype learning. Target-biased adversarial learning reduces style discrepancies between datasets, achieving 88.9% accuracy on BreakHis. This approach improves unsupervised classification while minimizing annotation dependency in breast cancer diagnosis [33].

2.3. Transfer learning-based studies

Transfer learning (TL) has emerged as a potent breast cancer (BC) histopathology technique. It allows models to utilize pre-trained deep learning (DL) architectures to compensate for limited annotated datasets. By applying knowledge learned from large-scale datasets, TL enhances feature extraction, improves diagnostic accuracy, and accelerates training times, making it particularly useful in breast histopathology and radiology imaging.

A study in [34] showed the applications of DL in BC Histopathological Imaging, which reviewed multiple CNN-based transfer learning models for extracting key features from whole-slide images (WSIs) and digital breast tomosynthesis (DBT). This study emphasized that transfer learning improves diagnostic accuracy significantly, particularly when datasets are limited, making it an essential tool for breast cancer histopathology analysis [34]. Beyond imaging, transfer learning (TL) enhances genomic data analysis, particularly in transcriptome-based breast cancer prognostication. GAN-based classifiers trained with TL enhance gene expression prediction, facilitating accurate cancer progression assessment and reinforcing the role of TL in omics-based diagnostics [35]. TL enhances AI-driven breast pathology by improving diagnostic accuracy and efficiency in histopathology image analysis. Fine-tuned pre-trained models outperform traditional approaches, excelling in tumor classification and segmentation [36]. Transfer learning (TL) enhances breast cancer subtype detection, particularly in identifying invasive ductal carcinoma (IDC). Pre-trained CNN models enhance accuracy in IDC classification from histopathological images, thereby strengthening the role of Transfer Learning in deep learning-based diagnostics, as reported in [37].

2.4. Explainable AI (XAI) and attention-based studies

Explainable AI (XAI) and attention-based deep learning (DL) play a critical role in enhancing the interpretability, transparency, and clinical reliability of AI-driven breast cancer (BC) diagnostics. Multi-modal approaches that integrate histopathology, genomics, and clinical data further enhance model explainability and decision-making support. XAI techniques, such as Grad-CAM, SHAP, LIME, Saliency Maps, and DeepLIFT, emphasize diagnostically relevant features, fostering clinician trust and enabling the transparent validation of model outputs [38–41]. Several CNN architectures, including VGG16, VGG19, and ResNet-50, have been successfully applied to BC histopathology image classification, with ResNet-50 achieving the highest accuracy in comparative studies. Attention-based models such as ESAE-Net and the Guided Soft Attention (GuSA) Network enhance classification performance by improving feature localization and reducing noise through adaptive filtering and region-supervised attention mechanisms [42,43]. Explainable AI frameworks that integrate deep learning with biomarker discovery enhance the classification of breast cancer subtypes by identifying relevant genomic patterns, such as DNA methylation features, thereby improving the accuracy of breast cancer subtype classification. Autoencoder-based models, combined with explainability techniques such as SHAP and DeepLIFT, improve interpretability, making these approaches more suitable for clinical decision-making [41].

2.5. Transformer and multimodal-based studies

Recent developments in transformer-based and multimodal deep learning (DL) approaches have significantly advanced cancer diagnosis and prognosis using histopathology images. The transformer-based ROAM model employs weakly supervised multiple-instance learning and multiscale self-attention mechanisms to achieve high diagnostic accuracy and enable biomarker discovery in glioma analysis [44]. In breast cancer (BC) classification, ResNet-18 with block-wise fine-tuning, combined with graph convolutional networks (GCN) and data augmentation, improves generalization across magnification levels in both binary and multiclass settings [45].

Vision Transformers (ViTs) are increasingly used in BC histopathology due to their ability to model global dependencies. They effectively select regions of interest from whole-slide images (WSIs), thereby reducing analysis time and improving diagnostic consistency, particularly in tasks such as HER2 grading [46]. OncoDHNet combines transformer-extracted histopathology features with clinicopathological data, offering improved recurrence risk prediction and a cost-effective alternative to genomic assays [47]. In data-scarce clinical environments, multimodal large language models such as GPT-4V demonstrate strong diagnostic performance through in-context learning [48]. ViT-based models enhance BC histopathology analysis by capturing local and global features using self-attention mechanisms [49,50]. They mitigate overfitting through techniques such as principal component analysis (PCA) and enhance generalization with data augmentation [49]. Lightweight designs, such as EMViT-BCC and MobileViT, enable efficient classification across magnifications, making them suitable for clinical and mobile applications [50,51]. Multi-resolution ViT frameworks aggregate features from multiple magnifications, improving

Table 1

Overview of conventional and CNN-based deep learning models used in breast cancer histopathology image analysis. It highlights model, datasets, evaluation metrics (accuracy and AUC), key strengths, and limitations. The average value of accuracy isa taken from results in [24], [25], [28], [45].

Study	Deep Learning Model	Primary Task	Dataset Used	Accuracy (%)	Key Strengths	Challenges & Limitations
[24]	CNN	BC histopathology image classification	BreakHis	91.51	Effective CNN topology, strong data augmentation, and ensemble improved accuracy.	Lower accuracy in multi-class, dataset size limits, overfitting risk
[25]	Google's Inception, Inception and ResNet	BC Classification	TCGA, BreastBiomarkers	89.90	High subtype classification accuracy, robust across datasets	High computational cost, dataset bias risk
[26]	CNN	4-class BC histopathology classification	Bioimaging 2015	80.55	Multi-scale features, augmentation, improved diagnostic reliability	Limited dataset, manual annotation variability, high computing needs
[27]	CNN, RF	Predicting breast density and cancer risk	BREAST-Stamp	92, 84	Key histological feature detection, correlation with mammographic density	Large annotated data needed, variable subgroup performance
[28]	ResHist	BC histopathology classification	BreakHis	92.52	Residual learning and augmentation improve generalization	High compute needs, overfitting without augmentation
[58]	Ensemble of Fine-Tuned VGG16 and VGG19	Binary classification (carcinoma vs. non-carcinoma)	Private dataset	95.29	High sensitivity, robust ensemble, effective on small data	Private limits reproducibility; only two classes
[37]	CNN	BC classification (IDC vs. non-IDC)	Private dataset	78.4	Improves early detection, reduces observer variability	Limited dataset, potential bias
[45]	ResNet-18	Magnification-dependent, independent BC classification	BreakHis	95.23	Block-wise tuning, GCN, and augmentation improve generalization	Lower accuracy at high magnification, 8-class lower performance

Table 2

Summary of advanced and emerging deep learning models applied to breast cancer histopathology image analysis, including GAN-based, transformer-based, and multimodal approaches. For studies reporting multiple accuracy values, a combined average accuracy is presented for consistency.

Study	Deep Learning Model	Primary Task	Dataset Used	Accuracy (%)	Key Strengths	Challenges & Limitations
[29]	HistoXGAN	Reconstruction of histopathological images from genomic, radiographic, and pathology data	BreaKHis	91.51	Accurate reconstruction supports virtual biopsy, improves explainability	Requires extensive computational resources, limited generalizability to rare cancer subtypes, challenges in training stability
[30]	NAS-SGAN	Nuclear atypia scoring for BC grading	MITOS-ATYPIA1	80.55	Effectively utilizes both labeled and unlabeled data; Reduces the need for large annotated datasets	Computationally expensive; Requires stable training for adversarial networks
[31]	ACGAN + ResNet-50 + XGBoost	BC histopathology image classification	BreakHis	88.24	Synthesizes realistic histopathology images; Improves dataset size and balance	Computationally intensive; Requires stable training of adversarial networks.
[32]	Hybrid CNN	Generating high-resolution synthetic pathology images for cancer diagnosis	TCGA, OVCARE	77.13	Generates high-resolution (1024 × 1024) synthetic pathology images, enhances classifier performance	Requires large training datasets, potential GAN artifacts, needs further clinical validation
[33]	PTGAN (Prototype Transfer GAN)	Unsupervised BC histology image classification	BreakHis	88.90	Effective domain adaptation Reduces labeling costs using adversarial learning.	Requires stable training; Style differences between source and target domains.
[35]	Trained GAN discriminator (T-GAN-D)	Prognostication of BC patients using transcriptome data	TCGA + Radboud + Karolinska	70.60	Enhances prognostic accuracy by integrating GAN-based synthetic data, improves patient risk classification	Requires large-scale transcriptomic data, potential biases in GAN-generated synthetic profiles

accuracy and consistency while allowing fine-grained analysis of tissue patterns [52,53]. Postprocessing methods, such as region-growing and fast marching level sets, further refine tumor boundaries, thereby enhancing interpretability [52,53]. These models also support early detection and robust feature learning. Ensemble strategies and hybrid designs strengthen performance across datasets [54,55]. RI-ViT, for example, integrates residual and Inception modules with transformers to address the class imbalance and improve accuracy across scales [55]. Pretraining on large datasets, combined with color normalization and augmentation, helps ViTs generalize better while maintaining lower complexity than CNN-based ensembles [56]. Several ViT models outperform conventional CNNs in patch- and slide-level classification tasks. They accurately identify cancer subtypes and demonstrate enhanced interpretability through visual explanations such as Grad-CAM, SHAP, LIME, and Score-CAM [49,56,57].

2.6. Comparative analysis of deep learning models in breast cancer histopathology

Table 1 and Table 2 provide a comprehensive comparison of deep-learning approaches applied to breast cancer histopathology. Table 1 focuses on conventional models, primarily convolutional neural networks (CNNs) such as VGG, Inception, and ResNet. These models have demonstrated strong performance in classification and risk prediction tasks across various datasets, using effective feature extraction, data augmentation, and transfer learning. However, they often face challenges such as overfitting, reliance on large annotated datasets, and reduced accuracy in multiclass settings.

In contrast, Table 2 highlights advanced and emerging models, including generative adversarial networks (GANs), transformer-based architectures, and multimodal frameworks that integrate histopathology with genomic and clinical data. These methods enhance performance

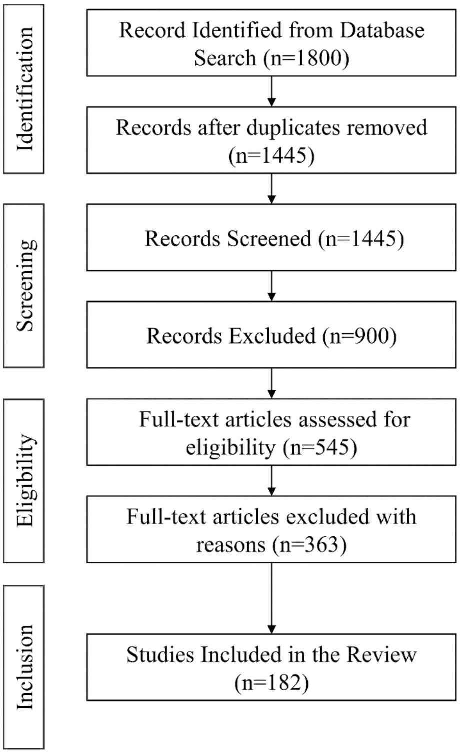

Fig. 5. PRISMA flow diagram illustrating the study selection process. The chart outlines the number of records identified, screened, assessed for eligibility, and ultimately included in the review based on predefined inclusion and exclusion criteria.

in complex tasks, such as image synthesis, unsupervised classification, and prognostic modeling. While they offer improved interpretability, domain adaptability, and biomarker discovery, their limitations include high computational demands, training instability, and regulatory complexities.

2.7. Research gap

Addressing the following gaps is essential to enhance the reliability, fairness, and clinical adoption of deep learning (DL) in breast cancer (BC) histopathology.

- Variability in staining, slide preparation, and imaging equipment across institutions limits the generalization of DL models, underscoring the need for standardized image capture and processing methods.
- The limited size, diversity, and representation of existing datasets restrict models from learning comprehensive diagnostic patterns, emphasizing the need for automated annotation methods and synthetic data generation techniques like GANs to address this limitation.
- The lack of transparency in DL models reduces clinician trust, underlining the need for research on improving interpretability through visualization techniques and explainable AI tools.
- Most studies focus solely on imaging data, overlooking the potential of integrating genetic, proteomic, and clinical data to improve diagnostic accuracy and treatment planning, with limited research on effective multimodal integration.
- Imbalanced training datasets introduce biases that reduce fairness and exacerbate healthcare disparities, necessitating efforts to identify, quantify, and mitigate these biases for broader applicability across diverse populations.

3. Research methodology

This study follows the PRISMA guidelines [59] to ensure methodological rigor and transparency in identifying and analyzing relevant research. The methodology includes structured steps in data identification, screening, eligibility assessment, and final inclusion, as illustrated in Fig. 5.

3.1. Literature search strategy

A comprehensive systematic literature search was conducted across eight major databases: PubMed, IEEE Xplore, Web of Science, Elsevier, Google Scholar, Springer, Wiley, and Scopus. The search targeted peer-reviewed studies published between 2014 and 2024 focusing on deep learning (DL) applications in breast cancer (BC) histopathology.

The search query was structured using controlled vocabulary (e.g., MeSH terms) and Boolean operators (AND/OR) to refine search precision. The search terms included:

(i). Primary Keywords: (“Breast Cancer” OR “Histopathology”) AND (“Deep Learning” OR “Artificial Neural Network” OR “CNN” OR “GAN” OR “Transformer”).
(ii). Task-Specific Terms: “Classification,” “Segmentation,” “Detection.”
(iii). Advanced Techniques: “Generative Adversarial Network,” “Transfer Learning,” “Explainable AI,” “Breast Cancer Imaging,” “Multimodal Learning,” “Self-Supervised Learning.”

Boolean operators (AND/OR) and controlled vocabulary (e.g., MeSH terms) were used to refine and expand the search scope.

3.2. Study selection process

The selection of studies was conducted in four sequential stages following the PRISMA framework:

3.2.1. Identification

A total of 1,800 studies were retrieved from the database search. After removing 355 duplicate records, 1,445 unique studies remained for screening.

3.2.2. Screening

Two independent reviewers conducted title and abstract screening. Studies that lacked relevance to deep learning-based breast cancer (BC) detection, classification, or segmentation were excluded. Moreover, 900 studies were removed, leaving 545 full-text studies for eligibility assessment.

3.2.3. Eligibility

Each full-text study was assessed based on methodological consistency, dataset quality, and clinical relevance. Studies were excluded if they:

- Focused only on traditional machine learning (ML) without deep learning (DL).
- Lacked reproducibility (e.g., used proprietary datasets without validation).
- Did not report performance metrics such as accuracy, Dice coefficient, or F1-score.
- Non-published or non-peer-reviewed articles are also excluded.

After this phase, 363 studies were excluded, leaving 182 for final inclusion.

3.2.4. Inclusion

A total of 182 peer-reviewed studies were selected for in-depth

Table 3
Inclusion and exclusion criteria are used to select studies for the literature review, outlining parameters such as time frame, study type, focus, techniques, data sources, metrics, and language.

Criteria	Inclusion	Exclusion
Time Frame	2014–2024	Studies before 2014
Publication Type	Peer-reviewed journal/conference papers	Preprints, book chapters, non-peer-reviewed sources
Study Focus	DL-based classification, segmentation, detection	Studies unrelated to breast cancer histopathology
Techniques	CNN, GAN, Vision Transformers, Explainable AI	Traditional ML without deep learning
Data Sources	Public datasets (CAMELYON, BACH, TCGA, BreakHis, BreCaHAD, MITOS-ATYPIA-14, Histopathologic Cancer Detection Dataset, IDC, TUPAC 2016, and SNOW) and a few studies employed private datasets.	Non-pathology datasets
Metrics	Accuracy, F1-score, Dice coefficient, AUC	Studies without performance metrics
Language	English Language	Non-English Language

analysis to explore deep learning (DL) advancements, challenges, and clinical applications in breast cancer (BC) histopathology. Furthermore, 17 references from reputable online sources were included to support statistical insights, dataset descriptions, and regulatory context, bringing the total number of references to 199. The study selection process is summarized in the PRISMA flow diagram presented in Fig. 5. Out of 1,800 records initially identified through database searches, 1,445 remained after duplicates were removed. Following title and abstract screening, 900 records were excluded. The remaining 545 full-text articles were assessed for eligibility, of which 363 were excluded based on predefined criteria, resulting in 182 studies being included in this review.

3.3. Inclusion and exclusion criteria

This review encompasses peer-reviewed journal articles and conference proceedings published between 2014 and 2024, ensuring that the most recent and high-quality research is considered. Only studies that applied deep learning (DL) techniques, such as CNNs, GANs, and ViTs, to breast cancer histopathology were included, focusing on critical tasks like classification, segmentation, and detection. Studies utilizing publicly available datasets (e.g., CAMELYON, BACH, TCGA, BreakHis) or well-documented reproducible private datasets were selected to ensure transparency and reliability. Furthermore, only studies reporting quantitative evaluation metrics such as accuracy, Dice coefficient, sensitivity, and specificity were included, as these metrics provide objective performance assessments crucial for comparing different deep learning models.

Studies relying solely on traditional machine learning (ML) approaches without incorporating deep learning (DL) components were excluded, as they do not align with the scope of this review. Papers that lacked performance evaluation metrics or validation methods were omitted, as they did not provide measurable insights into model effectiveness. Book chapters, preprints, and non-peer-reviewed sources were also excluded to maintain a focus on original, high-quality, peer-reviewed research. Published books and websites are explored to get statistics or dataset information. Furthermore, studies that did not specifically address breast cancer (BC) histopathology or medical imaging were removed to ensure the review remains relevant to its core objective. Non-English studies were excluded to ensure consistency in data extraction and interpretation, as translation inaccuracies could introduce biases or misinterpretations. Furthermore, most high-impact medical imaging and deep learning (DL) research is published in English, ensuring that the included studies are widely recognized and accessible within the global research community. Table 3 presents the eligibility criteria required for peer-reviewed studies from 2014 to 2024 that apply deep-learning methods (CNNs, GANs, Vision Transformers, and explainable AI) to breast cancer histopathology tasks.

Table 4
A quality assessment checklist is used to evaluate the methodological rigor and reporting standards of the included studies.

Checklist Question	Criteria
1. Is the study objective clearly defined?	Yes / No
2. Are dataset details (e.g., source, size) provided?	Yes / No
3. Is the deep learning architecture well-described?	Yes / No
4. Are performance metrics (e.g., accuracy, F1-score) reported?	Yes / No
5. Is a comparative analysis with other models included?	Yes / No
6. Are limitations and biases discussed?	Yes / No
7. Are validation methods (cross-validation, external testing) applied?	Yes / No
8. Is reproducibility ensured (open-source code, dataset accessibility)?	Yes / No
9. Does the study contribute new insights to the field?	Yes / No

3.4. Quality assessment

As stated by [60], only studies that provided a “yes” response to at least seven questions on the quality checklist were considered eligible for inclusion, as presented in Table 4. This meticulous quality assessment approach ensured that the selected studies made a meaningful and reliable contribution to the review. Data extraction and assessment procedures were employed to determine the significance and robustness of the findings.

3.5. Data extraction and synthesis

For each selected study, we extracted the following details:

- Publication details (year, authors, journal/conference).
- Deep learning techniques used (CNN, GAN, ViTs, multimodal learning).
- Performance metrics (accuracy, precision, recall, AUC, F1-score).
- Datasets used (CAMELYON, BACH, TCGA, BreakHis).

A comparative evaluation of DL models was conducted, and findings were visually represented using bar charts, pie charts, and histograms to illustrate:

- The distribution of DL techniques across studies.
- The prevalence of dataset usage in histopathology research.
- The impact of different architectures on model performance.

3.6. Research questions

The formulation of research questions guided the entire methodological process. For instance, questions related to model performance and generalizability (e.g., Q1 to Q3) informed the data search strategy through targeted terms such as "staining variability," "augmentation," and "domain adaptation." Studies reporting performance metrics (e.g., accuracy, Dice, AUC) were prioritized during selection to address technical and clinical questions (Q4-Q9). Furthermore, the quality assessment checklist ensured that included studies provided sufficient detail to support evidence-based responses, particularly on interpretability (Q5) and fairness (Q8).

(A) Technical Challenges in Breast Cancer Histopathology

1. How do staining variability, image resolution, and dataset bias impact deep learning (DL) model performance, and what strategies can mitigate these effects?

2. What data augmentation and domain adaptation techniques enhance the generalizability of DL models in BC histopathology?

3. What are the key challenges in deploying DL models in diverse

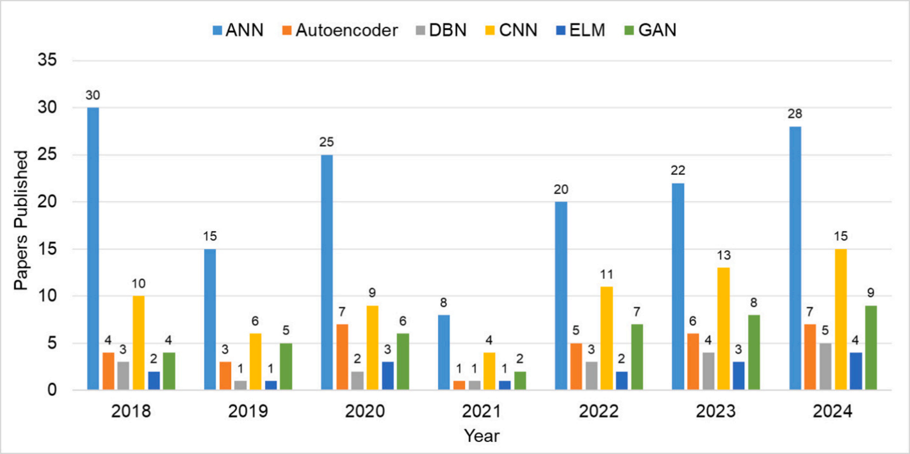

Fig. 6. Year-wise distribution of published studies (2018–2024) using various deep-learning techniques in breast cancer histopathology.

clinical settings, and how can they be addressed to improve robustness?

(B) Methodological Innovations in Deep Learning

4. What are the most effective CNN, GAN, and Transformer architectures for BC histopathology, and how do they compare regarding classification, segmentation, and prognostic tasks?

5. How does explainable AI (XAI), including Grad-CAM, SHAP, and LIME, improve clinical trust and interpretability in AI-driven histopathology models?

6. Can multimodal fusion improve diagnostic precision and patient stratification?

(C) Clinical Applications of AI in Histopathology

7. How can deep learning be integrated into real-world histopathology workflows while adhering to regulatory guidelines such as FDA's Good Machine Learning Practice (GMLP) and European CE-MDR?

To support real-world deployment, deep learning models must align with Good Machine Learning Practice (GMLP) principles set by the U.S. FDA and CE-MDR regulations in the European Union. Platforms such as Paige.AI and Ibex Medical Analytics have successfully integrated AI tools into clinical pathology workflows under regulatory supervision. Furthermore, federated learning approaches enable collaborative model development across institutions while preserving data privacy, thereby facilitating adoption within current legal and ethical frameworks.

8. What strategies ensure fairness, transparency, and inclusivity in AI-driven histopathology models across diverse patient demographics?

Ensuring fairness and inclusivity requires deliberate inclusion of diverse patient data during model development and validation. Bias mitigation frameworks such as IBM's AI Fairness 360 can guide the detection and correction of algorithmic bias. Moreover, reporting model performance across demographic subgroups and utilizing explainability techniques such as SHAP and LIME can enhance transparency, support equitable decision-making, and foster trust among clinicians and patients.

9. How do AI-driven diagnostic systems impact early detection, personalized treatment planning, and complete patient outcomes in breast cancer histopathology?

4. Innovative deep learning solutions for breast cancer histopathology image analysis

This section reviews the main deep learning architectures applied to breast cancer diagnosis namely autoencoders, artificial neural networks (ANNs), deep belief networks (DBNs), extreme learning machines (ELMs), generative adversarial networks (GANs), and convolutional neural networks (CNNs). Each architecture offers unique capabilities,

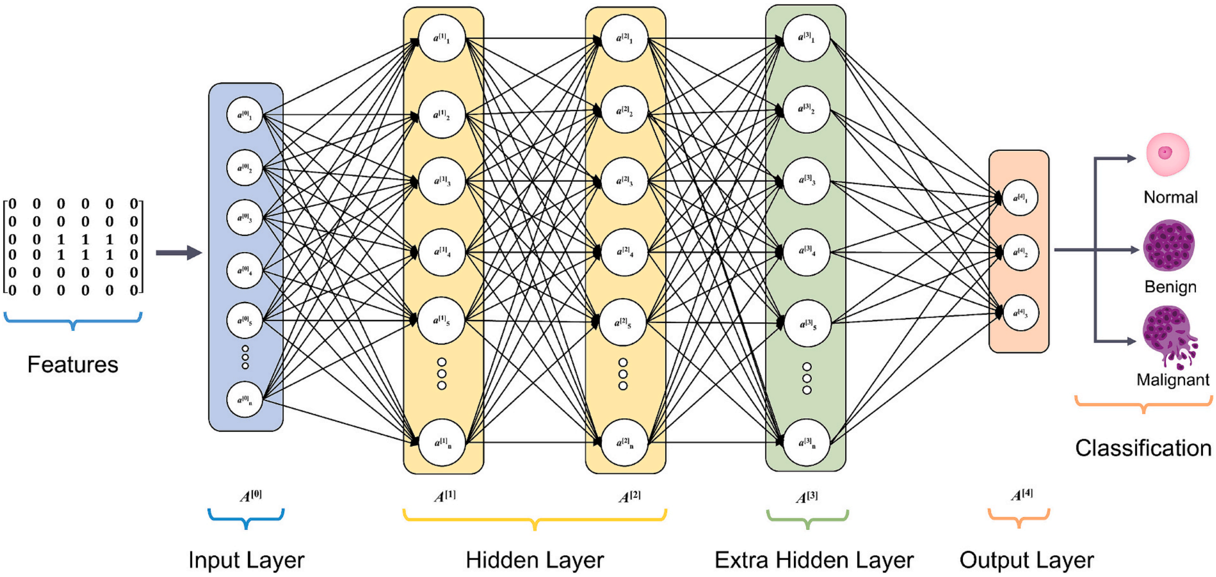

Fig. 7. Artificial Neural Network (ANN) architecture with multiple hidden layers for classifying breast cancer histopathology features into normal, benign, and malignant categories (modified from [67]).

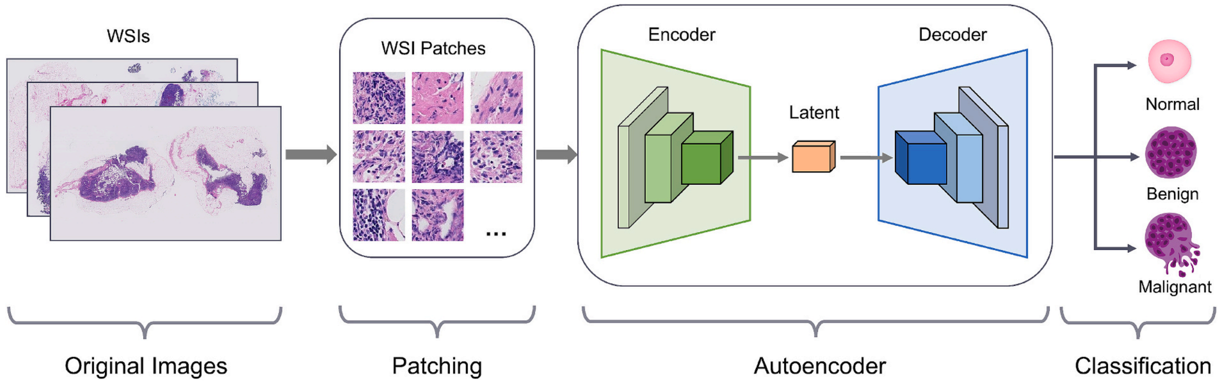

Fig 8. An example of a breast cancer detection approach using an autoencoder model. WSIs are patched and passed through an encoder-decoder structure to extract latent features for classifying tissue as normal, benign, or malignant [14].

contributing to diagnostic accuracy and efficiency advancements. Fig. 6 illustrates the annual progress in research publications focusing on these architectures, reflecting the increasing use of DL methods in addressing breast cancer (BC) detection and treatment challenges.

DL algorithms have significantly advanced in recent years, driving the widespread adoption of BC diagnosis. DL-based computer-aided diagnosis (CAD) systems modernize BC detection by automating tasks such as lesion segmentation, feature extraction, and feature selection, enabling the classification of breast masses as benign or malignant [61]. The efficiency and automation provided by DL have led to its increasing acceptance in medical diagnostics.

4.1. *Artificial neural networks (ANNs)*

Artificial Neural Networks (ANNs) are crucial in breast cancer (BC) histopathology image analysis, enhancing feature extraction, classification accuracy, and decision-making. Deep ANNs with multiple hidden layers effectively process high-dimensional histopathology images but require longer training times and significant computational resources, posing challenges in resource-limited settings [62,63]. In contrast, simpler ANNs offer faster training and ease of optimization but may struggle with complex datasets, necessitating further refinement for reliable diagnostic performance [34]. ANNs improve BC histopathology analysis by predicting nodal status reducing unnecessary sentinel lymph node biopsies (SLNB). They outperform traditional models by integrating patient and clinicopathological features for more accurate surgical decision-making. However, prospective validation is needed to ensure clinical reliability [64]. ANNs enhance BC histopathology analysis by improving tumor classification accuracy using Probabilistic Neural Networks (PNN). They outperform traditional methods but rely on high-quality image segmentation and feature extraction. However, their performance can be affected by noisy or low-contrast images, reducing reliability [65]. Further advancements, such as the life-sensitive self-organizing error-driven (LS-SOED) model, aimed to guide ANNs toward cancer-specific objectives, enhancing diagnostic precision. However, data quality remains a critical factor, as missing values in datasets significantly impacted ANN accuracy and reliability, strengthening the need for high-quality, complete datasets for effective BC histopathology analysis [66].

Fig. 7 (modified from [67]) illustrates the architecture of an Artificial Neural Network (ANN) used for breast cancer classification. The model processes feature vectors through an input, multiple hidden, and output layers. Intermediate layers capture complex nonlinear relationships within the data, enabling the network to distinguish between normal, benign, and malignant tissue types. This layered design enhances the model's learning capacity, making it suitable for high-dimensional histopathological image data.

4.2. *Autoencoders*

Autoencoders are essential for breast cancer (BC) histopathology image analysis, supporting dimensionality reduction and robust feature extraction. These models, comprising encoder, hidden (encoding), and decoder layers, effectively compress and reconstruct input data, improving data representation for medical imaging tasks [68–70]. Unlike linear methods like PCA, autoencoders handle linear and non-linear transformations, making them particularly suited for analyzing complex histopathology images. However, training autoencoders is computationally intensive, requiring large datasets and careful hyperparameter tuning [68–70]. Specialized autoencoders such as denoising autoencoders (DAE), sparse autoencoders (SAEs), and variational autoencoders (VAEs) improve breast cancer histopathology analysis by removing noise, enforcing compact feature representations, and generating probabilistic samples for robust analysis [71,72]. Stacked denoising autoencoders (SDAEs) further refine feature extraction, making them effective for processing high-dimensional histopathology slides [73]. These models extract meaningful latent features from WSIs, improving classification into normal, benign, or malignant categories [70,72].

The stacked sparse autoencoder (SSAE) framework enhances nuclei patch classification in BC histopathology by using two sparse autoencoders (SAEs) to learn high-level features, outperforming traditional techniques like PCA. Integrated with a Softmax classifier, SSAE improves classification performance but faces generalizability issues due to small dataset sizes [74]. Further studies expanded the dataset to 537 H&E-stained histopathology images to improve model robustness, though the absence of pre-processing techniques limited performance in handling image variations and noise [75]. Complementary research demonstrated that integrating SSAE with additional feature extraction methods enhanced classification accuracy [72]. Beyond imaging, autoencoders aid in gene expression analysis, offering deeper insights into biological data. They have been used to reconstruct datasets and extract meaningful gene expression patterns, improving cell nuclei classification in histopathology images, particularly when labeled data is scarce [76,77]. SDAEs have shown potential in extracting critical features from image patches, strengthening their potential in BC histopathology image analysis [73].

Fig. 8 illustrates a schematic for breast cancer detection based on an autoencoder architecture. The process begins with whole-slide images (WSIs), which are divided into smaller patches. These patches are passed through an autoencoder, which consists of an encoder that compresses the input into a latent representation and a decoder that reconstructs it. The learned latent features are then classified into normal, benign, or malignant categories. This approach enables unsupervised feature learning and facilitates the efficient representation of high-dimensional histopathology data [68–70].

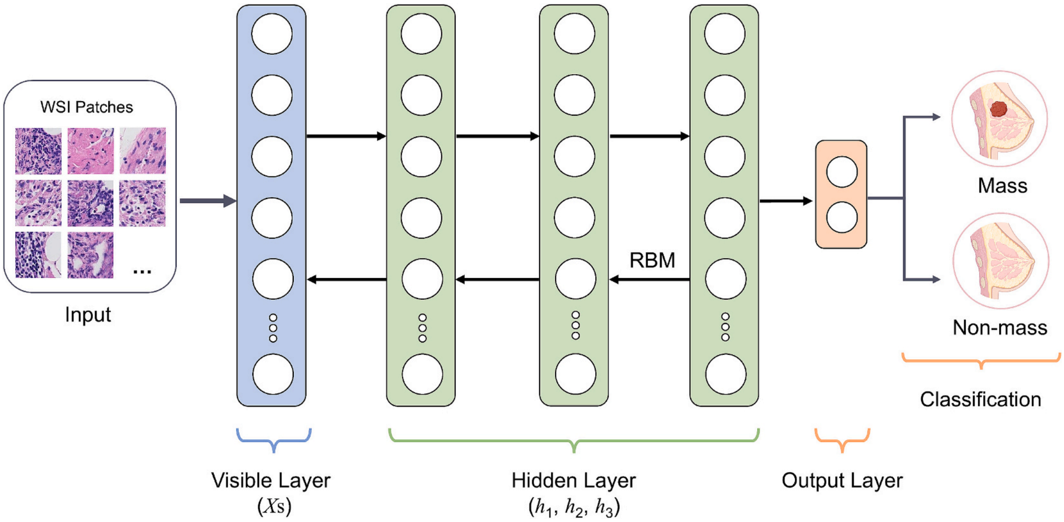

Fig 9. Example architecture of a Deep Belief Network (DBN) for breast cancer diagnosis. The model processes WSI patches through a visible layer and multiple hidden layers using Restricted Boltzmann Machines (RBMs), leading to classification as mass or non-mass lesions (illustration is modified from [80] and the WSI Patches [14]).

4.3. Deep belief networks (DBNs)

Deep Belief Networks (DBNs) have been utilized to enhance breast cancer (BC) histopathology analysis by improving feature learning and classification accuracy. However, they require careful hyperparameter tuning to ensure generalization and clinical applicability [78]. DBNs improve BC histopathology analysis by learning hierarchical features and enhancing classification accuracy. However, they are computationally demanding and sensitive to latent neuron activation, affecting performance [79]. Their multi-layered architecture reduces non-linear dimensionality, enhancing classification and segmentation precision after unsupervised pre-training and supervised fine-tuning [80]. Fig. 9 illustrates the structure of a Deep Belief Network (DBN) applied to breast cancer diagnosis. WSI patches are fed into a visible input layer, followed by stacked hidden layers trained using Restricted Boltzmann Machines (RBMs). The output layer performs classification to distinguish between mass and non-mass lesions. This hierarchical architecture enables unsupervised feature learning and supports efficient tissue representation for diagnostic decision-making [80]. Despite their advantages, DBNs are computationally intensive, requiring precise parameter adjustments for optimal training. Hybrid approaches combining DBNs with supervised backpropagation networks enhance classification accuracy but suffer from performance limitations due to the absence of noise reduction [80]. Similarly, integrating genetic and clinical data with DBNs enhances

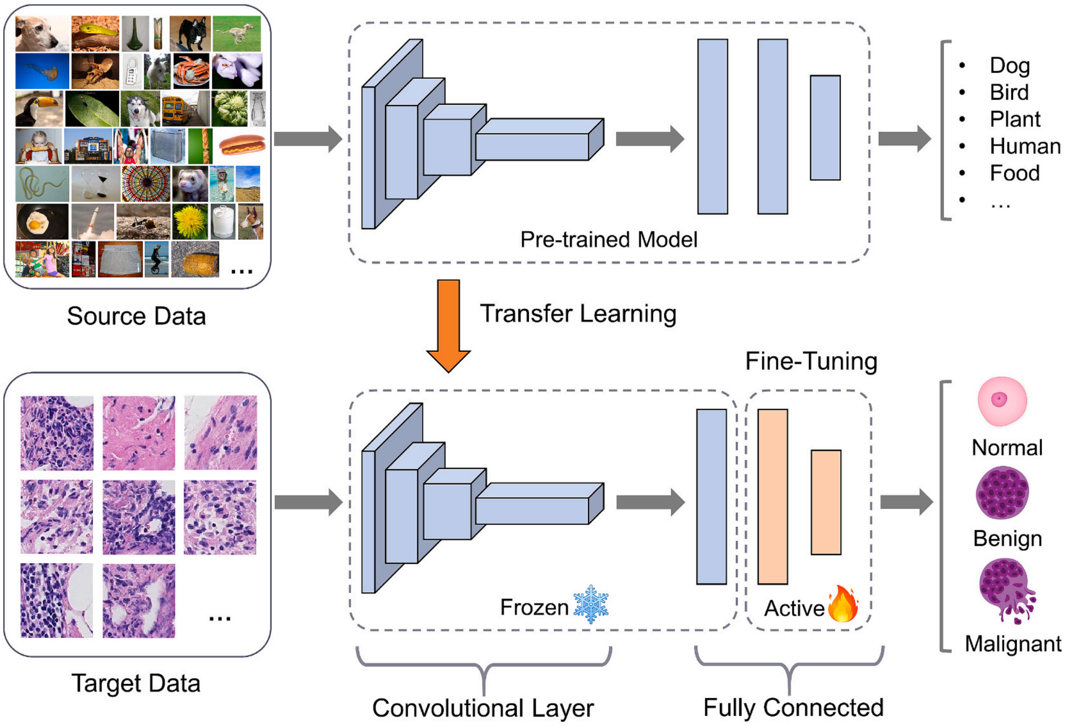

Fig. 10. Illustration of transfer learning in CNN-based breast cancer classification. A pre-trained model on source data is fine-tuned using histopathology images to classify tissue as normal, benign, or malignant (the WSI Patches retrieved from [14]).

disease progression analysis, although challenges in high-dimensional feature spaces and small datasets impact learning efficiency. Microarray-specific DBN techniques address these issues, optimizing performance for sparse data [81]. In breast mass segmentation, DBNs combined with conditional random fields (CRFs) and structured support vector machines (SSVMs) improve segmentation accuracy, with CRFs demonstrating faster computational efficiency [82]. DBNs improve BC histopathology analysis by capturing complex spatial relationships for better segmentation and classification. Their hierarchical structure enhances feature extraction, reducing diagnostic variability. DBNs improve histopathology image analysis by learning hierarchical features and reducing the need for large labeled datasets. They enhance tumor classification accuracy and segmentation [83]. Further improvements can be made by integrating hybrid models like attention mechanisms and self-supervised learning for enhanced accuracy and interpretability [83,84].

4.3.1. Convolutional neural networks (CNNs)

Convolutional Neural Networks (CNNs) are widely used in breast cancer (BC) histopathology due to their ability to extract multi-scale and hierarchical features directly from raw images, reducing reliance on manual feature engineering and improving diagnostic accuracy. Modern implementations often incorporate transfer learning, residual learning, and hybrid models with classifiers like SVMs or LSTMs to boost classification robustness [26,28,85–90].

As illustrated in Fig. 10, transfer learning involves adapting a pre-trained model (e.g., trained on ImageNet) to histopathology images. Convolutional layers are typically frozen to retain general feature extraction capabilities, while fully connected layers are fine-tuned to distinguish between normal, benign, and malignant tissue types using whole-slide image (WSI) patches [14].

Several techniques support CNN performance in histopathology. Data augmentation and stain normalization address limited data and staining variability, respectively, improving generalization [28,85,86,89,91,92]. Multi-scale learning and multi-magnification strategies enhance spatial context, while attention mechanisms improve interpretability and feature focus [87,90–95]. Ensemble approaches using architectures like GoogleNet, VGG11, and MobileNetV3_Small further increase robustness [87].

Transformer-based CNN variants and domain adaptation techniques have also improved generalization across datasets [28,94]. Specialized models such as DenseNet-201 and SEP networks offer efficient feature reuse and channel importance learning [91,92], while autoencoders assist with feature transformation [86,96]. Additional enhancements include spectral imaging, multimodal fusion, mitosis detection, and weak label refinement frameworks, all contributing to the effectiveness of CNNs in automated histopathological diagnosis [94,96].

4.4. Extreme learning machines (ELMs)

Extreme Learning Machines (ELMs), a specialized form of Artificial Neural Networks (ANNs), offer computational efficiency and adaptability for breast cancer (BC) classification. They use random weight initialization and analytical output calculations, allowing rapid training and retraining, making them suitable for real-time diagnostics [97]. However, ELMs often have lower accuracy than traditional neural networks, requiring optimization strategies to improve performance [98]. Hybrid ELM models integrate CNN-based feature extraction and genetic algorithms to enhance accuracy while maintaining efficiency, refining activation functions, and parameter selection [98,99]. CNN-ELM models, incorporating morphological, texture, and density features, have demonstrated higher accuracy in BC detection [100]. Similarly, interactive cross-task ELM (ICELM) uses deep transfer learning, improving diagnostic robustness in complex imaging tasks [101].

Further advancements involve combining ELM with Deep Belief Networks (DBNs) and backpropagation, addressing convergence issues, and optimizing weight assignment [102]. Feature reduction techniques like Moth Flame Optimization (MFO-ELM) balance computational efficiency with data integrity, ensuring optimal model performance in resource-sensitive applications [103]. Despite these enhancements, ELMs remain limited by lower accuracy and occasional longer training times than conventional deep learning methods. However, their speed, adaptability, and integration potential make them valuable BC histopathology imaging tools when optimized effectively [98]. As depicted in Fig. 11, the streamlined architecture of ELMs balances speed, efficiency, and accuracy, making them particularly suitable for applications where traditional neural networks (NNs) may be too resource-intensive [98].

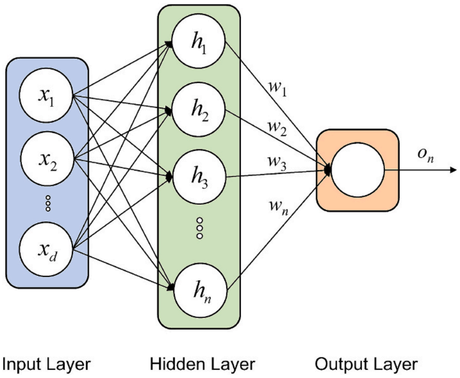

Fig. 11. The basic architecture of an Extreme Learning Machine (ELM) consists of an input layer, a single hidden layer with randomly assigned weights, and an output layer for final prediction [98].

4.5. Generative adversarial networks (GANs)

Generative Adversarial Networks (GANs) improve breast cancer (BC) classification in histopathology by enhancing image resolution, generating synthetic data for augmentation, and addressing class imbalance, eventually improving model robustness and accuracy [104–108]. They integrate with DL models such as CNNs and ELMs, supporting a multi-method approach in histopathology analysis [109]. GANs assist in data augmentation, segmentation, and feature extraction but require optimized loss functions, training stability techniques, and domain-specific constraints to prevent overfitting and mode collapse [105,106,108,109]. Advanced architectures, including hybrid Transformer-based models and normalization techniques, further enhance their performance in feature learning and predictive accuracy [105,107]. AnoGAN filters mislabeled patches in WSIs, improving classification reliability [110], while GAN-based segmentation models enhance mitosis and nuclei detection in H&E-stained images, reducing manual counting inefficiencies [111].

Fig. 12 illustrates a GAN framework applied to histopathology image analysis. The Generator $G(z)$ synthesizes histopathology images from a latent distribution $z_{\mu,\sigma}$, while the Discriminator $D(x,x')$ evaluates whether an image x' is real or synthetic.

5. Datasets for breast cancer histopathology images

Breast cancer research involving histopathology images can be conducted using public and private datasets. Public datasets are widely adopted due to their open accessibility, standardized annotations, and significant utility in benchmarking deep learning (DL) models. In contrast, private datasets, although often restricted in availability, provide high clinical relevance and are frequently employed in institution-specific studies.

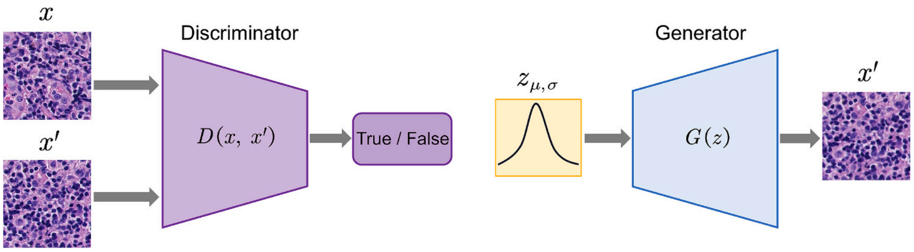

Fig. 12. Generative Adversarial Network (GAN) framework for the synthesis of histopathology images. The generator creates synthetic images from random noise, while the discriminator distinguishes between real and generated samples to improve the quality of the generated images.

5.1. Private datasets discussed in this study

The dataset characteristics of studies included in this review that employed private datasets are detailed below. (1). The dataset employed in [58] is from Colsanitas University Hospital in Bogotá, Colombia, and includes 845 H&E-stained image patches extracted from 544 whole slide images (WSIs) of 80 patients. It comprises 437 carcinoma and 408 non-carcinoma samples, scanned at 400× magnification and cropped at 200× (around 50 µm/pixel), with a resolution of 1278 × 760 pixels. Staining involved both H&E and immunohistochemical markers (ER, PR, HER2, Myosin, Ki-67), and all images were exported in PNG format using QuPath 0.1.2. Designed for binary classification, the dataset supported training an ensemble of VGG16 and VGG19 models.

(2). The dataset employed in [112] is sourced from Christian Medical College Hospital, India, and it comprises 196 mammography images categorized as BI-RADS 4 or higher. Tumor regions were annotated by radiologists and verified by oncology specialists. CycleGAN-based augmentation expanded the dataset to 1,184 images (947 for training and 237 for validation), resized to 256 × 256 pixels. It trained a 3D Connected-AUNets model for tumor segmentation in high-risk cases.

(3). The retrospective clinical dataset employed in [113] was collected from Karolinska University Hospital and Södersjukhuset in Stockholm, Sweden. This clinical dataset comprises 234 FFPE H&E-stained whole-slide images (WSIs) from early-stage, estrogen receptor (ER)- positive, Human Epidermal Growth Factor Receptor 2 (HER2)-Negative breast cancer patients diagnosed between 2020 and 2022. Slides were digitized at 40× magnification (0.226 µm/pixel) using a Hamamatsu NanoZoomer XR scanner. Associated metadata included histologic grade, Ki67 index, tumor size, and hormone receptor status. The dataset evaluated Stratipath Breast, an AI-based risk profiling tool, against the Prosigna (PAM50) assay.

(4). The Colsanitas Dataset (Extended) employed in [114] is an extended version of the Colsanitas dataset, including 2,250 H&E-stained image patches derived from the same 544 WSIs. It covers four diagnostic categories: normal (600), benign (250), in situ carcinoma (250), and invasive carcinoma (1,150). The images were scanned at 40× magnification using a Roche iScan HT scanner, saved in TIFF format (2048 × 1536 pixels, 0.46 µm resolution), and annotated by two pathologists. This dataset was used for multiclass classification using an Xception-based deep learning model.

(5). This private dataset employed in [115] is developed by the Histopathology Department of the National Institute of Oncology in Rabat, Morocco. This dataset includes 328 H&E-stained images from 116 surgical specimens of invasive carcinoma. Images were acquired at 200× magnification using an Olympus BX43 microscope and exported in JPEG format. The dataset is divided into three categories: benign/normal (152), in situ carcinoma (70), and invasive carcinoma (106). Extensive augmentation generated 750 × 750 patches, supporting multiclass classification using ResNet50 and Xception for feature extraction and XGBoost for classification.

While these private datasets offer high clinical validity, expert annotation, and diverse diagnostic content, their restricted availability limits reproducibility and broader comparative analysis.

5.2. Publicly accessible BC histopathology datasets

Fig. 13 presents a timeline of key publicly accessible datasets used in breast cancer histopathology research. Starting from TCGA-BRCA (2006–2012), the timeline includes widely used datasets such as MITOS-ATYPLA-14, BreakHis, CAMELYON16/17, TUPAC, and BACH. These datasets cover various tasks, including mitosis detection, tumor classification, metastasis detection, and pixel-level annotation. Their availability has significantly supported the development and benchmarking of deep learning models in this domain.

Table 5 presents a comparative summary of widely used public datasets in breast cancer histopathology. These datasets vary in size, magnification levels, imaging modalities, and intended tasks such as classification, segmentation, mitosis detection, and metastasis prediction. Each entry outlines key attributes, including data sources, sample sizes, image resolutions, application scope, and associated challenges such as staining variability, class imbalance, and annotation limitations. This overview serves as a reference for selecting suitable datasets based

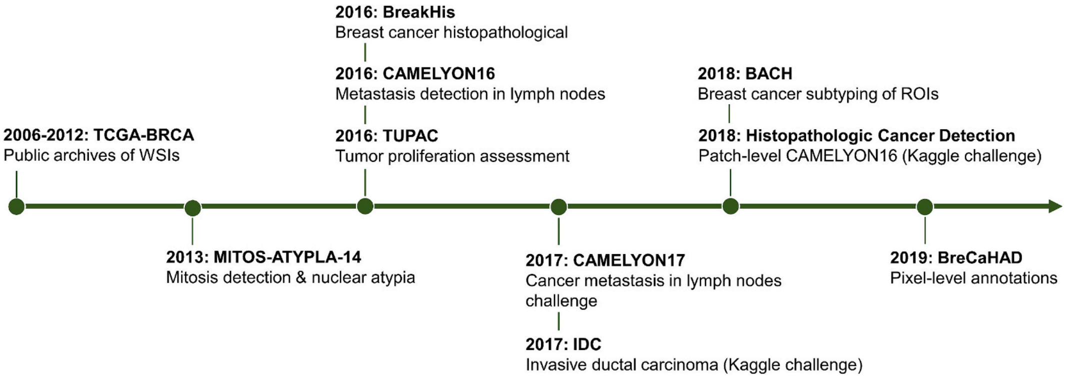

Fig. 13. Timeline of widely used and publicly accessible breast cancer histopathology datasets, highlighting major releases and dataset-specific challenges from 2006 to 2019.

Table 5
Overview of publicly available breast cancer histopathology datasets commonly used in deep learning research. Each entry includes dataset origin, composition, imaging specifications, intended applications, and known limitations.

Dataset Name	Source/Year/ Reference	Description	Statistics	Size	Magnification	Application	Data Challenges (Including Biases)
TCGA-BRCA (The Cancer Genome Atlas – Breast Cancer)	National Cancer Institute (NCI), NIH (2006-2012) [116]	A comprehensive dataset containing histopathology WSIs, genetic, and clinical data for BC research.	WSIs, mRNA expression, mutation profiles, and clinical data are available.	1,098 BC cases, including WSIs, genomic, and clinical data	Varies (dependent on sample source)	Multi-modal analysis: classification, prognosis, segmentation, and molecular subtyping	Variability in staining, scanner types, and institutional differences affect model generalization.
MITOS-ATYPIA-14 (MITOS-ATYPIA Grand Challenge Dataset)	MITOS and ATYPIA Grand Challenge (2013) [117]	Histopathological BC dataset for mitosis detection and nuclear atypia scoring	Annotations from multiple pathologists, detailed mitotic figures, and nuclear atypia scores	Train: 1,136 (40 ×), 284 (20 ×)	20 × (nuclear atypia), 40 × (mitosis)	Classification (nuclear atypia grading), segmentation (mitotic cell detection)	Variability in expert annotations, small dataset size, class imbalance
BreakHis (Breast Cancer Histopathological Image Database)	P&D Laboratory, Brazil (2016) [118]	A dataset of histopathological images of BC for classification tasks.	Images were captured using a 3CCD camera attached to an optical microscope.	7,909 images from 82 patients	40 ×, 100 ×, 200 ×, 400 ×	Binary (benign vs. malignant) and multi-class tumor subtype classification	Imbalance in benign vs. malignant cases, dataset limited to a single source
CAMELYON16	Radboud University Medical Center and UMC Utrecht (2016). Article: [119], Dataset Link: [14]	Annotated WSIs of lymph node sections for metastasis detection in BC patients	Contains WSIs from sentinel lymph node biopsies, with manually annotated metastases	400 WSIs (270 training, 130 testing)	RUMC: 20 ×, UMCU: 40 ×	Metastasis classification, tumor segmentation, cancer detection	High-resolution WSIs require large storage computational power, domain shifts across staining protocols, inter-observer variability in annotations
TUPAC 2016 (Tumor Proliferation Assessment Challenge 2016)	MICCAI (2016) [120]	WSIs of BC histopathology for assessing tumor proliferation through mitotic score prediction and gene expression-based PAM50 proliferation scores.	Training Set: 500 WSIs, Testing: 321 WSIs. Annotations: Mitotic scores and PAM50 proliferation scores for the training set.	821 WSIs	40 ×	Tumor proliferation assessment, mitosis detection	High Variability: Differences in staining and scanning protocols. Large Image Sizes: Challenges in processing gigapixel WSIs. Limited Annotations: Ground truth is provided only for the training set.
CAMELYON17	Radboud University Medical Center, UMC Utrecht (2017). Article: [121], Dataset Link: [122]	WSIs of lymph node sections for metastasis detection	Includes 1,000 WSIs from breast cancer patients, with annotations for metastasis	500 patients, 5 slides per patient (1,000 WSIs)	40 ×, 10 × (multi-resolution)	Metastasis detection, classification, segmentation	Data imbalance due to varying tumor sizes, staining variability, scanner differences, limited real-world generalizability
Invasive Ductal Carcinoma (IDC) Dataset	Kaggle (2017). Dataset Link: [123]	Histopathology images of BC specimens focusing on IDC detection. Each image is a 50x50 pixel patch extracted from WSIs.	RGB images (50x50 pixels), binary classification (0: non-IDC, 1: IDC), provided in .png format with filenames indicating patient ID, coordinates, and class.	277,524 images (198,738 IDC-negative; 78,786 IDC-positive)	40 ×	Classification (IDC detection)	Class imbalance, small patch size, lack of full context, variability in staining procedures.
BACH (BreAst Cancer Histology) Challenge Dataset	Organized by the International Conference on Image Analysis and Recognition (ICIAR) (2018). [124,125]	H&E-stained microscopy and WSIs of breast histology for classification into four categories: Normal, Benign, In Situ Carcinoma, and Invasive Carcinoma	Each image is 2048 × 1536 pixels (microscopy) + whole-slide images available	400 microscopy images (100 per class) + 30WSIs (unlabeled: 20, labeled: 10)	-	Classification and segmentation	Class imbalance, variability in staining, potential subjectivity in annotations
Histopathologic Cancer Detection Dataset	Uploaded on Kaggle (2018) [126]	Histopathologic images of lymph node sections for metastatic tissue detection. Each image is a 96x96 pixel patch from WSIs.	RGB images (96x96 pixels), binary classification (0: non-metastatic, 1: metastatic), provided in .tif format with train_labels.csv metadata.	220,025 images (110,000 labeled for training, 57,458 unlabeled for testing)	scanned at 40 ×, undersample at 10 × (PCam)	Classification (metastatic tissue detection)	Class imbalance, staining variability, small patch size, and lack of full context require data augmentation.

(continued on next page)

Table 5 (*continued*)

Dataset Name	Source/Year/ Reference	Description	Statistics	Size	Magnification	Application	Data Challenges (Including Biases)
BreCaHAD (Breast Cancer Histopathological Annotation Dataset)	Uploaded on Figshare [127]	histopathological image dataset designed for mitosis detection and annotation in BC tissue.	Contains H&E-stained histopathology slides with pixel-level annotations.	162 images with detailed mitosis annotations.	40 × (0.514 µm × 0.527 µm)	Mitosis detection, classification, and segmentation in BC histopathology images.	Limited diversity of staining techniques requires expert annotation for mitosis detection potential class imbalance.

Table 6
Performance comparison of deep learning models in breast cancer histopathology, covering segmentation and classification tasks across public and private datasets.

Model	Task	Performance	Dataset	Study
LinkNet	Segmentation	94.50%	SNOW	[21]
Feature Pyramid Network (FPN)	Segmentation	95.40%	SNOW	[21]
U-Net, ResNet-34	Segmentation	80.25%	SNOW	[131]
U-Net, DenseNet-121	Segmentation	79.90%	SNOW	[131]
U-Net, Xception	Segmentation	80.08%	SNOW	[131]
Deep WSI-Stroma	Segmentation	87.30%	TCGA-BRCA	[132]
3D Connected-AUNets	Segmentation	97.47%	Private	[112]
U-Net	Segmentation	83.05%	BCSS	[133,134]
U-Net++	Segmentation	86.46%	BCSS	[133,135]
MA-Net	Segmentation	88.48%	BCSS	[133,136]
DRD-UNet (Dilation, Residual, and Dense block)	Segmentation	85±8%	BCSS	[137]
Semi-supervised	Segmentation	83.40%	BCSS	[133]
U-Net	Segmentation	61.17%	ICPR2012, MITOS-ATYPIA-2014	[138]
VGG16	Classification	79.95%	BACH	[139]
ResNet-50	Classification	80.47%	BACH	[139]
Ensemble Model	Classification	92.58%	BACH	[139]
Hybrid DL Model	Diagnosis	88.07%	TCGA-BRCA	[140]
CNN	Classification	86.00%	PatchCamelyon 2016	[141]
CNN with Transfer Learning	Classification	94.4%	CAMELYON16	[142]
EfficientNet_B0	Classification	84.96%	BreakHis	[139]
VGG16	Classification	97.38%	BreakHis	[139]
ResNet-34	Classification	98.22%	BreakHis	[139]
ResNet-50	Classification	97.30%	BreakHis	[139]
Ensemble Models	Classification	98.43%	BreakHis	[139]
BreastNet	Classification	97.56%	BreakHis	[93]
DenseNet121-AnoGAN	Classification	99.13%	BreakHis	[110]
ResNet-18	Classification	92.15%	BreakHis	[45]
Inception-v3	Classification	98.97%	BreakHis	[128]
Pre-Trained Autoencoder	Classification	97.80%	BreakHis	[143]
Hybrid Model	Classification	92.57%	BreakHis	[129]
DRDA-Net	Classification	96.42%	BreakHis	[130]
DenTnet	Classification	99.28%	BreakHis	[144]
AlexNet-BC	Classification	86.31%	IDC	[145]
Fine-Tuned VGG16	Classification	91.67%	Private	[58]
Fine-Tuned VGG19	Classification	92.00%	Private	[58]

on research objectives.

6. Comparison of deep learning models in breast cancer histopathology datasets

Table 6 provides a comprehensive performance evaluation of deep learning (DL) models in breast cancer (BC) histopathology image analysis, highlighting segmentation results based on the Dice coefficient and classification outcomes using accuracy metrics. It summarizes the effectiveness of these models across various datasets, including BACH, BreakHis, TCGA-BRCA, and BCSS, demonstrating significant advancements in breast cancer (BC) detection and classification tasks. We noted the average values for the results of BreastNet [93], ResNet-18 [45], Inception-v3 [128], Hybrid Model [129], DRDA-Net [130], and Fine-Tuned VGG16, VGG19 [58].

Fig. 14 presents a comparative analysis of various deep learning (DL) models applied to four well-known histopathology datasets: BACH [124,125], TCGA [116,140,146–152], CAMELYON16 [14,119,153], and CAMELYON17 [121,122,154]. The results for the BACH and TCGA datasets are reported in terms of accuracy (percentage). However, the CAMELYON16 and CAMELYON17 datasets are evaluated using AUC (Area Under the Curve) and Cohen's Kappa, respectively, adhering to standard academic conventions. Fig. 14 depicts the performance variations of different models across datasets. In BACH and TCGA datasets (Fig. 14A and 14B), models such as Inception, ResNet, VGG16, and EfficientNet exhibit strong classification performance. KimiaNet achieves the highest accuracy (95%) on the TCGA dataset, while EfficientNet lags behind with 56% accuracy. In contrast, results from CAMELYON16 (Fig. 14C) demonstrate that ResNet attains the highest AUC (0.93), followed by VGGNet (0.87) and GoogLeNet (0.84), indicating robust performance in metastasis detection. Similarly, CAMELYON17 results (Fig. 14D) show that EfficientNet, PFA, and DeepLab achieve the highest Cohen's Kappa values (0.9), while EFCNN and DSNMF exhibit significantly lower agreement scores (0.07 and 0.09, respectively), reflecting weak classification performance.

The performance of DL models on these datasets is reported based on data obtained from their respective challenge websites and relevant studies used for comparison. The compiled data has been organized into Excel sheets and uploaded as supplementary files. The supplementary files are named "Supplementary_File1_BACH" for the BACH dataset, "Supplementary_File2_TCGA" for TCGA, "Supplementary_File3_CAMELYON16" for CAMELYON16, and "Supplementary_File4_CAMELYON17" for CAMELYON17. For consistency, in cases where different authors used various versions of a model (e.g., U-Net, U-Net++, U-Net3+, 3D U-Net), we unified them under a single category ("U-Net") in our analysis. This standardization ensured a fair comparative assessment. Fig. 14 presents average performance scores from multiple studies with complete methodologies sourced from public journals and dataset challenges.

7. Image processing and augmentation techniques

Data preprocessing prepares raw data for analysis by cleaning, standardizing, and transforming it to ensure consistent and accurate input for the model [155–157]. Data augmentation enhances dataset diversity by applying transformations like flipping, rotation, and

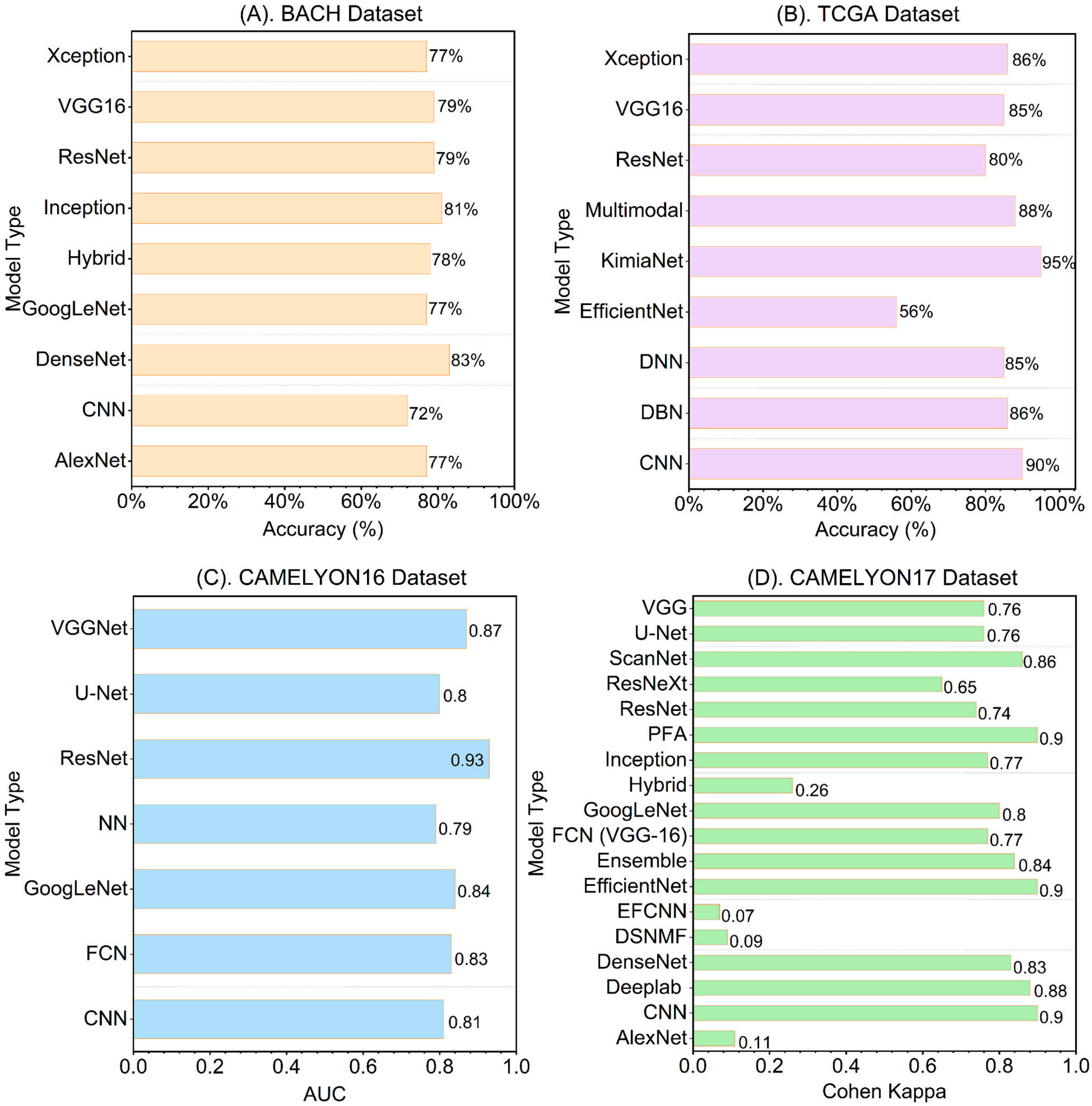

Fig. 14. Performance comparison of various DL Models on BC histopathology images: (A) represents accuracy results from the BACH dataset [124], while (B) presents accuracy findings from the TCGA dataset [116,140,146–152], compiled from multiple high-impact studies. Panel (C) and (D) illustrate the performance of models on the CAMELYON16 [153] and CAMELYON17 [154] challenge datasets, evaluated using AUC and Cohen's Kappa, respectively.

cropping, helping models generalize better and perform robustly under varied conditions [155–157]. Some data pre-processing and augmentation techniques are presented in Table 7 [155–157].

Fig. 15(A) [158] illustrates the pyramid representation of whole-slide images (WSIs), where WSIs are organized hierarchically across multiple zoom levels. The highest resolution is at the base, progressively decreasing towards the apex, with each level subdivided into fixed-size tiles for efficient processing. Fig. 15(B) depicts various dataset augmentation techniques applied to WSI patches, all aimed at enhancing model robustness and generalization [157]. The WSI patches are retrieved from CAMELYON16 [14].

8. Performance evaluation metrics

In medical imaging, the performance of deep learning models is often evaluated using different metrics depending on the task: detection, classification, or segmentation. Detection and classification metrics are presented in Eqs. (1) to (7) [159,160].

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2}$$

$$\text{Recall (Sensitivity or True Positive Rate)} = \frac{TP}{TP + FN} \tag{3}$$

$$\text{F1 Score} = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{4}$$

$$\text{Specificity (True Negative Rate)} = \frac{TN}{TN + FP} \tag{5}$$

Table 7
Data Pre-processing/augmentation techniques.

Techniques	Advantages
Adding Noise	It improves model robustness and reduces overfitting by simulating real-world variability.
Cropping	It enhances focus on relevant areas and handles object size and placement variations.
Flipping	It increases data diversity by introducing mirrored perspectives of the images.
Rescaling	It standardizes image sizes for consistent input to the model.
Brightness Adjustment	It adapts the model to varying lighting conditions in input images.
Rotation	It improves the ability to recognize objects in different orientations.
Translation	It enhances robustness by shifting objects to simulate spatial variations.
Contrast Adjustment	It improves feature visibility and adapts to varying image contrasts.
Saturation Adjustment	It adapts the model to changes in color intensity for better generalization.
Color Adjustment	It simulates various color properties to mitigate biases and improve adaptability.

$$AUC_{PR} = \sum_{i=1}^{n-1} \frac{(\text{Recall}_{i+1} - \text{Recall}_i) \times (\text{Precision}_{i+1} + \text{Precision}_i)}{2} \tag{7}$$

In Eqs. (1) to (5) above, TP (True Positives), TN (True Negatives), FP (False Positives), and FN (False Negatives) are used to describe classification outcomes. In equation (6), N shows the number of classes, AP_i is the average precision of the i^{th} class. AP is usually approximated using the precision values at discrete recall thresholds.

The performance evaluation metrics for segmentation tasks are presented in Eqs. (8) to (12) [159,160].

$$\textit{Dice Similarity Coefficient (DSC)} = \frac{2 \times |A \cap B|}{|A| + |B|} \tag{8}$$

$$\textit{Intersection over Union (IoU)} \Big/ \textit{Jaccard Index} = \frac{|A \cap B|}{|A \cup B|} \tag{9}$$

$$\text{Pixel Accuracy} = \frac{\text{Number of correctly classified pixels}}{\text{Total number of pixels}} \tag{10}$$

$$HD(A,B) = max\left\{ \sup_{a \in A} \inf_{b \in B} d(a,b), \sup_{b \in B} \inf_{a \in A} d\left(a,b\right) \right\} \tag{11}$$

$$\textit{Mean Squared Error (MSE)} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{12}$$

In Eq. (8), A is the set of predicted pixels, and B is the ground truth pixels. In Eq. (9), IoU measures the overlap between the predicted segmentation (A) and the ground truth (B). In Eq. (11), Hausdorff Distance measures the maximum distance between the predicted boundary and the ground truth boundary, where $d(a,b)$ is the Euclidean distance. For regression tasks involving image segmentation, MSE in Eq. (12) measures the average of the squared differences between predicted and actual pixel values.

9. Limitations of study

While this review comprehensively explores the role of DL in breast cancer histopathology, several limitations should be acknowledged as follows:

- Only peer-reviewed studies were included, excluding preprints and industry reports.
- The review considered only English-language studies, potentially omitting relevant non-English research.
- Many reviewed models rely on small, publicly available datasets that may not represent diverse populations.
- Variability in staining techniques, imaging resolutions, and annotation quality affects model robustness.
- Most AI models have not undergone large-scale clinical trials or regulatory approval.
- AI models require patient data, raising ethical and legal issues related to HIPAA and GDPR compliance.

10. Challenges in breast cancer histopathology image analysis

Deep learning (DL) has succeeded in breast cancer (BC) histopathology, but its application faces significant challenges in clinical settings. One major limitation is the scarcity of high-quality, annotated datasets for training robust models. Generating these datasets requires meticulous labeling by expert pathologists, which is time-consuming and expensive, further constrained by the limited availability of qualified professionals [161–164]. Variations in image quality, staining methods, and annotation standards also introduce inconsistencies that can negatively impact model performance and generalizability [161,165]. Another critical challenge is DL models' susceptibility to out-of-distribution (OOD) data. Models often excel on data resembling their

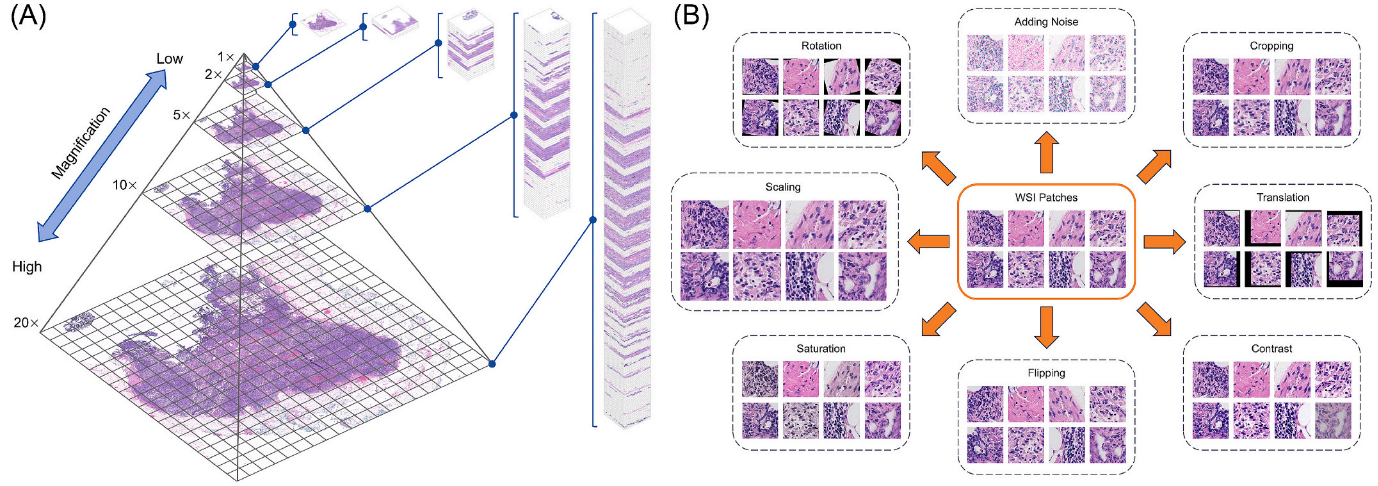

Fig. 15. WSI Pyramid representation and dataset augmentation: (A) WSI pyramid structure with multiple zoom levels, where images are tiled for efficient processing [158], (B) Dataset augmentation techniques applied to WSI patches.

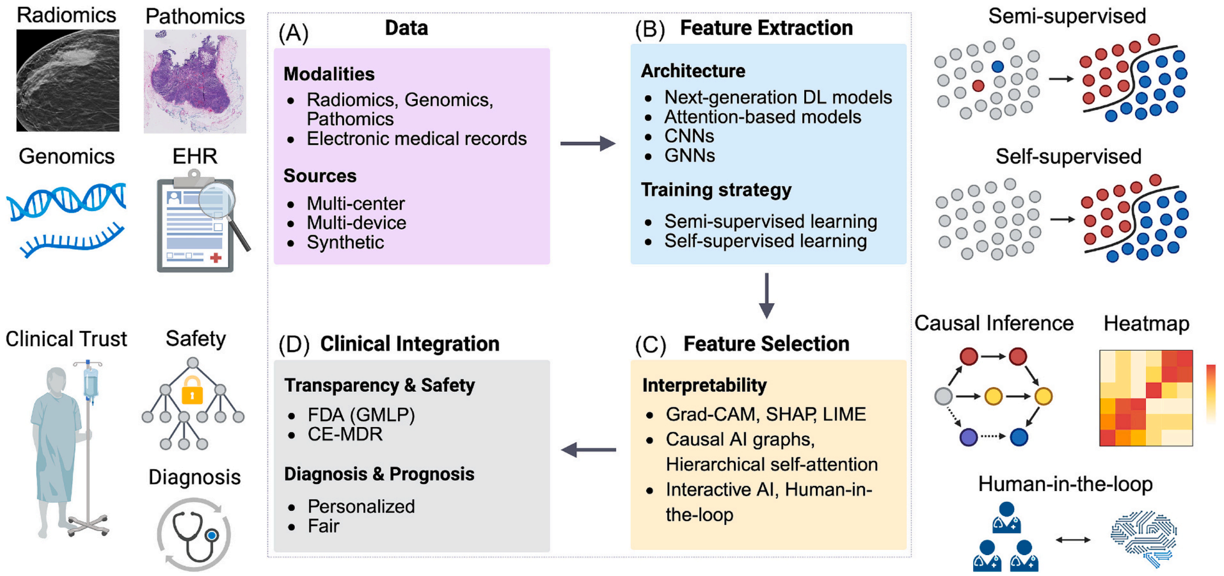

Fig. 16. A four-stage future directions roadmap: (A) gather diverse, multi-modal data; (B) apply advanced, semi- and self-supervised models for feature extraction; (C) use interpretable methods (e.g., Grad-CAM, SHAP, causal graphs) to select clinically salient features; and (D) integrate these insights under FDA/CE-MDR guidelines to enable fair, personalized diagnosis and prognosis.

training set but struggle with unfamiliar samples, leading to unreliable predictions. This is particularly problematic in safety-critical fields like medical diagnostics [166–168]. Developing uncertainty-aware models that recognize and communicate their limitations is vital to ensuring reliability and trustworthiness in clinical applications [169–171].

Class imbalance in medical datasets poses significant challenges, skewing predictions toward majority classes and undermining models' ability to generalize to minority or underrepresented groups. Addressing this imbalance is critical for improving the predictive accuracy of DL systems in clinical environments where accurate diagnosis is paramount [161,164]. Similarly, the opacity of DL models complicates their interpretability, making it challenging for clinicians to trust or adopt these tools without understanding their decision-making processes [172,173].

11. Future directions

Despite significant progress, deep learning (DL) in medical diagnostics requires further research to enhance robustness, interpretability, and integration for broader clinical applications.

11.1. Multimodal data integration and addressing dataset heterogeneity

Combining histopathological images with radiological scans, genomic data, and clinical information provides a comprehensive understanding of breast cancer, enhancing diagnostic precision and enabling personalized treatment strategies [34]. However, dataset heterogeneity due to variations in staining protocols, imaging resolutions, scanner types, and annotation quality across institutions remains a significant challenge. Strategies to mitigate these challenges include establishing standardized annotation protocols to ensure dataset consistency [124], applying domain adaptation methods like CycleGAN and adversarial networks to harmonize staining and imaging inconsistencies [174], utilizing federated learning to facilitate multi-institutional training without compromising data privacy, thus enhancing model robustness and generalization [175], and employing GAN-based synthetic data generation approaches to augment datasets and mitigate data imbalance [176].

11.2. Generalizable and robust models

Current DL models often face limitations in generalizability across diverse datasets and patient populations. Future research should focus on creating models that maintain high performance across various medical centers, imaging modalities, and demographic groups. Techniques such as cross-dataset training and domain adaptation are capable strategies to achieve this goal [177].

11.3. Enhancement of model interpretability

A significant barrier to the widespread clinical adoption of DL models in breast cancer (BC) histopathology is their lack of interpretability, commonly known as the "black-box" problem. Despite their high diagnostic accuracy, the opaque decision-making processes of DL models impede clinician trust and validation [178,179]. Explainable AI (XAI) techniques, such as Grad-CAM, SHapley Additive exPlanations (SHAP), Local Interpretable Model-Agnostic Explanations (LIME), and attention-based visual explanations, have been integrated into AI-driven histopathology workflows, enabling clinicians to visualize influential histopathological features underlying model predictions [179,180]. Recent advancements in causal inference-based XAI frameworks, including Causal AI Graphs and Hierarchical Self-Attention Mechanisms, have provided deeper insights into feature attribution, facilitating bias detection and enhancing model generalizability [179,181]. Interactive AI dashboards have been recommended to enhance interpretability, allowing pathologists real-time visualization, verification, and corrective feedback, thus improving clinical reliability and clinical usability of AI-driven models [182]. Human-in-the-loop approaches, combining AI outputs with pathologist-driven corrections, also improve accuracy by ensuring that AI predictions align closely with expert reasoning [183]. Regulatory frameworks such as the FDA's Good Machine Learning Practice (GMLP) guidelines and European CE-MDR regulations increasingly emphasize standardizing XAI methodologies to accelerate clinical adoption through enhanced transparency and safety [179,184,185]. Furthermore, recent advancements present the role of causal inference-based frameworks like Causal AI Graphs and Hierarchical Self-Attention Mechanisms in providing deeper insights into clinical reasoning and bias detection, further enhancing clinical trust and generalizability [179,184–186]. Future research should prioritize refining clinician-friendly and transparent XAI solutions to ensure DL augments rather than replaces expert decision-making in BC diagnostics.

11.4. Addressing data scarcity and quality

The effectiveness of DL models heavily depends on the availability of large, annotated datasets. Innovative data augmentation techniques, such as generative adversarial networks (GANs), can help mitigate data scarcity by generating synthetic yet realistic images for training purposes [187].

11.5. Exploration of novel deep learning architectures

Investigating and implementing advanced deep learning (DL) architectures, such as Vision Transformers and attention-based models, may offer improved performance in image analysis tasks. These architectures have shown the potential to capture complex patterns and contextual information within images [188]. Graph Neural Networks (GNNs) can model interactions among tissue elements to enhance tumor grading, molecular subtype prediction, and immune infiltration profiling in histopathology. Future work will utilize semi- and self-supervised learning to utilize vast amounts of unlabeled histopathology images, thereby enhancing model robustness and generalization across diverse staining protocols and scanners.

Fig. 16 illustrates a four-stage roadmap for future directions where stage (A) gathers diverse, multi-modal data; stage (B) applies advanced, semi- and self-supervised models for feature extraction; stage (C) uses interpretable methods (e.g., Grad-CAM, SHAP, causal graphs) to select clinically salient features; and stage (D) integrates these insights under FDA/CE-MDR guidelines to enable fair, personalized diagnosis and prognosis.

12. Clinical integration

Integrating deep learning (DL) models into clinical practice for breast cancer (BC) histopathology requires overcoming multiple real-world challenges. While AI models have demonstrated high accuracy in research settings, clinical adoption remains limited due to data heterogeneity, regulatory constraints, and workflow adaptation.

Enhancing model explainability through techniques like heatmaps, saliency maps, and feature attribution improves clinician trust by bridging the gap between AI predictions and human understanding [189]. However, trust alone is insufficient; models must undergo prospective validation through multi-center trials to assess their robustness across diverse populations and imaging protocols. AI must be tested in real-world histopathology settings, where variations in staining protocols, scanner resolutions, and patient demographics significantly impact performance [190]. Seamless workflow integration involves incorporating AI into whole-slide imaging systems and histopathology software through collaboration between digital histopathology providers and developers. This ensures that tools are user-friendly and do not disrupt existing workflows [191].

Furthermore, interoperability with electronic health records (EHRs) enables AI-assisted decision-making to be embedded in clinical practice, improving patient management and personalized treatment planning [192]. Regulatory approval remains a critical barrier. AI models must meet stringent FDA (GMLP) and CE-MDR standards, ensuring transparency, fairness, and robustness before being deployed in hospitals and diagnostic centers [193]. Prospective validation trials, real-world model auditing, and clinician-AI interaction frameworks are crucial for regulatory compliance and safe deployment. Addressing ethical concerns such as data privacy, algorithmic bias, and liability in AI-driven diagnostics is essential for widespread adoption [194]. Pathologists require specialized training via interactive modules and workshops to effectively interpret AI outputs and integrate them into diagnostics [195]. Continuous monitoring, feedback loops, and adaptation to

Table 8
Abbreviations and their definitions used in this study.

Abbreviations	Definition	Abbreviations	Definition
AI	Artificial Intelligence	HIPAA	Health Insurance Portability and Accountability Act
ANNs	Artificial Neural Networks	H&E	Hematoxylin and Eosin
BC	Breast Cancer	HRD	Homologous Recombination Deficiency
BRCA	Breast Cancer (specifically, Breast Invasive Carcinoma)	IDC	Invasive Ductal Carcinoma
BCCNN	Breast Cancer Convolutional Neural Network	LIME	Local Interpretable Model-agnostic Explanations
BCDR	Breast Cancer Digital Repository	ML	Machine Learning
BACH	BreAst Cancer Histology	MRI	Magnetic resonance Imaging
BMRI	Breast MRI	MAE	Mean Absolute Error
BUS	Breast Ultrasound	METABRIC	Molecular Taxonomy of Breast Cancer International Consortium
CAMELYON	Cancer Metastases in Lymph Nodes ChalleNge	OOD	out-of-distribution
CPTAC	Clinical Proteomic Tumor Analysis Consortium	PIK3CA	Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Alpha
CPTAC-BRCA	Clinical Proteomic Tumor Analysis Consortium`s dataset	PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
CAD	Computer-aided Diagnosis	PSP-Net	Pyramid Scene Parsing Network
CRF	Conditional Random Field	RF	Random Forest
CNNs	Convolutional Neural Networks	R-CNN	Region-based Convolutional Neural Network
DBNs	Deep Belief Networks	ROI	Region-of-Interest
DL	Deep Learning	ResNet	Residual Neural Network
DAEs	Denoising Autoencoders	RBM	Restricted Boltzmann Machine
DBT	Digital Breast Tomosynthesis	SHAP	SHapley Additive exPlanations
DM	Digital Mammography	SWE	Shear-wave elastography
XAI	Explainable AI	SAEs	Sparse Autoencoders
ELM	Extreme Learning Machines	SDAEs	Stacked Denoising Autoencoders
GANPIS	GAN-based Postcontrast Image Synthesis	SSVMs	Structured Support Vector Machines
GANs	Generative Adversarial Networks	SNOW	Synthetic Nuclei and Annotation Wizard
GMLP	Good Machine Learning Practice	TCGA	The Cancer Genome Atlas
Grad-CAM	Gradient-weighted Class Activation Mapping	OVCARE	Varian Cancer Research Centre
GNNs	Graph Neural Networks	VAEs	Variational Autoencoders
GuSA	Guided Soft Attention	WSIs	Whole-Slide Images

clinical settings ensure reliable performance while justifying AI biases and false positives. Incremental deployment in selected settings helps assess AI's impact on diagnostic accuracy and workflows, providing insights for optimization before broader adoption [196]. Furthermore, cost-benefit analyses are necessary to evaluate AI adoption's financial implications, including equipment costs, training investments, and maintenance expenses. These factors determine whether AI systems can provide long-term efficiency and sustainability in healthcare institutions [197]. Collaboration among pathologists, AI developers, regulatory bodies, and hospital administrators is essential to ensuring that AI-based histopathology tools are clinically viable, ethically sound, and aligned with patient care priorities [198].

13. Ethical considerations in AI-driven pathology

AI applications in pathology and laboratory medicine present opportunities and ethical challenges requiring careful oversight. Key ethical principles include patient autonomy, beneficence, non-maleficence, and justice, ensuring that AI-driven pathology systems respect patient rights, minimize harm, and provide equitable benefits [199].

i. Data Privacy and Security: AI systems rely on vast patient datasets, raising concerns about informed consent, data anonymization, and potential re-identification risks. Patients should have transparency and control over their medical data use [199].
ii. Bias and Fairness: AI models can inherit biases from historical medical data, leading to disparities in diagnosis and treatment recommendations. Ensuring diverse and representative datasets is crucial for equitable healthcare outcomes [199].
iii. Explainability and Trust: The black-box nature of many AI models complicates clinical adoption. To enhance transparency and clinician trust, AI models mustFig. integrate explainable AI (XAI) techniques [199].
iv. Clinical Accountability: While AI can assist pathologists, human experts' final diagnostic responsibility remains. AI-driven diagnoses should be well-documented, auditable, and subject to regulatory oversight [199].
v. Regulatory and Ethical Governance: Institutions and policymakers must establish formal guidelines for AI deployment in pathology, including performance validation, accountability mechanisms, and ethical safeguards to ensure responsible AI integration into clinical practice [199].

14. Definition of abbreviation used in the study

Table 8 contains the abbreviation and their full form used in this study.

15. Conclusion

This review systematically evaluates the advancements, challenges, and clinical implications of deep learning (DL) in breast cancer (BC) histopathology, synthesizing insights from 199 studies, including 182 peer-reviewed publications from 2014 to 2024 and 17 authoritative online sources. The study highlights how DL techniques, including Convolutional Neural Networks (CNNs), Generative Adversarial Networks (GANs), Transformers, and Transfer Learning, have significantly improved BC detection, classification, and segmentation. Nevertheless, several technical and clinical barriers hinder the widespread clinical adoption of DL models in histopathology.

To address these limitations, this review explores issues such as staining variability, dataset biases, and image resolution inconsistencies affecting model generalizability. Strategies like data augmentation (e.g., GAN-generated synthetic data) and domain adaptation have shown potential in enhancing robustness across diverse clinical datasets. Explainable AI (XAI) techniques, including Grad-CAM, SHAP, and LIME, improve model transparency and increase clinician trust in AI-driven workflows. From a methodological perspective, integrating multimodal data (histopathological, radiological, and genomic) holds significant potential for improving diagnostic precision and patient stratification. The review further emphasizes the need for rigorous validation using cross-validation and external datasets to ensure model reliability and reproducibility. Moreover, establishing standardized imaging protocols and curated benchmark datasets will facilitate fair model comparisons and accelerate regulatory approvals. In clinical practice, the deployment of DL models must align with evolving regulatory frameworks such as the FDA's Good Machine Learning Practice (GMLP) and the European CE-MDR. The review also underscores the importance of ensuring fairness and inclusivity in AI models to avoid biases in diagnostic performance across different demographic groups. Transparent training processes, bias mitigation strategies, and responsible AI development are crucial for building equitable and trustworthy clinical AI systems.

Therefore, this comprehensive review demonstrates that deep learning can improve diagnostic accuracy, early detection, and treatment planning in breast cancer histopathology. However, meaningful clinical integration demands collaborative efforts to expand diverse datasets, refine DL architectures, establish robust validation protocols, and ensure transparency and fairness.

Future Directions for Clinical Translation: To bridge the gap between research and clinical adoption, future efforts should focus on (1) validating models across multi-institutional and heterogeneous datasets, (2) embedding explainability into model design to support clinical decision-making, (3) ensuring compliance with regulatory standards, (4) using federated learning and privacy-preserving frameworks, and (5) developing user-centric tools that integrate seamlessly into digital pathology workflows. Interdisciplinary collaboration among AI researchers, clinicians, and regulatory bodies will translate DL innovations into reliable, safe, and impactful clinical solutions.

CRediT authorship contribution statement

Inayatul Haq: Writing – original draft, Methodology, Formal analysis, Conceptualization. **Zheng Gong:** Resources, Formal analysis, Data curation. **Haomin Liang:** Resources, Formal analysis, Data curation. **Wei Zhang:** Resources, Formal analysis, Data curation. **Rashid Khan:** Validation, Resources, Formal analysis. **Lei Gu:** Resources, Formal analysis, Data curation. **Roland Eils:** Visualization, Validation, Resources, Methodology, Data curation. **Yan Kang:** Writing – review & editing, Supervision, Project administration. **Bingding Huang:** Writing – review & editing, Supervision, Software, Project administration, Methodology, Funding acquisition.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used Grammarly to improve language and readability. The authors also used MS Visio, BioRender, and Canva to create diagrams. After using these tools, the authors reviewed and edited the content as needed and took full responsibility for the publication's content.

Declaration of competing interest

The authors declare no conflict of interest.

Acknowledgment

This study was supported by the Project of the Educational Commission of Guangdong Province, China (No. 2022ZDJS113), the Shenzhen Science and Technology Program (No.

KJZD20240903095605007), and the Higher Education Stability Support Program General Project (No. 20220715114836001).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.imavis.2025.105708.

Data availability

No data was used for the research described in the article.

References

[1] Organization, W.H, Breast Cancer [cited 2025 28 February]; Available from, https://www.who.int/news-room/fact-sheets/detail/breast-cancer?utm_source=chatgpt.com, 2024.
[2] Foundation, B.C.R, Breast Cancer Statistics And Resources [cited 2025 28 February]; Available from, https://www.bcrf.org/breast-cancer-statistics-and-resources/?utm_source=chatgpt.com, 2022.
[3] Society, A.C, Breast Cancer Statistics [cited 2025 28 February]; Available from, https://www.cancer.org/cancer/types/breast-cancer/about/how-common-is-breast-cancer.html?utm_source=chatgpt.com, 2025.
[4] Foundation, N.B.C, Breast Cancer Facts & Stats [cited 2025 28 February]; Available from, https://www.nationalbreastcancer.org/breast-cancer-facts/?utm_source=chatgpt.com, 2025.
[5] R.L. Siegel, A.N. Giaquinto, A. Jemal, Cancer statistics, 2024, CA 74 (1) (2024) 12–49.
[6] F. Bray, et al., Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA 68 (6) (2018) 394–424.
[7] Cancer, I.A.f.R.o, Absolute Numbers, Mortality, Both Sexes, in 2022_Breast_Continents [cited 2024 27 November 2024]; Mortality caused by breast cancer] Available from, https://gco.iarc.who.int/en, 2022.
[8] Organization, W.H Breats Cancer 2022 [cited 2025 28 February]; Available from https://www.who.int/news-room/fact-sheets/detail/breast-cancer?gad_source=1&gclid=EAIaIQobChMI1ou75cHFiAMVqJpoCR3Djzg9EAAYASAAEgIbNfD_BwE.
[9] G. Forrai, et al., Use of diagnostic imaging modalities in modern screening, diagnostics and management of breast tumours 1st central-eastern European professional consensus statement on breast cancer, Pathol. Oncol. Res. 28 (2022) 1610382.
[10] P. Mathur, et al., Cancer statistics, 2020: report from national cancer registry programme, India, JCO Glob. Oncol. 6 (2020) 1063–1075.
[11] R. Mehrotra, K. Yadav, Breast cancer in India: Present scenario and the challenges ahead, World J. Clin. Oncol. 13 (3) (2022) 209.
[12] C.H. Barrios, Global challenges in breast cancer detection and treatment, Breast 62 (2022) S3–S6.
[13] R.A. Dar, M. Rasool, A. Assad, Breast cancer detection using deep learning: datasets, methods, and challenges ahead, Comput. Biol. Med. 149 (2022) 106073.
[14] B. Ehteshami Bejnordi, V.M. Johannes, P. van Diest, B. van Ginneken, N. Karssemeijer, G. Litjens, van der Laak, The CAMELYON16 Challenge Has Ended in November 2016, 2016.
[15] P.H. Tan, et al., The 2019 WHO classification of tumours of the breast, Histopathology 77 (2) (2020).
[16] L. Lu, B.J. Daigle Jr., Prognostic analysis of histopathological images using pre-trained convolutional neural networks: application to hepatocellular carcinoma, PeerJ 8 (2020) e8668.
[17] S. Wang, et al., Pathology image analysis using segmentation deep learning algorithms, Am. J. Pathol. 189 (9) (2019) 1686–1698.
[18] M. Veta, et al., Breast cancer histopathology image analysis: a review, IEEE Trans. Biomed. Eng. 61 (5) (2014) 1400–1411.
[19] A.H. Song, et al., Artificial intelligence for digital and computational pathology, Nat. Rev. Bioeng. 1 (12) (2023) 930–949.
[20] M. Mubarak, Move from traditional histopathology to digital and computational pathology: are we ready? Indian J. Nephrol. 32 (5) (2022) 414–415.
[21] S. Schutte, J. Uddin, Deep segmentation techniques for breast cancer diagnosis, BioMedInformatics 4 (2) (2024) 921–945.
[22] G. Litjens, et al., A survey on deep learning in medical image analysis, Med. Image Anal. 42 (2017) 60–88.
[23] R. Aggarwal, et al., Diagnostic accuracy of deep learning in medical imaging: a systematic review and meta-analysis, NPJ Digit. Med. 4 (1) (2021) 65.
[24] D. Bardou, K. Zhang, S.M. Ahmad, Classification of breast cancer based on histology images using convolutional neural networks, IEEE Access 6 (2018) 24680–24693.
[25] P. Khosravi, et al., Deep convolutional neural networks enable discrimination of heterogeneous digital pathology images, EBioMedicine 27 (2018) 317–328.
[26] T. Araújo, et al., Classification of breast cancer histology images using convolutional neural networks, PLoS ONE 12 (6) (2017) e0177544.
[27] M. Mullooly, et al., Application of convolutional neural networks to breast biopsies to delineate tissue correlates of mammographic breast density, NPJ Breast Cancer 5 (1) (2019) 43.
[28] M. Gour, S. Jain, T. Sunil Kumar, Residual learning based CNN for breast cancer histopathological image classification, Int. J. Imaging Syst. Technol. 30 (3) (2020) 621–635.
[29] F.M. Howard, et al., Generative adversarial networks accurately reconstruct pan-cancer histology from pathologic, genomic, and radiographic latent features, Sci. Adv. 10 (46) (2024) eadq0856.
[30] A. Das, V.K. Devarampati, M.S. Nair, NAS-SGAN: a semi-supervised generative adversarial network model for atypia scoring of breast cancer histopathological images, IEEE J. Biomed. Health Inform. 26 (5) (2021) 2276–2287.
[31] S.H. Gheshlaghi, C.N.E. Kan, D.H. Ye, Breast cancer histopathological image classification with adversarial image synthesis, in: 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), IEEE, 2021.
[32] A.B. Levine, et al., Synthesis of diagnostic quality cancer pathology images by generative adversarial networks, J. Pathol. 252 (2) (2020) 178–188.
[33] D. Wang, Z. Chen, H. Zhao, Prototype transfer generative adversarial network for unsupervised breast cancer histology image classification, Biomed. Signal Process. Control 68 (2021) 102713.
[34] B. Jiang, et al., Deep learning applications in breast cancer histopathological imaging: diagnosis, treatment, and prognosis, Breast Cancer Res. 26 (1) (2024) 137.
[35] C. Guttà, C. Morhard, M. Rehm, Applying a GAN-based classifier to improve transcriptome-based prognostication in breast cancer, PLoS Comput. Biol. 19 (4) (2023) e1011035.
[36] A.M. Al Nemer, Application of artificial intelligence in the field of breast pathology diagnosis: narrative review, J. Med. Artif. Intell. (2024) 7.
[37] A. Choudhury, S. Perumalla, Detecting breast cancer using artificial intelligence: convolutional neural network, Technol. Health Care 29 (1) (2021) 33–43.
[38] F. Abdullakutty, et al., Histopathology in focus: a review on explainable multi-modal approaches for breast cancer diagnosis, Front. Med. 11 (2024) 1450103.
[39] S. Krishna, et al., An interpretable decision-support model for breast cancer diagnosis using histopathology images, J. Pathol. Inform. 14 (2023) 100319.
[40] T.K. Murugan, P. Karthikeyan, P. Sekar, Efficient breast cancer detection using neural networks and explainable artificial intelligence, Neural Comput. & Applic. (2024) 1–18.
[41] S. Rajpal, et al., XAI-MethylMarker: explainable AI approach for biomarker discovery for breast cancer subtype classification using methylation data, Expert Syst. Appl. 225 (2023) 120130.
[42] J. Peta, S. Koppu, Explainable Soft Attentive EfficientNet for breast cancer classification in histopathological images, Biomed. Signal Process. Control 90 (2024) 105828.
[43] H. Yang, et al., Guided soft attention network for classification of breast cancer histopathology images, IEEE Trans. Med. Imaging 39 (5) (2019) 1306–1315.
[44] R. Jiang, et al., A transformer-based weakly supervised computational pathology method for clinical-grade diagnosis and molecular marker discovery of gliomas, Nat. Mach. Intell. 6 (8) (2024) 876–891.
[45] S. Boumaraf, et al., A new transfer learning based approach to magnification dependent and independent classification of breast cancer in histopathological images, Biomed. Signal Process. Control 63 (2021) 102192.
[46] M.S. Hossain, et al., Region of interest (ROI) selection using vision transformer for automatic analysis using whole slide images, Sci. Rep. 13 (1) (2023) 11314.
[47] M. Goyal, et al., A multi-model approach integrating whole-slide imaging and clinicopathologic features to predict breast cancer recurrence risk, NPJ Breast Cancer 10 (1) (2024) 93.
[48] D. Ferber, et al., In-context learning enables multimodal large language models to classify cancer pathology images, Nat. Commun. 15 (1) (2024) 10104.
[49] H.H. Luong, et al., Principal component analysis and fine-tuned vision transformation integrating model explainability for breast cancer prediction, Vis. Comput. Indus. Biomed. Art 8 (1) (2025) 5.
[50] J. Potsangbam, S.S. Devi, EMViT-BCC: enhanced mobile vision transformer for breast cancer classification, Int. J. Imaging Syst. Technol. 35 (2) (2025) e70053.
[51] M.L. Abimouloud, et al., Vision transformer based convolutional neural network for breast cancer histopathological images classification, Multimed. Tools Appl. 83 (39) (2024) 86833–86868.
[52] M. Sabry, et al., Enhancing breast cancer diagnosis with multi-resolution vision transformers and robust decision-making, IEEE Access 13 (2025) 89704–89722.
[53] M. Sabry, et al., A vision transformer approach for breast cancer classification in histopathology, in: 2024 IEEE International Symposium on Biomedical Imaging (ISBI), IEEE, 2024.
[54] R.H.M. Alkhater, et al., Breast cancer detection based on histopathological images using vision transformers, in: 2024 International Conference on Computer and Applications (ICCA), IEEE, 2024.
[55] E. Monjezi, G. Akbarizadeh, K. Ansari-Asl, RI-ViT: a multi-scale hybrid method based on vision transformer for breast cancer detection in histopathological images, IEEE Access 12 (2024) 186074.
[56] G.L. Baroni, et al., Optimizing vision transformers for histopathology: pretraining and normalization in breast cancer classification, J. Imaging 10 (5) (2024) 108.
[57] I. Jahan, et al., Deep learning and vision transformers-based framework for breast cancer and subtype identification, Neural Comput. & Applic. (2025) 1–20.
[58] Z. Hameed, et al., Breast cancer histopathology image classification using an ensemble of deep learning models, Sensors 20 (16) (2020) 4373.
[59] M.J. Page, et al., The PRISMA 2020 statement: an updated guideline for reporting systematic reviews, BMJ (2021) 372.
[60] L. Tran, et al., Quality assessment tools used in systematic reviews of in vitro studies: a systematic review, BMC Med. Res. Methodol. 21 (1) (2021) 101.

[61] R. Sutjiadi, et al., Deep learning for segmentation and classification in mammograms for breast cancer detection: a systematic literature review, Adv. Ultrasound Diagn. Ther. 8 (3) (2024) 94–105.
[62] M.M. Abdelsamea, et al., A survey on artificial intelligence in histopathology image analysis, Wiley Interdisc. Rev. Data Mining Knowl. Discov. 12 (6) (2022) e1474.
[63] C. Li, et al., A survey for breast histopathology image analysis using classical and deep neural networks, in: Information Technology in Biomedicine, Springer, 2019.
[64] L. Dihge, et al., Artificial neural network models to predict nodal status in clinically node-negative breast cancer, BMC Cancer 19 (2019) 1–12.
[65] Y.A. Hamad, K. Simonov, M.B. Naeem, Breast cancer detection and classification using artificial neural networks, in: 2018 1st Annual International Conference on Information and Sciences (AiCIS), IEEE, 2018.
[66] R. Jafari-Marandi, et al., An optimum ANN-based breast cancer diagnosis: bridging gaps between ANN learning and decision-making goals, Appl. Soft Comput. 72 (2018) 108–120.
[67] M.F. Mridha, et al., A comprehensive survey on deep-learning-based breast cancer diagnosis, Cancers 13 (23) (2021) 6116.
[68] M.D. Tissera, M.D. McDonnell, Deep extreme learning machines: supervised autoencoding architecture for classification, Neurocomputing 174 (2016) 42–49.
[69] K. Berahmand, et al., Autoencoders and their applications in machine learning: a survey, Artif. Intell. Rev. 57 (2) (2024) 28.
[70] A. Radhakrishnan, et al., Cross-modal autoencoder framework learns holistic representations of cardiovascular state, Nat. Commun. 14 (1) (2023) 2436.
[71] L. Hou, et al., Sparse autoencoder for unsupervised nucleus detection and representation in histopathology images, Pattern Recogn. 86 (2019) 188–200.
[72] V.J. Kadam, S.M. Jadhav, K. Vijayakumar, Breast cancer diagnosis using feature ensemble learning based on stacked sparse autoencoders and softmax regression, J. Med. Syst. 43 (8) (2019) 263.
[73] Y. Feng, L. Zhang, Z. Yi, Breast cancer cell nuclei classification in histopathology images using deep neural networks, Int. J. Comput. Assist. Radiol. Surg. 13 (2018) 179–191.
[74] J. Xu, et al., Stacked Sparse Autoencoder (SSAE) based framework for nuclei patch classification on breast cancer histopathology, in: 2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI), IEEE, 2014.
[75] J. Xu, et al., Stacked sparse autoencoder (SSAE) for nuclei detection on breast cancer histopathology images, IEEE Trans. Med. Imaging 35 (1) (2015) 119–130.
[76] M. Toğaçar, B. Ergen, Z. Cömert, Application of breast cancer diagnosis based on a combination of convolutional neural networks, ridge regression and linear discriminant analysis using invasive breast cancer images processed with autoencoders, Med. Hypotheses 135 (2020) 109503.
[77] D. Zhang, et al., Integrating feature selection and feature extraction methods with deep learning to predict clinical outcome of breast cancer, IEEE Access 6 (2018) 28936–28944.
[78] S. Archana, Detection of breast cancer by deep belief network with improved activation function, Int. J. Adapt. Control Signal Process. 38 (9) (2024) 3074–3101.
[79] R. Karthiga, et al., Automatic approach for breast cancer detection based on deep belief network using histopathology images, Multimed. Tools Appl. (2024) 1–18.
[80] A.M. Abdel-Zaher, A.M. Eldeib, Breast cancer classification using deep belief networks, Expert Syst. Appl. 46 (2016) 139–144.
[81] M. Khademi, N.S. Nedialkov, Probabilistic graphical models and deep belief networks for prognosis of breast cancer, in: 2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA), IEEE, 2015.
[82] N. Dhungel, G. Carneiro, A.P. Bradley, Deep learning and structured prediction for the segmentation of mass in mammograms, in: International Conference on Medical image Computing and Computer-Assisted Intervention, Springer, 2015.
[83] P. Singh, et al., A critical analysis and classification of breast cancer using histopathology images, in: 2024 International Conference on Knowledge Engineering and Communication Systems (ICKECS), IEEE, 2024.
[84] N. Hatipoglu, G. Bilgin, Cell segmentation in histopathological images with deep learning algorithms by utilizing spatial relationships, Med. Biol. Eng. Comput. 55 (2017) 1829–1848.
[85] F.A. Spanhol, et al., Breast cancer histopathological image classification using convolutional neural networks, in: 2016 International Joint Conference on Neural Networks (IJCNN), IEEE, 2016.
[86] J. Xie, et al., Deep learning based analysis of histopathological images of breast cancer, Front. Genet. 10 (2019) 80.
[87] S. Majumdar, P. Pramanik, R. Sarkar, Gamma function based ensemble of CNN models for breast cancer detection in histopathology images, Expert Syst. Appl. 213 (2023) 119022.
[88] D. Albashish, et al., Deep CNN model based on VGG16 for breast cancer classification, in: 2021 International Conference on Information Technology (ICIT), IEEE, 2021.
[89] C. Hu, et al., Classification of breast cancer histopathological image with deep residual learning, Int. J. Imag. Syst. 31 (3) (2021) 1583–1594.
[90] M.M. Srikantamurthy, et al., Classification of benign and malignant subtypes of breast cancer histopathology imaging using hybrid CNN-LSTM based transfer learning, BMC Med. Imaging 23 (1) (2023) 19.
[91] C. Zhu, et al., Breast cancer histopathology image classification through assembling multiple compact CNNs, BMC Med. Inform. Decis. Making 19 (2019) 1–17.
[92] G. Wadhwa, A. Kaur, A deep cnn technique for detection of breast cancer using histopathology images, in: 2020 Advanced Computing and Communication Technologies for High Performance Applications (ACCTHPA), IEEE, 2020.
[93] M. Toğaçar, et al., BreastNet: a novel convolutional neural network model through histopathological images for the diagnosis of breast cancer, Phys. A Stat. Mech. Appl. 545 (2020) 123592.
[94] S. Albarqouni, et al., Aggnet: deep learning from crowds for mitosis detection in breast cancer histology images, IEEE Trans. Med. Imaging 35 (5) (2016) 1313–1321.
[95] A. Sohail, et al., A multi-phase deep CNN based mitosis detection framework for breast cancer histopathological images, Sci. Rep. 11 (1) (2021) 6215.
[96] S. Dabeer, M.M. Khan, S. Islam, Cancer diagnosis in histopathological image: CNN based approach, Inform. Med. Unlock. 16 (2019) 100231.
[97] V. Lahoura, et al., Cloud computing-based framework for breast cancer diagnosis using extreme learning machine, Diagnostics 11 (2) (2021) 241.
[98] A. Toprak, Extreme learning machine (elm)-based classification of benign and malignant cells in breast cancer, Med. Sci. Monit. 24 (2018) 6537.
[99] M. Nemissi, H. Salah, H. Seridi, Breast cancer diagnosis using an enhanced extreme learning machine based-neural network, in: 2018 International Conference on Signal, Image, Vision and Their Applications (SIVA), IEEE, 2018.
[100] Z. Wang, et al., Breast cancer detection using extreme learning machine based on feature fusion with CNN deep features, IEEE Access 7 (2019) 105146–105158.
[101] P. Wang, et al., Cross-task extreme learning machine for breast cancer image classification with deep convolutional features, Biomed. Signal Process. Control 57 (2020) 101789.
[102] S. Ronoud, S. Asadi, An evolutionary deep belief network extreme learning-based for breast cancer diagnosis, Soft. Comput. 23 (24) (2019) 13139–13159.
[103] D. Muduli, R. Dash, B. Majhi, Automated breast cancer detection in digital mammograms: a moth flame optimization based ELM approach, Biomed. Signal Process. Control 59 (2020) 101912.
[104] A. Juhong, et al., Super-resolution and segmentation deep learning for breast cancer histopathology image analysis, Biomed. Opt. Exp. 14 (1) (2022) 18–36.
[105] G.D. Zion, B. Tripathy, Amalgamation of GAN and ResNet methods in accurate detection of breast cancer with histopathological images, Int. J. Adv. Soft Comput. Appl. 16 (2) (2024).
[106] S. Ghose, et al., Predicting breast Cancer events in ductal carcinoma in situ (DCIS) using generative adversarial network augmented deep learning model, Cancers 15 (7) (2023) 1922.
[107] F. Shahidi, Breast cancer histopathology image super-resolution using wide-attention gan with improved wasserstein gradient penalty and perceptual loss, IEEE Access 9 (2021) 32795–32809.
[108] M. Ryspayeva, Generative adversarial network as data balance and augmentation tool in histopathology of breast cancer, in: 2023 IEEE International Conference on Smart Information Systems and Technologies (SIST), IEEE, 2023.
[109] E. Strelcenia, S. Prakoonwit, Improving cancer detection classification performance using GANs in breast cancer data, IEEE Access 11 (2023) 71594–71615.
[110] R. Man, P. Yang, B. Xu, Classification of breast cancer histopathological images using discriminative patches screened by generative adversarial networks, IEEE Access 8 (2020) 155362–155377.
[111] S. Razavi, et al., MiNuGAN: Dual segmentation of mitoses and nuclei using conditional GANs on multi-center breast H&E images, J. Pathol. Inform. 13 (2022) 100002.
[112] B. Kannappan, et al., Data augmentation guided breast tumor segmentation based on generative adversarial neural networks, Eng. Appl. Artif. Intell. 125 (2023) 106753.
[113] Y. Wang, et al., Clinical evaluation of deep learning-based risk profiling in breast cancer histopathology and comparison to an established multigene assay, Breast Cancer Res. Treat. 206 (1) (2024) 163–175.
[114] Z. Hameed, et al., Multiclass classification of breast cancer histopathology images using multilevel features of deep convolutional neural network, Sci. Rep. 12 (1) (2022) 15600.
[115] H. El Agouri, et al., Assessment of deep learning algorithms to predict histopathological diagnosis of breast cancer: first Moroccan prospective study on a private dataset, BMC Res. Notes 15 (1) (2022) 66.
[116] Institute, N.C, TCGA-BRCA, Available from, https://portal.gdc.cancer.gov/projects/TCGA-BRCA, 2012.
[117] F. Capron, MITOS-ATYPIA-14: Detection of Mitosis and Evaluation of Nuclear Atypia on Breast Cancer H&E Stained Images, h.b.t.I.C.o.P.R.I. MITOS & ATYPIA 14 Contest, Editor, Paris, France, 2014.
[118] A. Rocha, L.O. Guevara, T.J. de Carvalho, et al., Breast Cancer Histopathological Database (BreakHis), F.U.o.P.U. Vision and Robotics Laboratory, Editor, UFPR - Vision and Robotics Laboratory: Brazil, 2016.
[119] B.E. Bejnordi, et al., Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer, JAMA 318 (22) (2017) 2199–2210.
[120] M. Veta, et al., Predicting breast tumor proliferation from whole-slide images: the TUPAC16 challenge, Med. Image Anal. 54 (2019) 111–121.
[121] P. Bandi, et al., From detection of individual metastases to classification of lymph node status at the patient level: the camelyon17 challenge, IEEE Trans. Med. Imaging 38 (2) (2018) 550–560.
[122] G. Challenge, CAMELYON17, cited 2024 1 December Available from, https://camelyon17.grand-challenge.org/, 2017.
[123] P. Mooney, Breast Histopathology Images (IDC Dataset), 2017.
[124] G. Aresta, et al., Bach: grand challenge on breast cancer histology images, Med. Image Anal. 56 (2019) 122–139.
[125] A. Polónia, Catarina Eloy, Paulo Aguiar, BACH Dataset : Grand Challenge on Breast Cancer Histology images, Zenodo, Portugal, 2019.

[126] H.C. Detection, Histopathologic Cancer Detection, 2018. Online platform (Kaggle).
[127] D.J.D. Alper Aksac, Tansel Özyer, Reda Alhajj, BreCaHAD: A Dataset for Breast Cancer Histopathological Annotation and Diagnosis, 2019.
[128] M.F.I. Soumik, A.Z.B. Aziz, M.A. Hossain, Improved transfer learning based deep learning model for breast cancer histopathological image classification, in: 2021 International Conference on Automation, Control and Mechatronics for Industry 4.0 (ACMI), IEEE, 2021.
[129] H. Zerouaoui, A. Idri, Deep hybrid architectures for binary classification of medical breast cancer images, Biomed. Signal Process. Control 71 (2022) 103226.
[130] S. Chattopadhyay, et al., DRDA-Net: Dense residual dual-shuffle attention network for breast cancer classification using histopathological images, Comput. Biol. Med. 145 (2022) 105437.
[131] K. Ding, et al., A large-scale synthetic pathological dataset for deep learning-enabled segmentation of breast cancer, Sci Data 10 (1) (2023) 231.
[132] T. Huang, H. Yin, X. Huang, Deep learning and multiscale analysis for epithelial-mesenchyme segmentation and classification in breast cancer histological images, SIViP 18 (11) (2024) 7741–7754.
[133] K. Sun, et al., Semi-supervised breast cancer pathology image segmentation based on fine-grained classification guidance, Med. Biol. Eng. Comput. 62 (3) (2024) 901–912.
[134] O. Ronneberger, P. Fischer, T. Brox, U-net: convolutional networks for biomedical image segmentation, in: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, part III 18, Springer, 2015.
[135] Z. Zhou, et al., Unet++: A nested u-net architecture for medical image segmentation, in: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4, Springer, 2018.
[136] T. Fan, et al., Ma-net: A multi-scale attention network for liver and tumor segmentation, IEEE Access 8 (2020) 179656–179665.
[137] M.A. Ortega-Ruíz, et al., DRD-UNet, a UNet-like architecture for multi-class breast cancer semantic segmentation, IEEE Access 12 (2024) 40412–40424.
[138] G. Jiménez, D. Racoceanu, Deep learning for semantic segmentation vs. classification in computational pathology: application to mitosis analysis in breast cancer grading, Front. Bioeng. Biotechnol. 7 (2019) 145.
[139] A.A. Balasubramanian, et al., Ensemble deep learning-based image classification for breast cancer subtype and invasiveness diagnosis from whole slide image histopathology, Cancers 16 (12) (2024) 2222.
[140] T. Liu, et al., A hybrid deep learning model for predicting molecular subtypes of human breast cancer using multimodal data, IRBM 43 (1) (2022) 62–74.
[141] N. Aparicio, I.D. Chacón, D. Ruiz, CNN for breast cancer metastases classification, in: 2021 IEEE 2nd International Congress of Biomedical Engineering and Bioengineering (CI-IB&BI), 2021.
[142] Y.-G. Kim, et al., Effectiveness of transfer learning for enhancing tumor classification with a convolutional neural network on frozen sections, Sci. Rep. 10 (1) (2020) 21899.
[143] M. Liu, et al., Breast histopathological image classification method based on autoencoder and siamese framework, Information 13 (3) (2022) 107.
[144] M.A. Wakili, et al., Classification of breast cancer histopathological images using DenseNet and transfer learning, Comput. Intell. Neurosci. 2022 (1) (2022) 8904768.
[145] M. Liu, et al., A deep learning method for breast cancer classification in the pathology images, IEEE J. Biomed. Health Inform. 26 (10) (2022) 5025–5032.
[146] I. Hirra, et al., Breast cancer classification from histopathological images using patch-based deep learning modeling, IEEE Access 9 (2021) 24273–24287.
[147] A. Riasatian, et al., Fine-tuning and training of densenet for histopathology image representation using tcga diagnostic slides, Med. Image Anal. 70 (2021) 102032.
[148] N.N. Phan, et al., Predicting breast cancer gene expression signature by applying deep convolutional neural networks from unannotated pathological images, Front. Oncol. 11 (2021) 769447.
[149] Y. Ektefaie, et al., Integrative multiomics-histopathology analysis for breast cancer classification, NPJ Breast Cancer 7 (1) (2021) 147.
[150] O. Ciga, T. Xu, A.L. Martel, Self supervised contrastive learning for digital histopathology, Mach. Learn. Appl. 7 (2022) 100198.
[151] M. Mohammed, et al., A stacking ensemble deep learning approach to cancer type classification based on TCGA data, Sci. Rep. 11 (1) (2021) 15626.
[152] R.K. Mondol, et al., hist2rna: an efficient deep learning architecture to predict gene expression from breast cancer histopathology images, Cancers 15 (9) (2023) 2569.
[153] Radboud University Medical Center U.M.C.U, CAMELYON16: Grand Challenge on Cancer Metastasis Detection, Radboud University Medical Center, Netherlands, 2016.
[154] R.U.M.C.U. Utrecht, CAMELYON17: Grand Challenge on Lymph Node Metastasis Detection in Breast Cancer, U.U. Radboud University Medical Center, Editor, CAMELYON Grand Challenge Platform: Netherlands, 2017.
[155] C. Shorten, T.M. Khoshgoftaar, A survey on image data augmentation for deep learning, J. Big Data 6 (1) (2019) 1–48.
[156] A.J. Plompen, et al., The joint evaluated fission and fusion nuclear data library, JEFF-3.3, Eur. Phys. J. A 56 (2020) 1–108.
[157] S. Kumar, et al., Deep learning-based computer-aided diagnosis model for the identification and classification of mammography images, SN Comput. Sci. 4 (5) (2023) 502.
[158] D.S. McClintock, J.T. Abel, T.C. Cornish, Whole slide imaging hardware, software, and infrastructure, in: Whole Slide Imaging: Current Applications and Future Directions, Springer, 2021, pp. 23–56.
[159] M. Owusu-Adjei, et al., Imbalanced class distribution and performance evaluation metrics: a systematic review of prediction accuracy for determining model performance in healthcare systems, PLOS Digital Health 2 (11) (2023) e0000290.
[160] A.A. Taha, A. Hanbury, Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool, BMC Med. Imaging 15 (2015) 1–28.
[161] M. Salmi, et al., Handling imbalanced medical datasets: review of a decade of research, Artif. Intell. Rev. 57 (10) (2024) 273.
[162] R. Alizadehsani, et al., Model uncertainty quantification for diagnosis of each main coronary artery stenosis, Soft. Comput. 24 (2020) 10149–10160.
[163] R. Alizadehsani, et al., Hybrid genetic-discretized algorithm to handle data uncertainty in diagnosing stenosis of coronary arteries, Expert. Syst. 39 (7) (2022) e12573.
[164] J.H. Joloudari, et al., Application of artificial intelligence techniques for automated detection of myocardial infarction: a review, Physiol. Meas. 43 (8) (2022) 08TR01.
[165] K. Kakhi, et al., The internet of medical things and artificial intelligence: trends, challenges, and opportunities, Biocybern. Biomed. Eng. 42 (3) (2022) 749–771.
[166] R. Alizadehsani, et al., Handling of uncertainty in medical data using machine learning and probability theory techniques: A review of 30 years (1991–2020), Ann. Oper. Res. (2021) 1–42.
[167] N. Ayoobi, et al., Time series forecasting of new cases and new deaths rate for COVID-19 using deep learning methods, Results Phys 27 (2021) 104495.
[168] H. Asgharnezhad, et al., Objective evaluation of deep uncertainty predictions for covid-19 detection, Sci. Rep. 12 (1) (2022) 815.
[169] J. Liu, et al., Simple and principled uncertainty estimation with deterministic deep learning via distance awareness, Adv. Neural Inf. Proces. Syst. 33 (2020) 7498–7512.
[170] K. Chua, et al., Deep reinforcement learning in a handful of trials using probabilistic dynamics models, Adv. Neural Inf. Proces. Syst. 31 (2018).
[171] K. Wickstrøm, et al., Uncertainty-aware deep ensembles for reliable and explainable predictions of clinical time series, IEEE J. Biomed. Health Inform. 25 (7) (2020) 2435–2444.
[172] M. Ilse, J. Tomczak, M. Welling, Attention-based deep multiple instance learning, in: International Conference on Machine Learning, PMLR, 2018.
[173] M. Xin, et al., miR-22 inhibits tumor growth and metastasis by targeting ATP citrate lyase: evidence in osteosarcoma, prostate cancer, cervical cancer and lung cancer, Oncotarget 7 (28) (2016) 44252.
[174] M.W. Lafarge, et al., Domain-adversarial neural networks to address the appearance variability of histopathology images, in: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: Third International Workshop, DLMIA 2017, and 7th International Workshop, ML-CDS 2017, Held in Conjunction with MICCAI 2017, Québec City, QC, Canada, September 14, Proceedings 3, Springer, 2017.
[175] N. Rieke, et al., The future of digital health with federated learning, NPJ Digit. Med. 3 (1) (2020) 119.
[176] W. Dee, R. Alaaeldin Ibrahim, E. Marouli, Histopathological domain adaptation with generative adversarial networks: bridging the domain gap between thyroid cancer histopathology datasets, PLoS ONE 19 (12) (2024) e0310417.
[177] L. Luo, et al., Deep learning in breast cancer imaging: a decade of progress and future directions, IEEE Rev. Biomed. Eng. 18 (2024) 130–151.
[178] W. Samek, et al., Explainable AI: Interpreting, Explaining and Visualizing Deep Learning Vol. 11700, Springer Nature, 2019.
[179] A. Carriero, et al., Deep learning in breast cancer imaging: state of the art and recent advancements in early 2024, Diagnostics 14 (8) (2024) 848.
[180] A. Holzinger, et al., Causability and explainability of artificial intelligence in medicine, Wiley Interdisc. Rev. Data Mining Knowl. Discov. 9 (4) (2019) e1312.
[181] E. Tjoa, C. Guan, A survey on explainable artificial intelligence (xai): toward medical xai, IEEE Trans. Neural Netw. Learn. Syst. 32 (11) (2020) 4793–4813.
[182] R. Caruana, et al., Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission, in: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2015.
[183] S. Kumar, et al., Applications, challenges, and future directions of human-in-the-loop learning, IEEE Access 12 (2024) 75735–75760.
[184] (FDA), U.S.F.a.D.A, Artificial Intelligence/Machine Learning-Based Software as a Medical Device (SaMD) Action Plan [cited 2025 4 March]; FDA White Paper on AI/ML in medical devices]. Available from, https://www.fda.gov/media/153486/download, 2021.
[185] S.M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, Adv. Neural Inf. Proces. Syst. 30 (2017).
[186] D. Ahmedt-Aristizabal, et al., Graph-based deep learning for medical diagnosis and analysis: past, present and future, Sensors 21 (14) (2021) 4758.
[187] T.-C. Phan, H.N. Huu, Enhancing breast cancer detection with advanced deep learning techniques for Ki-67 nuclear protein analysis, SN Comput. Sci. 5 (6) (2024) 663.
[188] A.M. Sharafaddini, K.K. Esfahani, N. Mansouri, Deep learning approaches to detect breast cancer: a comprehensive review, Multimed. Tools Appl. (2024) 1–112.
[189] Y. Brima, M. Atemkeng, Saliency-driven explainable deep learning in medical imaging: bridging visual explainability and statistical quantitative analysis, BioData Mining 17 (1) (2024) 1–33.
[190] S. Shafi, A.V. Parwani, Artificial intelligence in diagnostic pathology, Diagn. Pathol. 18 (1) (2023) 109.

[191] D. Bhati, F. Neha, M. Amiruzzaman, A survey on explainable artificial intelligence (XAI) techniques for visualizing deep learning models in medical imaging, J. Imaging 10 (10) (2024) 239.

[192] F. Tucci, et al., The digital revolution in pathology: towards a smarter approach to research and treatment, Tumori J 110 (4) (2024) 241–251, https://doi.org/10.1177/0300891624123103503008916241231035.

[193] A. Rao, O. Aalami, Towards improving the visual explainability of artificial intelligence in the clinical setting, BMC Digit. Health 1 (1) (2023) 23.

[194] J. Webster, R. Dunstan, Whole-slide imaging and automated image analysis: considerations and opportunities in the practice of pathology, Vet. Pathol. 51 (1) (2014) 211–223.

[195] M.G. Hanna, et al., Integrating digital pathology into clinical practice, Mod. Pathol. 35 (2) (2022) 152–164.

[196] A. Aljuhani, et al., Whole slide imaging: deep learning and artificial intelligence, in: Whole Slide Imaging: Current Applications and Future Directions, 2022, pp. 223–236.

[197] V. Baxi, et al., Digital pathology and artificial intelligence in translational medicine and clinical practice, Mod. Pathol. 35 (1) (2022) 23–32.

[198] N. Sahni, et al., The Potential Impact of Artificial Intelligence on Healthcare Spending, National Bureau of Economic Research, Cambridge, MA, USA, 2023.

[199] B.R. Jackson, et al., The ethics of artificial intelligence in pathology and laboratory medicine: principles and practice, Acad. Pathol. 8 (2021) 2374289521990784.