

## Review

**Cite this article:** Javaeed A and Schuh A (2025). Artificial intelligence in breast cancer diagnosis: A systematic literature review. *Cambridge Prisms: Precision Medicine*, 3, e7, 1–19  
<https://doi.org/10.1017/pcm.2025.10006>

Received: 31 January 2025  
Revised: 06 November 2025  
Accepted: 09 November 2025

### Keywords:

artificial intelligence; neural networks; breast cancer; cancer diagnosis; histopathology

### Corresponding author:

Arslaan Javaeed;  
Email: [arslaanjaveed@yahoo.com](mailto:arslaanjaveed@yahoo.com)

© The Author(s), 2025. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.



# Artificial intelligence in breast cancer diagnosis: A systematic literature review

Arslaan Javaeed  and Anna Schuh 

Oncology, University of Oxford, UK

## Abstract

Breast cancer is the second leading cause of cancer-related deaths among women globally and the most prevalent cancer in women. Artificial intelligence (AI)-based frameworks have shown great promise in correctly classifying breast carcinomas, particularly those that may have been difficult to discern through routine microscopy. Additionally, mitotic number quantification utilizing AI technology is more accurate than manual counting. With its many advantages, such as improved accuracy, efficiency and consistency as shown in this literature review, AI has promise for significantly enhancing breast cancer diagnosis in the clinical world despite the paramount obstacles that must be addressed. Ongoing research and innovation are essential for overcoming these challenges and effectively harnessing AI's transformative potential in breast cancer detection and assessment.

## Impact statement

This article explores the impact of artificial intelligence (AI) on breast cancer (BC) diagnosis within the field of pathology. It examines several applications of AI in BC pathology and provides a succinct summary of the principal findings from multiple investigations. Incorporating AI with conventional pathology methods may enhance diagnostic accuracy and reduce preventable errors. Studies have shown the efficacy of AI in detecting invasive breast tumors by rapidly analyzing extensive whole-slide images. Advanced convolutional neural networks underpin these discoveries. AI-driven quantitative analysis has facilitated the assessment of an individual's hormonal status, which is crucial for determining the appropriate BC treatment. This is due to its facilitation of consensus among many observers regarding their findings. AI has the potential to become essential for assessing BC and quantifying mitotic cells, as it can accurately classify moderate-grade breast carcinomas. Furthermore, the utilization of AI for measuring mitotic numbers has proven to be more precise and sensitive than manual methods, resulting in enhanced predictive outcomes. In the context of triple-negative BC, to maximize the benefits of AI in BC pathology, it is essential to address issues such as the necessity for comprehensive annotations and the challenges of differentiation. Despite existing challenges, AI's numerous contributions to BC pathology indicate a promising future characterized by enhanced accuracy, efficiency and uniformity. It is imperative that we continue researching and developing novel approaches to address these challenges and fully harness AI's promise to revolutionize BC detection and assessment.

## Introduction

### History of artificial intelligence

Artificial intelligence (AI) refers to the utilization of technology and computers to imitate human-like cognitive processes and intelligent actions (Försch et al., 2021; Briganti, 2023). The history of computers can be traced back over 200 years, marked by several significant advancements. The exact year of the first computer's invention is uncertain, but it is commonly attributed to 1822, when Charles Babbage introduced a design for a functional computer on paper (Grzybowski et al., 2024). The history of AI spans several decades (Muthukrishnan et al., 2020), beginning in the 1950s with Alan Turing's research on the feasibility of intelligent machines, culminating in his landmark 1950 article (Turning, 1950; Kaul et al., 2020). In 1956, the term "AI" was coined by John McCarthy (Anyoha et al., 2017) during a conference at Dartmouth College, where the first AI program, "Logic Theorist," was introduced (Moor, 2006).

### Concepts in AI

AI applications in medicine have evolved significantly due to advancements in machine learning (ML) and deep learning (DL) (Lanzagorta-Ortega et al., 2022). AI-based models can assist in

diagnosing diseases, forecasting therapy responses and promoting preventive medicine (Kaul et al., 2020; Pettit et al., 2021).

### Machine learning

The term “ML” was first introduced by Arthur Samuel in 1959, with applications in medicine emerging in the 1980s and 1990s (Brown, 2021). As a subset of AI, ML uses algorithms to build models that can learn from data and help in segmentation, classification or make predictions (Jiang et al., 2022). It is classified into three categories: supervised, unsupervised and reinforcement learning (Hosny et al., 2018; Ono and Goto, 2022). Supervised learning trains models with input and output data to predict outcomes and unsupervised learning analyzes unannotated data to discover patterns without predefined results; similarly, reinforcement learning involves learning through interactions with an environment, receiving rewards or penalties based on actions taken (Jovel and Greiner, 2021; Lee et al., 2022; Jiang et al., 2022; Al-Hamadani et al., 2024).

### Deep learning

DL is a subset of ML using artificial neural networks (ANNs) with multiple layers (Sarker, 2021) effective in complex tasks and large datasets (Sidey-Gibbons and Sidey-Gibbons, 2019; Birhane et al., 2023). Neural networks are designed like biological neurological systems based on the fundamental unit perceptron or neuron, and usually comprise of an input, hidden and output layers (Kriegeskorte and Golan, 2019). Deep neural networks (DNNs) are advanced models with multiple hidden layers used in healthcare for medical imaging and diagnostics (Baji’ et al., 2022; Egger et al., 2022). Some ANNs have no hidden layer while DNNs have multiple, enabling them to understand complex behaviors (Kufel et al., 2023). Convolutional neural networks (CNNs) are specifically designed for image recognition and classification tasks (Alajanbi et al., 2021). Recently, these models have shown great potential for accurate diagnoses such as diabetic retinopathy from retinal images (Ragab et al., 2022).

### AI in medicine

ML, a key technology in AI, is used across various medical specialties, including oncology, cardiology and neurology (Bitkina et al., 2023). AI applications include screening, diagnosis, treatment, drug development (Xu et al., 2023), genomic analysis, patient monitoring and wearable health technology (Shajari et al., 2023). Additionally, AI enhances doctor–patient interactions, enables remote therapy and manages large datasets (Shajari et al., 2023; Chen and Decary 2020; Basu et al. 2020). Integrating AI into healthcare can significantly improve the effectiveness, accuracy and personalization of medical diagnoses and treatments (Alowais et al., 2023).

### AI in cancer diagnosis

AI has the potential to significantly advance cancer diagnosis by using annotated medical data, advanced ML techniques and enhanced processing power (Sufyan et al., 2023; Alshuhri et al., 2024). These developments are expected to transform patient care by improving efficiency, accuracy and customization in diagnoses and treatments (Chen and Decary, 2020). Over the past decade, DL architectures have outperformed traditional ML methods in cancer diagnosis, effectively utilizing genomic and phenotype data for

cancer classification and treatment (Miotto et al., 2018). Computer-aided detection and diagnosis (CADx) are playing important roles in clinical imaging and are expected to further improve (He et al., 2020; Jairam and Ha, 2022). AI technology has the potential to improve the accuracy of clinical image analysis for identifying cancer progression, aiding in the early detection and diagnosis, while medical imaging remains essential for early identification and monitoring of cancer (Suberi et al., 2017; Liu et al., 2020; Lathwal et al., 2020).

### AI in breast cancer pathology

Breast cancer (BC) is the most diagnosed cancer among women and the second leading cause of cancer-related deaths worldwide (Watkins et al., 2019), with approximately 2.3 million new cases and 6,85,000 fatalities reported in 2020 cancer (Hanna et al., 2017; Nardin et al., 2020). It represents 25% of all newly diagnosed cancer cases, and projections suggest an increase to nearly 2.96 million cases by 2040 (Sedeta et al., 2023). Accurate histopathological diagnosis is crucial, as it confirms the presence of tumor cells and helps classify the type and grade of cancer (Cardoso et al., 2019). Discrepancies in diagnoses can significantly affect treatment decisions, highlighting the need for precise pathologic assessments (Soliman et al., 2024).

Despite advancements in imaging-based diagnostics and therapies, the field of histopathology has been slow to digitize, beginning this process only about two decades ago (Hanna et al., 2019; Försch et al., 2021). Histopathological diagnosis has remained largely unchanged, still relying on microscopic evaluations by pathologists. This reliance can lead to errors, such as false positives or negatives, especially under stress (Morelli et al., 2013; Cohen et al., 2022). Studies have shown significant variability in pathologists’ assessments, with a concordance rate of only 75.3% overall and a particularly low rate of 48% for ductal carcinoma in situ (DCIS) and atypical hyperplasia, indicating ambiguity in pathology interpretations (Elmore et al., 2015).

The use of AI in cancer diagnosis is essential for improving diagnostic processes and addressing the shortage of pathologists alongside the increasing number of cancer cases (Robboy et al., 2020). AI in pathology relies on whole-slide imaging (WSI) technology (Niazi et al., 2019; Ahn et al., 2023), which converts physical pathological slides into high-resolution digital images. These images are an abundant source of information – with sizes up to 1,00,000 × 1,00,000 pixels and are the first step in creating AI-assisted models (Mukhopadhyay et al., 2018; Niazi et al., 2019). WSI aids in the easy sharing and consultation of images, reducing interpretation errors and enhancing the analysis of complex cases (Hanna et al., 2019; Jones et al., 2015). Overall, WSI presents a promising alternative to conventional microscopic examination, which has limitations due to its ephemeral nature (Mukhopadhyay et al., 2018; Tizhoosh and Pantanowitz, 2018; Hanna et al., 2019).

Digitalization in pathology has been slower compared to other medical specialties, largely due to pathologists’ reluctance to abandon traditional methods and the presence of barriers like regulatory and cost issues as well as “pathologist technophobia” (Hanna et al., 2019; Hanna and Pantanowitz 2019; Moxley-Wyles et al., 2020; Försch et al., 2021). Despite these challenges, AI has shown promise in enhancing diagnostic capabilities. Studies indicate that digital slide reviewing is as effective as manual methods (Loughrey et al., 2015; Elmore et al., 2017; Tabata et al., 2017). AI algorithms have been developed for detecting and classifying BC, achieving high

("Artificial intelligence" OR "Machine Learning" OR "Deep Learning" OR "Sentiment Analysis") AND ("breast cancer" OR "Breast Neoplasm" OR "Breast Tumor" OR "Mammary Neoplasm" OR "Mammary Carcinoma") AND (histo\* OR cyto\*).

**Figure 1.** Boolean search with keywords and their synonyms.

accuracy in differentiating between benign and malignant tumors (Cruz-Roza et al., 2017). A DL model differentiated benign and malignant tumors when tested on eight categories of images, four benign and four malignant, with an accuracy of 93.2% (Han et al., 2017). The breast cancer histology (BACH) challenge demonstrated that AI could achieve accuracy levels comparable to pathologists, improving overall diagnostic performance and interobserver concordance (Polonia et al., 2021). Additionally, DL models have successfully identified markers in BC and utilized nuclear characteristics to predict risk categories for patients (Romo-Bucheli et al., 2016; Lu et al., 2018; Whitney et al., 2018).

For example, the visual assessment of mitotic figures in BC histological sections stained with hematoxylin and eosin (H&E), referred to as the mitotic score, serves as the gold standard method for evaluating the proliferative activity of BC (Aleskandarany et al., 2012; van Dooijeweert et al., 2021) and pathologists face challenges in manually counting mitosis in histopathology slides, a process that is time-consuming (Cree et al., 2021). To address this, various contests such as the MITOSIS detection contest and others have facilitated advancements in automated counting methods (Aubreville et al., 2024). Recent studies have demonstrated that DL models can accurately count mitotic figures from H&E-stained slides of early stage ER-positive breast tumors, significantly reducing the time required for pathologists to read slides (Roux et al., 2013; ICPR, 2014; Romo-Bucheli et al., 2017; Veta et al., 2019). Notably, algorithms utilizing advanced architectures like ResNet-101, and Faster R-CNN have shown high accuracy, with one approach reducing reading time by 27.8% (Pantanowitz et al., 2020). Furthermore, a comprehensive automated system developed by Nateghi et al. (2021) can identify regions of interest for high mitotic activity, count mitoses from WSI, and predict tumor proliferation scores, outperforming previous methods. Although these AI models have yet to be deployed in formal clinical practice, the integration of digitalization and AI in pathology has the potential to enhance accuracy, reduce human errors and optimize the time needed for pathologists to review slides, ultimately benefiting both pathologists and patients (Aeffner et al., 2019; Kim et al., 2022).

This systematic review aims to examine AI models and their effectiveness in diagnosing BC, taking into account existing problems in the field of pathological diagnosis as well as the potential advantages of incorporating AI. Additionally, it investigates the potential of AI models to offer second opinions and their integration into the pathology workflow.

## Methodology

### Preferred Reporting Items for Systematic Reviews and Meta-Analyses statement

This review process follows the guidelines set out in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement, which was first developed in 2009 and updated in 2020. PRISMA functions as a framework specifically

created to standardize the process of conducting systematic reviews and improve the thoroughness of their reporting (Page et al., 2021).

### Search strategy

A comprehensive literature search was carried out across three electronic databases: PubMed, EMBASE and Cochrane Library, to identify original articles that met the specified inclusion and exclusion criteria and were published up to April 2024. The search involved the use of keywords, their synonyms and Boolean operators, as illustrated in Figure 1. Additionally, the bibliographies of all relevant articles were meticulously reviewed to review additional studies that could potentially be included in the analysis. The titles were meticulously screened using the predetermined inclusion criteria, focusing on studies that assessed the application of AI in BC diagnosis. Notably, no restrictions were placed on publication year, country of origin or age of the studies. Subsequently, a full-text screening was carried out to include the most pertinent research papers for subsequent data extraction and analysis.

### Selection process

The literature search results were screened in a two-step process. Initially, the titles and abstracts of all articles were assessed for eligibility. After identifying relevant articles, full-text screening was conducted for the studies that met the eligibility criteria. The screening was done by applying the predefined inclusion criteria.

### Eligibility criteria

All research papers using AI in BC diagnosis or staging in comparison with pathologists' report or known datasets as a reference test were included. All types of original studies (either prospective or retrospective) containing their own data on AI validation or development and validation were included. The studies published in English language were included. We further excluded other types of publications such as reviews, single cases, editorial material, books, comments and papers in languages other than English, Articles dealing with other malignancies, articles that used AI to analyze data other than histological or cytological images (e.g., MRI, mammography) and articles with non-AI approaches for diagnosis (i.e., slide flow).

### Inclusion criteria

The study included

- BC diagnosis based on histopathology.
- AI models and their diagnostic accuracy, sensitivity, specificity and area under the curve (AUC) in BC diagnosis.
- Potential of AI models to be integrated in regular practice, provide second opinion in BC diagnosis.

### Exclusion criteria

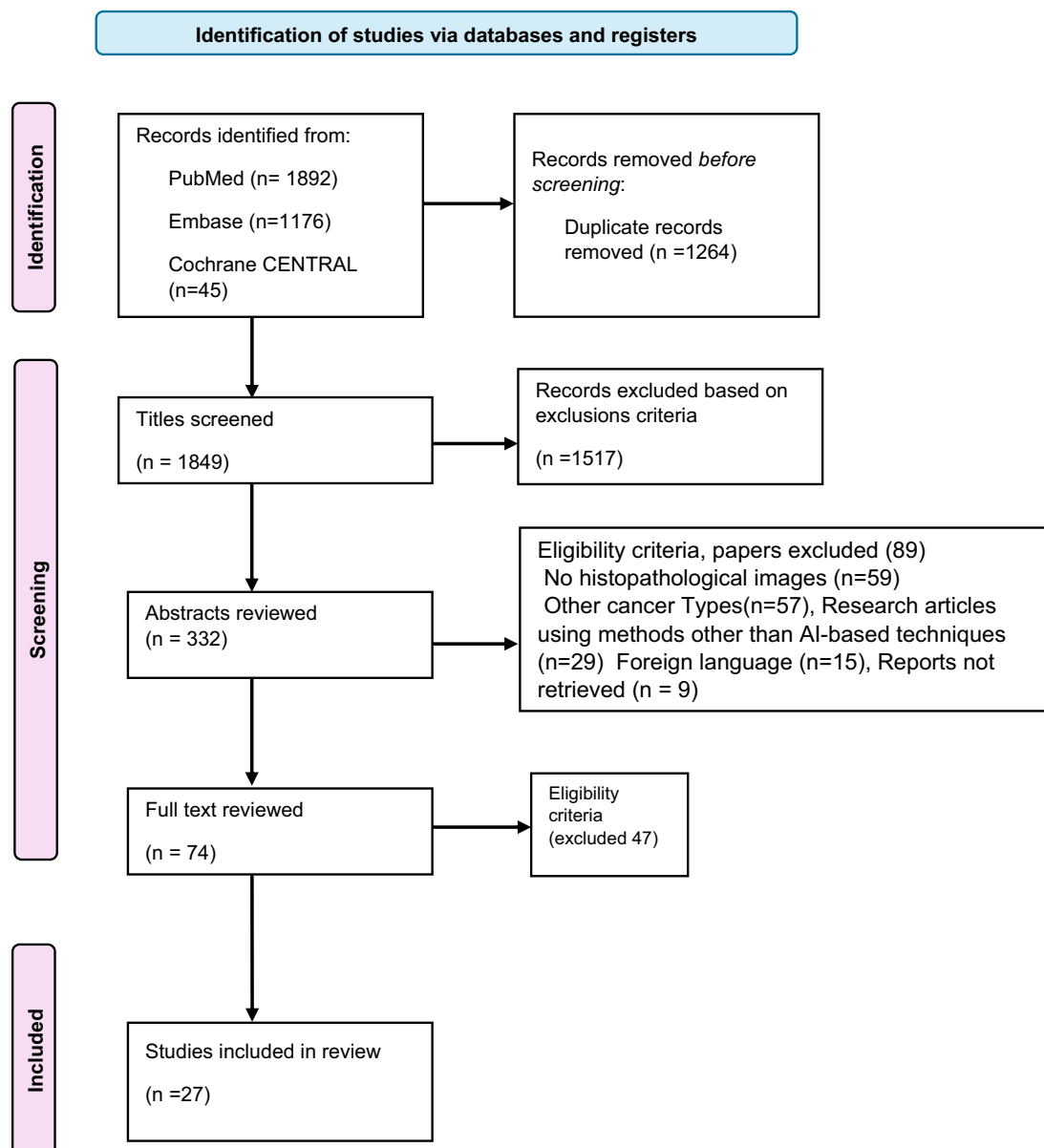
- Following studies were excluded.
- Articles used AI to analyze data other than histological or cytological images (e.g., MRI, mammography).
- Articles with non-AI approaches for diagnosis (i.e., slide flow).
- Review papers, case studies, editorials, book chapters and commentary.
- Papers in languages other than English.
- Articles dealing with other malignancies.

Data from the included studies was extracted and recorded in a standardized data extraction sheet. The extracted data encompassed two main categories: (1) characteristics of the included studies and (2) outcome measures.

### Results

#### Literature search and screening

A thorough search of the three main databases, namely PubMed, Cochrane CENTRAL and EMBASE, resulted in a total of 3113 records. After removing duplicates and irrelevant records, 1849 unique studies were assessed for eligibility based on their titles and abstracts. Over 1500 (1517) studies that did not meet our inclusion criteria were excluded. The abstracts of the remaining 332 articles were obtained for further evaluation. Application of predefined criteria led to the exclusion of 258 studies. The full text of 74 articles was reviewed out of which 47 were excluded due to the lack of relevant outcomes, inappropriate study design or poor quality. Finally, 27 studies were included in the systematic review. The flow diagram of the literature search and screening process is shown in Figure 2.



**Figure 2.** Flowchart describing the literature inclusion process.

### Characteristics of included studies

Twenty-seven research papers were analyzed in a thorough literature analysis. These studies specifically investigated the development of AI algorithms for diagnosing BC using histopathology images. The training data for histopathology images used in these research studies showed substantial variety, with data obtained from 21 to more than 400 patients. The literature evaluation includes papers published from 2017 to 2024, originating from various geographical regions including Korea, India, Egypt, the Netherlands, Brazil, Spain, Australia, Turkey, Poland, Japan, China and Malaysia. The research employed a range of AI techniques, with CNNs being particularly prominent. These CNNs were typically used in combination with other methods such as support vector machines (SVMs) or ensemble learning approaches. The datasets used for training and validation purposes included BreakHis (Spanhol et al., 2016), BACH (Aresta et al., 2019) and (ICAR, 2018 – Grand Challenge), which are well-known. The pixel resolutions of the images varied in the investigations, which affected the level of detail and the potential effectiveness of the models. In addition, the performance metrics in the included studies showed variation, with accuracy, F1-score and AUC being the most reported. Table 1 provides a concise summary of the fundamental features of the research.

### Summary of findings

#### Breast lesion categorization

Researchers have developed innovative AI models for classifying breast lesions as benign or malignant, with the potential to identify histological subtypes (Singh et al., 2024). The SegEIR-Net model combines segmentation and classification techniques using EfficientNet, InceptionNet and ResNet, achieving high accuracies on the BreakHis dataset (up to 98.66%) and strong results on the BACH and UCSB datasets (93.33% and 96.55%, respectively) (Singh et al., 2024). Additionally, the Multilevel Context and Uncertainty aware (MCUa) model categorizes breast histology images into four types: normal tissue, benign lesion, in situ carcinoma and invasive carcinoma. The MCUa model demonstrated impressive performance, with static ensemble accuracy reaching 95.75% and dynamic accuracy reaching 98.11% on the BACH dataset, and outstanding results on the BreakHis dataset (up to 100% accuracy) (Senousy et al., 2022).

The context-aware stacked CNN (CAS-CNN) has demonstrated strong performance in classifying breast WSIs, achieving an AUC of 0.962 in distinguishing normal or benign slides from malignant ones (DCIS and IDC). The system exhibited a precision of 89.1% for categorizing WSIs and an overall accuracy of 81.3% in a three-class classification involving normal/benign, DCIS and IDC categories. While it effectively differentiated between normal/benign and IDC slides, it faced challenges in distinguishing between normal/benign and DCIS slides, as well as between DCIS and IDC slides (Bejnordi et al., 2017). Additional details on studies classifying breast lesions are provided in Tables 1–3.

#### Molecular subtyping

After the identification of a malignant breast lesion, immunohistochemistry is commonly performed to ascertain the molecular subtype. This procedure involves analyzing the levels of ER, PR, Her2 and the Ki67 mitotic index to determine the subtype, which can include luminal A, luminal B, Her2-enriched or triple negative, among other possibilities. The results of the studies regarding the

use of AI models for molecular subtyping, with or without Ki67 measurement, have been summarized in Table 4. The AI models have exhibited an impressive accuracy rate (AUC of 0.75–0.91 vs. 0.67–0.8) when compared with conventional multiple instance learning models (Bae et al., 2023) and approximately 90% for an automated BC classification system utilizing SVM (Aswathy and Jagannath 2021).

#### Mitotic index assessment and quantification

Several studies have aimed to create models for identifying the mitotic proliferation index (Ki67) in BC. A notable approach is the FMDet method, designed to detect mitotic rates in breast histopathology images while addressing the domain shift problem caused by variability across different scanners (Wang et al., 2023). To enhance model applicability, two key strategies were implemented:

1. A novel data augmentation technique using Fourier analysis was introduced, which alters the frequency characteristics of training images to generate diverse samples that reflect real-world datasets. This involves replacing the low-frequency components of a source image with those from a reference image from a different domain.

2. Pixel-level annotations, specifically “instance masks,” were utilized for identifying mitotic figures. These masks, derived from the Mitosis Domain Generalization challenge (MIDOG) 2021 bounding box data and a pretrained nucleus segmentation model (HoVer-Net), improved detection accuracy by allowing the network to capture subtle morphological differences in mitotic cells, surpassing traditional bounding box annotations (Wang et al., 2023). The model outperformed all other submissions in the challenge, with an F1 score of 0.77 (Wang et al., 2023).

A DL model specifically targeted for identifying mitotic proliferation in breast histopathology images has shown considerable potential (Saha et al., 2018). The precision, recall and *F*-score measures of this model were 92%, 88% and 90%, respectively. Significantly, the model’s performance improved when hand-crafted features were integrated into its DL architecture. The model was trained and tested using datasets from MITOS-ATYPIA-14, ICPR-2012 and AMIDA-13. The study’s findings highlight the model’s high true positive rate, which indicates its precise ability to identify mitotic cells. Details of included studies focusing on training models for detecting and quantifying mitotic proliferation index have been tabulated in Table 5.

#### Obstacles to widespread adoption of AI

Despite the remarkable outcomes demonstrated by the examined AI models in research contexts, huge challenges and constraints persist that hinder the extensive use of AI on a broader scale (Soliman et al., 2024). A primary procedural drawback of supervised AI models is their need of extensive, annotated datasets for training. Manual annotation is labor-intensive and exhibits variability both among and across pathologists, undermining the fundamental objective of AI models. Similarly, the intraclass heterogeneity and dependence on binary categorization during training, as previously mentioned, constitute additional barriers that may impact the efficacy of AI models. Furthermore, the lack of a defined area size for assessing Ki67 may lead to either an overestimation or underestimation of Ki67 activity. Moreover, certain preanalytic variables, including suboptimal sample quality, air bubbles, staining artifacts, unexpected staining patterns and discrepancies in interlaboratory sample preparation and staining

**Table 1.** Summary of studies categorizing breast lesions (i.e., benign vs. malignant)

Author	Year	Country	AI model	Training and validation dataset	Model output	No. patients/ images used in training	Pixel level	Performance of the model	External validation dataset	Results of external validation
Amin et al.	2023	Korea	FabNet	BreakHis dataset	1) Benign vs. malignant 2) Benign/malignant lesion histological classification	58 patients	700 × 460 pixels	99% Accuracy and a 98.9% F1 score for binary classification at 40× magnification	NR	NR
Bejnordi et al.	2017	Netherlands	Context-aware stacked CNN	Dataset of 157 WSIs (118 for training and 39 for validation) of breast tissue sections from the pathology archive of Radboud University Medical Center	1) Benign vs. malignant 2) Normal/benign vs. in situ vs. invasive samples	118 WSIs	0.243 μm × 0.243 μm pixel size	Model achieved an accuracy of 81.3% and a kappa value of 0.7	Digitized WSIs selected for validation	Maximum accuracy of 0.9135 for the CAS-CNN 0.9241 for the WRN-4-2
Gandomkar et al.	2018	Australia	ResNet-152	BreakHis dataset	1) Benign vs. malignant 2) Benign/malignant lesion histological classification	70 patients	700 × 460 pixels	Achieved an average CCR of 98.10% for classifying images as benign or malignant and 95.15% for classifying images into eight classes (four benign and four malignant subtypes)	BreakHis	NR
Kolla et al.	2024	India	Modified pretrained Tiny Swin-Transformer V2	BreakHis dataset	1) Benign vs. malignant 2) Benign/malignant lesion histological classification	NR	Resized to 256 × 256 pixels during preprocessing	Accuracy of 98.27%, 97.95%, 98.97% and 99.16% in the eight-groups, malignant, benign and binary, respectively	BreakHis	Best validation accuracy of 98.27%
Murthy and Balaji et al.	2022	NR	CNN	NR	1) Benign vs. malignant 2) Benign/malignant lesion histological classification	NR	220 × 220 pixels	1) For benign class ACC = 0.7866, TPR = 0.7921, TNR = 0.7837, FPR = 0.2163, PPV = 0.6597, NPV = 0.8769 2) For malignant class ACC = 0.7849, TPR = 0.788, TNR = 0.7832, FPR = 0.2168, PPV = 0.673, NPV = 0.8671	NR	NR
Nahid et al.	2018	Brazil	CNN, LTSM, combined CNN and LTSM	BreakHis	Benign vs. malignant	82 patients	760 × 460 pixels	Accuracy of 77.4% for model 3, with a sensitivity of 95% and specificity of 59% – on using both the MS cluster algorithm and the SVM classifier together at 40× dataset	NR	NR

(Continued)

Table 1. (Continued)

Author	Year	Country	AI model	Training and validation dataset	Model output	No. patients/ images used in training	Pixel level	Performance of the model	External validation dataset	Results of external validation
Srikantamurthy et al.	2023	Korea	Hybrid CNN-LSTM ImageNet model	BreakHis dataset	1) Benign vs. malignant 2) Benign/malignant lesion histological classification	82 patients	NR	Hybrid CNN-LSTM model achieved 99% accuracy for binary classification (benign vs. malignant) and 92.5% accuracy for multiclass classification (subtypes of benign and malignant)	NR	Accuracy of 99.03% at 40× images for binary classification, and 96.5% for multiclass classification
Stanitias et al.	2020	NR	WAID built upon a CKD	FABCD	1) Benign vs. malignant 2) Benign/malignant lesion histological classification	30 Patients	10,000 × 9,000 pixels that are further divided into 150 × 150 pixel patches	Using LERM for similarity calculations, achieved the highest performance on the FABCD: 92.83% accuracy and 0.98 AUC	BreakHis	92% CCM rate at the image level and 91.27% CCM at the patient level
Xu et al.	2022	China	MFSCNet, which is an improved DenseNet driven by an attention mechanism	BreakHis dataset	1) Benign vs. malignant 2) Benign/malignant lesion histological classification	82 patients	224 × 224 pixels	The best performing MFSCNet model (MFSCNet A) achieved 99.05%–99.89% accuracy in binary classification (benign or malignant). In multiclass classification (across 8 subtypes), the accuracy ranged from 94.36% to 98.41%	NR	NR
Zhu et al.	2019	China	CNNs	BreakHis, BACH	Benign (including normal/benign lesions) vs. cancerous (in situ/invasive lesions)	20 WSIs in the BACH WSI dataset	40,000 × 60,000 for the BACH WSI dataset	87.5% patient-level and 84.4% image-level accuracy	BACH WSI dataset	BACH WSI dataset, the multimodel ensemble achieved 87.5% patient-level accuracy and 84.4% image-level accuracy

AUC, area under the curve; BACH, breast cancer histology challenge; BreakHis, breast cancer histopathological database; CCM, correctly classified malignant; CCR, correct classification rate; CKD, covariance-kernel descriptor; CNNs, convolutional neural networks; DCNNs, deep CNNs; DenseNet, dense neural network; FABCD, fully annotated breast cancer database; LERM, log-Euclidean Riemannian metric; MFSCNet, classification of mammary cancer fusing spatial and channel features network; MLP, multilayer perceptron; QDA, quadratic discriminant analysis; ResNet-152, deep residual network with 152 layers; RF, random forest; SVM, support vector machine; WAID, weakly annotated image descriptor; WSIs, whole-slide images.

**Table 2.** Summary of studies categorizing breast lesions (i.e., normal/benign/in situ/invasive)

Author	Year	Country	AI model	Training and validation dataset	Model output	No. patients/images used in training	Pixel level	Performance of the model	External validation dataset	Results of external validation
Attallah et al.	2021	Egypt	Histo-CADx, a system that utilizes multiple AI models: Five CNNs: AlexNet, GoogleNet, ResNet-50, ShuffleNet and Inception-ResNet V2 Autoencoder (AE): For feature fusion and dimensionality reduction Classifiers: SVM, RF and QDA	BreakHis and ICIAR 2018 datasets	Normal vs. benign vs. in situ vs. invasive samples	82 patients for BreakHis dataset and 400 WSIs for the ICIAR 2018 dataset	2048 × 1536 pixels	Accuracy up to 99.54%	ICIAR 2018 dataset	Accuracy up to 97.93%
Bejnordi et al.	2017	Netherlands	Context-aware stacked CNN	Dataset of 157 WSIs (118 for training and 39 for validation) of breast tissue sections from the pathology archive of Radboud University Medical Center	1) Benign vs. malignant 2) Normal/benign vs. in situ vs. invasive samples	118 WSIs	0.243 μm × 0.243 μm pixel size	Model achieved an accuracy of 81.3% and a kappa value of 0.7	Digitized WSIs selected for validation	Maximum accuracy of 0.9135 for the CAS-CNN 0.9241 for the WRN-4-2
Carvalho et al.	2020	Brazil	Several machine learning models: SVM RF, MLP, Xgboost	BACH 2018 dataset	Normal vs. benign vs. in situ vs. invasive samples (with different subgroup analyses)	400 WSIs	2040 × 1536 pixels	SVM performed best, with accuracy of 97.5% in differentiating normal vs. abnormal samples, and 100% in differentiating between benign vs. malignant samples	BACH 2018 dataset	Best performing model was SVM with PDIs achieving 95% accuracy and 96% precision
Fondón et al.	2018	Spain	SVM classifier with a quadratic kernel	High-resolution images from anonymous breast tumor biopsies	Normal vs. Benign vs. In Situ vs. Invasive Samples	120 WSIs	2048 × 1536 pixels	75.8% accuracy using fivefold cross-validation on the training dataset	20 images of similar complexity to the training dataset, and 16 images considered “extremely difficult” due to artifacts and	1) Dataset with 20 images: 75% accuracy 2) Dataset with the 16 extremely difficult images: 61.11% accuracy

(Continued)

Table 2. (Continued)

Author	Year	Country	AI model	Training and validation dataset	Model output	No. patients/ images used in training	Pixel level	Performance of the model	External validation dataset	Results of external validation
									ambiguous malignancy	
Gecer et al.	2018	Turkey	FCNs followed by CNN	240 digital breast histopathology images collected as part of NIH-sponsored projects	Benign vs. in situ vs. invasive samples	180 WSIs	535 × 416 pixels	Accuracy of 55%	External validation dataset consisted of independent diagnoses from 45 pathologists who evaluated the same 60 test slides used to assess the model's performance	55% accuracy was comparable to the average accuracy of the 45 pathologists
Kanavati et al.	2022	Japan	EfficientNetB1 and EfficientNetB3	Biopsies collected from three different hospitals and assessed by surgical pathologist	Benign vs. in situ vs. invasive samples	3304 WSIs	Tile size of 224 × 224 pixels and a stride of 112 × 122 pixels	AUCs ranging from 0.95 to 0.99.8	Biopsies diagnosed with pathologist from hospitals	Hospital 1 (Biopsy): AUC = 0.980, Log loss = 0.269 Hospital 1 (Surgical): AUC = 0.958, Log loss = 0.377, Hospital 2 (Surgical): AUC = 0.994, Log loss = 0.180, TCGA (Surgical): AUC = 1.000, Log loss = 0.274
Laxmisagar et al.	2022	India	MobileNet AI architecture model	ICIAR BACH challenge dataset	Normal vs. benign vs. in situ vs. invasive samples	NR	2040 × 1536 pixels	87.5% accurate	NR	NR
Liu et al.	2022	China	MVMS-PFENet	BACH	Normal vs. benign vs. in situ vs. invasive samples	400 for training, and 100 for testing	2048 × 1536 pixels	f1 higher than 92.8%	Hybrid deep neural network	f1 score higher than 92% and accuracy reached 95%
Murtaza et al.	2021	Malaysia	EBrT-Net, which is based on AlexNet, but with modifications	BCBH dataset	Normal vs. benign vs. in situ vs. invasive samples	249 WSIs	258 × 258 pixels	Ranging from 87.50% to 100% for the four subtypes of BrT	BreakHis dataset	For the NC <sub>2</sub> Net model with the best performance: Least validation loss achieved was 0.1281, corresponding to a

(Continued)

Table 2. (Continued)

Author	Year	Country	AI model	Training and validation dataset	Model output	No. patients/ images used in training	Pixel level	Performance of the model	External validation dataset	Results of external validation
										validation accuracy of 95.66%
Senousy et al.	2022	NR	MCUa dynamic deep learning ensemble model	BACH dataset	Normal vs. benign vs. in situ vs. invasive samples	39 patients	2048 × 1536 pixels	98.11% accuracy using the dynamic ensemble method	BreakHis	99.80% accuracy using the static ensemble technique. The model achieved 100%, 99.95% and 99.90% accuracy using dynamic ensemble on $\delta$ values of 0.001, 0.003 and 0.03
Singh et al.	2024	India	SegEIR-Net	BreakHis, BACH and UCSB	Breast cancer histopathological classification	NR	NR	1) BACH dataset: 93.33%, UCSB dataset: 96.55% 2) BreakHis dataset: <ul style="list-style-type: none"> <li>• 40X Magnification: 98.66%,</li> <li>• 100X Magnification: 98.39%,</li> <li>• 200X Magnification: 97.52%,</li> <li>• 400X Magnification: 95.22%"</li> </ul>	NR	NR
Yang et al.	2019	China	Used an ensemble of three pretrained DCNNs: DenseNet–161, ResNet–152 and ResNet–101. This ensemble model is referred to as EMS-Net	* Training: BACH dataset * Validation: For validation, 40 images were randomly selected from the training dataset	Normal vs. benign vs. in situ vs. invasive samples	400 WSIs	2048 × 1536 pixels	EMS-Net algorithm achieved an overall accuracy of $91.75 \pm 2.32\%$ in a fivefold cross-validation using the training images from the BACH dataset	100 microscopy images from the BACH dataset	EMS-Net model achieved an accuracy of 90%
Zhu et al.	2019	China	CNNs	BreakHis, BACH	Benign (including normal/benign lesions) vs. cancerous (in situ/invasive lesions)	20 WSIs in the BACH WSI dataset	40,000 × 60,000 for the BACH WSI dataset	87.5% patient-level and 84.4% image-level accuracy	BACH WSI dataset	BACH WSI dataset, the multimodel ensemble achieved 87.5% patient-level accuracy and 84.4% image-level accuracy

AUC, area under the curve; BACH, breast cancer histology challenge; BCBH, bioimaging challenge 2015 breast histology; BreakHis, breast cancer histopathological database; CCM, correctly classified malignant; CCR, correct classification rate; CKD, covariance-kernel descriptor; CNNs, convolutional neural networks; DCNNs, deep CNNs; DenseNet, dense neural network; FABCD, fully annotated breast cancer database; FCNs, fully convolutional networks; LERM, log-Euclidean Riemannian Metric; MCUa, multilevel context and uncertainty aware; MFSCNet, classification of mammary cancer fusing spatial and channel features network; MLP, multilayer perceptron; PDIs, phylogenetic diversity indices; QDA, quadratic discriminant analysis; ResNet-152, deep residual network with 152 layers; RF, random forest; SimCLR, simple framework for contrastive learning of visual representations; SOA, state-of-the-art; SSL, self-supervised learning; SVM, support vector machine; UCSB, University of California Santa Barbara; WAID, weakly annotated image descriptor; WSIs, whole-slide images; Xgboost, eXtreme Gradient Boost.

**Table 3.** Summary of studies assessing different histopathological subtypes of both benign and malignant breast lesions

Author	Year	Country	AI model	Training and validation dataset	Model output	No. patients/ images used in training	Pixel level	Performance of the model	External validation dataset	Results of external validation
Amin et al.	2023	Korea	FabNet	BreakHis dataset	1) Benign vs. malignant 2) Benign/ malignant lesion histopathological classification	58 patients	700 × 460 pixels	99% accuracy and a 98.9% F1 score for binary classification at 40× magnification	NR	NR
Gandomkar et al.	2018	Australia	ResNet–152	BreakHis dataset	1) Benign vs. malignant samples 2) Benign/ malignant lesion histopathological classification	70 patients	700 × 460 pixels	Achieved an average CCR of 98.10% for classifying images as benign or malignant and 95.15% for classifying images into eight classes (four benign and four malignant subtypes)	BreakHis	NR
Kolla et al.	2024	India	Modified pretrained tiny swin-transformer V2	BreakHis dataset	1) Benign vs. malignant samples 2) Benign/ malignant lesion histopathological classification	NR	Resized to 256 × 256 pixels during preprocessing	Accuracy of 98.27%, 97.95%, 98.97% and 99.16% in the eight-groups, malignant, benign and binary, respectively	BreakHis	Best validation accuracy of 98.27%
Murthy and Balaji	2022	NR	CNN	NR	1) Benign vs. malignant samples 2) Benign/ malignant lesion histopathological classification	NR	220 × 220 pixels	1) For benign class ACC = 0.7866, TPR = 0.7921, TNR = 0.7837, FPR = 0.2163, PPV = 0.6597, NPV = 0.8769 2) For malignant class ACC = 0.7849, TPR = 0.788, TNR = 0.7832, FPR = 0.2168, PPV = 0.673, NPV = 0.8671	NR	NR
Singh et al.	2024	India	SegEIR-Net	BreakHis, BACH and UCSB	Breast cancer histopathological classification	NR	NR	1) BACH dataset: 93.33%, UCSB dataset: 96.55% 2) BreakHis dataset: • 40X Magnification: 98.66%, • 100X Magnification: 98.39%, • 200X Magnification: 97.52%, 400× magnification: 95.22%”	NR	NR
Srikantamurthy et al.	2023	Korea	Hybrid CNN-LSTM ImageNet model	BreakHis dataset	1) Benign vs. malignant samples 2) Benign/ malignant lesion histopathological classification	82 patients	NR	Hybrid CNN-LSTM model achieved 99% accuracy for binary classification (benign vs. malignant) and 92.5% accuracy for multiclass classification (subtypes of benign and malignant)	NR	Accuracy of 99.03% at 40× images for binary classification, and 96.5% for multiclass classification

(Continued)

Table 3. (Continued)

Author	Year	Country	AI model	Training and validation dataset	Model output	No. patients/ images used in training	Pixel level	Performance of the model	External validation dataset	Results of external validation
Stanitsas et al.	2020	NR	WAID built upon a CKD	FABCD	1) Benign vs. malignant samples 2) Benign/ malignant lesion histopathological classification	30 patients	10,000 × 9000 pixels that are further divided into 150 × 150-pixel patches	Using LERM for similarity calculations, achieved the highest performance on the FABCD: 92.83% accuracy and 0.98 AUC	BreakHis	92% CCM rate at the image level and 91.27% CCM at the patient level
Umer et al.	2022	NR	6B-Net a novel deep CNN mode, Additionally, they use a pretrained RESNET-50 model for feature extraction	BreakHis dataset	Benign/malignant lesion histopathological classification	82 patients in the BreakHis dataset and 3771 WSIs the breast cancer pathology dataset	2048 × 1536 pixels	1) BreakHis dataset achieves a multiclass average accuracy of 90.10% with a training time of 147 s using the ESKNN classifier. 2) Breast cancer pathology dataset method achieves a multiclass average accuracy of 94.20% with a training time of 226 s using the ESD classifier"	Breast cancer pathology dataset	A maximum 93.60% accuracy using the ESD classifier and a minimum 83.00% accuracy using the EBT classifier
Xu et al.	2022	China	MFSCNet, which is an improved DenseNet driven by an attention mechanism	BreakHis dataset	1) Benign vs. malignant samples 2) Benign/ malignant lesion histopathological classification	82 patients	224 × 224 pixels	The best performing MFSCNet model (MFSCNet A) achieved 99.05–99.89% accuracy in binary classification (benign or malignant). In multiclass classification (across 8 subtypes), the accuracy ranged from 94.36% to 98.41%	NR	NR

AUC, area under the curve; BACH, breast cancer histology challenge; BCBH, bioimaging challenge 2015 breast histology; BreakHis, breast cancer histopathological database; CCM, correctly classified malignant; CCR, correct classification rate; CKD, covariance-kernel descriptor; CNNs, convolutional neural networks; DCNNs, deep CNNs; DenseNet, dense neural network; FABCD, fully annotated breast cancer database; FCNs, fully convolutional networks; LERM, log-Euclidean Riemannian Metric; MUCa, multilevel context and uncertainty aware; MFSCNet, classification of mammary cancer fusing spatial and channel features network; MLP, multilayer perceptron; PDIs, phylogenetic diversity indices; QDA, quadratic discriminant analysis; ResNet-152, deep residual network with 152 layers; RF, random forest; SimCLR, simple framework for contrastive learning of visual representations; SOA, state-of-the-art; SSL, self-supervised learning; SVM, support vector machine; UCSB, University of California Santa Barbara; WAID, weakly annotated image descriptor; WSIs, whole-slide images; Xgboost, eXtreme Gradient Boost.

**Table 4.** Summary of studies assessing breast cancer molecular subtyping (i.e., according to estrogen receptors (ER), progesterone receptors (PR) and Her2 – with or without ki67 mitotic index analysis)

Author	Year	Country	AI model	Training and validation dataset	Model output	No. patients/ images used in training	Pixel level	Performance of the model	External validation dataset	Results of external validation
Aswathy et al.	2021	India	SVM	UCSB dataset and validated by BreakHis dataset	Breast cancer molecular classification (ER, PR, Her2)	58 WSIs	896 × 768 pixels	Accuracy: 91%, Sensitivity: 92%, Specificity: 90.3%, F-score: 89.9%, Precision: 88.9%, Balanced Accuracy: 91.2%	BreakHis dataset	SVM model achieved an average accuracy of 89.1% on the BreakHis dataset
Bae et al.	2023	Korea	3DHistoNet, which utilizes SSL approach called SimCLR	Patients with primary breast carcinoma diagnosed between 2018 and 2020 at the Department of Pathology, National Cancer Center, South Korea	Breast cancer molecular classification (including Ki67 mitotic index)	401 patients	0.25 μm	AUC ranging from 0.75 to 0.91	NR	NR

AUC, area under the curve; BCBH, bioimaging challenge 2015 breast histology; BreakHis, breast cancer histopathological database; MLP, multilayer perceptron; SVM, support vector machine; UCSB, University of California Santa Barbara; WSIs, whole-slide images; Xgboost, xExtreme.

techniques, remain unavoidable to this day. Consequently, each AI tool necessitates validation and verification under these specific conditions. The identified variables may influence the accuracy of specimen analysis conducted by the AI model, resulting in incorrect outcomes. Additionally, significant economic and regulatory challenges persist regarding the implementation of AI technology in pathology laboratories on a global or national scale (van Diest et al., 2024). The concern regarding AI's potential to replace pathologists raises ethical issues, as noted by Moxley-Wyles et al. (2020) and van Diest et al. (2024). Critics, comprising both individuals and governmental bodies, articulate concerns regarding the notion that a patient's treatment may be dictated by an AI analysis. This is particularly concerning as most existing algorithms remain in the early stages of development, having been tested on limited populations and lacking adequate safety data. The broad adoption of this technology faces obstacles stemming from the diverse staining techniques employed across laboratories, the necessity for full digitalization prior to implementation, and the partial integration of Picture Archiving and Communication Systems with AI algorithms (van Diest et al., 2024). Addressing these concerns is crucial prior to the deployment of AI models in clinical practice.

### Addressing limited sample size

A study aimed at improving BC histopathology image classification addressed the challenge of limited slide image datasets through data augmentation and transfer learning (Zhu et al., 2019). The researchers developed a hybrid CNN that combines local and global visual inputs to capture detailed features and structure. They introduced a Squeeze-Excitation-Pruning block to reduce the model size without sacrificing accuracy. To enhance generalization, they used a multimodel assembly technique, training various models on different data subsets and merging their predictions. This approach

proved more effective than a single-model systems and outperformed existing methods on the BACH dataset (87.5% patient-level and 84.4% image-level accuracy), indicating its potential for real-world clinical use (Zhu et al., 2019).

Recent studies have explored the development of advanced models for classifying BC subtypes using histopathology images. A hybrid CNN-LSTM model achieved high accuracy in binary and multiclass classifications, with a binary accuracy ranging from 98.07% to 99.75% and multiclass accuracy from 88.04% to 96.5% (Srikantamurthy et al., 2023). Another study utilized a pretrained DenseNet-169 model on the BreakHis dataset, achieving 98.73% accuracy on the validation set and 94.55% on the test set, highlighting the importance of compatible-domain transfer learning – a method through which histological images are used in the pretraining of the model and then fine-tuned on a finite cytological target dataset (Shamshiri et al., 2023). Additionally, various studies emphasize the importance of feature selection and fusion to improve classification accuracy, noting that integrating DL features with handcrafted attributes can lead to better outcomes. Multiscale analysis, incorporating different image patch scales, also contributes to enhanced accuracy in classification tasks (Attallah et al., 2021; Liu et al., 2022).

### Training datasets

Another important consideration in the included studies is the choice of dataset and the use of transfer learning. Multiple studies have utilized the BreakHis dataset, which contains images of varying magnification levels, to advance BC diagnosis, despite challenges like small sample sizes and data variability (Gandomkar et al., 2018; Nahid et al., 2018; Carvalho et al., 2020). To address these issues, researchers have applied transfer learning techniques, initially training models on larger datasets like ImageNet before

**Table 5.** Summary of studies assessing the Ki67 mitotic index

Author	Year	Country	AI model	Training and validation dataset	Model output	No. patients/ images used in training	Pixel level	Performance of the model	External validation dataset	Results of external validation
Bae et al.	2023	Korea	3DHistoNet, which utilizes SSL approach called SimCLR	Patients with primary breast carcinoma diagnosed between 2018 and 2020 at the Department of Pathology, National Cancer Center, South Korea	Breast cancer molecular classification (including Ki67 mitotic index)	401 patients	0.25 $\mu$ m	AUC ranging from 0.75 to 0.91	NR	NR
Saha et al.	2018	NR	NR	MITOS-ATYPIA-14, ICPR-2012 and AMIDA-13	1) Mitotic detection 2) Mitotic figure counting	11,921 training WSIs (256 batch size)	71 $\times$ 71 pixels	Average precision of 0.92, a recall of 0.88, and an <i>F</i> -score of 0.90	NR	0.92% precision, 0.87 recall and 0.89 at fivefold cross-validation
Wang et al.	2023	China	FMDet, a novel mitosis detection algorithm based on a semantic segmentation approach. The core architecture is a U-Net framework	Training: MIDOG 2021 challenge training set, comprising 200 WSIs from four different scanners. Validation: The paper used an out-of-domain validation set extracted from the MIDOG21 training set	Mitotic figure counting	200 WSIs	8000 $\times$ 8000 pixels	F1 score = 0.7773	Four external datasets were used: AMIDA13, MITOSIS14, TUPAC-auxiliary, MIDOG22	AMIDA13: F1 score = 0.6791, Recall = 0.7405, Precision = 0.6271. TUPAC-auxiliary (Centers 2 and 3): F1 score = 0.7458, Recall = 0.8018, Precision = 0.6971. TUPAC-auxiliary (Centers 1, 2 and 3): F1 score = 0.6946, Recall = 0.7738, Precision = 0.6301. MITOSIS14: F1 score = 0.4901, Recall = 0.5564, Precision = 0.4380 MIDOG22: F1 score = 0.7389, Recall = 0.7796, Precision = 0.7022

AUC, area under the curve; SimCLR, simple framework for contrastive learning of visual representations; SSL, self-supervised learning; WSIs, whole-slide images.

fine-tuning them on BreakHis (Kanavati et al., 2022; Laxmisagar and Hanumantharaju, 2022; Liu et al., 2022; Xu et al., 2022). Additionally, compatible-domain transfer learning has been shown to boost model performance (Shamshiri et al., 2023). Various approaches have been explored to make up for insufficient training datasets, including the combination of multiple classifiers and a self-trained AI algorithm utilizing a three-stage analysis technique with a WSI stacking system (Bae et al., 2023). Overall, these efforts highlight the potential of DL and image analysis in enhancing BC diagnosis and prognosis.

## Discussion

Commercially available AI models specifically designed for BC detection exist (Soliman et al., 2024). The mentioned entities comprise Mindpeak, Owkin, Visiopharm, Paige Breast Suite and IBEX Galen Breast. Mindpeak is located in Hamburg, Germany (Abele et al., 2023); Owkin is based in Paris, France; Visiopharm is situated in Hovedstaden, Denmark (Shafi et al., 2022); Paige Breast Suite is established in New York, United States and IBEX Galen Breast is

positioned in Tel Aviv, Israel. These algorithms improve pathologists' consistency, precision and sensitivity while decreasing time demands (Soliman et al., 2024). Nonetheless, various limitations persist in obstructing its extensive implementation on a larger scale. AI-driven specimen analysis now demonstrates specific procedural drawbacks, as observed by Soliman et al. (2024). They also evaluated the contribution of AI in enhancing histological analysis of BC and to collectively assess the efficacy of each AI model, while also emphasizing potential limitations and downsides associated with each model.

## Tissue classification

Most studies in this review evaluated the efficacy of AI in accurately classifying breast tissue specimens, demonstrating significant precision in distinguishing normal (or benign) tissues from malignant ones (Amin et al., 2023; Attallah et al., 2021; Gandomkar et al., 2018; Singh et al., 2024). In the research conducted by Singh et al. (2024), the SegEIR-Net model achieved an accuracy surpassing 98% on the BreakHis dataset. Likewise, the Histo-CADx model

introduced by Attallah et al. (2021) has attained an accuracy of 99.54%. The initial study presenting the BreakHis dataset (Spanhol et al., 2016), aimed at evaluating AI model efficacy in tissue classification, attained an accuracy of approximately 97%. Han et al. (2017) conducted a study that attained an average accuracy of approximately 93% for eight-class classification (comprising four benign and four malignant categories) of the BreakHis dataset, in contrast to the traditional binary or ternary classification systems evaluated in most studies. The findings from these studies demonstrate that AI can continuously prove to be a dependable instrument for breast specimen classification.

Nonetheless, despite these encouraging results, many issues persist regarding tissue classification. For instance, there remains a notable heterogeneity in the datasets employed and the outcomes evaluated. Should research evaluate the efficacy of AI in distinguishing between normal and malignant? distinguishing benign from malignant? Perhaps, DCIS from invasive carcinoma (Amin et al., 2023)? In-class heterogeneity exists, and binary classification tasks for AI models (e.g., normal vs. abnormal) are frequently insufficient due to the presence of gray zones and unusual findings in clinical practice (Amin et al., 2023; Bejnordi et al., 2017). Furthermore, certain unusual subtypes exist but may not be included in training and/or testing datasets (Hatta et al., 2023). The presence of in-class heterogeneity and imbalanced datasets complicates the assessment of AI's accuracy in classifying each category, particularly the rare ones (Amin et al., 2023; Hatta et al., 2023). Similarly, Bejnordi et al. (2017) reported that although their model excelled in distinguishing benign from malignant tumors, it encountered difficulties with borderline categories such as DCIS. Additionally, Amin et al. (2023) also documented similar findings.

### Molecular subtyping

Current issues associated with manual molecular subtyping include frequent interpathologist discrepancies, particularly with Her2 status (Robbins et al., 2023). In contrast to specimen categorization, which relies mostly on visible microscopic features, subtyping is contingent upon supplementary technical factors, such as the extent of staining and/or manual enumeration in the computation of Ki67. Therefore, several studies are being conducted using different AI models to address the challenge of inconsistency (Abele et al., 2023).

Similar to the reported findings about AI's performance in tissue classification, the studies assessing molecular subtyping in this review have likewise yielded promising outcomes. The data from the studies reported in this review indicate that AI excelled in both molecular subtyping and Ki67 computation. The research conducted by Bae et al. (2023) attained a molecular subtyping accuracy of 91%. Aswathy et al. (2021) found a balanced accuracy of 91.2% for their SVM model in predicting ER, PR and Her2 status. Consistent with findings in this review, several other studies have documented significantly enhanced interpathologist agreement rates following the implementation of AI aid. AI enhanced interpathologist agreement rates from approximately 88% to 96% for the Ki67 score and from roughly 89% to 93% for the ER/PR status (Abele et al., 2023). Similarly, concerning Her2 status, a study by Jung indicated a significant rise in interpathologist agreement rates from approximately 49.3% to 74.1% ( $p < 0.001$ ), with the use of AI driven ER/PR and Her2 analyzers. Additionally, the concordance rates for ER and PR status improved, albeit to a lower degree (93.0–96.5%  $p = 0.096$  for ER, 84.6–91.5%,  $p = 0.006$  for PR) (Jung et al., 2024).

### Conclusion and future directions

Various AI tools may rapidly become an effective aid to histopathologists facing increasing demands for precise and speedy BC diagnosis. The identified drawbacks are paramount and need to be effectively addressed before we can reap the true benefits of this technology. It is recommended that future research endeavors focus on the following key areas to improve the performance and validity of existing AI models in the context of histopathological evaluation:

1. **Establishment of Standardized Datasets:** The creation of standardized, multi-institutional datasets that adhere to consistent preanalytic methodologies – such as sample preparation and tissue staining – should be prioritized to improve the generalizability of AI models across diverse clinical settings.
2. **Integration of Multimodal Data:** To enhance the predictive performance of AI systems, it is imperative to incorporate additional diagnostic modalities, including but not limited to imaging techniques and molecular profiling, into the analysis of histopathological specimens. This multimodal approach can offer a more comprehensive understanding of disease characteristics.
3. **Development of Ethical Protocols:** The formulation of robust ethical guidelines is essential to ensure the responsible application of AI technologies. This includes strategies for mitigating biases inherent in data and algorithms and enhancing transparency in the decision-making processes of AI systems.
4. **Improvement in Economic Viability:** It is crucial to explore cost-effective strategies for the implementation of AI solutions within clinical practice. An analysis of economic sustainability will ultimately support the broader adoption and integration of AI technologies in the healthcare sector.

By addressing these areas, future research can significantly contribute to the advancement of AI methodologies, ensuring they are both practical and ethically sound in clinical applications.

**Open peer review.** To view the open peer review materials for this article, please visit <http://doi.org/10.1017/pcm.2025.10006>.

**Data availability statement.** Data sharing not applicable – no new data generated.

**Author contribution.** AJ and AS made substantial contributions to the conception and design of the work. A.J. drafted the work and revised it critically for important intellectual content. A.S. was responsible for final revision and approval of the version to be published.

**Financial support.** This research received no specific grant from any funding agency, commercial or not-for-profit sectors.

**Competing interests.** The authors declare none.

### References

- Abele N, Tiemann K, Krech T, Wellmann A, Schaaf C, Länger F, Peters A, Donner A, Keil F, Daifalla K, Mackens M, Mamilos A, Minin E, Krümmelbein M, Krause L, Stark M, Zapf A, Papper M, Hartmann A and Lang T (2023) Noninferiority of artificial intelligence-assisted analysis of Ki-67 and estrogen/progesterone receptor in breast cancer routine diagnostics. *Modern Pathology: An Official Journal of the United States and Canadian Academy of Pathology Inc* 36(3), 100033. <https://doi.org/10.1016/J.MODPAT.2022.100033>.
- Aeffner F, Zarella MD, Buchbinder N, Bui MM, Goodman MR, Hartman DJ, Lujan GM, Molani MA, Parwani AV, Lillard K, Turner OC, Vemuri VNP, Yuil-Valdes AG and Bowman D (2019) Introduction to digital image analysis in whole-slide imaging: A white paper from the digital pathology

- association. *Journal of Pathology Informatics* 10(1), 1–19. [https://doi.org/10.4103/JPI.JPI\\_82\\_18](https://doi.org/10.4103/JPI.JPI_82_18).
- Ahn JS, Shin S, Yang SA, Park EK, Kim KH, Cho SI, Ock CY and Kim S (2023) Artificial intelligence in breast cancer diagnosis and personalized medicine. *Journal of Breast Cancer* 26(5), 405–435. <https://doi.org/10.4048/jbc.2023.26.e45>.
- Al-Hamadani MN, Fadhel MA, Alzubaidi L and Harangi B (2024) Reinforcement learning algorithms and applications in healthcare and robotics: A comprehensive and systematic review. *Sensors* 24(8), 2461. <https://doi.org/10.3390/s24082461>.
- Alajanbi M, Malerba D and Liu H (2021) Distributed reduced convolution neural networks. *Mesopotamian Journal of Big Data* 20, 25–28. <https://doi.org/10.58496/MJBD/2021/005>.
- Aleskandarany MA, Green AR, Benhasouna AA, Barros FF, Neal K, Reis-Filho JS, Ellis IO and Rakha EA (2012) Prognostic value of proliferation assay in the luminal, HER2-positive, and triple-negative biologic classes of breast cancer. *Breast Cancer Research* 14(1), R3.
- Alowais SA, Alghamdi SS, Alsuhebany N, Alqatani T, Alshaya AI, Almo-hareb SN, Aldairem A, Alrashed M, Bin Saleh K, Badreldin HA and Al Yami MS (2023) Revolutionizing healthcare: The role of artificial intelligence in clinical practice. *BMC Medical Education* 23(1), 689. <https://doi.org/10.1186/s12909-023-04698-z>.
- Alshuhri MS, Al-Musawi SG, Al-Alwany AA, Uinarni H, Rasulova I, Rodrigues P, Alkhafaji AT, Alshanberri AM, Alawadi AH and Abbas AH (2024) Artificial intelligence in cancer diagnosis: Opportunities and challenges. *Pathology- Research and Practice* 253, 154996. <https://doi.org/10.1016/j.prp.2023.154996>.
- Amin MS and Ahn H (2023) FabNet: A features agglomeration-based convolutional neural network for multiscale breast cancer histopathology images classification. *Cancers* 15(4), 1013. <https://doi.org/10.3390/cancers15041013>.
- Anyoha R (2017) The history of artificial intelligence. *Science in the News-Harvard Graduate School of Arts and Sciences*. Available at <https://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/>.
- Aswathy MA and Jagannath M (2021) An SVM approach towards breast cancer classification from H&E-stained histopathology images based on integrated features. *Medical & Biological Engineering & Computing* 59(9), 1773–1783. <https://doi.org/10.1007/s11517-021-02403-0>.
- Aresta G, Araújo T, Kwok S, Chennamsetty SS, Safwan M, Alex V, Marami B, Prastawa M, Chan M, Donovan M, Fernandez G, Zeineh J, Kohl M, Wlaz C, Ludwig F, Braunewell S, Baust M, Vu QD, To MNN, Kim E, Kwak JT, Galal S, Sanchez-Freire V, Brancati N, Frucci M, Riccio D, Wang Y, Sun L, Ma K, Fang J, Kone I, Boulmane L, Campilho A, Eloy C, Polónia A and Aguiar P (2019) Bach: Grand challenge on breast cancer histology images. *Medical Image Analysis* 56, 122–139. <https://doi.org/10.1016/j.media.2019.05.010>.
- Attallah O, Anwar F, Ghanem NM and Ismail MA (2021) Histo-CADx: Duo cascaded fusion stages for breast cancer diagnosis from histopathological images. *PeerJ Computer Science* 7, e493. <https://doi.org/10.7717/PEERJ-CS.493>.
- Aubreville M, Stathonikos N, Donovan TA, Klopffleisch R, Ammeling J, Ganz J, Wilm F, Veta M, Jabari S, Eckstein M, Annuscheit J, Krumnow C, Bozaba E, Çayır S, Gu H, Chen X'A', Jahanifarf M, Shephard A, Kondo S, Kasai S, Kotte S, Saipradeep VG, Lafarge MW, Koelzer VH, Wang Z, Zhang Y, Yang S, Wang X, Breininger K and Bertram CA (2024) Domain generalization across tumor types, laboratories, and species – Insights from the 2022 edition of the mitosis domain generalization challenge. *Medical Image Analysis* 94, 103155. <https://doi.org/10.1016/j.media.2024.103155>.
- Bae K, Jeon YS, Hwangbo Y, Yoo CW, Han N and Feng M (2023) Data-efficient computational pathology platform for faster and cheaper breast cancer subtype identifications: Development of a deep learning model. *JMIR Cancer* 9(1), e45547. <https://doi.org/10.2196/45547>.
- Bajić F, Orel O and Habijan M (2022) A multi-purpose shallow convolutional neural network for chart images. *Sensors* 22(20), 7695. <https://doi.org/10.3390/s22207695>.
- Basu K, Sinha R, Ong A and Basu T (2020) Artificial intelligence: How is it changing medical sciences and its future? *Indian Journal of Dermatology* 65(5), 365–370. [https://doi.org/10.4103/ijd.IJD\\_421\\_20](https://doi.org/10.4103/ijd.IJD_421_20).
- Bejnordi BE, Veta M, Van Diest PJ, Van Ginneken B, Karssemeijer N, Litjens G, Van Der Laak JA, Hermesen M, Manson QF, Balkenhol M and Geessink O (2017) Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA* 318(22), 2199–2210. <https://doi.org/10.1001/JAMA.2017.14585>.
- Bejnordi BE, Zuidhof G, Balkenhol M, Hermesen M, Bult P, van Ginneken B, Karssemeijer N, Litjens G and van der Laak J (2017) Context-aware stacked convolutional neural networks for classification of breast carcinomas in whole-slide histopathology images. *Journal of Medical Imaging* 4(4), 044504. <https://doi.org/10.1117/1.JMI.4.4.044504>.
- Birhane A, Kasirzadeh A, Leslie D and Wachter S (2023) Science in the age of large language models. *Nature Reviews Physics* 5(5), 277–280. <https://doi.org/10.1038/s42254-023-00581-4>.
- Bitkina OV, Park J and Kim HK (2023) Application of artificial intelligence in medical technologies: A systematic review of main trends. *DIGITAL HEALTH* 9, 20552076231189331. <https://doi.org/10.1177/20552076231189331>.
- Briganti G (2023) Intelligence artificielle: Une introduction pour les cliniciens. *Revue des Maladies Respiratoires* 40(4), 308–313. <https://doi.org/10.1016/j.rmr.2023.02.005>.
- Brown S (2021) Machine learning, explained. MIT Sloan. Ideas Made to Matter. Available at <https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained> (accessed 4 May 2023).
- Cardoso F, Kyriakides S, Ohno S, Penault-Llorca F, Poortmans P, Rubio IT, Zackrisson S and Senkus E, ESMO Guidelines Committee. Electronic address: [clinicalguidelines@esmo.org](mailto:clinicalguidelines@esmo.org) (2019) ESMO guidelines committee. Early breast cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. *Annals of Oncology* 30(8), 1194–220. <https://doi.org/10.1093/annonc/mdz173>.
- Carvalho ED, Antonio Filho OC, Silva RR, Araujo FH, Diniz JO, Silva AC, Paiva AC and Gattass M (2020) Breast cancer diagnosis from histopathological images using textural features and CBIR. *Artificial Intelligence in Medicine* 105, 101845. <https://doi.org/10.1016/J.ARTMED.2020.101845>.
- Chen M and Decary M (2020) Artificial intelligence in healthcare: An essential guide for health leaders. *Healthcare Management Forum*, 18 33(1), 10. <https://doi.org/10.1177/0840470419873123>.
- Cohen MB, Martin M, Gross DJ, Johnson K, Robboy SJ, Wheeler TM, Johnson RL and Black-Schaffer WS (2022) Features of burnout amongst pathologists: A reassessment. *Academic Pathology* 9(1), 100052. <https://doi.org/10.1016/J.ACPATH.2022.100052>.
- Cree IA, Tan PH, Travis WD, Wesseling P, Yagi Y, White VA, Lokuhetty D and Scolyer RA (2021) Counting mitoses: SI (ze) matters! *Modern Pathology* 34(9), 1651–1657. <https://doi.org/10.1038/s41379-021-00825-7>.
- Cruz-Roa A, Gilmore H, Basavanthally A, Feldman M, Ganesan S, Shih NN, Tomaszewski J, González FA and Madabhushi A (2017) Accurate and reproducible invasive breast cancer detection in whole-slide images: A deep learning approach for quantifying tumor extent. *Scientific Reports* 7(1), 46450. <https://doi.org/10.1038/srep46450>.
- Egger J, Gsaxner C, Pepe A, Pomykala KL, Jonske F, Kurz M, Li J and Kleesiek J (2022) Medical deep learning – A systematic meta-review. *Computer Methods and Programs in Biomedicine* 221, 106874. <https://doi.org/10.1016/j.cmpb.2022.106874>.
- Elmore JG, Barnhill RL, Elder DE, Longton GM, Pepe MS, Reisch LM, Carney PA, Titus LJ, Nelson HD, Omega T, Tosteson AN, Weinstock MA, Knezevich SR and Piepkorn MW (2017) Pathologists' diagnosis of invasive melanoma and melanocytic proliferations: Observer accuracy and reproducibility study. *BMJ* 357, j2813. <https://doi.org/10.1136/bmj.j2813>.
- Elmore JG, Longton GM, Carney PA, Geller BM, Omega T, Tosteson AN, Nelson HD, Pepe MS, Allison KH, Schnitt SJ, O'Malley FP and Weaver DL (2015) Diagnostic concordance among pathologists interpreting breast biopsy specimens. *JAMA* 313(11), 1122–1132. <https://doi.org/10.1001/JAMA.2015.1405>.
- Försch S, Klauschen F, Hufnagl P and Roth W (2021) Artificial intelligence in pathology. *Deutsches Ärzteblatt International* 118(12), 199–204. <https://doi.org/10.3238/ARZTEBL.M2021.0011>.
- Gandomkar Z, Brennan PC and Mello-Thoms C (2018) MuDeRN: Multi-category classification of breast histopathological image using deep residual networks. *Artificial Intelligence in Medicine* 88, 14–24. <https://doi.org/10.1016/J.ARTMED.2018.04.005>.

- Grzybowski A, Pawlikowska-Lagód K and Lambert WC (2024) A history of artificial intelligence. *Clinics in Dermatology* 42(3), 221–229. <https://doi.org/10.1016/j.clindermatol.2023.12.016>.
- Han Z, Wei B, Zheng Y, Yin Y, Li K and Li S (2017) Breast cancer multi-classification from histopathological images with structured deep learning model. *Scientific Reports* 7(1), 4172. <https://doi.org/10.1038/s41598-017-04075-z>.
- Hanna MG, Pantanowitz L (2017) Why is digital pathology in cytopathology lagging behind surgical pathology? *Cancer Cytopathol.* 125(7):519–520. <https://doi.org/10.1002/cncy.21855>.
- Hanna MG, Reuter VE, Hameed MR, Tan LK, Chiang S, Sigel C, Hollmann T, Giri D, Samboy J, Moradel C, Rosado A, Otilano JR, III, England C, Corsale L, Stamelos E, Yagi Y, Schöffler PJ, Fuchs T, Klimstra DS and Sirintrapun SJ (2019) Whole slide imaging equivalency and efficiency study: Experience at a large academic center. *Modern Pathology* 32(7), 916–928. <https://doi.org/10.1038/s41379-019-0205-0>.
- Hatta S, Ichiuji Y, Mabu S, Kugler M, Hontani H, Okoshi T, Fuse H, Kawada T, Kido S, Imamura Y, Naiki H and Inai K (2023) Improved artificial intelligence discrimination of minor histological populations by supplementing with color-adjusted images. *Scientific Reports* 13(1), 19068. <https://doi.org/10.1038/s41598-023-46472-7>.
- Hosny A, Parmar C, Quackenbush J, Schwartz LH and Aerts HJ (2018) Artificial intelligence in radiology. *Nature Reviews Cancer* 18(8), 500–510. <https://doi.org/10.1038/s41568-018-0016-5>.
- He Z, Chen Z, Tan M, Elingarami S, Liu Y, Li T, Deng Y, He N, Li S, Fu J and Li W (2020) A review on methods for diagnosis of breast cancer cells and tissues. *Cell Proliferation* 53(7), e12822. <https://doi.org/10.1111/cpr.12822>.
- ICIAIR 2018 - Grand Challenge. Available at <https://iciar2018-challenge.grand-challenge.org/> (accessed 09 July 2025).
- Jiang Y, Luo J, Huang D, Liu Y and Li DD (2022) Machine learning advances in microbiology: A review of methods and applications. *Frontiers in Microbiology* 13, 925454. <https://doi.org/10.3389/fmicb.2022.925454>.
- Jairam MP and Ha R (2022) A review of artificial intelligence in mammography. *Clinical Imaging* 88, 36–44. <https://doi.org/10.1016/j.clinimag.2022.05.005>.
- Jones NC, Nazarian RM, Duncan LM, Kamionek M, Lauwers GY, Tambouret RH, Wu CL, Nielsen GP, Brachtel EF, Mark EJ, Sadow PM, Grabbe JP and Wilbur DC (2015) Interinstitutional whole slide imaging teleconsultation service development: Assessment using internal training and clinical consultation cases. *Archives of Pathology and Laboratory Medicine* 139(5), 627–635. <https://doi.org/10.5858/ARPA.2014-0133-OA>.
- Jovel J and Greiner R (2021) An introduction to machine learning approaches for biomedical research. *Frontiers in Medicine* 8, 771607. <https://doi.org/10.3389/fmed.2021.771607>.
- Jung M, Song SG, Cho SI, Shin S, Lee T, Jung W, Lee H, Park J, Song S, Park G, Song H, Park S, Lee J, Kang M, Park J, Pereira S, Yoo D, Chung K, Ali SM and Kim SW (2024) Augmented interpretation of HER2, ER, and PR in breast cancer by artificial intelligence analyzer: Enhancing interobserver agreement through a reader study of 201 cases. *Breast Cancer Research* 26(1), 31. <https://doi.org/10.1186/s13058-024-01784-y>.
- Kanavati F, Ichihara S and Tsuneki M (2022) A deep learning model for breast ductal carcinoma in situ classification in whole slide images. *Virchows Archiv.* 480(5), 1009–1022. <https://doi.org/10.1007/S00428-021-03241-Z>.
- Kaul V, Enslin S and Gross SA (2020) History of artificial intelligence in medicine. *Gastrointestinal Endoscopy* 92(4), 807–812. <https://doi.org/10.1016/j.gie.2020.06.040>.
- Kim I, Kang K, Song Y and Kim TJ (2022) Application of artificial intelligence in pathology: Trends and challenges. *Diagnostics* 12(11), 2794. <https://doi.org/10.3390/diagnostics12112794>.
- Kolla B (2024) An integrated approach for magnification independent breast cancer classification. *Biomedical Signal Processing and Control* 88, 105594. <https://doi.org/10.1016/j.bspc.2023.105594>.
- Kriegeskorte N and Golan T (2019) Neural network models and deep learning. *Current Biology* 29(7), R231–R236. <https://doi.org/10.1016/j.cub.2019.02.034>.
- Kufel J, Bargiel-Łączek K, Kocot S, Kozlik M, Bartnikowska W, Janik M, Czogalik Ł, Dudek P, Magiera M, Lis A, Paszkiewicz I, Nawrat Z, Cebula M and Gruszczynska K (2023) What is machine learning, artificial neural networks and deep learning?—examples of practical applications in medicine. *Diagnostics* 13(15), 2582. <https://doi.org/10.3390/diagnostics13152582>.
- Lanzagorta-Ortega D, Carrillo-Pérez DL and Carrillo-Esper R (2022) Inteligencia artificial en medicina. *presente y futuro .Gaceta medica de Mexico* 158(Supplement 1), 17–21. <https://doi.org/10.24875/GMM.M22000688>.
- Lathwal A, Kumar R, Arora C and Raghava GPS (2020) Identification of prognostic biomarkers for major subtypes of non-small-cell lung cancer using genomic and clinical data. *Journal of Cancer Research and Clinical Oncology* 146(11), 2743–2752. <https://doi.org/10.1007/s00432-020-03318-3>.
- Laxmisagar HS and Hanumantharaju MC (2022) Detection of breast cancer with lightweight deep neural networks for histology image classification. *Critical Reviews in Biomedical Engineering* 50(2), 1–19. <https://doi.org/10.1615/CRITREVBIOENGENG.2022043417>.
- Lee J, Warner E, Shaikhouni S, Bitzer M, Kretzler M, Gipson D, Pennathur S, Bellovich K, Bhat Z, Gadegbeku C, Massengill S, Perumal K, Saha J, Yang Y, Luo J, Zhang X, Mariani L, Hodgins JB, Rao A and C-PROBE Study (2022) Unsupervised machine learning for identifying important visual features through bag-of-words using histopathology data from chronic kidney disease. *Scientific Reports* 12(1), 4832. <https://doi.org/10.1038/s41598-022-08974-8>.
- Liu B, Chi W, Li X, Li P, Liang W, Liu H, Wang W and He J (2020) Evolving the pulmonary nodules diagnosis from classical approaches to deep learning-aided decision support: Three decades' development course and future prospect. *Journal of Cancer Research and Clinical Oncology* 146(1), 153–185. <https://doi.org/10.1007/s00432-019-03098-5>.
- Liu L, Feng W, Chen C, Liu M, Qu Y and Yang J (2022) Classification of breast cancer histology images using MSMV-PFNet. *Scientific Reports* 12(1), 1–10. <https://doi.org/10.1038/s41598-022-22358-y>.
- Liu M, Hu L, Tang Y, Wang C, He Y, Zeng C, Lin K, He Z and Huo W (2022) A deep learning method for breast cancer classification in the pathology images. *IEEE Journal of Biomedical and Health Informatics* 26(10), 5025–5032. <https://doi.org/10.1109/JBHI.2022.3187765>.
- Loughrey MB, Kelly PJ, Houghton OP, Coleman HG, Carson A, Salto-Tellez M and Hamilton PW (2015) Digital slide viewing for primary reporting in gastrointestinal pathology: A validation study. *Virchows Archiv* 467(2), 137–144. <https://doi.org/10.1007/s00428-015-1780-1>.
- Lu C, Romo-Bucheli D, Wang X, Janowczyk A, Ganesan S, Gilmore H, Rimm D and Madabhushi A (2018) Nuclear shape and orientation features from H&E images predict survival in early-stage estrogen receptor-positive breast cancers. *Laboratory Investigation* 98(11), 1438–1448. <https://doi.org/10.1038/s41374-018-0095-7>.
- Miotto R, Wang F, Wang S, Jiang X and Dudley JT (2018) Deep learning for healthcare: Review, opportunities and challenges. *Briefings in Bioinformatics* 19(6), 1236–1246. <https://doi.org/10.1093/bib/bbx044>.
- Mitos-ATYPIA-14 Grand Challenge. <https://mitos-atypia-14.grand-challenge.org/>.
- Moor J (2006) The Dartmouth College artificial intelligence conference: The next fifty years. *AI Magazine* 27(4), 87–91. <https://www.aaai.org/ojs/index.php/aimagazine/article/view/1904/1802>.
- Morelli P, Porazzi E, Ruspini M, Restelli U and Banfi G (2013) Analysis of errors in histology by root cause analysis: A pilot study. *Journal of Preventive Medicine and Hygiene* 54(2), 90.
- Moxley-Wyles B, Colling R and Verrill C (2020) Artificial intelligence in pathology: An overview. *Diagnostic Histopathology* 26(11), 513–520. <https://doi.org/10.1016/j.mpdhp.2020.08.004>.
- Mukhopadhyay S, Feldman MD, Abels E, Ashfaq R, Beltaifa S, Cacciabeve NG, Cathro HP, Cheng L, Cooper K, Dickey GE, Gill RM, Heaton RP, Jr, Kerstens R, Lindberg GM, Malhotra RK, Mandell JW, Manlucu ED, Mills AM, Mills SE, Moskaluk CA and Taylor CR (2018) Whole slide imaging erus microscopy for primary diagnosis in surgical pathology: A multicenter blinded randomized noninferiority study of 1992 cases (pivotal study). *The American Journal of Surgical Pathology* 42(1), 39–52. <https://doi.org/10.1097/PAS.0000000000000948>.
- Muthukrishnan N, Maleki F, Ovens K, Reinhold C, Forghani B and Forghani R (2020) Brief history of artificial intelligence. *Neuroimaging Clinics of North America* 30(4), 393–399. <https://doi.org/10.1016/j.nic.2020.07.004>.
- Nahid AA, Mehrabi MA and Kong Y (2018) Histopathological breast cancer image classification by deep neural network techniques guided by local

- clustering. *BioMed Research International* 18(1), 2362108. <https://doi.org/10.1155/2018/2362108>.
- Nardin S, Mora E, Varughese FM, D'Avanzo F, Vachanaram AR, Rossi V, Saggia C, Rubinelli S and Gennari A** (2020) Breast cancer survivorship, quality of life, and late toxicities. *Frontiers in Oncology* 10, 864. <https://doi.org/10.3389/fonc.2020.00864>.
- Nateghi R, Danyali H and Helfroush MS** (2021) A deep learning approach for mitosis detection: Application in tumor proliferation prediction from whole slide images. *Artificial Intelligence in Medicine* 114, 102048. <https://doi.org/10.1016/j.artmed.2021.102048>.
- Niazi MKK, Parwani AV and Gurcan MN** (2019) Digital pathology and artificial intelligence. *The Lancet Oncology* 20(5), e253–e261. [https://doi.org/10.1016/S1470-2045\(19\)30154-8](https://doi.org/10.1016/S1470-2045(19)30154-8).
- Ono S and Goto T** (2022) Introduction to supervised machine learning in clinical epidemiology. *Annals of Clinical Epidemiology* 4(3), 63–71. <https://doi.org/10.37737/ace.22009>.
- Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE, Chou R, Glanville J, Grimshaw JM, Hróbjartsson A, Lalu MM, Li T, Loder EW, Mayo-Wilson E, McDonald S, McGuinness L, Stewart LA, Thomas J, Tricco AC, Welch VA, Whiting P and Moher D** (2021) The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* 372, n71. <https://doi.org/10.1136/bmj.n71>.
- Pantanowitz L, Hartman D, Qi Y, Cho EY, Suh B, Paeng K, Dhir R, Michelow P, Hazelhurst S, Song SY and Cho SY** (2020) Accuracy and efficiency of an artificial intelligence tool when counting breast mitoses. *Diagnostic Pathology* 15(1), 80. <https://doi.org/10.1186/s13000-020-00995-z>.
- Pettit RW, Fullem R, Cheng C and Amos CI** (2021) Artificial intelligence, machine learning, and deep learning for clinical outcome prediction. *Emerging Topics in Life Sciences* 5(6), 729–745. <https://doi.org/10.1042/ETLS20210246>.
- Polónia A, Campelos S, Ribeiro A, Aymore I, Pinto D, Biskup-Fruzynska M, Veiga RS, Canas-Marques R, Aresta G, Araújo T, Campilho A, Kwok S, Aguiar P and Eloy C** (2021) Artificial intelligence improves the accuracy in histologic classification of breast lesions. *American Journal of Clinical Pathology* 155(4), 527–536. <https://doi.org/10.1093/ajcp/aqaa151>.
- Ragab M, Al-Ghamdi ASA, Fakieh B, Choudhry H, Mansour RF and Koundal, D** (2022) Prediction of diabetes through retinal images using deep neural network. *Computational Intelligence and Neuroscience* 7887908. <https://doi.org/10.1155/2022/7887908>.
- Robbins CJ, Fernandez AI, Han G, Wong S, Harigopal M, Podoll M, Singh K, Ly A, Kuba MG, Wen H, Sanders MA, Brock J, Wei S, Fadare O, Hanley K, Jorns J, Snir OL, Yoon E, Rabe K, Soong TR, Reisenbichler ES and Rimm DL** (2023) Multi-institutional assessment of pathologist scoring HER2 immunohistochemistry. *Modern Pathology* 36(1), 100032.
- Robboy SJ, Gross D, Park JY, Kittle E, Crawford JM, Johnson RL, Cohen MB, Karcher DS, Hoffman RD, Smith AT and Black-Schaffer WS** (2020) Reevaluation of the US pathologist workforce size. *JAMA Network Open* 3(7), e2010648. <https://doi.org/10.1001/JAMANETWORKOPEN.2020.10648>.
- Romo-Bucheli D, Janowczyk A, Gilmore H, Romero E and Madabhushi A** (2016) Automated tubule nuclei quantification and correlation with onco-type DX risk categories in ER+ breast cancer whole slide images. *Scientific Reports* 6, 32706. <https://doi.org/10.1038/srep32706>.
- Romo-Bucheli D, Janowczyk A, Gilmore H, Romero E and Madabhushi A** (2017) A deep learning based strategy for identifying and associating mitotic activity with gene expression derived risk categories in estrogen receptor positive breast cancers. *Cytometry Part A* 91(6), 566–573. <https://doi.org/10.1002/cyto.a.23065>.
- Roux L, Racoceanu D, Loménie N, Kulikova M, Irshad H, Klossa J, Capron F, Genestie C, Le Naour G and Gurcan MN** (2013) Mitosis detection in breast cancer histological images; An ICPR 2012 contest. *Journal of Pathology Informatics* 4, 8. <https://doi.org/10.4103/2153-3539.112693>.
- Saha M, Chakraborty C and Racoceanu D** (2018) Efficient deep learning model for mitosis detection using breast histopathology images. *Computerized Medical Imaging and Graphics* 64, 29–40. <https://doi.org/10.1016/j.COMPMEDIMAG.2017.12.001>.
- Sarker IH** (2021) Deep learning: A comprehensive overview on techniques, taxonomy, applications and research directions. *SN Computer Science* 2(6), 420. <https://doi.org/10.1007/s42979-021-00815-1>.
- Sedeta ET, Jobre B and Avezbakiyev B** (2023) Breast cancer: Global patterns of incidence, mortality, and trends. *Journal of Clinical Oncology* 41(16\_suppl), 10528–10528. [https://doi.org/10.1200/jco.2023.41.16\\_suppl.10528](https://doi.org/10.1200/jco.2023.41.16_suppl.10528).
- Senousy Z, Abdelsamea MM, Gaber MM, Abdar M, Acharya UR, Khosravi A and Nahavandi S** (2022) MCUa: Multi-level context and uncertainty aware dynamic deep ensemble for breast cancer histology image classification. *IEEE Transactions on Biomedical Engineering* 69(2), 818–829. <https://doi.org/10.1109/TBME.2021.3107446>.
- Shafi S, Kellough DA, Lujan G, Satturwar S, Parwani AV and Li Z** (2022) Integrating and validating automated digital imaging analysis of estrogen receptor immunohistochemistry in a fully digital workflow for clinical use. *Journal of Pathology Informatics* 13, 100122. <https://doi.org/10.1016/J.JPI.2022.100122>.
- Shajari S, Kuruvinashetti K, Komeili A, Sundararaj U** (2023) The emergence of AI-based wearable sensors for digital health technology: A review. *Sensors* 29(23), 9498. <https://doi.org/10.3390/s23239498>.
- Shamshiri MA, Krzyżak A, Kowal M and Korbicz J** (2023) Compatible-domain transfer learning for breast cancer classification with limited annotated data. *Computers in Biology and Medicine* 154, 106575. <https://doi.org/10.1016/j.COMPBIMED.2023.106575>.
- Sidey-Gibbons JAM and Sidey-Gibbons CJ** (2019) Machine learning in medicine: A practical introduction. *BMC Medical Research Methodology* 19(1), 64. <https://doi.org/10.1186/s12874-019-0681-4>.
- Singh P, Kumar R, Gupta M and Al-Turjman F** (2024) SegEIR-net: A robust histopathology image analysis framework for accurate breast cancer classification. *Current Medical Imaging* 20, e15734056278974. <https://doi.org/10.2174/0115734056278974231211102917>.
- Soliman A, Li Z and Parwani AV** (2024) Artificial intelligence's impact on breast cancer pathology: A literature review. *Diagnostic Pathology* 19(1), 1–18. <https://doi.org/10.1186/S13000-024-01453-W/FIGURES/2>.
- Spanhol FA, Oliveira LS, Petitjean C and Heutte L** (2016) A dataset for breast cancer histopathological image classification. *IEEE Transactions on Biomedical Engineering* 63(7), 1455–1462. [10.1109/TBME.2015.2496264](https://doi.org/10.1109/TBME.2015.2496264).
- Srikantamurthy MM, Rallabandi VPS, Dudekula DB, Natarajan S and Park J** (2023) Classification of benign and malignant subtypes of breast cancer histopathology imaging using hybrid CNN-LSTM based transfer learning. *BMC Medical Imaging* 23(1), 19. <https://doi.org/10.1186/S12880-023-00964-0>.
- Suberi AA, Zakaria WN and Tomari R** (2017) Dendritic cell recognition in computer aided system for cancer immunotherapy. *Procedia Computer Science* 105, 177–182. <https://doi.org/10.1016/j.procs.2017.01.201>.
- Sufyan M, Shokat Z and Ashfaq UA** (2023) Artificial intelligence in cancer diagnosis and therapy: Current status and future perspective. *Computers in Biology and Medicine* 165, 107356. <https://doi.org/10.1016/j.compbiomed.2023.107356>.
- Tabata K, Mori I, Sasaki T, Itoh T, Shiraishi T, Yoshimi N, Maeda I, Harada O, Taniyama K, Taniyama D, Watanabe M, Mikami Y, Sato S, Kashima Y, Fujimura S and Fukuoka J** (2017) Whole-slide imaging at primary pathological diagnosis: Validation of whole-slide imaging-based primary pathological diagnosis at twelve Japanese academic institutes. *Pathology International* 67(11), 547–554. <https://doi.org/10.1111/PIN.12590>.
- Tizhoosh HR and Pantanowitz L** (2018) Artificial intelligence and digital pathology: Challenges and opportunities. *Journal of Pathology Informatics* 9, 38. [https://doi.org/10.4103/jpi.jpi\\_53\\_18](https://doi.org/10.4103/jpi.jpi_53_18).
- Turning A** (1950) I.—Computing machinery and intelligence. *Mind* 236, 433–460. <https://doi.org/10.1093/mind/lix.236.433>.
- Umer MJ, Sharif M, Kadry S and Alharbi A** (2022) Multi-class classification of breast cancer using 6B-net with deep feature fusion and selection method. *Journal of Personalized Medicine* 12(5), 683. <https://doi.org/10.3390/JPM12050683/S1>.
- van Diest PJ, Flach RN, van Dooijeweert C, Makineli S, Breimer GE, Stathonikos N, Pham P, Nguyen TQ and Veta M** (2024) Pros and cons of artificial intelligence implementation in diagnostic pathology. *Histopathology* 84(6), 924–934. <https://doi.org/10.1111/HIS.15153>.

- Veta M, Heng YJ, Stathonikos N, Bejnordi BE, Beca F, Wollmann T, Rohr K, Shah MA, Wang D, Rousson M, Hedlund M, Tellez D, Ciompi F, Zerhouni E, Lanyi D, Viana M, Kovalev V, Liauchuk V, Phoulady HA, Qaiser T and Pluim JPW (2019) Predicting breast tumor proliferation from whole-slide images: The TUPAC16 challenge. *Medical Image Analysis* **54**, 111–121. <https://doi.org/10.1016/j.media.2019.02.012>.
- van Dooijeweert C, van Diest PJ and Ellis IO (2021) Grading of invasive breast carcinoma: The way forward. *Virchows Archiv* **480**(1), 33–43. <https://doi.org/10.1007/s00428-021-03141-2>.
- Wang X, Zhang J, Yang S, Xiang J, Luo F, Wang M, Zhang J, Yang W, Huang J and Han X (2023) A generalizable and robust deep learning algorithm for mitosis detection in multicenter breast histopathological images. *Medical Image Analysis* **84**, 102703. <https://doi.org/10.1016/j.MEDIA.2022.102703>.
- Watkins EJ (2019) Overview of breast cancer. *Journal of the American Academy of Physician Assistants* **32**(10), 13–17. <https://doi.org/10.1097/01.JAA.0000580524.95733.3d>.
- Whitney J, Corredor G, Janowczyk A, Ganesan S, Doyle S, Tomaszewski J, Feldman M, Gilmore H and Madabhushi A (2018) Quantitative nuclear histomorphometry predicts oncotype DX risk categories for early stage ER+ breast cancer. *BMC Cancer* **18**(1), 610. <https://doi.org/10.1186/s12885-018-4448-9>.
- Xu N, Yang D, Arikawa K and Bai C (2023) Application of artificial intelligence in modern medicine. *Clinical eHealth* **6**, 130–137. <https://doi.org/10.1016/j.ceh.2023.09.001>.
- Xu X, An M, Zhang J, Liu W and Lu L (2022) A high-precision classification method of mammary cancer based on improved denseNet driven by an attention mechanism. *Computational and Mathematical Methods in Medicine* **22**, 8585036. <https://doi.org/10.1155/2022/8585036>.
- Zhu C, Song F, Wang Y, Dong H, Guo Y and Liu J (2019) Breast cancer histopathology image classification through assembling multiple compact CNNs. *BMC Medical Informatics and Decision Making* **19**(1), 1–17. <https://doi.org/10.1186/S12911-019-0913-X/TABLES/9>.