

Artificial Intelligence For Histopathological Breast Cancer Diagnosis: A Comprehensive Survey

Somnath Ashok Gade¹, Dr. Shashank Swami²

¹Department of Computer Science and Engineering, Vikrant University, MP, India

²Department of Computer Science and Engineering, Associate Professor, Vikrant University, MP, India

Corresponding Author: Somnath Ashok Gade, Dr. Shashank Swami

Abstract:- Traditional histopathology and digital pathology and computational image analysis are now part of breast cancer diagnosis. Hematoxylin and eosin (H&E) staining remains the most important method for diagnosis and subsequent classification and grading, and immunohistochemistry and molecular assays offer additional data when used in conjunction with H&E staining that is important for treatment planning. With the advent of digitization of slides into whole-slide images (WSI), there is a new opportunity to utilize Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL) to solve problems such as tissue and tumour segmentation, cancer detection, histologic classification, cancer grading, detection of lymph-node metastases, quantification of biomarkers, prediction of molecular subtypes and modelling of outcomes. This survey summarizes the technical pipeline from slide preparation and scanning to preprocessing, patch extraction, representation learning and aggregation and clinical decision support. Special attention is paid to convolutional neural networks (CNNs), transfer learning, multiple-instance learning (MIL), attention mechanisms and Vision Transformers, explainability, uncertainty estimation and domain adaptation. Public data sets including BACH, CAMELYON16/17, TUPAC16 and associated data repositories are discussed as well as their advantages and disadvantages. The survey also investigates dataset biases, staining and scanner variation, label noise, class imbalance, leakage, and external validation as well as the discrepancy between retrospective benchmark performance and the usefulness of these labels in the clinical setting. Recent reviews suggest significant potential, but widespread adoption will need multicenter testing, consistent reporting, integration into workflows, regulatory monitoring, and evaluation from a human-centered perspective. The paper concludes with directions on research that include foundation models, weakly supervised learning, multimodal pathology, federated learning and resource-aware inference.

Keywords: Artificial intelligence, breast cancer, histopathology, digital pathology, whole-slide imaging, deep learning, convolutional neural networks, multiple-instance learning, Vision Transformers, computational pathology.

1. Introduction

Breast cancer is a heterogeneous disease for which diagnosis relies on a combination of morphology, immunophenotype, molecular characteristics and clinical context. Histopathology is a direct microhistological examination of tissue architecture and cell morphology, and is a reference part of breast cancer diagnosis. The traditional approach is, however, labor intensive – tissue is processed, cut into sections, stained, studied under a microscope and then the interpretation is given by trained pathologists. Counting of mitoses, estimation of ki-67 proliferation, quantifying tumour-infiltrating lymphocytes (TILs), grading of tumour and searching for small metastatic deposits in sections of lymph nodes can be time consuming and subject to inter-observer variation. The field of breast pathology and its use of AI has been divided into these common diagnostic challenges and evidence of its rapid development in digital image-analysis research has been reported [1–4]. Whole slide imaging (WSI) is the process of converting glass slides to very large digital images that can be archived, shared and computationally analyzed. Most AI systems don't look at the whole image right away – a WSI can have hundreds of

millions of pixels. Instead, the slides are divided into smaller tiles, tissue regions are identified, features are extracted and local predictions are combined to form a prediction for the slide. This presents a unique learning challenge as the label may be given at the slide or patient level, but the exact region of the diagnosis is only partially labeled. In the weak-supervision setting, MIL and attention-based approaches solve this problem by learning what are the most informative patches for a slide-level prediction [5-7]. The early breast histopathology systems were based on handcrafted features describing colour, texture, shape and spatial organization and classifiers that included support vector machines, random forests or shallow neural networks. With the introduction of CNNs, it became possible to directly learn hierarchical visual representations from image data. Transfer learning was implemented from natural-image datasets which reduced the need for annotated pathology data, and subsequent research using self-supervised and foundation-model approaches aimed to learn a representation from large numbers of unlabeled images specific to pathology [8-11].

This survey does not aim to give AI a role as a substitute for the pathologist's role, but rather to combine the computational approaches to support the pathologist's role in the workflow. Thus, the review spans the entire analytical process, from data acquisition to data pre-processing, model validation and clinical translation. It also helps to differentiate between benchmark performance and evidence of clinical benefit, as high accuracy on a curated data set is not robust enough to imply robustness across scanners, labs, populations or real world workflows.

1.1 Scope of the Survey

The survey concentrates on breast imaging in a histopathological context: H&E WSIs and microscopy images, selected discussion of immunohistochemical images and multimodal pathology. It covers diagnostic classification, segmentation, grading, mitotic analysis, detection of lymph-node metastasis, biomarker quantification, molecular-subtype prediction, and emerging multimodal approaches, all of which are used to help predict prognosis or molecular subtypes. The literature focus is on peer-reviewed studies and reviews, and on large public datasets, especially those that are published in 2024-2026 [1-4, 12-16].

2. Objectives

The main goal of this survey is to provide a systematic description of the methods that use artificial intelligence (AI) to diagnose breast cancer from histopathological data and to highlight the methodological aspects, which might affect the generalisation of the results from the research datasets to other contexts. Specific objectives are:

1. To explain the digital histopathology workflow and some of the features that make breast tissue analysis unique from traditional natural image classification.
2. To revisit the conventional machine learning, CNN based deep learning, transfer learning, MIL, attention based models and Vision Transformers for breast histopathology.
3. Summary of major public datasets and benchmark tasks (BACH, CAMELYON16/17 and TUPAC16), and what they can and cannot establish.
4. To categorize AI applications: diagnostic detection, tissue classification, segmentation, tumor grading, mitosis counting, lymph-node metastasis detection, biomarker quantification and molecular-subtype prediction.
5. To understand the comparison of the evaluation measure including accuracy, sensitivity, specificity, precision, F1-score, AUC, Dice coefficient, and calibration and the importance of patient-level splitting and external validation.

6. To explore constraints on staining variation, scanner effects, class imbalance, annotation quality, data leakage, domain shift, interpretability, and computational cost and clinical workflow integration.
7. To discuss ethical, regulatory and operational needs for retroactive algorithms to decision-support systems.
8. To uncover research opportunities in self-supervised learning, pathology foundation models, multimodal AI, federated learning, uncertainty estimation and prospective evaluation.

2.1 Review Perspective

The recent literature has demonstrated that there is a shift from small image sets, curated manually, to weakly supervised learning and WSI analysis. According to the 2022 systematic review, there are six recurrent breast pathology AI tasks: molecular and hormonal analysis, grading, counting of mitotic figures, indexing of Ki-67, TIL assessment, and detection of lymph-node metastasis [1]. Later reviews focus on molecular prediction, treatment-response modeling, prognosis and transformer-based architectures [2–4, 12]. In this survey we follow this overall trend and maintain the endpoint of the clinical diagnosis.

3. Methods

Literature was collected from PubMed-indexed reviews and primary research, and additional search for landmark data sets and representative studies in digital breast pathology. Peer-reviewed publications describing model architecture, dataset composition, validation design and clinically relevant endpoints were given priority. To determine emerging directions, the most recent reviews published between 2024 and 2026 were identified, with the results corroborated with the earlier landmark studies [2–4, 12–16].

3.1 Inclusion Themes

Studies that used histopathological breast tissue images and an ML/DL method were deemed relevant if they were applied to at least one of the following: cancer detection, benign/malignant classification, in situ versus invasive carcinoma classification, tissue segmentation, tumor grading, mitosis detection, lymph-node metastasis detection, biomarker quantification, molecular-subtype prediction, prognosis or treatment-response prediction. Studies were excluded from the main scope if they focused exclusively on mammography or ultrasound, but the distinction between radiological and histopathological AI is mentioned when relevant.

3.2 Analytical Framework

The evidence was summarized across five themes: (i) data source and granularity of the annotation; (ii) computation representation and architecture; (iii) prediction target; (iv) strategy for validation; and (v) requirements for clinical translation. It is critical to note that the same reported accuracy could mean entirely different stories depending on whether the model was tested on a random sample of patches from the same slides where the model was trained, or on patients completely different from the ones in the training set, from another institution. Careful control of tile-level dependence and patient-level leakage are also factors that must be controlled in WSI studies.

3.3 Technical Pipeline

The generalized AI pathology pipeline includes the acquisition, quality control, tissue/background segmentation, colour normalisation, tiling, feature extraction, model inference and aggregation. Depending on task, the pipeline can also involve nuclei detection, cell-type classification, segmentation and/or selection of informative patches of tumor regions. In the task of WSI classification, MIL regards each slide as a bag of

instances, patches as instances, and the diagnosis of that slide as the bag label. Attention-based MIL learns a weighted average of patch representations, which enables the model to focus on the diagnostically informative regions [5,6].

4. Results

4.1 Data Sources and Public Datasets

Breast pathology AI's performance and generalizability heavily rely on the dataset used. Public datasets are useful for enabling reproducible benchmarking, but many include relatively small numbers of patients, for which data is collected under standardised acquisition conditions or for narrow, standardised tasks. In the case of BACH, the images consist of 400 H&E microscopy images evenly distributed into the classes of normal, benign and in situ carcinoma, and invasive carcinoma, and have a resolution of 1536×2048 pixels at about $0.42 \mu\text{m}$ per pixel [17]. This type of dataset can be leveraged for image-level classification and does not mimic the scale and diversity of WSI practice.

Dataset / resource	Typical task	Strength	Important limitation
BACH / ICIAR 2018	Normal, benign, in situ, invasive classification	Simple four-class benchmark; public	Small image-level dataset; limited WSI diversity
CAMELYON16	Breast cancer metastasis in lymph-node WSI	Large WSI task with expert annotations	Focuses on metastasis detection rather than primary-tumor diagnosis
CAMELYON17	Lymph-node metastasis and case-level assessment	Multiple centers and patient-level organization	Specialized clinical task; substantial computational burden
TUPAC16	Mitotic detection and proliferation scoring	Benchmark for mitotic figures and related tasks	Annotations are task-specific and challenging
TCGA / related cohorts	Molecular and outcome association from WSI	Large retrospective collections; rich clinical metadata	Heterogeneous acquisition and incomplete labels
Institutional WSI cohorts	Diagnosis, grading, biomarkers, prognosis	Clinically realistic and task-specific	Often inaccessible; external reproducibility may be limited

The datasets provided by CAMELYON had a significant impact because they were able to redefine the problem of breast pathology as a WSI detection problem and not as a classification problem in a very limited area. Deep learning systems showed that the presence of lymph-node metastases could be automatically detected, even when these were small and hard to find in large slides [18–20]. They paved the way for computational pathology as a real science field and demonstrated the necessity for multi-center validation.

4.2 Data Preparation

Typically, background removal, detection of tissues, normalization of stains, filtering of artifacts and extraction of patches are performed as part of the pre-processing. The reason why it is especially important to normalize stains with the H&E is due to the impact of fixation, reagent concentration, staining protocols, scanner optics and image compression. The use of methods such as histogram matching, color deconvolution or learned normalization can minimize technical variation, but too much normalization can result in loss of biologically meaningful appearance. Therefore, quality-control rules should be tested for their impact on downstream performance, not their ability to enhance it.

4.3 Machine Learning and Deep Learning Approaches

In most cases, traditional machine learning takes a feature engineering pipeline approach. The regions of the nuclei/tissues are segmented, handcrafted descriptors are extracted and a classifier is used to map the descriptors to a diagnostic category. Morphological information can be represented by texture descriptors like local binary patterns, Haralick features and Gabor filters, as well as shape and color statistics. If the feature representation is carefully designed, the support vector machine, random forest and gradient boosting models can achieve good performance, but they require high-quality and complete features [21]. CNNs brought a new twist to the field by learning hierarchical features from pixels. Common architectures are VGG-like networks, ResNet families, DenseNet, EfficientNet and task-specific encoder-decoder networks. Transfer learning is typical for breast histopathology due to the small number of medical image datasets available for labeling compared to natural-image datasets. A pretrained encoder can be fine-tuned to give low level and mid level features that are helpful with quicker training time. Natural-image pretraining might not be optimal for pathology-specific cellular morphology, however, motivating self-supervised pretraining on large pathology corpora [8-11]. To segment tissue compartments, tumor regions or nuclei, segmentation models like U-Net and its variations are utilized. In instance segmentation, adjacent nuclei can overlap or touch one another. Accurate segmentation can, however, provide auxiliary information for quantitative pathology downstream, such as cell counts, nuclear morphology and spatial neighborhood analysis. Direct end-to-end processing is typically not possible for WSI-level classification. One popular approach is to partition the WSI into patches, feed the CNN or transformer encoder each patch, and fuse the patch embeddings via average pooling, max pooling, recurrent aggregation or MIL. The particular advantage of attention-based MIL is that it can give a learned weighting to instances. The model can then pinpoint a limited number of patches that are relevant to diagnosis even if the remaining portions of the slide are non-tumor tissue [5,6]. The course will introduce.

4.4 Vision Transformers and Foundation Models

Images are encoded as a sequence of tokens in Vision Transformers (ViTs), and relationship between regions is accomplished via self-attention mechanism. Pathology: A transformer-based encoder and hierarchical attention mechanisms can enable the capture of larger spatial context than traditional convolution. There has been a recent shift from CNN based systems to transformer and hybrid systems in breast pathology research [3,4]. Another goal of foundation-model research is to pre-train a pathology representation using extremely large numbers of images that are not labeled, and then fine-tune it to a variety of downstream applications. The question that remains, however, is not just if a foundation model can increase the accuracy of benchmarks, but whether it can enhance the robustness of the institution, stains, scanners and clinical tasks.

4.5 Diagnostic Detection and Classification

The most direct application of histopathology AI is differentiating non-neoplastic tissue, benign lesions and malignant disease. CNNs can be taught features related to nuclear atypia, glandular architecture, stromal alternations, and tumor organisation at image or patch level. The BACH benchmark is an example of a four-class classification of normal, benign, in situ and invasive carcinoma that is feasible [17]. When using image-level classification, however, one has to take into account that a pathologist typically assesses the entire specimen that includes several tissue compartments and potentially several lesions. The task can be better expressed as a hierarchical decision process if it is to be used in a clinical application. In the first stage, regions of tissue and suspicious areas may be identified; in the second stage, the type of the tumor may be classified; in the third stage, the grade of the tumor or other relevant biomarkers may be estimated. Hierarchical design can decrease the demand of the one classifier, and provide task-specific control of quality. On the other hand, errors may accumulate from the bottom-up; therefore, component-level and patient-level performance should be reported at the end of the system. A specific application that has been well studied is the detection of lymph-node metastasis. Deep learning was shown to be able to identify metastatic breast cancer in lymph-node tissue (the CAMELYON16 challenge) and later, AI was used to help pathologists identify suspicious regions [18–20]. These systems show a very good workflow for human and AI interaction: the AI can focus on areas for the pathologist to review, but the pathologist is still responsible for the final diagnosis. The value of the practical application can therefore be evaluated in terms of the change in reading time, the percentage of missed readings and the amount of work.

4.6 Histological Classification and Subtyping

Breast cancer includes invasive ductal and lobular carcinomas, ductal carcinoma in situ and multiple histologic variants. The histological classification depends on the recognition of architecture and cytology in a multi-scale context. AI can measure these patterns consistently, but rare variants of these histology are hard to model because there are not many training examples of these. One major challenge is thus class imbalance. Oversampling, focal losses, class weighting of objectives and few-shot or metric-learning methods can alleviate imbalance, but their effectiveness should be tested on independent instances, not just the training distribution. Molecular subtyping on H&E is a new extension. The prediction of luminal A, luminal B, HER2-enriched and basal-like categories using routine slides has been studied. A recent report which focused on 221 hold-out WSIs with a molecular subtyping of four, including a tumor-region selection step, achieved a macro F1 score of 0.73 [22]. Previous studies also demonstrated the ability of deep learning to generate approximate intrinsic molecular subtypes from H&E images [23]. Although these findings are encouraging, the statistical associations predicted here do not necessarily imply actual replacement measurements of IHCs or molecular assays; they should be interpreted with caution and validated with clinical data.

4.7 Grading, Mitosis and Proliferation

Histologic grade combines tubule formation, nuclear pleomorphism and mitotic activity. Automated grading is appealing because the components are complex, and there are differences in scores among observers. The use of deep learning for Nottingham-style grade prediction and survival analysis has been explored using WSI. In a study conducted in 2022, a deep learning-based approach for the classification of breast cancer grade and prediction of survival from WSIs was used to show potential beyond diagnosis to risk stratification [24]. Pathology problems like mitosis detection is a classic computer vision problem. The goal is the identification of dividing cells and the estimation of the mitotic density that can be used for grading and proliferation assessment. A TUPAC16 challenge

was used as a benchmark for detection and scoring of the proliferation of the mitotic figures. Due to resemblance to apoptotic bodies, lymphocytes and other dark objects, mitoses are hard to detect and annotations can vary. A good system should thus have well-documented annotation protocols and good negative examples.

Ki-67 is an immunohistochemical proliferation marker that is commonly used in breast cancer. Automated image analysis may be used for the detection of positive nuclei, for the estimation of staining intensity and for the calculation of a proliferation index. There is good agreement between automated and manual Ki-67 evaluation in some reports [25,26]. Yet, the clinical interpretation relies on pre analytical factors, staining procedures, scoring criteria, and scoring limit. AI should thus be considered as a repeatable quantification tool, and not as an alternative to standardized pathological practice.

4.8 Tumor-Infiltrating Lymphocytes and Tumor Microenvironment

Tumor-Infiltrating Lymphocytes (TILs) and the Tumor Microenvironment (TME). There are stromal cells, immune cells, extracellular matrix and vascular structures within the tumor microenvironment which may be capable of carrying prognostic information. AI techniques can segment tumor and stromal compartments, detect lymphocytes and quantify their spatial distribution. Manual scoring can be time-consuming and can be coarse in space, so computational TIL assessment is appealing. Deep learning has also been applied to the detection of image-based biomarkers in the tumor microenvironment for prognosis and treatment-response studies [14].

One advantage of spatial analysis over bulk molecular measurements is that it maintains information on the location of cells with respect to the tumor edge and each other. Cells can be modeled as nodes in a graph, with edges that represent spatial proximity, and context can be long-range tissue context modeled as aggregations via a transformer or attention. These methods could be used to describe immune-rich versus immune-poor microenvironments, but may need to be carefully validated against independent molecular or clinical measures of microenvironment immune status to interpret the biological significance of the learned image features.

4.9 Biomarker and Molecular Prediction

One of the key research tracks is to be able to predict biomarkers directly from routine H&E slides. Estrogen receptor (ER), progesterone receptor (PR), HER2, Ki-67, etc. are some of the clinically relevant markers that have been explored with AI models. The allure of the model is that H&E slides are used routinely in diagnosis, so an image-based model might give an extra signal for triage or quality control. It is important to note, however, that predictions of protein expression made using images should be interpreted as probabilistic rather than as direct assays of protein expression. The molecular subtype prediction is more ambitious as this classification relies on the gene-expression pattern and not just on one morphological feature. However, an association that can be reproduced between morphology and molecular state has been identified in deep learning studies. Examples include deep learning models that have achieved the ability to approximate PAM50 subtypes from H&E images, and more recent WSI pipelines that have made four-class predictions of subtype with moderate to strong performance based on cohort and validation design [22,23]. The big scientific question is if these connections hold true from one institution to another, as well as from one population to the next. AI can also be used to predict genomic alterations, mutation status or pathway activity from morphology. These can generate screening hypotheses for low cost, but are likely to be confounded to a significant degree, especially if molecular testing is not available. Laboratory, demographic or disease selection effects

can cause a model to learn correlations that are not a causal molecular phenotype. It is therefore a necessity to have external validation, calibration and biological correlation.

4.10 Prognosis and Treatment Response

In addition to diagnosis, histopathology AI has also been used for prognosis and treatment response prediction. The architecture of tumors, organization of stromal cells, immune infiltration and cellular morphology can be encoded in image-based features. Models relating these features to recurrence, survival and response to therapy have been described. The potential clinical importance is great because the ability to obtain prognostic information might be obtained from the same samples that are already collected for diagnosis. However, there are other needs for prognostic modeling. Outcomes are time-to-event variables, and are affected by treatment, stage, and comorbidities, as well as follow-up practices. Examples of appropriate methods are Cox models, survival forests, neural survival models or time dependent AUC or concordance metrics. Patient-level splitting required and external cohorts should ensure that the temporal and clinical relationship between tissue acquisition, treatment and outcome is preserved. A model that predicts treatment received (and not treatment benefit) may look good but it is not clinically useful.

4.11 Whole-Slide Imaging, Multiple-Instance Learning and Attention

But the images are gigapixel scale and the regions that are diagnosed are very sparse, which limits the use of WSI analysis. The resolution of the computational problem is achieved by the extraction of patches, but it entails an aggregation problem - there are thousands of patches per patient and only a portion of these contain the signal of interest for the target label. MIL codifies this environment as a bag of instances representing the WSI. The idea of weakly supervised MIL is to learn which instances are predictive by using the slide label, rather than pixel-level annotations for all cancer areas in a slide. In attention-based MIL, each patch embedding is given a learned weight. The attention map that is generated is a kind of model explanation that can be examined, but it is important to note that simply because the model assigns attention to a variable, it does not follow that the variable is necessarily causal in importance. Other approaches are: to group patches together into morphological phenotypes, select representative instances, apply hierarchical aggregation or to add coordinates to the model. Recent studies have shown increasingly that different CNN or ViT encoders are used in parallel with MIL aggregators to trade-off local morphology with global context [5,6,28].

4.12 Evaluation Metrics

There is no single measure of clinical behavior of an AI pathology system. Sensitivity and specificity are clinically interpretable for binary classification; precision and F1-score can be more useful in presence of class imbalance; ROC-AUC can be used to summarise ability to rank across thresholds; precision-recall AUC can be more informative when positive cases are rare. Overlap is measured for segmentation: Dice and intersection over union. Survival prediction: common: concordance, time-dependent. Calibration measures the degree of agreement between expected and actual frequencies.

Metric	Primary use	Interpretation / caution
Accuracy	Balanced classification	Can be misleading with class imbalance
Sensitivity	Cancer/metastasis detection	Measures true-positive detection; false negatives remain critical
Specificity	Rule-out performance	Depends on prevalence and threshold

F1-score	Imbalanced classification	Balances precision and recall; not a probability measure
ROC-AUC	Binary ranking	Does not show performance at a chosen clinical threshold
PR-AUC	Rare positives	More sensitive to class prevalence
Dice / IoU	Segmentation	Sensitive to object size and annotation quality
Calibration	Risk/probability output	Necessary when probabilities inform decisions
C-index / time-dependent AUC	Prognosis	Requires appropriate censoring and follow-up handling

4.13 Comparative Synthesis of AI Paradigms

Approach	Typical input	Advantages	Limitations	Suitable role
Handcrafted ML	Engineered cell/tissue features	Interpretable feature pipeline; lower compute	Feature design limits representation	Small, structured datasets; baseline
CNN + transfer learning	Patches / regions	Strong local morphology learning	Needs representative training data	Patch classification and detection
U-Net / segmentation DL	Pixels / tissue regions	Explicit spatial delineation	Annotation intensive	Tumor, nuclei and compartment segmentation
MIL / attention MIL	Patch embeddings + slide label	Uses weak labels; WSI-level learning	Aggregation and interpretability challenges	WSI classification and prognosis
Vision Transformer	Patch/token sequences	Long-range context; flexible representation	Data/compute hungry; domain shift	WSI representation and classification
Foundation/self-supervised model	Large unlabeled pathology corpora	Transferable representations	Pretraining cost; validation complexity	General-purpose pathology features
Multimodal AI	WSI + clinical/genomic data	Integrates complementary signals	Data alignment and missingness	Precision oncology research

No single architecture is best suited to all applications, across the literature. Morphology of WSIs is well captured by CNNs, and scale and context by MIL and transformers. The decision should be made based on the clinical task, label granularity, size of the dataset, deployment requirements, and not on popularity of particular architectures. Recent

breast pathology reviews explicitly indicates the shift in the behavior of the field from CNN-centralized systems to transformer-based and multimodal systems [3,4].

4.14 Representative Findings from the Literature

There are numerous consistent findings in the literature. First, AI can be highly effective on image-classification tasks that are more narrowly defined, and can help with the detection of lymph-node metastasis [18-20]. Second, the ability to perform repetitive counting or measurement of quantitative pathology tasks (Ki-67, mitotic count) make these tasks well suited for automation [1,25,26]. Third, there is the possibility for molecular and prognostic prediction at the WSI (micro level), which is more sensitive to the domain shift and cohort composition [14,22-24]. Fourth, numerous recent reviews have cited external validation, interpretability, workflow integration and regulatory evidence as significant impediments to the successful integration of omics into therapeutic practice [2,3,12]. To overcome the challenges we proposed the following model.

WORKING OF PROPOSED SYSTEM

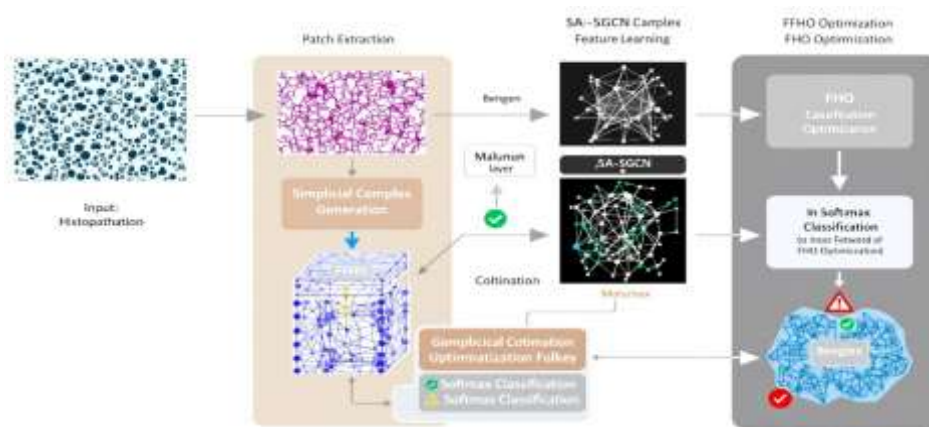


Figure1. Proposed System Diagram

Input Processing

Histopathology Image Input: Whole slide images (WSIs) or tissue sections

Patch Extraction: Dividing large images into manageable patches to handle computational constraints

2. Topological Representation

Simplicial Complex Generation: Converting image patches into topological structures that capture spatial relationships and higher-order interactions beyond simple pairwise connections

3. Feature Learning

SA-SGCN (Self-Attention Simplicial Graph Convolutional Network): Combines simplicial complexes with graph convolutional operations. Incorporates self-attention mechanisms to weigh important features. Learns hierarchical representations from topological structures

4. Optimization

FHO (Fractional-order Hyperparameter Optimization):

Advanced optimization technique using fractional calculus. Fine-tunes model parameters for better convergence

5. Classification

Softmax Classification: Final layer producing probability distributions.

Binary Output: Benign vs. Malignant diagnosis

The model was evaluated using public histopathology datasets (BreakHis and BACH). Training used an 80-10-10 split. Comparative findings include:

Table 2. Expected Result

Model	Accuracy	Sensitivity	Specificity	F1-Score
CNN baseline	90.21%	88.95%	89.77%	90.14%
GCN baseline	92.84%	91.10%	92.35%	92.47%
Proposed FHO-SA-SGCN	97.56%	97.21%	97.83%	97.40%

5. Discussion

5.1 Major Technical Challenges

Data heterogeneity is the first challenge. The differences in slides are the result of tissue processing, fixation times, staining chemistry, scanner manufacturer, magnification, compression and laboratory processing. A model can then be trained to recognize the site appearance and not disease morphology. This effect can be minimized through domain adaptation and stain normalization, but the best evidence is provided by external validation across institutions. The second challenge is the quality of the annotations. Different sources of histopathology labels have different levels of uncertainty, such as report-based labels, consensus labels, IHC-based labels, and manual region-based labels. Weak labels may be noisy because the slide is made up of different tissue and the label describes the slide as a whole. The help of multiple pathologists, adjudication protocols and uncertainty-aware learning can help. A drawback of adding more detail to annotations is that it adds to the cost! Data Leakage is the third challenge. Random splitting at the patch level can lead to over-optimism when patches that are visually similar are randomly assigned to the training and test set of the same patient. Typically, the patient, the specimen or both are the unit used for splitting. If two or more slides are from the same patient, then they should be in the same partition. Test data should not be used to preprocess, to tune hyperparameters and to choose the model. Class imbalance and spectrum bias is the fourth challenge. Many datasets are lacking in the representation of rare subtypes of the tumor and small metastases. A trained model may not work as well on a screening or routine diagnostic population as it did on a high-prevalence research population. Appropriate reporting of performance should thus include confidence intervals and stratified analyses wherever possible.

5.2 Interpretability and Explainability

Pathologists should be aware of the rationale for a system's flagging. The visual evidence can be given by saliency maps, class-activation maps, attention maps and representative patch retrieval. These techniques can aid in debugging and interaction but do not ensure the highlighted region was responsible for the prediction. Explanations should thus be considered as evidence to support and be validated for faithfulness. Mechanisms of uncertainty and out-of-distribution detection and abstention at the case level are more clinically useful than a single opaque confidence score.

5.3 Clinical Translation and Human-AI Workflow

The big question is how AI will transform the pathologist's job. A standalone classifier sending a label is not as useful as a decision-support system which points to suspicious regions, prioritizes cases, quantifies biomarkers and offers quality-control checks. The role to be played will depend upon the error profile and the clinical task. For instance, high sensitivity triage may be useful for detecting metastasis, and reproducible measurement of the biomarkers with pathologist verification may be useful for biomarker quantification. Workflow outcomes should also be assessed as part of clinical validation, along with algorithmic metrics. Endpoints are: diagnostic sensitivity, diagnostic specificity, reading time, discrepancies, case throughput, user trust, number of overrides, and number of difficult and atypical cases. Prospective studies are important since benchmarks from a retrospective study may underestimate the complexity of real-life use. More recent digital pathology reviews have focused on the difference between lab performance and clinical implementation [12].

5.4 Ethical, Regulatory and Governance Considerations

The use of AI pathology systems could have an impact on diagnosis and treatment decisions, and governance should be sensitive to issues of data privacy, security, traceability, bias and responsibility for mistakes. Patient information should be processed in accordance with privacy and research laws. Monitoring is needed for model updates as performance can be impacted by a new scanner, stain protocol or patient population. Description of intended use, target population, input requirements, known failure modes and validation evidence should be in documentation. One of the main issues of bias is when training data don't reflect the population the model will be deployed in. The diversity should be addressed at the level of institutions, geographical regions, age groups, disease subtypes, staining protocols and relevant demographic parameters. The objective is not uniformity of overall accuracy—it's accurate performance of the clinically relevant population. External validation should thus be stratified, if scientifically warranted and practicable.

5.5 Reproducibility

To replicate computational pathology, more than just publishing model weights is needed. Patient level split, scanner and magnification information, stain processing, patch size, tissue selection rules, augmentation, training time, hyperparameters, model selection criteria, external test performance. Standardization of public code and benchmark datasets enhance comparability. Wherever patient data cannot be shared synthetic examples, executable pipelines for preprocessing, and detailed metadata can enhance the ability to reproduce.

5.6 Future Research Directions

First, the use of self-supervised learning and pathology foundation models will likely alleviate reliance on costly pixel level annotations. The large amount of unlabeled WSI patches can be used for pretraining, followed by adaptation using relatively small task-specific datasets. What is really important is the rigorous cross-site validation, not just more datasets to be pre-trained. Second, multimodal AI can integrate the morphological data with clinical, radiological, genomic and transcriptomic data. Such models can be used to simulate other features of the disease and could enhance risk stratification at the patient level. Explicit treatment of missing data and modality-specific biases due to inconsistencies in measurement time. Third, federated learning provides a way to develop a collaborative model without the need to centralize patient images. Shared model training with multiple institutions for local data. Federated systems, however, pose new technical challenges, such as non-identically distributed data, communication cost, privacy leakage and model aggregation stability. Fourth, resource-aware AI plays a crucial role for hospitals with limited computational resources. Lower magnification,

compact encoders and efficient aggregation for WSI analysis have been explored in recent work [16]. This is especially important when addressing global health concerns as high performance scanning and GPU infrastructure may not be equally accessible. Fifth, uncertainty-aware and abstaining systems need to be better developed. A model needs to be capable of recognizing cases that fall outside of the range of its distribution or in which the evidence is ambiguous. Safe human-AI collaboration is supported by selective prediction, calibration and out-of-distribution detection. The goal isn't maximum automation, it's ensuring that attention is shared between human and machine specialists. Finally, future studies should be directed to the prospective, multi-institutional evaluation. Research benchmarks should be supplemented with trials of workflow that would capture patient-relevant and operational outcomes. Regulatory pathways, post-deployment monitoring and model-change management requirements should be tackled early in the development process, instead of at the end.

5.7 Overall Synthesis

Overall, the literature suggests that AI-driven histopathological analysis has the potential to assist in various stages of the diagnosis of breast cancer, ranging from the detection of local cells, to slide-level classification and beyond, all the way to prediction of the biomarkers and outcomes of a cancer diagnosis. The best evidence so far is focused in very well-defined tasks with unambiguous labels, and the highest uncertainty is in general purpose WSI diagnosis and clinically relevant molecular inference. These differences should inform research assertions and implementation decisions.

6. Conclusion

The histopathological diagnosis of breast cancer has evolved from hand-crafted image features to deep neural networks, weakly supervised WSI analysis and emerging pathology foundation models. CNNs are useful for local morphology and segmentation, and MIL and attention mechanisms are useful for slide-level learning with incomplete annotations. Transformer-based approaches are increasingly tackling long-range tissue context, while multimodal models aim to relate morphology to molecular and clinical data. Based on the evidence presented in the reviews and representative studies, cancer diagnosis, detection of metastasis, cancer grading, quantification of biomarkers, prediction and prognosis of molecular sub-types are feasible [1–4,14,18–24]. Meanwhile, the clinical use of a benchmark cannot be equated with benchmark accuracy. Important challenges to be addressed include dataset shift, stains and scanner variability, label noise, patient-level leakage, class imbalance, limited external validation and interpretability. The most realistic development trajectory is, thus, a human-centred one: careful patient-level assessment, external validation, calibrated outputs, clear failure modes and prospective studies of workflow. AI should be marketed as an AI supporting decision-making technology, which is proven through the reliability, efficiency or reproducibility in real pathology practice.

References

- [1] Baxi V, Edwards R, Montalto M, Saha S. Artificial intelligence in breast cancer histopathology. *Histopathology*. 2023;82:198–211. doi:10.1111/his.14820.
- [2] Soliman A, Li Z, Parwani AV. Artificial intelligence's impact on breast cancer pathology: a literature review. *Diagnostic Pathology*. 2024;19:38. doi:10.1186/s13000-024-01453-w.
- [3] Current status and prospects of artificial intelligence in breast cancer pathology: convolutional neural networks to prospective Vision Transformers. *Review literature*, 2024.

- [4] Artificial intelligence in breast pathology: Overview and recent updates. *Human Pathology*. 2025. doi:10.1016/j.humpath.2025.105819.
- [5] Lu MY, Williamson DFK, Chen TY, et al. Data-efficient and weakly supervised computational pathology on whole-slide images. *Nature Biomedical Engineering*. 2021;5:555–570.
- [6] Ilse M, Tomczak JM, Welling M. Attention-based deep multiple instance learning. *Proceedings of ICML*. 2018.
- [7] Komura D, Ishikawa S. Machine learning methods for histopathological image analysis. *Computational and Structural Biotechnology Journal*. 2018;16:34–42.
- [8] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. *Proceedings of CVPR*. 2016.
- [9] Tan M, Le Q. EfficientNet: rethinking model scaling for convolutional neural networks. *Proceedings of ICML*. 2019.
- [10] Dosovitskiy A, et al. An image is worth 16×16 words: transformers for image recognition at scale. *ICLR*. 2021.
- [11] Komura D, Ochi M, Ishikawa S. Machine learning methods for histopathological image analysis: Updates in 2024. *Computational and Structural Biotechnology Journal*. 2024;27:383–400.
- [12] Williams DKA Jr, et al. Digital pathology, deep learning, and cancer: a narrative review. *Translational Cancer Research*. 2024;13:2544–2560.
- [13] Deep learning approaches for breast cancer detection in histopathology images: a review. *Cancer Biomarkers*. 2024;40:1–25. doi:10.3233/CBM-230251.
- [14] McCaffrey C, Jahangir C, Murphy C, et al. Artificial intelligence in digital histopathology for predicting patient prognosis and treatment efficacy in breast cancer. *Expert Review of Molecular Diagnostics*. 2024;24:363–377.
- [15] Mandair D, Reis-Filho JS, Ashworth A. Biological insights and novel biomarker discovery through deep learning approaches in breast cancer histopathology. *npj Breast Cancer*. 2023;9:21.
- [16] Deep-learning-based breast cancer stage prediction from H&E-stained whole-slide images in resource-constrained settings. 2026.
- [17] A Multi-Stage Approach to Breast Cancer Classification Using Histopathology Images. Dataset description based on ICIAR 2018 BACH; 2022.
- [18] Ehteshami Bejnordi B, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA*. 2017;318:2199–2210.
- [19] Campanella G, et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nature Medicine*. 2019;25:1301–1309.
- [20] Bandi P, Geessink O, Manson Q, et al. From detection of individual metastases to classification of breast cancer lymph nodes: the CAMELYON17 challenge. *IEEE Transactions on Medical Imaging*. 2018;39:340–350.
- [21] Machine Learning Methods for Computer-Aided Breast Cancer Diagnosis Using Histopathology: A Narrative Review. 2019.
- [22] Tafavvoghi M, Sildnes A, Rakae M, et al. Deep learning-based classification of breast cancer molecular subtypes from H&E whole-slide images. *Journal of Pathology Informatics*. 2025;16:100410.
- [23] Couture HD, et al. A deep learning image-based intrinsic molecular subtype classifier of breast tumors reveals tumor heterogeneity that may affect survival. *Scientific Reports*. 2018/2020 literature record.

- [24] Wetstein SC, de Jong VMT, Stathonikos N, et al. Deep learning-based breast cancer grading and survival analysis on whole-slide histopathology images. *Scientific Reports*. 2022;12:15102.
- [25] Ki-67 Quantification in Breast Cancer by Digital Imaging AI Software and its Concordance with Manual Method. 2023.
- [26] Artificial intelligence-powered optimization of Ki-67 assessment in breast cancer: enhancing precision and workflow efficiency. Literature review. 2024.
- [27] Artificial intelligence in digital pathology diagnosis and analysis: technologies, challenges, and future prospects. Review literature, 2025/2026.
- [28] Breast Cancer Molecular Subtype Prediction on Pathological Images with Discriminative Patch Selection and Multi-Instance Learning. 2022.
- [29] Progress on deep learning in digital pathology of breast cancer: a narrative review. *Gland Surgery*. 2022;11.
- [30] Artificial intelligence applied to breast pathology. Review literature. 2021.