

Deep Learning-Based Multi-Class Breast Cancer Histopathological Image Classification: A Review

Aastha Mishra¹, Prof. Brajesh Patel²

¹Research Scholar, Shri Ram Institute of Technology

²Professor, Shri Ram Institute of Technology

Published date: 05/10/2026

Abstract

Breast cancer histopathology is a challenging computer-vision domain because diagnostically relevant evidence is distributed across cells, tissue architecture, staining patterns, and multiple microscopic scales. The availability of digital histopathology benchmarks has enabled a transition from handcrafted image descriptors toward machine learning, convolutional neural networks, transfer learning, attention mechanisms, and increasingly multimodal artificial intelligence. This review examines the development of multi-class breast histopathological image classification with particular attention to the problem addressed by an InceptionResNetV2-CBAM framework using the BreakHis benchmark. The review organizes the field across dataset design, class taxonomy, image preprocessing, class balancing, data augmentation, convolutional feature extraction, transfer learning, attention-based feature refinement, evaluation methodology, explainability, and translational considerations. It also considers how recent breast-cancer AI research is expanding into ensemble learning, multimodal fusion, subgroup analysis, biosensing, spectroscopy, and specialized architectures. The literature indicates that high internal accuracy should not be interpreted independently from the experimental protocol. Patient-wise partitioning, training-only resampling, external validation, calibration, probability-based ROC/AUC, and component-level ablation are increasingly important for determining whether a high-performing benchmark system generalizes beyond its development data. The reviewed evidence suggests that the next generation of breast histopathology systems will need to combine strong representation learning with methodological rigor, interpretable error analysis, and validation across heterogeneous institutions and acquisition conditions. The review concludes by defining a practical research agenda for robust multi-class breast histopathological image classification and by positioning the InceptionResNetV2-CBAM design as one representative integrated architecture within that broader research landscape.

Keywords: Breast Cancer, Histopathology, BreakHis, Deep Learning, Transfer Learning, InceptionResNetV2, CBAM, CLAHE, Multi-Class Classification, Medical Image Analysis, Explainable AI.

Introduction

Breast cancer represents a heterogeneous group of diseases in which histological subtype, cellular morphology, and tissue architecture contribute to diagnostic interpretation and downstream management. Histopathology images preserve microscopic evidence that is not captured completely by screening or imaging modalities alone, which is why computational analysis of stained tissue has become a major research area in medical artificial intelligence. Reviews of breast

cancer detection technologies have documented a broad transition from conventional diagnostic approaches toward computer-assisted analysis and data-driven methods. [1]

Artificial intelligence has expanded this transition by allowing image-classification systems to learn patterns directly from large collections of labeled examples. Systematic reviews of AI for breast cancer detection emphasize both the potential of automated image analysis and the challenges that accompany dataset variation, validation, reproducibility, and clinical translation. [2] The resulting research landscape includes mammography, ultrasound, MRI, histopathology, spectroscopy, biosensing, and multimodal systems. Histopathological classification occupies a distinctive position because the signal of interest is microscopic and can involve subtle differences in cellular organization rather than large-scale anatomical structures.

One of the most important changes in the field has been the move from binary classification toward fine-grained multi-class recognition. A binary task asks whether an image belongs to a broad benign or malignant category, whereas a multi-class task requires the model to distinguish among specific tumor subtypes. Systematic evaluations of AI in breast cancer have shown that task definition strongly affects both the apparent difficulty of a problem and the interpretation of performance results. [3] Mammographic AI literature provides a related lesson: the reported value of a model depends not only on architecture, but also on the population, acquisition protocol, label definition, and intended use. [4]

The present review is centered on the eight-class histopathological problem represented in the supplied work: adenosis, fibroadenoma, phyllodes tumor, tubular adenoma, ductal carcinoma, lobular carcinoma, mucinous carcinoma, and papillary carcinoma. These classes include four benign and four malignant categories, creating a classification space in which visually similar categories can occur on both sides of the benign-malignant divide. The review therefore treats multi-class classification as a representation-learning and evaluation problem rather than as a simple accuracy-optimization task.

Breast cancer AI is broader than histopathology alone. Reviews of electromagnetic detection illustrate that different sensing mechanisms capture different disease-related signals and operate at different stages of diagnosis. [5] Broader technological reviews likewise show that breast cancer detection has developed as a multimodal field rather than as a single-modality competition. [6] This wider context is relevant because future histopathology systems may operate as one component within a multimodal diagnostic workflow rather than as isolated image classifiers.

Deep learning provides the principal computational foundation for the work reviewed here. CNN-based systems can learn hierarchical representations that progress from local edges and texture toward increasingly abstract structural patterns. Reviews of deep learning in breast cancer detection and classification describe the growing use of learned feature representations in automated analysis. [7] However, performance alone is not sufficient: the broader clinical context includes accessibility, reproducibility, interpretability, and the risk of deploying a model outside the distribution on which it was developed. [8]

The integrated framework examined in the underlying research combines four major design ideas: class-distribution management, contrast enhancement through CLAHE, transfer learning with InceptionResNetV2, and feature refinement using CBAM. The architecture can therefore be viewed as a representative example of a modern pipeline in which data preparation and model

architecture are designed together. This review uses that pipeline as a focal case for understanding broader methodological trends, not as evidence that any individual component is universally optimal.

The objectives of this review are fourfold. First, it summarizes the evolution of breast histopathological image classification and the role of public datasets. Second, it examines the major families of preprocessing, machine learning, deep learning, transfer-learning, and attention-based methods. Third, it synthesizes recent work across histopathology and related breast-cancer AI modalities. Finally, it identifies unresolved challenges and proposes a research agenda focused on leakage-resistant evaluation, interpretability, calibration, multimodal learning, and clinically meaningful validation.

I. Background and Motivation

Histopathology converts biological tissue into a two-dimensional visual representation through fixation, sectioning, staining, microscopy or scanning, and digital image storage. The resulting images contain information at several scales. At high magnification, nuclei, cellular boundaries, mitotic structures, and local tissue patterns dominate the field of view. At lower magnification, glandular organization, stromal context, and broader tissue architecture become more visible. A computational system must therefore learn features that remain useful despite changes in spatial scale, staining appearance, and tissue composition.

The visual heterogeneity of histopathology creates two distinct statistical problems: intra-class variability and inter-class similarity. Images within the same pathological class may differ because of magnification, tissue region, staining intensity, section characteristics, or acquisition conditions. At the same time, two different classes may share color, texture, and cellular patterns. A model that learns only broad appearance cues may therefore achieve good performance on a benchmark while failing when the acquisition environment changes.

The BreakHis benchmark became important precisely because it provided a structured collection of breast histopathology images spanning benign and malignant subtypes and multiple magnification levels. The original dataset paper describes the benchmark as a resource designed to support reproducible experiments in breast histopathological image classification. [9] The existence of such a benchmark has allowed the research community to compare architectures and preprocessing choices, but it has also exposed the importance of patient-level independence and dataset construction when interpreting apparently high accuracy.

Multi-class classification introduces multiple competing decision boundaries. The error between two malignant subtypes may have different implications from an error between two benign classes, and a model can have a high overall accuracy while still underperforming systematically on one subtype. For this reason, precision, recall, F1-score, confusion matrices, and class-specific analysis are necessary complements to aggregate accuracy.

The need for class-aware evaluation also follows from the natural imbalance of many medical datasets. A class with many training samples can dominate gradient updates, while a less frequent class can receive insufficient representation. This creates a distinction between optimizing a model and evaluating it: the training distribution may be intentionally altered through oversampling or weighting, whereas the evaluation distribution should preserve the intended meaning of the test population.

The focal framework uses a mean-count path-level balancing operation before partitioning, followed by augmentation during training. This is useful for studying the effect of balanced learning, but it also illustrates a central methodological risk: resampling before splitting can create overlap between partitions. A review of breast histopathology methods should therefore treat dataset construction as part of the model, not as a minor preprocessing detail.

Recent breast cancer AI research increasingly combines several components rather than relying on a single backbone. Dynamic model-selection approaches have been proposed to adapt model complexity to image or tumor characteristics, illustrating a shift toward task-aware architecture design. [10] Multimodal systems have also explored combinations of histopathology, mammography, and ultrasound, suggesting that complementary modalities can reduce reliance on a single visual representation. [11]

Contemporary reviews describe a wider move toward integrated AI systems that include advanced architectures, multimodal learning, explainability, and stronger validation protocols. [12] Ensemble systems for histopathological images have similarly emphasized patient-wise splitting, augmentation, and feature integration, highlighting that experimental protocol is increasingly treated as a core part of model design rather than as a reporting afterthought. [13]

II. Breast Histopathological Image Classification Techniques and Taxonomy

Breast histopathological image classification can be organized into a taxonomy that follows the degree of automation and representation learning used by the system. The main families are handcrafted-feature methods, conventional machine learning, deep convolutional learning, transfer learning, attention-enhanced deep learning, multimodal methods, and explainability-aware frameworks. The boundaries between these families are not absolute; modern systems often combine several of them.

A. Handcrafted Features and Classical Computer-Aided Analysis

Early computer-aided histopathology systems commonly relied on engineered descriptors of color, texture, edges, morphology, and spatial arrangement. The appeal of these methods was interpretability: a researcher could define which visual characteristics were extracted and then train a conventional classifier on the resulting feature vectors. Their limitation is that histopathology is sufficiently heterogeneous that a fixed descriptor set may fail to capture all relevant morphological variation.

Classical methods remain useful as baselines because they provide a reference against which learned representations can be compared. They can also be computationally efficient when the feature set is compact. Nevertheless, the broader literature on breast cancer detection has increasingly shifted toward automated feature learning because manual feature design can constrain the representational capacity of the classifier.

B. Conventional Machine Learning

Machine learning methods such as support vector machines, random forests, gradient boosting, and ensemble classifiers can operate on handcrafted descriptors or deep features extracted from a pretrained network. In breast histopathology, this hybrid pattern remains relevant because deep networks can be used as feature extractors while lightweight classifiers are used for the final decision. The approach can reduce the amount of end-to-end optimization needed, but the feature representation remains the critical determinant of performance.

A representative example is the work of Sharmin et al., which combines ResNet50V2 feature extraction with ensemble machine-learning classifiers for breast cancer detection. [14] The study illustrates a broader transition in which transfer-learned representations are separated from the downstream decision layer. Such hybrids can be effective, but differences in dataset, task definition, and evaluation protocol make direct numerical comparison difficult.

C. Deep Convolutional Learning

Deep CNNs learn hierarchical features directly from images. Lower layers generally capture local patterns such as edges and simple textures, while deeper layers represent combinations of structures and increasingly semantic visual patterns. The major advantage for histopathology is that the feature representation is learned jointly with the classification task rather than being fixed in advance.

However, deep models introduce a different challenge: the more expressive the architecture becomes, the more important dataset size, regularization, initialization, and validation become. In medical imaging, where labels may be expensive and images from the same patient may be highly related, a large model can learn dataset-specific structure that does not generalize to new clinical populations.

D. Transfer Learning and Inception-ResNet

Transfer learning addresses the mismatch between the scale of modern CNNs and the size of many medical datasets. Instead of training from random initialization, a model pretrained on a large natural-image corpus can be adapted to the target domain. Inception-style networks use multi-scale processing, while residual connections support optimization of deeper architectures. The Inception-ResNet family explicitly combines these ideas and was introduced as a way to accelerate training while retaining strong representational capacity. [15]

The InceptionResNetV2-based framework reviewed here uses the pretrained backbone as a feature extractor and adapts it to eight histopathological classes. The motivation is consistent with a wider transfer-learning principle: the low-level visual primitives learned from large datasets can provide a useful starting point, while domain-specific optimization reshapes deeper representations toward tissue morphology.

Transfer learning should not be confused with guaranteed domain transfer. Histopathology differs substantially from natural imagery in color distributions, textures, spatial scales, and semantic structure. Therefore, the quality of the adapted representation must still be evaluated under the target domain, and the final performance should not be attributed to pretraining alone.

E. Attention-Enhanced Classification

Attention mechanisms attempt to address the fact that not every feature channel or spatial location contributes equally to every prediction. Channel attention asks which feature dimensions are most informative, while spatial attention asks where the informative evidence is located. CBAM is a lightweight module that sequentially infers channel and spatial attention maps and applies them to intermediate feature maps. [16]

In a breast histopathology setting, this idea is attractive because diagnostic evidence can be localized. A model may need to emphasize nuclear arrangements, glandular structures, stromal boundaries, or other local patterns rather than treating the entire image uniformly. CBAM therefore

provides a mechanism for feature reweighting after the backbone has produced a rich convolutional representation.

The use of CBAM should nevertheless be evaluated experimentally rather than assumed to be beneficial. If the complete system is trained only with CBAM included, the observed performance cannot identify the marginal contribution of attention. A rigorous review of attention-based histopathology research therefore distinguishes between studies demonstrating strong integrated performance and studies that perform controlled ablations.

Method Family	Typical Representation	Main Strength	Main Limitation
Handcrafted features	Color, texture, morphology	Compact and interpretable	Limited representation capacity
Classical ML	Engineered or extracted features	Simple downstream optimization	Depends on feature quality
CNNs	Learned hierarchical features	End-to-end visual learning	Data and validation demands
Transfer learning	Pretrained CNN + target adaptation	Useful with limited medical data	Domain mismatch remains
Attention-based CNNs	CNN features + channel/spatial weights	Focuses on informative representations	Needs ablation to establish contribution
Multimodal AI	Histopathology + imaging/clinical data	Complementary evidence	Higher integration and validation burden

Table 1. Taxonomy of major methodological families in breast histopathological classification.

III. Deep Learning-Based Histopathological Image Analysis

A. Dataset Design and Benchmarking

Dataset design determines what a model can learn and what a reported metric means. In addition to class labels, breast histopathology datasets contain information about patient identity, specimen source, magnification, staining, scanner characteristics, and image-generation procedures. If those factors are not considered during partitioning, a model can indirectly benefit from shared acquisition patterns across training and test images.

BreakHis remains useful because it offers a structured benchmark with multiple tumor categories and magnification levels. [9] Its value is greatest when experiments preserve the distinction between image-level and patient-level independence. A benchmark score obtained from image-wise partitioning is informative as an internal result but answers a different question from a score obtained on completely unseen patients.

Class definition is another source of variation. Some research addresses binary benign-versus-malignant recognition, while other studies distinguish several benign and malignant subtypes. As the number of classes grows, the model must represent increasingly fine-grained differences. This makes class-wise metrics and confusion analysis more important than accuracy alone.

B. Preprocessing and Stain or Contrast Variation

Histopathology images can contain local illumination and staining differences that modify the appearance of otherwise similar structures. Contrast enhancement methods are therefore frequently considered as a preprocessing stage. CLAHE improves local contrast while limiting the amplification of noise, which makes it attractive when tissue structures are visible at different local intensity levels. The original CLAHE method was introduced as a contrast-limited extension of adaptive histogram equalization. [17]

The framework reviewed here applies CLAHE in the LAB color space, enhances the luminance channel, then resizes the image to 224×224 pixels and normalizes pixel values. The advantage of separating luminance from chromatic information is that local contrast can be modified without deliberately rewriting the full color composition. However, preprocessing can also introduce a model-dependent bias: an enhancement that helps one dataset may amplify acquisition artifacts in another.

For that reason, preprocessing should be evaluated as a controlled component. A useful review protocol compares the full pipeline against the same architecture without CLAHE, using the same patient-wise split and evaluation metrics. Without this control, improvements cannot be attributed to contrast enhancement alone.

C. Class Balancing and Data Augmentation

Medical image datasets often have uneven class frequencies. Common responses include oversampling, undersampling, class weighting, synthetic image generation, focal loss, or combinations of these methods. Oversampling increases the frequency of minority examples, while undersampling reduces majority examples. Augmentation is conceptually different because it changes image appearance through transformations such as flipping, rotation, cropping, or intensity changes.

The reviewed framework uses mean-count path-level balancing and random horizontal and vertical flips. This is computationally simple and produces a more uniform training distribution. The critical issue is not the use of balancing itself but where it occurs in the pipeline. When resampling happens before partitioning, duplicate image paths may appear across train and test sets. A stronger protocol creates the patient-wise or specimen-wise split first and balances only the training partition.

Augmentation also has domain-specific constraints. A transformation is helpful only if it preserves the class semantics. Horizontal and vertical flips can often be reasonable for tissue patches because histological orientation is not fixed, while aggressive geometric transformations may distort structures whose orientation carries diagnostic meaning. Thus, augmentation policy should be justified according to the morphology of the target task rather than copied from natural-image classification.

D. Optimization and Regularization

Transfer-learning pipelines typically combine a pretrained backbone with a task-specific classification head. The reviewed framework uses a dense head of 512, 256, and 128 units followed by an eight-class softmax output, together with batch normalization, dropout, label smoothing,

early stopping, and learning-rate reduction. These controls are intended to stabilize optimization and reduce overfitting.

Label smoothing changes the target distribution so that the model is discouraged from becoming excessively confident during training. Batch normalization standardizes intermediate activations during optimization, while dropout introduces stochastic feature suppression. Early stopping and learning-rate scheduling use validation behavior to regulate training. These mechanisms operate at different levels and should therefore be analyzed separately in ablation studies.

A recurrent theme in the literature is that training performance should not be used as the sole model-selection criterion. A network can continue improving its training accuracy even after validation accuracy has stopped improving. For medical imaging, the preferred checkpoint should be selected using a predefined validation criterion, and the final test set should be used only after model development is complete.

E. Attention and Feature Refinement

CBAM provides two sequential operations. Channel attention aggregates feature maps through global average and maximum pooling, passes the resulting descriptors through a shared transformation, and produces channel-wise weights. Spatial attention then aggregates along the channel dimension, combines average and maximum projections, and generates a spatial attention map. The sequential design lets the network adapt both which features matter and where they matter. [16]

The value of attention in histopathology is tied to the localized nature of visual evidence. When a slide patch contains large background or non-diagnostic tissue, a spatial attention mechanism can in principle reduce the influence of less informative locations. Channel attention can also reduce the influence of feature dimensions that are less relevant to a particular subtype.

Attention should, however, be distinguished from explainability. An attention map can show which features were reweighted by the network, but this is not automatically equivalent to a faithful causal explanation of the prediction. Explainability research should therefore combine attention analysis with external explanation methods and, when possible, expert review.

Architecture Family	Representative Use	Advantages	Considerations
CNN / ResNet	General image classification	Strong hierarchical representations	May need careful regularization
Inception	Multi-scale vision	Parallel receptive-field processing	Architectural complexity
Inception-ResNet	Deep transfer learning	Multi-scale features + residual optimization	High capacity and compute
Vision Transformers	Global context modeling	Long-range feature interactions	Data and compute demands
CNN + Attention	Histopathology refinement	Channel/spatial focus	Requires component ablation
Multimodal / Ensemble	Integrated diagnosis	Combines complementary evidence	Complex validation and deployment

Table 2. Representative deep-learning architecture families relevant to breast cancer image analysis.

IV. Related Work

Recent literature demonstrates that breast cancer AI is no longer limited to single-model image classification. Research now spans specialized deep networks, ensembles, multimodal fusion, subgroup analysis, biosensors, spectroscopy, and computer-aided diagnosis. This section organizes representative work from the supplied literature into thematic groups and identifies where histopathological classification fits within the broader research landscape.

Architecture specialization is an active direction in breast cancer AI. Khan et al. describe PRMS-Net for thermal-image analysis, illustrating how residual architectures can be adapted to a particular sensing modality and task. [18] The significance for histopathology is not the thermal modality itself, but the broader lesson that architecture design is increasingly tailored to the information content of the input data.

Akbari et al. propose dynamic model scaling based on segmented tumor size, showing a different strategy for adapting computational complexity to the target image characteristics. [19] This type of conditional architecture selection suggests future opportunities for histopathology models that dynamically adjust feature extraction depth according to magnification, tissue complexity, or uncertainty.

Cross-patient multimodal fusion has also been studied by Akbari et al., who combine information from complementary breast-cancer modalities through a virtual-patient formulation. [20] Such work points toward a future in which histopathology is not treated as an isolated source, but as one stream within a multimodal representation.

Mondal et al. review trends and challenges across artificial intelligence for breast cancer detection and highlight the increasing breadth of the field. [21] Their perspective is useful for situating histopathology inside a wider ecosystem of screening, diagnosis, and future multimodal systems.

Ahmad et al. report an ensemble deep-learning framework for histopathological images and emphasize patient-wise splitting, augmentation, and feature integration. [18] The emphasis on patient-wise validation is particularly important because several image patches or magnification variants can originate from the same biological source.

Gao et al. provide a systematic review of deep learning in breast cancer segmentation, detection, and classification and emphasize dataset characteristics, methodological quality, and translation into clinically meaningful systems. [19] This reinforces the principle that a review of reported accuracy should also examine how the dataset was partitioned, how labels were formed, and how the evaluation protocol maps to the intended use.

Larsen et al. examine AI-based mammographic density assessment in BreastScreen Norway, illustrating the importance of evaluating an AI method within a realistic screening environment rather than only on a static research benchmark. [20] Although the imaging modality differs from histopathology, the methodological message is transferable: performance depends on the deployment population and workflow.

Owotogbe et al. synthesize a decade of machine-learning and deep-learning research across breast cancer detection, classification, prognosis, and explainability. [21] Their broader review helps connect image classification to explainable and clinically integrated AI.

Brown-Mulry et al. investigate subgroup performance of a commercial digital breast tomosynthesis model and demonstrate the value of checking whether overall performance is consistent across clinically relevant groups. [22] This principle motivates subgroup-aware evaluation in histopathology, where magnification level, specimen characteristics, institution, scanner, and staining protocol may act as sources of distribution shift.

Breast cancer detection is also expanding through biosensing and physical measurement systems. Toha et al. study a graphene-assisted multilayer surface plasmon resonance biosensor for breast cancer cell detection. [23] Wertani et al. investigate machine-learning-assisted optimization of a terahertz biosensor. [24][25] These studies demonstrate that AI is increasingly being integrated with engineered sensing systems rather than used only for image classification.

Hlali et al. explore deep-learning-based optimization of two-dimensional material biosensors, and Appadurai et al. investigate an AI-enhanced terahertz metasurface for breast cancer biomarker detection. [26] [27] The importance of these studies for histopathology is strategic: future diagnostic systems may integrate tissue imaging with biochemical or physical measurements to compensate for the limitations of a single modality.

Hoseini and Kheyri examine a hybrid capsule-network computer-aided diagnosis framework for mammographic images, showing continued experimentation with architectures designed to capture spatial relationships beyond conventional CNN representations. [28] Van Der Waal et al. focus on recall-rate optimization in breast cancer screening, underscoring the consequences associated with sensitivity and false-positive decisions. [29] [30] Nakul et al. review machine-learning-enhanced optical spectroscopy for breast cancer diagnosis and illustrate another path toward multi-signal diagnostic intelligence. [31] [32].

The integrated framework underlying this review occupies the intersection of these research directions. It combines a high-capacity transfer-learning backbone with attention, preprocessing, class balancing, and explicit multi-class evaluation. The architecture therefore addresses several known challenges simultaneously, but the literature also indicates that simultaneous integration does not automatically establish causality. A system can perform well because of one dominant component, because several components complement one another, or because the evaluation protocol permits related images to occur across partitions. [33]

The contribution of an integrated framework is therefore best understood at two levels. At the engineering level, it provides a coherent pipeline that can be executed and extended. At the scientific level, it creates a platform for ablation, patient-wise validation, and external testing. A review paper should distinguish these two levels rather than treating a strong benchmark result as proof that every component is necessary.

The reviewed literature also suggests a progression from isolated classification toward richer systems: stronger feature extractors, attention mechanisms, multimodal fusion, subgroup evaluation, sensing integration, and explainability. The opportunity for histopathological image classification is to combine these advances without losing methodological discipline. The next

generation of research should therefore prioritize validation design as aggressively as architecture design.

Study	Modality / Task	Approach	Review Relevance
Khan et al.	Thermal breast imaging	PRMS-Net	Optimized residual learning
Akbari et al.	Histopathology	Dynamic model scaling	Complexity adapts to tumor size
Akbari et al.	Multimodal breast AI	Virtual-patient fusion	Cross-patient multimodal integration
Mondal et al.	Breast AI overview	Review	Trends, challenges, future directions
Ahmad et al.	Histopathology	Ensemble CNN/transformer	Patient-wise splitting and feature integration
Gao et al.	Breast AI	Systematic review	Methodological quality and clinical translation
Larsen et al.	Mammography	AI density model	Screening-context evaluation
Owotogbe et al.	Breast AI	Systematic review	Detection, prognosis, explainability
Brown-Mulry et al.	Digital breast tomosynthesis	Commercial AI	Subgroup performance
Hoseini & Kheyri	Mammography	Hybrid capsule network	Computer-aided diagnosis

Table 3. Representative recent studies spanning breast cancer AI modalities and methodological themes.

Component	Purpose	Typical Benefit	Key Risk / Research Need
CLAHE	Local contrast enhancement	Improved visibility of local structures	May amplify artifacts
Oversampling	Increase minority representation	More balanced training exposure	Duplication before split can leak information
Undersampling	Reduce majority representation	Controls class dominance	Can discard informative examples
Augmentation	Increase appearance diversity	Potentially better robustness	Must preserve morphology
CBAM	Channel + spatial attention	Feature reweighting	Benefit must be isolated by ablation
Calibration	Align confidence with correctness	More reliable decision support	Rarely reported in benchmark papers

Patient-wise split	Protect biological independence	More realistic generalization estimate	Requires patient/specimen metadata
--------------------	---------------------------------	--	------------------------------------

Table 4. Preprocessing, attention, and evaluation components in breast histopathology AI.

V. Challenges and Future Research Directions

One of the most important challenges in breast histopathological image classification is preventing information leakage and ensuring patient-independent evaluation. Histopathology datasets may contain multiple image crops, fields of view, or magnification levels derived from the same patient or specimen. When images are randomly divided at the image level, related observations may appear in both training and testing partitions, allowing the model to learn source-specific characteristics rather than disease-related patterns. Resampling can introduce an additional risk when oversampling is performed before dataset partitioning because duplicated image paths may subsequently occur in different subsets. A more reliable protocol is therefore to establish patient-wise or specimen-wise training, validation, and test partitions first and then perform balancing exclusively within the training set. Future studies should report patient, specimen, and image counts together with magnification distributions and duplicate-detection procedures so that the true unit of independence is clear. Such reporting would make it easier to determine whether a system is being evaluated on unseen images, unseen specimens, or genuinely unseen patients.

Another important challenge is domain shift caused by magnification, staining, scanners, and image-acquisition conditions. Histopathological images obtained at different magnification levels can contain substantially different spatial information, and a model that performs strongly at one scale may not maintain the same behavior at another. Staining variation and differences between imaging devices or institutions can further alter color, contrast, and tissue appearance. Although preprocessing methods such as CLAHE can improve local contrast, excessive enhancement may also modify diagnostically relevant information. Future research should therefore evaluate performance separately across magnification levels and investigate multi-scale representations, stain normalization, domain-adaptation techniques, and cross-institution validation. Such experiments would help determine whether learned features correspond to genuine morphological characteristics rather than to acquisition-specific visual patterns.

Class imbalance and unequal error consequences also require greater attention. A balanced training distribution does not mean that all classification errors have the same clinical importance. A model may achieve high overall accuracy while retaining a specific sensitivity weakness in an individual tumor category. The reported framework illustrates this issue through the lower recall observed for ductal carcinoma compared with its precision. Future research should therefore examine class-wise sensitivity, precision, recall, F1-score, calibration, and clinically meaningful error patterns rather than relying solely on aggregate accuracy. Cost-sensitive optimization, threshold adjustment, targeted augmentation, and hard-negative mining may provide useful strategies for addressing difficult classes, but their effects should be tested using patient-independent evaluation so that apparent improvements are not caused by data leakage or changes in partition composition.

Trustworthy artificial intelligence also requires careful consideration of explainability, calibration, uncertainty, and confidence. Attention mechanisms such as CBAM refine internal feature representations, but attention should not automatically be interpreted as proof that the model has

learned biologically meaningful evidence. Future studies should combine attention or gradient-based explanations with confidence estimates, representative-image retrieval, and expert review. Explanations should also be compared across correct and incorrect predictions so that researchers can determine whether the model is focusing on plausible tissue morphology or on irrelevant background and acquisition patterns. At the same time, classification accuracy does not indicate whether model confidence is well calibrated or whether the system can recognize images that fall outside its training distribution. Probability-based ROC/AUC analysis, reliability diagrams, calibration error, Brier score, uncertainty estimation, selective prediction, and out-of-distribution detection can therefore provide a more complete assessment of model reliability. A reject or abstention mechanism may also be useful so that uncertain cases can be referred for human assessment rather than forcing every image into a categorical prediction.

The broader research direction is toward multimodal, reproducible, and systematically validated systems. Histopathology can potentially be combined with mammography, ultrasound, clinical information, molecular data, biosensing, or spectroscopy to provide complementary evidence, but multimodal models introduce additional requirements for patient-level alignment, consistent labeling, missing-data handling, and external validation. Such systems should therefore be developed incrementally, with each additional modality evaluated through controlled ablation to determine its actual contribution. Reproducibility should also be treated as a core research requirement. Future studies should provide the exact dataset version, class definitions, patient-level split strategy, image counts, resampling and augmentation procedures, model initialization, optimizer configuration, checkpoint-selection criteria, deterministic class-to-index mapping, and complete class-wise evaluation results. Most importantly, researchers should clearly distinguish observed performance from explanations about why that performance occurred. A high score demonstrates the behavior of a model under a particular experimental protocol; it does not by itself demonstrate that a specific component caused the improvement or that the model will generalize to independent clinical environments. Overall, future breast histopathological image-classification research should progress from internally evaluated high-performing models toward patient-independent, externally validated, calibrated, interpretable, reproducible, and clinically meaningful systems.

VI. Conclusion

This review has examined the development of deep learning-based multi-class breast histopathological image classification from the perspectives of dataset design, image preprocessing, class balancing, conventional machine learning, transfer learning, attention mechanisms, multimodal AI, explainability, and evaluation methodology. The central observation is that the field has moved beyond the question of whether deep learning can classify histopathology images. The more important questions now concern how reliably the learned representation generalizes, which components contribute to the observed performance, and whether the resulting predictions can be trusted outside the original benchmark environment.

Public datasets such as BreakHis have enabled progress by providing standardized multi-class image collections and multiple magnification levels. The benchmark remains valuable, but its scientific value depends on how researchers partition and interpret it. Patient-wise or specimen-wise separation, training-only resampling, and explicit reporting of acquisition and magnification conditions should become routine elements of evaluation.

Transfer learning and attention provide complementary architectural strategies. InceptionResNetV2 offers a high-capacity feature representation that combines multi-scale processing with residual optimization, while CBAM provides channel and spatial feature refinement. CLAHE and class-balancing operations address data-level variability, and together these components form a coherent integrated framework. However, the literature shows that integrated performance should not be confused with component-level causality. Ablation studies are required to establish the independent value of each module.

The broader literature also demonstrates that breast cancer AI is expanding beyond histopathology into multimodal fusion, screening, biosensing, spectroscopy, subgroup evaluation, and explainability. This expansion creates an opportunity for future systems in which histopathology is one source of evidence within a larger diagnostic ecosystem. It also increases the need for robust interoperability, privacy-preserving data sharing, and external validation.

Finally, methodological rigor should progress alongside architectural innovation. A model that reaches high internal accuracy on a benchmark is useful, but a clinically meaningful system must additionally demonstrate patient-independent generalization, calibration, uncertainty awareness, interpretable failure modes, and performance across heterogeneous acquisition environments. The research agenda developed in this review therefore places experimental validity before deployment claims and controlled evidence before architectural complexity.

The InceptionResNetV2-CBAM framework reviewed in this work can be positioned as a representative integrated architecture within this broader evolution. Its value is not limited to its particular network combination; it provides a testable platform for future patient-wise validation, component ablation, probability-based evaluation, explanation analysis, and multimodal extension. These steps can help transform strong benchmark classification into a more reproducible and trustworthy research direction for multi-class breast histopathology analysis.

References

- [1] Jaglan, Poonam, Rajeshwar Dass, and Manoj Duhan. "Breast cancer detection techniques: issues and challenges." *Journal of The Institution of Engineers (India): Series B* 100.4 (2019): 379–386.
- [2] Díaz, Oliver, Alejandro Rodríguez-Ruíz, and Ioannis Sechopoulos. "Artificial Intelligence for breast cancer detection: Technology, challenges, and prospects." *European Journal of Radiology* 175 (2024): 111457.
- [3] Nassif, Ali Bou, et al. "Breast cancer detection using artificial intelligence techniques: A systematic literature review." *Artificial Intelligence in Medicine* 127 (2022): 102276.
- [4] Gardezi, Syed Jamal Safdar, et al. "Breast cancer detection and diagnosis using mammographic data: Systematic review." *Journal of Medical Internet Research* 21.7 (2019): e14464.
- [5] Hassan, Ahmed M., and Magda El-Shenawee. "Review of electromagnetic techniques for breast cancer detection." *IEEE Reviews in Biomedical Engineering* 4 (2011): 103–118.
- [6] Nover, Adam B., et al. "Modern breast cancer detection: a technological review." *International Journal of Biomedical Imaging* 2009.1 (2009): 902326.

- [7] Hamed, Ghada, et al. "Deep learning in breast cancer detection and classification." The International Conference on Artificial Intelligence and Computer Vision. Cham: Springer International Publishing, 2020.
- [8] Barrios, Carlos H. "Global challenges in breast cancer detection and treatment." *The Breast* 62 (2022): S3–S6.
- [9] Spanhol, Fabio A., Luiz S. Oliveira, Caroline Petitjean, and Laurent Heutte. "A Dataset for Breast Cancer Histopathological Image Classification." *IEEE Transactions on Biomedical Engineering* 63.7 (2016): 1455–1462. doi:10.1109/TBME.2015.2496264.
- [10] Khan, Mudassir, et al. "Enhancing breast cancer detection through optimized thermal image analysis using PRMS-Net deep learning approach." *Journal of Imaging Informatics in Medicine* 39.1 (2026): 864–883.
- [11] Akbari, Younes, et al. "Dynamic model scaling based on segmented tumor size for breast cancer detection." *Biomedical Signal Processing and Control* 113 (2026): 109118.
- [12] Akbari, Younes, et al. "A novel virtual patient approach for cross-patient multimodal fusion in enhanced breast cancer detection." *Computerized Medical Imaging and Graphics* 127 (2026): 102687.
- [13] Mondal, Jyotirmoy, et al. "Artificial intelligence (AI) for breast cancer detection: Trends, challenges, and future directions." *Nano Theragnostics in Breast Cancer: Advances, Challenges, and Future Prospects*. Singapore: Springer Nature Singapore, 2026. 707–750.
- [14] Sharmin, Selina, et al. "A hybrid dependable deep feature extraction and ensemble-based machine learning approach for breast cancer detection." *IEEE Access* 11 (2023): 87694–87708.
- [15] Szegedy, Christian, Sergey Ioffe, Vincent Vanhoucke, and Alex Alemi. "Inception-v4, Inception-ResNet and the impact of residual connections on learning." arXiv:1602.07261, 2016.
- [16] Woo, Sanghyun, Jongchan Park, Joon-Young Lee, and In So Kweon. "CBAM: Convolutional Block Attention Module." *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 3–19.
- [17] Zuiderveld, Karel J. "Contrast Limited Adaptive Histogram Equalization." *Graphics Gems IV*, Academic Press, 1994, pp. 474–485. doi:10.1016/B978-0-12-336156-1.50061-6.
- [18] Ahmad, Faizan, et al. "Ensemble Deep Learning-Based High-Precision Framework for Breast Cancer Detection from Histopathological Images." *Diagnostics* 16.5 (2026): 653.
- [19] Gao, Shuo, et al. "Application of deep learning technology in breast cancer: a systematic review of segmentation, detection, and classification approaches." *BioMedical Engineering OnLine* 25.1 (2026): 19.
- [20] Larsen, Marthe, et al. "Mammographic density assessment by an artificial intelligence model for breast cancer detection in BreastScreen Norway." *European Radiology* (2026): 1–12.
- [21] Owotogbe, Joshua, et al. "Machine learning and deep learning for breast cancer: A decade systematic review of detection, classification, prognosis, and explainability." *Informatics in Medicine Unlocked* (2026): 101756.

- [22] Brown-Mulry, Beatrice, et al. "Subgroup performance of a commercial digital breast tomosynthesis model for breast cancer detection." *Nature Communications* 17.1 (2026): 4249.
- [23] Toha, Tasnim Ferdous, et al. "Graphene assisted multilayer SPR biosensor with superior specificity for breast cancer cell detection." *Results in Engineering* (2026): 109739.
- [24] Wertani, Hamza, et al. "Advanced optimization of THz biosensor for breast cancer detection through machine learning." *IEEE Transactions on Plasma Science* 54.4 (2026): 1750–1761.
- [25]. Suryawanshi, Pooja, and Pankaj Richhariya. "Maximizing Accuracy in Image Classification using Transfer Learning and Random Forest." *International Journal of Innovations in Science, Engineering And Management* (2023): 39-48.
- [26] Hlali, Aymen, et al. "Deep learning-based performance optimization of 2D material biosensors for breast cancer detection." *Microsystem Technologies* 32.5 (2026): 53.
- [27] Appadurai, Jothi Prabha, et al. "Artificial intelligence-enhanced terahertz metasurface biosensor for breast cancer biomarker detection." *Plasmonics* 21.1 (2026): 605–618.
- [28] Hoseini, Farnaz, and Masume Kheyri. "A Hybrid Capsule Network–Based Computer-Aided Diagnosis Framework for Breast Cancer Detection in Mammographic Images." *Intelligence-Based Medicine* (2026): 100381.
- [29]. Bhardwaj, Kumar Kinjalk, and Vishal Pathak. "Nanostructured Materials and Their Physical Applications." *International Journal of Innovations in Science, Engineering And Management* (2025): 413-420.
- [30] Van Der Waal, Danielle, et al. "Finding the optimal recall rate in breast cancer screening: results from the ROCS study." *European Radiology* 36.7 (2026): 5595–5603.
- [31] Nakul, Mihir, et al. "Machine learning enhanced optical spectroscopy for breast cancer diagnosis: A review." *Lasers in Medical Science* 41.1 (2026): 86.
- [32]. Tiwari, Sanjay Kumar, and Shashwat Tiwari. "A Philosophical Evaluation of the Role of Yoga in the Treatment of Cancer." *International Journal of Innovations in Science Engineering And Management* (2026): 610-619.
- [33]. Sarasija, M. "Applications of Nanotechnology in Medicinal Chemistry Impact." *AG Volumes* (2025): 57-63.