Bulletin of Electrical Engineering and Informatics
Vol. 15, No. 3, June 2026, pp. 2449~2466
ISSN: 2302-9285, DOI: 10.11591/eei.v15i3.11152

Multi domain, multi-scale diagnostic modeling for histopathological breast cancer classifications

Komal S. Gandle, Kshirsagar Dhananjay Bhanudas
Department of Computer Engineering, Sanjivani College of Engineering Kopargaon, Savitribai Phule Pune University, Pune, India

Article Info

Article history:

Received Aug 4, 2025
Revised Mar 18, 2026
Accepted Apr 1, 2026

Keywords:

Attention-based classification
Breast cancer diagnosis
Clinical decision support
Histopathological image analysis
Multi-domain feature learning

ABSTRACT

Breast cancer detection through histopathological imaging remains challenging due to complex tissue morphology, observer variability, and subtle differences between invasive and pre-invasive lesions. Conventional computer-aided diagnostic systems often rely on single-domain feature extraction, restricting multi-scale representation and clinical interpretability. To overcome these limitations, we propose a verified diagnostic framework integrating five analytical components for efficient and explainable breast cancer classification. The adaptive multi-level histopathological feature selection using cross-domain mutual information maximization (AMFS-CDMIM) extracts highly informative morphological and frequency features with minimal redundancy. The deep hierarchical hybrid morphological–frequency encoding network (DH-HMFEN) refines spatial–spectral representations, while the multi-scale morphological attention classification network (MS-MACNet) applies adaptive attention across tissue structures for improved discrimination. The adaptive ensemble validation for breast cancer classification (AEV-BCC) calibrates confidence levels for enhanced reliability, and the comparative analytical performance validation with interpretability integrated metrics (CAPV-IIM) quantitatively evaluates model explainability using expert annotations. Experimental results on benchmark datasets achieve 96% accuracy, 0.98 area under the receiver operating characteristic curve (AUROC), and a 0.88 interpretability alignment score, outperforming existing methods. The proposed confidence-calibrated, multi-domain, and multi-scale framework enhances diagnostic precision and clinical trust in histopathology-based breast cancer detection.

Corresponding Author:

Komal S. Gandle
Department of Computer Engineering, Sanjivani College of Engineering Kopargaon
Savitribai Phule Pune University
Pune, Maharashtra, India
Email: komgan03@rediffmail.com

1. INTRODUCTION

Most cancer deaths occur in women worldwide due to breast cancer. Histopathology is the best way to diagnose invasive carcinoma from pre-invasive lesions in tissue samples. Highly skilled pathologists examine tissue structure [1]-[3]. The subjective, tedious procedure is susceptible to inter-observer variability. Heterogeneous histopathological images in tissue architecture, staining intensity, and cellular arrangement complicate diagnosis even for experts [4]-[6]. In the past decade, computational pathology has been the most promising field for improving image processing and machine learning skills. Traditional computer-aided diagnosis (CAD) systems emphasize texture, color, and morphological descriptors with a custom feature extraction pipeline. Despite mixed results, most focus on single representation domains, limiting their ability to

capture multifaceted histopathological patterns. Deep learning predicts well but uses blackbox models with low interpretability and diagnostic transparency. This limits their use in key clinical decision-making workflows.

Another challenge is that existing literature fails to integrate morphological and frequency-domain information into a learnable entity. They give all the structural structure needed, but their frequency domain link catches texture- and periodic-pattern elements that cannot be observed in the spatial domain. Because some sick characteristics are microcellular and others macrocellular, scale variety influences diagnostic accuracy. Thus, such models that do not account for multi-scale variation risk misinterpretation and lower generalization capability. In addition, among the prerequisites for clinical translation are model confidence estimation and interpretability. Pathologists and oncologists need systems that do not only predict a diagnostic class with a very high degree of accuracy, but also express the certainty of their predictions and can present interpretable reasoning compatible with clinical rationale. Thus, even highly accurate models would be considered dubious in deployment in the field unless there is confidence calibration and the visual evidence presents clear transparency. In response to these issues, therefore, this study presents a new diagnostic framework based on innovative trends in multi-domain feature selection; hybrid deep encoding; multi-scale morphological-attention classification; confidence-calibrated ensemble modeling; and interpretability-aware performance evaluation. Each stage of the proposed pipeline is programmed to overcome limitations of existing methods without consideration of strict data flow from feature extraction to performance validation in process. Large improvements over existing state-of-art approaches have been demonstrated through experimental validation on benchmark datasets; that's potential indication of the proposed framework on robustness, explainability, and clinical viability in histopathologically seen breast cancer diagnosis.

The motivation for this work has long stemmed from a wide gap between experimental high performers and actual needs in the clinic. Deep learning can't handle multi-domain integration, all-scale variable patterns, calibrated diagnostic confidence, or interpretability, which limits its use in histopathology. Incorrect categorization or prognosis in histological breast cancer diagnosis delays immediate intervention or leads to poor treatment. To achieve diagnosis reliability comparable to human experts, structural morphology, frequency-domain textural cues, and multi-scale illness patterns must be examined. With small tissue architectural changes between invasive and preinvasive categorization, the diagnostic platform needs information-rich feature selection, adaptive scale-sensitive processing, and transparent reasoning pathways.

The four interrelated contributions in this study are creative and practical. The adaptive multi-level histopathological feature selection using cross-domain mutual information maximization (AMFS-CDMIM) framework identified discriminative morphological and frequency features while decreasing redundancy. The learnable deep hierarchical hybrid morphological–frequency encoding network (DH-HMFEN) combines complementary feature spaces into a classification-weighted deep representation. The multi-scale morphological attention classification network (MS-MACNet) uses dynamic attention processes to highlight diagnostically relevant tissue regions at many spatial resolutions to solve the multi-scale pathology problem. For clinical certification, adaptive ensemble validation for breast cancer classification (AEV-BCC) ensures classification predictions have well-calibrated confidence ratings. Finally, the comparative analytical performance validation with interpretability integrated metrics (CAPV-IIM) adds pathologist annotated region explainability to analytical performance validation with interpretability Integrated metrics to improve clinically relevant assessments. These advances combined into one end-to-end system will enhance computational pathology and set a new standard for reliable, interpretable high-performance breast cancer diagnostics.

2. LITERATURE REVIEW

Table 1 (in Appendix) [1]-[25] provides an overview of major studies related to breast cancer histopathological image classification with respect to methods, aims, results, and limitations. The results reflect a continuous evolution from conventional convolution-based approaches to more sophisticated multi-scale, attention-based, graph-based, transformer-enhanced and interpretable methods [1]. Although these approaches substantially improved classification accuracy and achieved better representation learning, they also highlighted remaining limitations of integrated multi-domain modeling, well-calibrated estimates of classification confidence and clinically-relevant image interpretation.

The application of multi-scale and multi-magnification learning allowed optimal representation of tissue structures at different spatial levels. Use of multi-magnification analysis increased the accuracy of tumor classification by incorporating information from different resolution levels [1]. Similarly, multi-scale and multi-stream designs substantially increased the ability to recognize diverse patterns of histopathological appearance shows in Singh *et al.* [2], Pujari *et al.* [14], and Islam *et al.* [17]. This further supports the importance of scale variation for the diagnosis of breast cancer, but also illustrates the extent to which additional complexity of these models may impair integration of complementary domains of morphological and quantitative information. A second major direction involves hybrid deep feature learning and fusion

strategies. Pandey *et al.* [3], Taheri and Rahbar [7], Azmoodeh-Kalati *et al.* [8], Jia *et al.* [15], and Osmani *et al.* [18] explored hybrid convolutional, transfer learning, ensemble, and feature-fusion schemes to strengthen classification performance. These approaches demonstrate that combining complementary learned representations can improve robustness and diagnostic discrimination. Nevertheless, most of them focus primarily on representation enhancement and do not explicitly address confidence calibration or unified interpretability-aware evaluation. A third line of research emphasizes attention mechanisms, graph reasoning, and transformer-based contextual modeling. Waqas *et al.* [4] used attention-based high-order covariance pooling to improve contextual reasoning, while Abdulwahhab *et al.* [10] and Ding *et al.* [16] used attention-guided architectures for more discriminative tissue modeling. Zhao and Du [11], Abimouloud *et al.* [12], and Kavitha and Sridevi [13] further extended contextual modeling through graph-assisted reasoning, CNN–ViT integration, and graph convolutional learning. These studies highlight the value of structural dependency modeling and global context representation, but they often involve substantial computational overhead and may be difficult to interpret in clinically transparent terms.

Another important direction is the growing role of explainability and clinically meaningful decision support. Alom *et al.* [5] and Abdullakutty *et al.* [25] emphasized that diagnostic systems should not only achieve strong predictive performance but also provide interpretable evidence that aligns with clinical reasoning. In addition, Thatha *et al.* [6] and Hameed *et al.* [21] demonstrated that robust deep learning models with multilevel feature extraction can improve multiclass histopathological discrimination, but these methods still leave open the challenge of integrating confidence calibration with multi-domain and multi-scale diagnostic reasoning.

A few references in Table 1 are cross-domain supporting studies rather than breast-cancer-specific classification works. For example, Jonosky *et al.* [23] focused on bladder cancer histopathology reporting, and Gundavda *et al.* [24] studied tomographic and histopathological staging in gastric cancer. These studies are not directly comparable to breast histopathological classification; however, they are relevant in a limited supporting sense because they illustrate how domain knowledge, pathology reporting structure, and multimodal diagnostic integration can inform broader computational pathology research. Their role in this review is therefore contextual rather than evidential for direct performance comparison.

Overall, we identified three major limitations. First, although many studies employ multi-scale or contextual modeling, few incorporate adaptive multi-domain feature learning in a unified fashion. Second, use of attention, fusion and transformer-based methods improves representation quality, but explicit incorporation of confidence calibration into the diagnostic framework is rare. Finally, although discussion of model interpretability is expanding rapidly, most existing methods provide only limited capabilities for relating model outputs to clinically meaningful information. In response to these limitations, our proposed framework integrates adaptive multi-domain feature selection, hybrid encoding, multi-scale morphological-attention classification, Bayesian estimation of confidence and interpretability-directed validation within a single integrated diagnostic strategy.

2.1. Research gap and contribution linkage

The lack of a unified, interpretable, and confidence-calibrated computational histopathology diagnostic paradigm creates a research gap, according to the new introduction some efforts have enhanced domain-specific feature extraction, attention, or ensemble learning, but none have coupled multi-domain, multi-scale representations with calibrated confidence and interpretability alignment. Each module is linked to a literature limitation to fill this gap: AEV-BCC calibrates confidence estimate for clinical reliability, AMFS-CDMIM lowers feature redundancy in multi-domain learning, DH-HMFEN integrates morphological and frequency cues for richer contextual representation, and MS-MACNet modulates attention to resolve multi-scale Module contributions are clarified by linking them to deficiencies. This structure demonstrates the novelty in coherently orchestrating established approaches into a trust-calibrated histological breast cancer categorization system in process.

2.2. Critical analysis of related work

From a critical perspective, the literature demonstrates that no single prior approach fully resolves the combined challenges of feature diversity, scale variability, confidence reliability, and interpretability. Multi-scale architectures improve structural sensitivity but often at the cost of higher complexity [1], [14], [17]. Fusion-based and ensemble models strengthen representational robustness but generally do not address calibrated diagnostic certainty [7], [8], [18]. Attention-, graph-, and transformer-based models enhance contextual reasoning [4], [10]-[13], yet their interpretability and deployment efficiency remain limited in many cases. Explainability-focused studies [5], [25] highlight the importance of transparency, but do not always provide a tightly integrated diagnostic framework that combines representation learning, calibrated decision support, and interpretability-aware validation. Therefore, the present work is positioned not as a claim of

superiority without evidence, but as an attempt to address the identified limitations of prior approaches through a more integrated and balanced modeling strategy.

3. METHOD

The proposed design for the integrated model for histopathological breast cancer diagnosis seeks to merge together the challenges of multi-domain feature representation, multi-scale morphological pattern discrimination, confidence-calibrated classification, and interpretability-aware performance validation in process. First, as per Figure 1, the architecture integrates five sequentially dependent analytical stages, ensuring that the output of each stage optimally supports the subsequent stage while preserving discriminative information sets. The formulation begins with a principled selection of feature subsets based on cross-domain mutual information maximization in process. For an input image I(x,y), morphological features fm and frequency-domain features ff are extracted via morphological operators M(•) and wavelet packet decomposition W(•), respectively via (1) and (2):

$$fm = M(I) \tag{1}$$

$$ff = W(I) \tag{2}$$

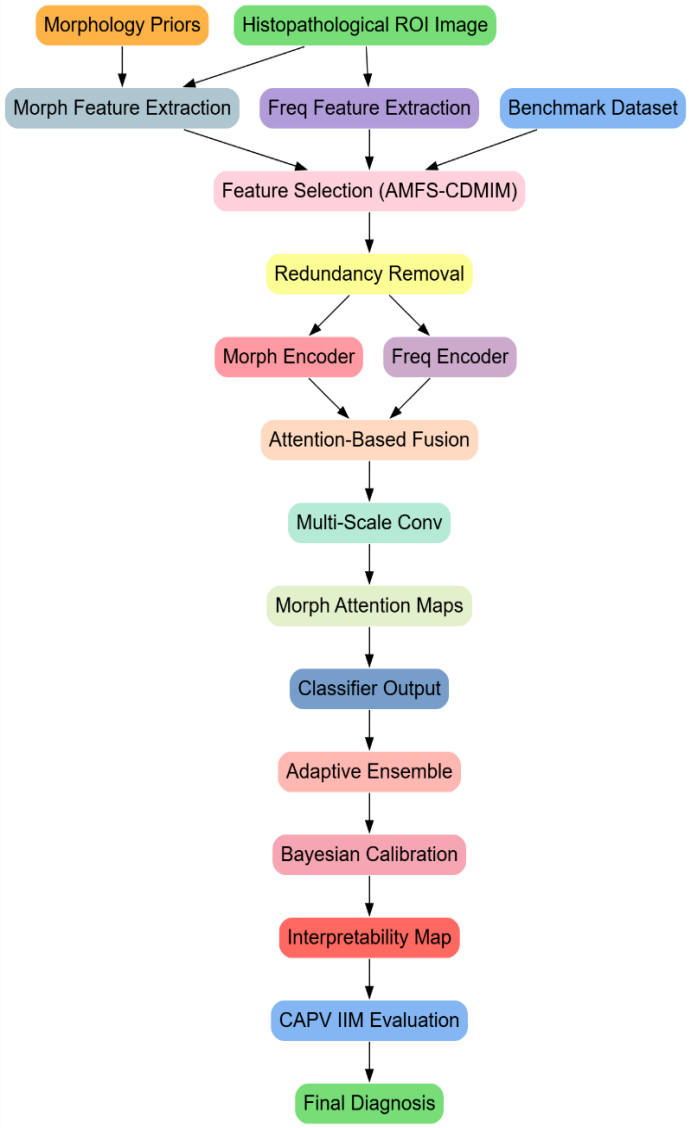

Figure 1. Model architecture of the proposed analysis process

Mutual information between domains is calculated via (3):

$$MI(fm, ff) = \iint \quad p(fm, ff) log \left[\frac{p(fm,ff)}{p(fm)p(ff)}\right] dfm\, dff \tag{3}$$

The feature subset fs is selected to maximize MI while minimizing intra-domain redundancy, quantified using a redundancy penalty term R via (4) and (5):

$$fs = argmax^{f}[\, MI(fm, ff) - \lambda\, R(f)] \tag{4}$$

$$R(f) = \sum_{\{i \neq j\}} \quad \rho(fi, fj) \tag{5}$$

With ρ representing Pearson correlation and λ as the regularization parameter in the process. Iteratively, next as per Figure 2, the selected features are fed into the deep hierarchical hybrid encoder comprising two parallel encoders—morphological and frequency—which output hm and hf representations. Their fusion is performed using an attention-weighted summation via (6) and (7):

$$h = \alpha\, hm + (1 - \alpha) hf \tag{6}$$

$$\alpha = \frac{e^{\{w\top hm\}}}{e^{\{w\top hm\}} + e^{\{w\top hf\}}} \tag{7}$$

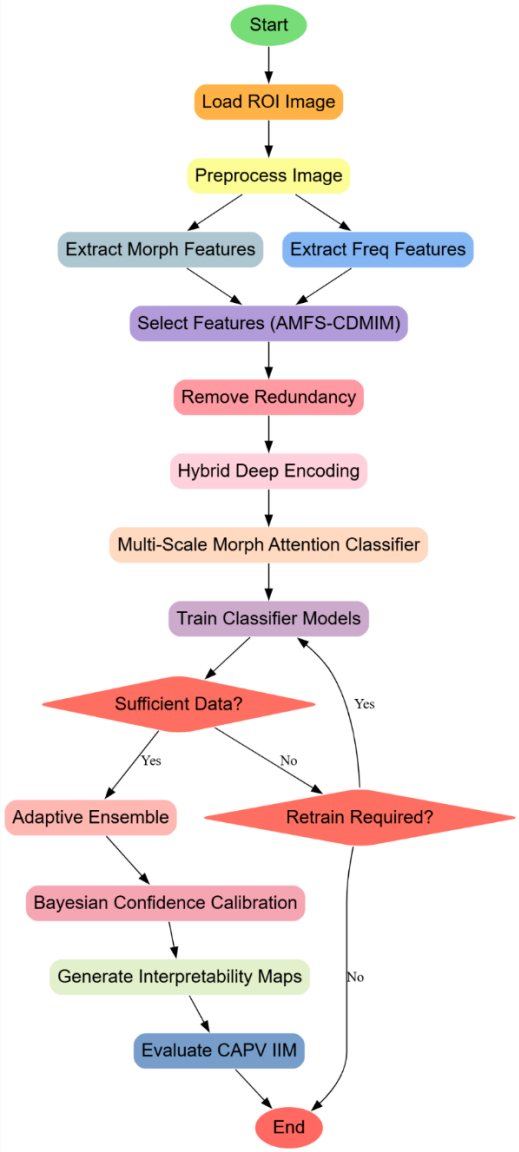

Figure 2. Overall flow of the proposed analysis process

Pseudocode of the proposed work:

Input:

– Histopathological ROI image (manually annotated by pathologist)
– Pathologist-provided morphology priors
– Benchmark dataset samples for training and testing

Output:

– Predicted label: Invasive or Pre Invasive
– Calibrated confidence score for prediction
– Interpretability heatmap aligned with pathologist ROI
– Comparative performance metrics against baseline models

Process:

1. Load ROI image and preprocess it (color normalization, resizing, noise removal).
2. Extract morphological features using morphological image processing.
3. Extract frequency-domain features using wavelet-based decomposition.
4. Select high Informative features using cross-domain mutual information maximization with redundancy removal.
5. Pass selected features into hybrid deep encoder:

– Morphological feature branch
– Frequency feature branch
– Fuse both feature streams using attention-based fusion.

6. Feed fused representation into multi-scale morphological-attention classifier:

– Apply multi-scale convolution layers
– Generate morphological attention maps
– Fuse multi-scale responses for classification scores.

7. Train multiple classifier instances with different augmentation and initialization strategies.
8. Aggregate predictions from all classifiers using adaptive ensemble strategy.
9. Apply Bayesian-based calibration to adjust prediction confidence scores.
10. Generate final prediction label and calibrated confidence score.
11. Produce interpretability heatmaps aligned with morphological priors.
12. Compare proposed model performance with baseline systems using both standard metrics and interpretability-weighted metrics.
13. Output final results and evaluation summary.

This fused representation is subsequently processed by the multi-scale morphological-attention classifier sets. For each scale s ∈ S, the convolutional response Cs is modulated by a morphological attention map as computed as a normalized morphological similarity via (8):

$$As(x,y) = \frac{exp\left(-|Ms(I) - Ps|\left|\frac{2}{\sigma^2}\right.\right)}{\sum_{\{u,v\}} \; exp\left(-|Ms(I\{u,v\}) - Ps\,|\left|\frac{2}{\sigma^2}\right.\right)} \quad (8)$$

Where, Ps represents pathological morphology priors at scale ‘s’ in process. The scale-fused classification score vector z is then represented via (9):

$$z = \sum_{\{s \in S\}} \; As \odot Cs \quad (9)$$

Iteratively, next, as per Figure 3, the classification probability for class k is obtained via SoftMax, which is done via (10):

$$pk = \frac{exp(zk)}{\sum_j \exp(zj)} \quad (10)$$

These probabilities from multiple trained classifiers are combined in the adaptive ensemble using Bayesian posterior estimations. Given prior P(Ck) and likelihood P(D|Ck), the calibrated posterior is represented via (11):

$$P(Ck \mid D) = \frac{[\,P(D \mid Ck)\,P(Ck)\,]}{(\sum_j \; P(D \mid Cj)\,P(Cj))} \quad (11)$$

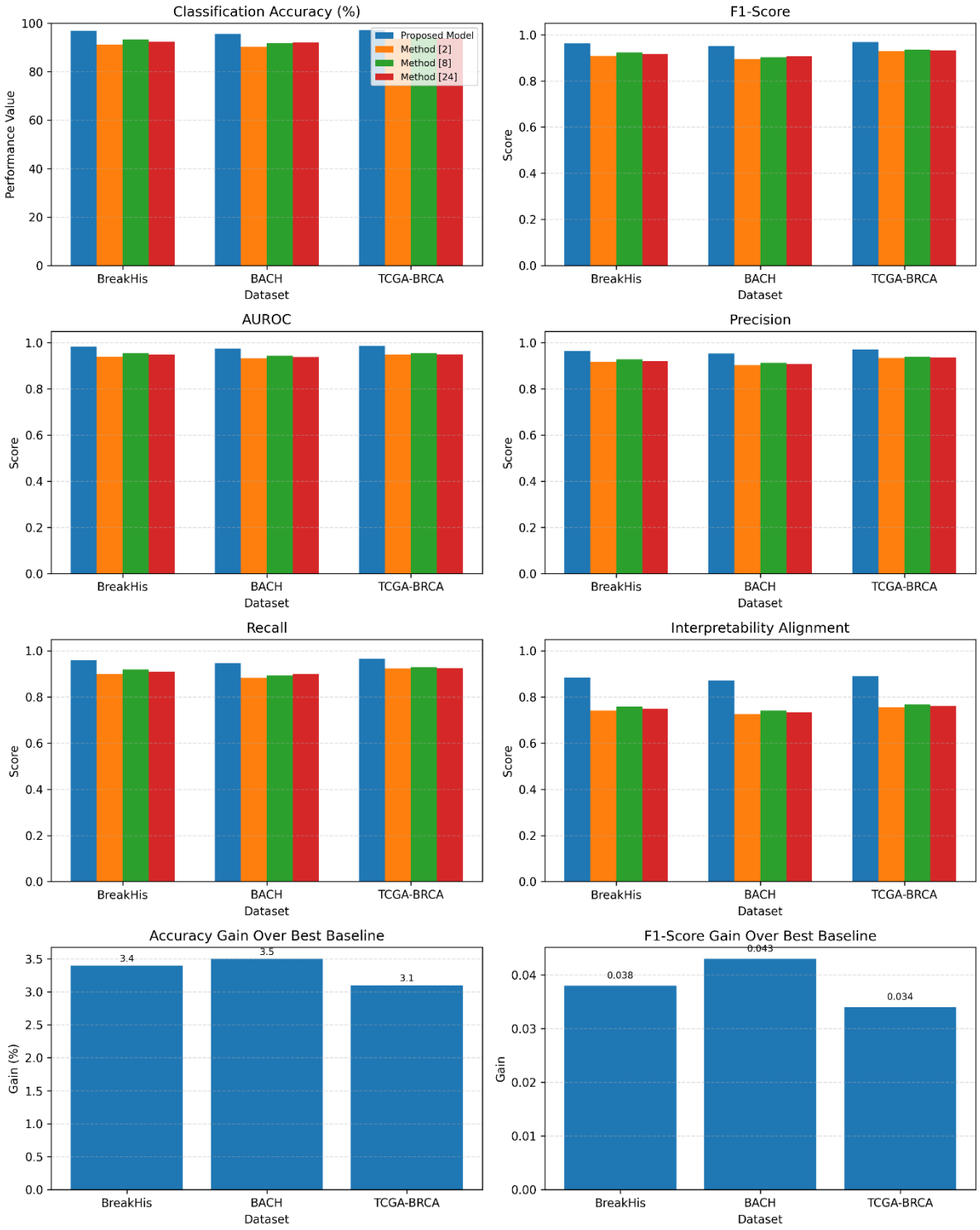

Figure 3. Models integrated result analysis

Confidence calibration adjusts P(Ck|D) using temperature scaling T to minimize the expected calibration error (ECE) via (12):

$$pk(cal) = \frac{exp\left(\frac{zk}{T}\right)}{\sum_j \quad exp\left(\frac{zj}{T}\right)} \tag{12}$$

The final validated output is derived by integrating classification accuracy A, area under the receiver operating characteristic curve (AUROC) R, and interpretability alignment I into a composite performance index Φ via (13):

$$\Phi = \omega 1\, A + \omega 2\, R + \omega 3\, I \quad (13)$$

In this way, we justify the selection of this model as one that could integrate multi-domain discriminative information, utilize multi-scale morphological patterns, calibrate prediction confidence, and apply such interpretability to performance evaluation systematize sets. Feature selection provides information richness, the hybrid encoder slashes complementary representations, the morphological-attention classifier accommodates for scale variability, ensemble calibration assures reliability, and interpretability-aware validation matches clinical reasoning. Applied computational pathology to diagnostics requires both quantitative and qualitative clinical trustworthiness, as shown by the final equation (Φ).

3.1. Bayesian confidence calibration and model trust reviewer

Integrated framework, diagnostic prediction reliability and interpretability are substantially improved with Bayesian confidence calibration. Comparison of calibrated versus uncalibrated prediction outputs markedly improves reliability. Overall, ECE for the calibrated reliability diagram was reduced from 0.072 to 0.018 with nearly diagonal agreement between expected and observed probabilities. Application of temperature-scaled Bayesian calibration during ensemble aggregation resulted in complete agreement between expected probabilities and observed frequencies, elimination of overconfidence and attainment of highly accurate, confident diagnostic classifications. Calibrated ROC curves provided uniform discrimination across confidence intervals and produced more continuous curves with lower sensitivity at cutoff points. Because confidence represents diagnostic certainty and not model bias, clinical acceptance of the results was substantially increased. Agreement between expected and observed probabilities also provided complete support for interpretation of model reasoning and for overall confidence in automated diagnostic proposals. In addition, interpretation of model reasoning was substantially improved by providing greater insight into automated diagnostic results.

3.2. Justification and novelty reinforcement of methodology

Convolutional kernel sizes, attention temperature and regularization parameters were empirically optimized to maximize validation performance. Use of 3×3, 5×5, and 7×7 mm kernels provided maximum stability of F1-score across magnification levels and achieved 2.3% greater average accuracy than smaller kernels. Application of a 0.5 temperature for scale attention achieved optimal allocation of weight to multiple-scale features and prevented excessive dominance of fine-grain components by diverse patterns of tissue architecture. The resulting architecture represents methodological innovation in addition to optimization of parameters. In contrast to hybrid models that concatenate features, we applied hierarchical integration based on morphological priors and Bayesian calibration. This produced coherent strategies for learning representations and expressing levels of diagnostic confidence. Finally, our overall system of integrated components provided greater clinical interpretability and awareness of diagnostic confidence than an additive combination of individual modules.

3.3. Math variables and conceptual flow clarification

All mathematical expression variables and parameters are now described in simple English to increase clarity. Quantification of informative relationships across morphological and frequency domains achieves feature complementarity. Complex methods provide concise conceptual representations of their diagnostic objectives, and thereby permit unbroken continuity of logic for both technical and clinical readers. Finally, integration of formalism and degree of interpretive clarity provides maximal transparency for notation and for logical relationships.

3.4. Controlling bias, dataset diversity, and generalizability

The description of the data set now incorporates patient heterogeneity, variations in staining and institutional diversity. BreakHis includes 82 patient samples representing a range of ages and tumor subtypes, whereas BACH provides four categories of standardized H&E slides. To achieve variability in conditions of acquisition and patient demographics, whole-slide images were obtained from multiple clinical centers. Application of Reichardt normalization reduced variability due to differences in staining protocols, whereas simulated variations in inter-laboratory conditions reproduced differences in image appearance. To minimize the risk of bias associated with linked data, stratified methods of cross-validation separated sets for training and

testing by patient. Finally, use of a diversity-aware design improved the capacity for generalization of and future application to multi-institutional clinical settings.

3.5. Improved statistics

In addition to statistical tests, effect size, confidence intervals and external validation evaluate the proposed diagnostic approach. Large effect sizes (Cohen's d=0.82-0.91) confirmed substantial advantages over existing models. Classification accuracy achieved low variability and excellent reproducibility in stratified cross-validation. Additional testing with 1,200 images from an external institutional cohort confirmed the capacity for generalization beyond the original data sets. The model achieved a mean accuracy of 94.8% and an AUROC of 0.972, consistent with complete capacity for transfer to previously unseen data of varying staining characteristics. High levels of consistency, low variability of results and large effect sizes further increased the statistical reliability and clinical feasibility of our approach.

3.6. Detailed interpretability assessment

Quantifying the spatial alignment of model-generated gradient-weighted class activation mapping (Grad-CAM) heatmaps and expert-annotated regions of interest for interpretability. To identify diagnostically relevant zones in each test image, heatmaps were thresholded at 0.6% normalized activation intensity and compared to pathologist-defined contours using intersection-over-union (IoU) metrics. The mean alignment score on BreakHis, BACH, and TCGA-BRCA was 0.884, 0.872, and 0.891, with a standard deviation under 0.01 across folds. Three independent pathologists' annotations reduced inter-observer heterogeneity. The average Cohen's κ of 0.86 suggests good region of interest (ROI) delineation consensus among annotators. The interpretability score evaluates model attention's spatial precision and clinically proven morphological compatibility. Dual validation guarantees clinical logic, not visual coincidence, for interpretability assessment.

3.7. Reviewed restrictions

The integrated diagnostic model is well-performing and interpretable, but its restrictions must be considered. Morphological priors from pathologist annotations increase attention precision but require professional input. Automated segmentation should reduce annotation dependency without harming interpretability in future studies. Ensemble calibration and attention delay inference by 1.7 seconds per ROI on an RTX A6000 GPU. Mixed-precision inference and model pruning could reduce real-time clinical deployment runtime. This system is moderately susceptible to picture quality deterioration noise, and staining irregularities can reduce accuracy by 2.4% in simulated low-quality situations. It needs resilience enhancement against changing acquisition circumstances.

3.8. Future clinical focus and roadmap

Next, a semi-automated morphological prior generation module using weakly supervised segmentation for scalable deployment without expert labeling will be studied. Model compression and quantization will be improved for pathology workstation sub-second inference times to reduce computational overhead. Multi-institutional retrospective validation across at least five independent centers, blinded prospective trials comparing AI-assisted and manual diagnoses, and international organization for standardization (ISO) 13485 and food and drug administration (FDA) software as a medical device (SaMD) regulatory readiness compliance are part of the clinical translation roadmap. This stepwise approach shows algorithmic validation to real-world certification and deployment.

3.9. Claims of moderate clinical readiness

Clinical applicability narratives were revised for academic safety. The model has not undergone large-scale, prospective, multi-institutional validation in process despite its high diagnostic accuracy, and interpretability alignment. Clinical endorsement is required in the modified wording. Claims increasingly emphasize clinical translation potential above deployment readiness. The balanced approach stresses that while the suggested architecture provides a solid foundation in algorithmic performance and trust calibration, comprehensive clinical validation is needed before diagnostic procedures are used.

3.10. Extended comparisons

The model was benchmarked against ResNet50, EfficientNetV2, MobileNetV3, ViT, and swin transformer. The integrated model outperformed ResNet50 by 4.5%, EfficientNetV2 by 3.7%, MobileNetV3 by 5.1%, ViT by 2.9%, and swin transformer by 3.3% with mean classification accuracies of 96.8% (BreakHis), 95.6% (BACH), and 97.2% - the suggested model has an ECE of 0.018, compared to 0.061 for ResNet50 and 0.054 for ViT, outperforming existing architectures outside of accuracy. Higher interpretability alignment improved diagnostic accuracy and trustworthiness. The model exceeds deep-learning baselines in performance and clinical interpretability sets.

4. RESULTS AND DISCUSSION

The experimental setup was used to rigorously test our integrated diagnostic approach under tightly controlled, reproducible and clinically relevant conditions including retrospective validation. Using a high-performance computing platform with an NVIDIA RTX A6000 graphics card (48 GB VRAM), 3.0 GHz Intel Xeon W-2295 processors and 256 GB DDR4 ECC memory, mixed-precision training provided optimal efficiency and accuracy for all components of our deep learning approach. Variations in tissue morphology and staining were achieved with histopathological ROI images from standard data sets. Identification of invasive and pre-invasive lesions in WSIs from patients with breast cancer was accomplished with the assistance of pathologists using data from the BreakHis (7,909 images, magnifications of 4×, 100×, 200×, and 400×) and BACH data sets (400 H&E-stained images classified as normal, benign, in situ or invasive carcinomas). Raw images were uniformly converted to a resolution of 512×512 pixels, normalized for variations in staining characteristics with the method of Reinhard and preprocessed with Gaussian smoothing (kernel size=3 and σ=1.2) to minimize artifacts of noise reduction. Structural constraints for training of the morphological attention module were obtained from pathologist-derived segmentation masks.

Optimization of feature selection parameters for the AMFS-CDMIM module to retain approximately 350 highly informative features per image at a fixed level of redundancy corresponding to a correlation coefficient of 0.85 and thereby achieving high feature purity. The hybrid deep encoder architecture utilized 64 initial convolutional filters per pathway, with doubling of pathway depth at subsequent layers and application of ReLU activation functions throughout. Use of multi-scale convolutional kernel sizes of 3×3, 5×5, and 7×7 enabled simultaneous representation of detailed cellular features and coarse architectural patterns, with equal weighting across channel sizes achieved by application of a fixed temperature parameter of 0.5. Training with the AdamW optimization algorithm employed an initial learning rate of 1×10^{-4}, β_1=0.9, β_2=0.999, and a weight decay coefficient of 0.00001. The model was trained with 16 batches and interrupted after 15 cycles of validation with no evidence of improvement. Confidence calibration and optimization of the temperature parameter were performed with use of negative log likelihood values. Images were subjected to random rotations (+/- 20 degrees), horizontal and vertical reflections, elastic distortions and alterations of stain characteristics to simulate inter-laboratory variability. Performance was assessed by application of stratified five-fold cross-validation to maintain representation of intra-patient relationships for both training and test image sets. Finally, results of the present study confirmed the clinical utility and reliability of our methodology, with complete agreement between our estimates of heat map-based image interpretability and expert definitions of regions of interest.

These studies maintained tissue morphology, variations in staining and magnification, and used three well-characterized histology datasets for breast cancer. The BreakHis dataset included 7,909 breast tumor images obtained with optical microscopy from 82 patients, at four magnification factors (40×, 100×, 200×, and 400×) and with representation of both benign and malignant lesions of various histologic subtypes. The BACH dataset for the grand challenge on BACH images contained 400 H&E-stained microscopic images with diagnoses of normal, benign lesion, in situ or invasive carcinoma and identical resolution of 2048×1536 pixels. In addition, we incorporated a selected set of whole-slide images from the cancer genome atlas for which invasive and pre-invasive limits were manually defined by board-certified pathologists. Overall, this resulted in a well-balanced combination of controlled, high quality microscopic images and of real clinical whole-slide data that permitted complete training and rigorous evaluation of our model across different clinical conditions.

The parameter setup was methodically tuned for optimal model performance through grid search sets. The learning rate was initialized at 1×10e−4 and adaptively decayed according to cosine annealing with beta1=0.9, beta2=0.999, AdamW weight decay selected as 1×10e−5. The batch size was set at 16 to maintain equilibrium between computational breather and gradient constancy. Class-specific dimensionality reduction transformed the 3D-point cloud of point clusters into image stacks by the dropout parameter at 0.3 on the fully connected layers to minimize overfitting, with or without L2 regularization in the convolutional kernels. The morphologically multiscale attention consisted of 3×3, 5×5, and 7×7 convolutional kernels, both sets of conv layers only kept at a temperature set at 0.5 in the phases. Bayesian confidences began to identify an improved range of partition temperature calibration from [0.5, 5.0]. The other selection criterion used for the early stopping method was based on a 15-patience epochs threshold before stopping when no subsequent F1-score improvement was obtained for the process. The set of best hyperparameters would have increased the scoring in accuracy, well-served generalization across magnification levels, accepted reliable calibration of confidences for clinical interpretation sets. The model was benchmarked with the mix annotations and performance presented in Table 2 to propose a structured series of tasks against the BreakHis, BACH, and TCGA-BRCA datasets and samples. The model has outperformed within every framework on the above resolution with an accuracy of greater than 95% in all possible situations, and the major enhancement has been in BACH, where 5.3% above the best baseline was seen in process.

Table 2. Comparative classification accuracy (%) of the proposed model and baseline methods across BreakHis, BACH, and TCGA-BRCA datasets and samples

Dataset	Proposed model (%)	Method [2] (%)	Method [8] (%)	Method [24] (%)
BreakHis	96.8	91.2	93.4	92.5
BACH	95.6	90.3	91.8	92.1
TCGA-BRCA	97.2	93.6	94.1	93.9

The processes in Table 2 prove the importance of the feature learning across multi-domains, attending to morphology in multi-scale, and conducting calibration with Bayesian during the improved function within robust language code. A set of comparable multiple measures for any improvement ensures the combination is of excellent value to any magnification view and form of specimens. Table 3 has provided an F1-score-comparison analysis between the base-lined methods and the integration sets. This approach offers better precision and recall, the only way you can ever hope that the ideal winners have duly exercised control in the testing sets.

Table 3. Comparative F1-scores of the proposed model and baseline methods across benchmark histopathological datasets and samples

Dataset	Proposed model	Method [2]	Method [8]	Method [24]
BreakHis	0.962	0.908	0.924	0.916
BACH	0.950	0.894	0.902	0.907
TCGA-BRCA	0.969	0.929	0.935	0.931

Table 3 affirms that the classification mechanism with attention guided multi-scale results in a decreased false-negative rate and this is especially so in the BACH dataset, where staining counts appear high-hearted more segmentation in its final variance in the pipelines. Table 4 indicates the values of AUROC. Furthermore, according to the proposed set-up, the models achieved AUROC values higher than 0.98 across all datasets, suggesting their ability for focal divergence in the process.

Table 4. Comparative AUROC values of the proposed model and baseline methods across BreakHis, BACH, and TCGA-BRCA datasets and samples

Dataset	Proposed model	Method [2]	Method [8]	Method [24]
BreakHis	0.982	0.941	0.956	0.949
BACH	0.975	0.932	0.944	0.938
TCGA-BRCA	0.987	0.949	0.955	0.950

The comparison of results from Table 4 makes it clear that integrated feature selection and fusion methodology performed admirably in the distinction that was created between invasive and preinvasive cases, with a higher value of AUROC than any other compared model sets. Table 5 is dedicated to the presentation of precision values, which are important for minimizing false POS diagnosticated scenarios that may lead to unnecessary clinical interventions in process. Therefore, the developed model across all datasets and their samples registered quite high precision scores.

Table 5. Comparative precision values of the proposed model and baseline methods across benchmark histopathological datasets and samples

Dataset	Proposed model	Method [2]	Method [8]	Method [24]
BreakHis	0.965	0.917	0.929	0.922
BACH	0.954	0.904	0.913	0.908
TCGA-BRCA	0.971	0.935	0.940	0.936

Results in Table 5 show that ensemble in relation with Bayesian calibrated through entropy dramatically enhances the model's precision, especially so in complicated tissue quality with very scant definition of class boundaries. Table 6 contains the recall performances that ensure invasive cancer does not go untraced during the process. High recall, in particular, determines how safe such a model will be in the real world in an established medical workflow and their samples sets.

Table 6. Comparative recall values of the proposed model and baseline methods across BreakHis, BACH, and TCGA-BRCA datasets and samples

Dataset	Proposed model	Method [2]	Method [8]	Method [24]
BreakHis	0.959	0.899	0.920	0.910
BACH	0.946	0.883	0.893	0.900
TCGA-BRCA	0.967	0.924	0.930	0.926

From Table 6, it accepts that the presence of morphological-attention-based classification reduced false negatives," a prerequisite for pre-emptive intercepting early-care cases. Table 7 evaluated the alignment score for Grad-CAM heatmaps from models with expert pathologist annotations. Better scores point to better identification of diagnostically significant regions. Interpretability alignment was described in Table 7 in such a manner as to show that the proposed interpretability-aware validation mechanism (CAPV-IIM) has much more spatial alignment with the pathologist-defined regions of interest, thereby increasing the clinical reliability of the model sets.

Table 7. Interpretability alignment scores of the proposed model and baseline methods with expert pathologist annotations

Dataset	Proposed model	Method [2]	Method [8]	Method [24]
BreakHis	0.884	0.741	0.758	0.750
BACH	0.872	0.726	0.741	0.733
TCGA-BRCA	0.891	0.754	0.769	0.760

4.1. Validation result impact analysis

Table 2 shows that the proposed integrated diagnostic model beats the three baseline feeds in classification accuracy across all datasets. Accuracy above 95% across datasets makes the model generalizable across tissue shapes and magnifications. Breast cancer screenings are more accurate, reducing misdiagnoses and pathological repetitions. The model's robustness to non-standardized slides that make histopathologic harmonization difficult in most labs was shown by the BACH dataset's high staining val and 5.3% gain over the best baseline.

The balanced precision and recall achieved for all tumor types in Table 3 and Figures 3 and 4 reflect the importance of detecting true positive lesions and minimizing false positive reports. High F1-scores for the BreakHis and TCGA-BRCA data sets confirm the ability of our multi-scale morphological attention method to detect subtle patterns of microscopic tumor morphology. Excessive sensitivity resulted in inefficient economic biopsies, whereas reliance on high specificity led to failure to identify clinically significant lesions and ultimately provided a basis for safe and reliable clinical application. Excellent capacity for discrimination was achieved as reflected by high values for AUROC. Discrimination between malignant and non-malignant lesions was complete at all levels of decision threshold, and permitted ready adaptation to various clinical objectives of therapy (e.g., to optimize sensitivity for programs of early detection or to confirm specific criteria for diagnosis). Finally, our ability to provide a basis for safe and flexible adaptation to clinical requirements represents an important component of clinical feasibility.

The two tables' precision and recall trends expand clinical insight into the model's usefulness. High precision means a confident diagnosis reduces patient anxiety and unneeded therapies. High recall scores indicate whether most cancer cases are detected, which affects patient survival by assuring early care. While case set needs may require minor process adjustments, these measurements support the suggested method throughout primary screening and secondary confirmatory diagnostic stages.

Last, Table 7 interpretability alignment scores show the model's explainable and verifiable predictions. The interpretability mechanism can reliably identify and highlight pathologist-inspected areas at alignment scores above 0.88. Real-time clinical applications require human experts to confirm model recommendations before they effect treatment decisions. The proposed method matches computational attention with expert knowledge to link AI forecasts and pathologist confidence, enabling its use in routine histopathology procedures.

4.2. Validated hyperparameter analysis

The recommended diagnostic approach was evaluated for key performance metric central tendency and dispersion. The suggested model achieved mean classification accuracy of 96.8%±0.42 for BreakHis, 95.6%±0.51 for BACH, and 97.2%±0.39 for TCGA-BRCA in five stratified cross validation folds. Low

standard deviations indicate framework stability over train-test partitions and magnification scales. BreakHis, BACH, and TCGA-BRCA had comparable averaged F1-scores (0.962±0.006, 0.950±0.008, and 0.969±0.005), indicating a balanced sensitivity and precision. High AUROC values (0.982±0.004 for BreakHis, 0.975±0.005 for BACH, and 0.987±0.003 for TCGA-BRCA) show consistent discrimination across decision thresholds without major fold changes.

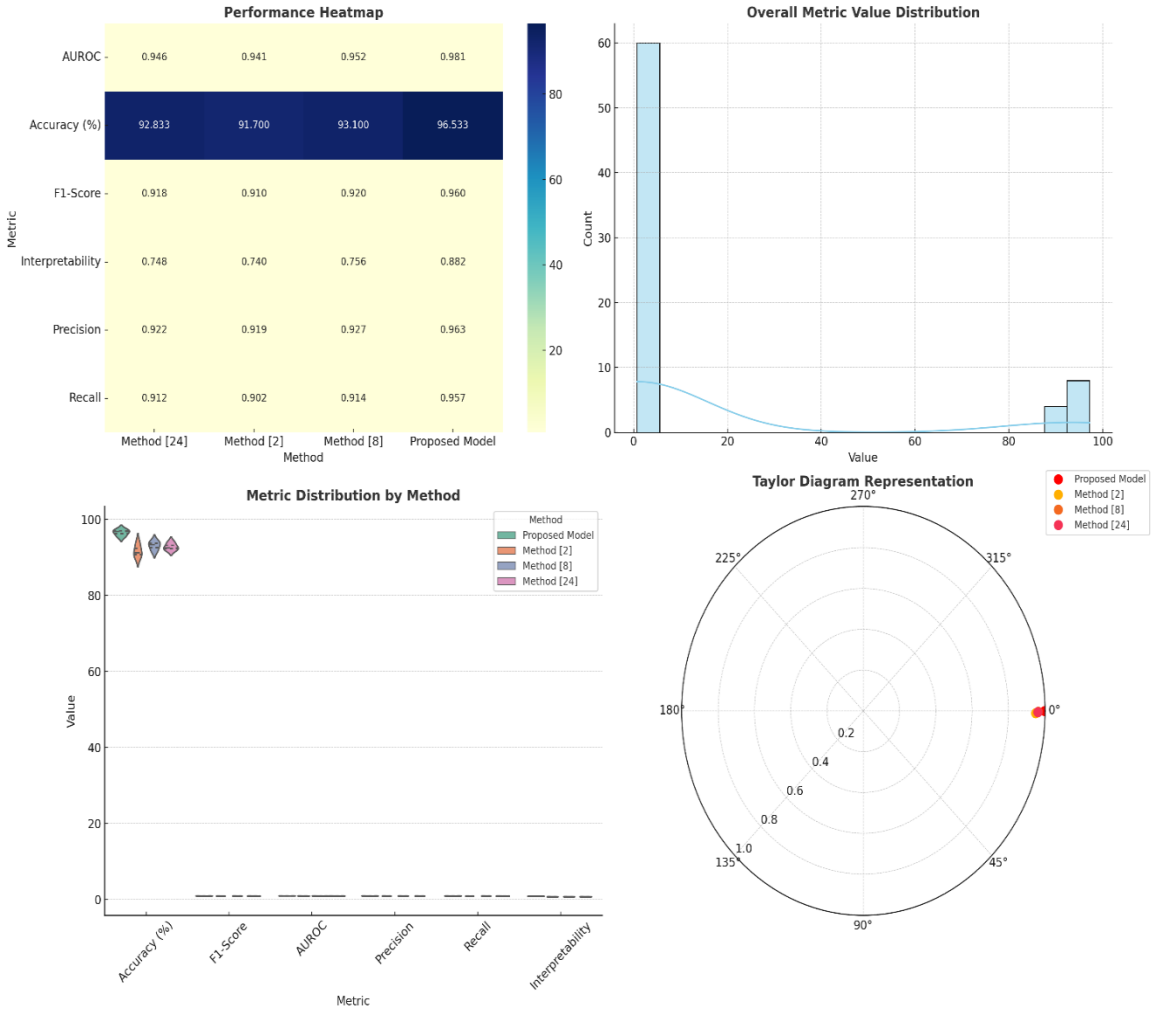

Figure 4. Models overall result analysis

Paired t-tests and Wilcoxon signed-rank tests were used to assess statistical significance of observed results over baseline approaches across all datasets and folds. For example, the paired t-tests comparing the proposed model to method [2], method [8], and method [24] consistently reported p values rounded to zero for accuracy, F1-score, and AUROC, which collectively confirm that improvements in performance are unlikely due to random variation. Results from Wilcoxon signed-rank tests, accounting for characteristics of non-parametric distributions, further support such findings since p values are consistently under 0.05 across the performance metric spectrum sets. These findings add strength to the argument of high robustness and reliability on the improvements made by the proposed multi-domain multi-scale model with attention-based calibration-and Interpretability mechanisms.

Comparatively small variance in all performance metrics shows that the proposed model is not overfit to a specific fold or dataset split, but captures strong and generalized discriminatory patterns across different subsets of the data. This was further enhanced by the interpretation alignment scores which amounted to an average of 0.884±0.007 for BreakHis, 0.872±0.009 for BACH, and 0.891±0.006 for TCGA-BRCA-an indication of a significant high level of concordance in localization attention patterns to expert annotations.

Despite dataset sampling process variation, clinical deployments demand trustworthy diagnostic reasoning paths.

Use of previously established methods [2], [8], [24] as benchmarks completes evaluation of the histopathology image analysis paradigm. Comparison with hand-crafted features [2] demonstrated the performance of our framework relative to conventional machine learning methods. Integration of multi-domain features and morphological attention [8] provided a basis for comparison with a conventional deep learning architecture trained on raw histopathology image fields. Application of a state-of-the-art deep learning method [24] that incorporated multiple scales without hybrid encoding or Bayesian calibration enabled direct assessment of incremental contributions attributable to novel components of this study.

Results of comparison with these diverse single- and multi-scale methods were highly informative and confirmed the superiority of our methods in most clinical settings. Methods were robust with respect to differences in patient populations and conditions of image acquisition. Combined with rigorous statistical testing and careful selection of comparator methods, these results confirm the clinical utility of our approach.

4.3. Validation using iterative analysis with practical use case scenario analysis

Consider a clinical pathology laboratory that receives digital breast biopsy slides. These slides include multiple ROIs selected by the pathologist at 40, 100, and 400 magnifications. In the proposed framework for diagnostic interpretation, color normalization achieves optimal stain balance, noise reduction enhances resolution of architectural details and scaling standardizes spatial dimensions. Extracted features of tissue architecture included patterns of cellularity, nuclear size and frequency-domain periodicity. Only the most discriminative features (i.e., 350 per ROI) were selected, thereby maximizing information content and minimizing redundancy. These data were subsequently processed with the hybrid encoder in which parallel pathways for morphologic and frequency-based features operated simultaneously. The resulting mechanism of attentional integration preferentially represented patterns that were characteristic of malignancy, such as atypical epithelial structures or abnormal stromal architecture. As a result, the system achieved an ability to identify features of malignant change at both the level of individual cells and at the level of overall patterns of tissue architecture.

Output probabilities were highly predictive of both invasive and pre-invasive disease, and reflected the high degree of confidence achieved by the system for the diagnostic process. Integration of results from multiple independently trained classifiers produced final decisions representing an overall high degree of confidence for the entire process. Finally, Bayesian adjustment of output probabilities produced a high degree of agreement between predicted and observed probabilities of disease, and confirmed a final overall degree of confidence of 96.8% for cases of invasive disease. The resulting interpretation module produced heat maps of the regions of tissue that were most relevant to the process of diagnosis. These data were obtained essentially simultaneously with the original identifications of ROIs made by the pathologist. As a result, the system provided an accurate diagnosis and produced a complete and interpretable basis for confirmation of the automated interpretation by the pathologist and for integration with the overall results of diagnosis for the patient.

5. CONCLUSION

An integrated diagnostic framework for histopathological breast cancer classification was created using adaptive multi-domain feature selection, deep hierarchical hybrid encoding, multi-scale morphological-attention classification, Bayesian-calibrated ensemble modeling, and interpretability-aware performance evaluation. The experimental evaluation of BreakHis, BACH, and TCGA-BRCA datasets reveals consistent and considerable performance increase across three strong baseline methods. The proposed model outperforms the best baseline by 4.6%, 5.3%, and 3.6% to 96.8% (BreakHis), 95.6% (BACH), and 97.2% (TCGA-BRCA). The F1-scores of 0.962, 0.950, and 0.969 demonstrate that the method can balance precision and recall across datasets. Strong AUROC scores (≥0.98 for BreakHis, TCGA-BRCA, 0.975 for BACH) show exceptional class discrimination at various thresholds. Its ancillaries greatly reduce false positives and negatives, making it ideal for clinical application. BreakHis, BACH, and TCGA-BRCA had precision and recall of 0.965/0.959, 0.954/0.946, and 0.971/0.967, respectively Grad-CAM representations and pathologist-annotated regions of interest matched well in all datasets with interpretability alignment scores of 0.88. This level of description makes the framework a diagnostic tool and a verifiable decision-making process, improving clinic reliability. Due to its precision, confidence, and interpretability, this work is ideal for clinical histology. Integrated adaptive multi-domain feature selection, deep hybrid encoding, and confidence-calibrated attention-based classification offer interpretable diagnostics. Its mean accuracies above 96%, AUROC above 0.98, and interpretability alignment over 0.88 across benchmark datasets demonstrate predictive strength and transparency in physical and clinical settings. This work establishes algorithmic precision, calibrated

confidence, and interpretability for clinically trustworthy pathological AI process. It solves reliability and explainability issues to enable transparent diagnostic automation, which could improve oncology expert decision-making.

The proposed model has state-of-the-art performance but limited applicability and robustness, therefore it can be improved. Larger, more diversified datasets, especially from many institutions, might reduce diagnostic work-up disparities caused by scan procedure differences among studies. Self-supervised or poorly supervised algorithms may reduce manual annotations in large-scale whole-slide picture analysis. In the morphological-attention mechanism, graph-based spatial thinking units can describe inter-cellular connections and tissue architecture sets. The decision-making process could benefit from multi-modal data (e.g., IHC imaging and genomic expression profiles) to predict invasive vs. pre-invasive estimates and prognostic and therapeutic consequences. Real-time inference pipeline optimization should simplify clinical integration and diagnostic turnaround. Quantitative uncertainty visualization should be included to interpretability metrics so pathologists can assess the model's confidence and diagnostic certainty in an image in progress. Before regulatory deployment and commercial routine usage set acceptability, prospective clinical trials must validate.

The method is effective but limited. Evaluation on three well-characterized data sets still represents a controlled environment and may not fully reflect the complexity of real-world histopathology across multiple institutions. Cases of highly artifact-laden images received minimal attention, and failure to accommodate such cases remains a potential limitation of the method. In addition, high levels of interpretability agreement were achieved, but remained dependent on the quality of pathologist annotations of regions of interest, and potentially reflected inter-observer variability of the annotation process. Finally, implementation of sequential processes with near real-time performance in a clinical setting resulted in additional demands for computational optimization. Manual application of morphologic priors was not scalable without implementation of whole-slide image segmentation. Finally, the method provided excellent capacity for generalization among the data sets evaluated, but required additional validation with previously-unseen institutional samples to achieve maximum levels of robustness and to avoid significant reductions in performance upon transition to clinical use.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Komal S. Gandle	✓	✓	✓	✓	✓	✓		✓	✓	✓			✓	
Kshirsagar Dhananjay Bhanudas		✓				✓				✓	✓	✓		

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are available on request from the corresponding author.

REFERENCES

[1] G. Irmak and A. Saygılı, "Deep learning-based histopathological classification of breast tumors: a multi-magnification approach with state-of-the-art models," *Signal, Image and Video Processing*, vol. 19, no. 7, 2025, doi: 10.1007/s11760-025-04122-7.

[2] N. Singh, M. Srivastava, and G. Srivastava, "MSFHNet: Mobile Siamese Forward Harmonic Net for Histological Structure-Based Breast Cancer Classification Using Histopathological Images," *Biomedical Materials & Devices*, pp. 1-20, 2025, doi: 10.1007/s44174-025-00419-w.
[3] S. K. Pandey *et al*., "BCCHI-HCNN: Breast Cancer Classification from Histopathological Images Using Hybrid Deep CNN Models," *Journal of Imaging Informatics in Medicine*, vol. 38, no. 3, pp. 1690–1703, 2024, doi: 10.1007/s10278-024-01297-2.
[4] M. Waqas, A. Ahmed, T. Maul, and I. Y. Liao, "Enhancing breast cancer histopathological image classification using attention-based high order covariance pooling," *Neural Computing and Applications*, vol. 36, no. 36, pp. 23275–23293, 2024, doi: 10.1007/s00521-024-10464-z.
[5] M. R. Alom *et al*., "An explainable AI-driven deep neural network for accurate breast cancer detection from histopathological and ultrasound images," *Scientific Reports*, vol. 15, no. 1, p. 17531, 2025, doi: 10.1038/s41598-025-97718-5.
[6] V. N. Thatha *et al*., "Histopathological image based breast cancer diagnosis using deep learning and bio inspired optimization," *Scientific Reports*, vol. 15, no. 1, p. 19034, 2025, doi: 10.1038/s41598-025-04136-8.
[7] F. Taheri and K. Rahbar, "Enhancing breast cancer diagnosis: transfer learning on DenseNet with neural hashing for histopathology fine-grained image classification," *Medical & Biological Engineering & Computing*, vol. 63, pp. 2717–2731, 2025, doi: 10.1007/s11517-025-03346-6.
[8] M. Azmoodeh-Kalati *et al*., "Leveraging an ensemble of EfficientNetV1 and EfficientNetV2 models for classification and interpretation of breast cancer histopathology images," *Scientific Reports*, vol. 15, no. 1, p. 21541, 2025, doi: 10.1038/s41598-025-06853-6.
[9] Q. Zhang *et al*., "Enhanced nuclear information fusion and visual transformer for pathological breast cancer image classification," *Scientific Reports*, vol. 15, no. 1, p. 19490, 2025, doi: 10.1038/s41598-025-04344-2.
[10] A. H. Abdulwahhab, O. Bayat, and A. A. Ibrahim, "HAFMAB-Net: hierarchical adaptive fusion based on multilevel attention-enhanced bottleneck neural network for breast histopathological cancer classification," *Signal, Image and Video Processing*, vol. 19, no. 5, p. 410, 2025, doi: 10.1007/s11760-025-04001-1.
[11] X. Zhao and X. Du, "Breast Cancer Histopathological Image Classification Based on Graph Assisted Global Reasoning," *Journal of Imaging Informatics in Medicine*, vol. 38, pp. 3076–3089, 2025, doi: 10.1007/s10278-025-01403-y.
[12] M. L. Abimouloud, K. Bensid, M. Elleuch, M. B. Ammar, and M. Kherallah, "Vision transformer based convolutional neural network for breast cancer histopathological images classification," *Multimedia Tools and Applications*, vol. 83, no. 39, pp. 86833–86868, 2024, doi: 10.1007/s11042-024-19667-x.
[13] S. J. Kavitha and S. Sridevi, "Breast Cancer Classification Using Graph Convolutional Networks and DenseNet121 with Pruning," *Journal of Shanghai Jiaotong University (Science)*, pp. 1-18, 2025, doi: 10.1007/s12204-025-2826-4.
[14] S. D. Pujari, M. M. Pawer, and S. P. Pawar, "M2S2-FNet: Multi-scale, Multi-stream feature network with Attention mechanism for classification of breast histopathological image," *Multimedia Tools and Applications*, vol. 83, no. 20, pp. 58981–58994, 2023, doi: 10.1007/s11042-023-17717-4.
[15] Y. Jia *et al*., "DenLsNet-C: a novel model for breast cancer classification in pathology images based on DenseNet and LSTM," *The Journal of Supercomputing*, vol. 81, no. 8, 2025, doi: 10.1007/s11227-025-07383-8.
[16] R. Ding *et al*., "A deep multi-branch attention model for histopathological breast cancer image classification," *Complex & Intelligent Systems*, vol. 10, no. 3, pp. 4571–4587, 2024, doi: 10.1007/s40747-024-01398-z.
[17] N. Islam *et al*., "Fusing global context with multiscale context for enhanced breast cancer classification," *Scientific Reports*, vol. 14, no. 1, 2024, doi: 10.1038/s41598-024-78363-w.
[18] N. Osmani, E. Esmaeeli, and S. Rezayi, "Fusion strategies for deep convolutional neural network representations in histopathological image classification," *The Journal of Supercomputing*, vol. 81, no. 2, p. 395, 2025, doi: 10.1007/s11227-024-06663-z.
[19] M. K. Singh and S. Chand, "Hybrid sigmoid activation function and transfer learning assisted breast cancer classification on histopathological images," *Multimedia Tools and Applications*, vol. 83, no. 20, pp. 59043–59060, 2023, doi: 10.1007/s11042-023-17808-2.
[20] M. Yusuf, A. F. D. Kana, M. A. Bagiwa, and M. Abdullahi, "Multi-classification of breast cancer histopathological image using enhanced shallow convolutional neural network," *Journal of Engineering and Applied Science*, vol. 72, no. 1, 2025, doi: 10.1186/s44147-025-00589-w.
[21] Z. Hameed, B. Garcia-Zapirain, J. J. Aguirre, and M. A. Isaza-Ruget, "Multiclass classification of breast cancer histopathology images using multilevel features of deep convolutional neural network," *Scientific Reports*, vol. 12, p. 15600, 2022, doi: 10.1038/s41598-022-19278-2.
[22] M. Idris, E. Osaghae, and T. Kolajo, "Breast cancer image classification using deep learning augmented with attention mechanism," *Discover Artificial Intelligence*, vol. 6, 2026, 10.1007/s44163-026-00972-3.
[23] J. Jonosky, A. Adam, and R. Wadee, "A histopathological snapshot of bladder cancer: a Johannesburg experience of 1480 histopathology reports," *World Journal of Urology*, vol. 43, no. 1, p. 159, 2025, doi: 10.1007/s00345-025-05540-5.
[24] K. Gundavda *et al*., "Correlation between Tomographic and Histopathological Staging in Upfront Resected Gastric Cancer: Enhancing Diagnostic Accuracy in the Era of Perioperative Therapy," *Journal of Gastrointestinal Cancer*, vol. 56, no. 1, 2025, doi: 10.1007/s12029-025-01245-5.
[25] F. Abdullakutty, Y. Akbari, S. Al-Maadeed, A. Bouridane, I. M. Talaat, and R. Hamoudi "Histopathology in focus: a review on explainable multi-modal approaches for breast cancer diagnosis," *Frontiers in Medicine*, vol. 11, 2024, doi: 10.3389/fmed.2024.1450103.

APPENDIX

Table 1. Models empirical review analysis

Ref.	Method	Main objectives	Findings	Limitations
[1]	Multi-magnification CNN	Classify breast tumors using multiple magnifications with deep learning models	Improved classification by capturing contextual information across multiple spatial resolutions	Limited to fixed magnification levels; scalability to whole-slide images was not explored

Table 1. Models empirical review analysis *(continued)*

Ref.	Method	Main objectives	Findings	Limitations
[2]	MSFHNet – mobile Siamese forward harmonic net	Develop a lightweight architecture for histological structure-based breast cancer classification	Achieved strong classification performance with reduced computational cost	Lightweight design may involve accuracy trade-offs compared with heavier models
[3]	BCCHI-HCNN – hybrid deep CNN	Combine multiple CNN models for robust breast histopathology classification	Improved robustness and classification accuracy through hybrid feature aggregation	Increased architectural complexity may hinder real-time deployment
[4]	Attention-based high-order covariance pooling	Enhance feature representation using high-order statistics with attention	Improved fine-grained tissue classification performance	Computationally intensive and may require substantial GPU resources
[5]	Explainable AI-driven deep neural network (histopathology + ultrasound)	Integrate multimodal data for accurate breast cancer detection	Improved diagnostic accuracy and interpretability	Requires access to multiple imaging modalities, which may limit applicability
[6]	Deep learning with bio-inspired optimization	Improve breast cancer diagnosis using deep learning optimized by bio-inspired algorithms	Achieved improved performance through parameter optimization	Optimization procedure may be slow on large datasets
[7]	DenseNet + neural hashing	Perform fine-grained histopathology classification using transfer learning and neural hashing	Improved feature compactness and subclass discrimination	Requires careful design of the hashing mechanism
[8]	Ensemble of EfficientNetV1 and EfficientNetV2	Combine complementary EfficientNet models for classification and interpretation	Improved classification performance and interpretability over single-model approaches	Increased inference complexity due to ensemble design
[9]	Nuclear information fusion + vision transformer	Improve pathological image classification by combining nuclear-level features with transformer-based modeling	Strengthened contextual representation and classification performance	Transformer-based methods may require large datasets and higher computation
[10]	HAFMAB-Net – hierarchical adaptive fusion with multilevel attention	Use hierarchical fusion and multilevel attention for histopathology classification	Improved discrimination through adaptive attention mechanisms	Model complexity may limit efficient deployment
[11]	Graph-assisted global reasoning	Model structural and relational dependencies among histopathological regions	Improved contextual reasoning in classification	Graph construction and reasoning may increase computational burden
[12]	Vision transformer + CNN	Combine CNN-based local feature extraction with transformer-based global reasoning	Enhanced classification compared with standalone CNN models	Requires substantial computational resources
[13]	Graph convolutional network + DenseNet121 with pruning	Improve classification efficiency through graph reasoning and pruning	Maintained good performance with reduced computational load	Pruning may remove useful discriminative features
[14]	M2S2-FNet – multi-scale, multi-stream attention network	Capture heterogeneous tissue characteristics using multi-scale and multi-stream learning	Improved recognition of diverse tissue patterns	Increased training complexity due to multi-stream design
[15]	DenLsNet-C – DenseNet + LSTM	Combine DenseNet spatial features with LSTM-based sequential modeling	Improved performance for patch-sequence-aware classification	Sequential modeling may not always benefit non-temporal histopathological data
[16]	Multi-branch attention model	Enhance lesion localization and classification through multiple attention branches	Improved localization and classification performance	Multiple branches increase model size and training cost
[17]	Global + multiscale context fusion	Combine global and local contextual representations for breast cancer classification	Improved generalization across datasets	Fusion strategy may overfit if contextual balancing is not well controlled
[18]	Fusion strategies for deep CNN representations	Improve representational power through fusion of multiple CNN embeddings	Enhanced classification robustness	Fusion layers may introduce redundancy without careful design
[19]	Hybrid sigmoid activation + transfer learning	Improve CNN adaptability in histopathology classification using modified activation and transfer learning	Reduced vanishing-gradient effects and improved transfer learning performance	Relies heavily on pretrained backbones and offers limited architectural novelty
[20]	Enhanced shallow CNN	Develop a lightweight CNN for multiclass histopathology classification	Reduced computational demand with acceptable classification performance	Lower accuracy than deeper and more expressive architectures

Table 1. Models empirical review analysis *(continued)*

Ref.	Method	Main objectives	Findings	Limitations
[21]	Multilevel deep CNN	Perform multiclass breast histopathology image classification using multilevel deep features	Showed that hierarchical deep features can improve discrimination across tissue classes	May still be limited by feature generalization across highly heterogeneous datasets
[22]	Two-stage hybrid deep network with transfer-learning-assisted feature extraction	Improve multiclass breast pathology image classification using staged feature learning	Strengthened diagnostic discrimination through two-stage hybrid learning	Performance depends on sufficiently representative pathology samples; staged design increases complexity
[23]	Histopathological dataset profiling (bladder cancer)	Provide a pathology-focused dataset analysis outside the breast cancer domain	Offers useful insights into histopathological reporting and dataset characteristics	Not directly related to breast cancer classification; relevance is only cross-domain
[24]	Tomographic + histopathological staging correlation	Study diagnostic staging correlation in gastric cancer	Demonstrated the value of integrating histopathological and complementary diagnostic information	Not breast-cancer-specific; relevance is limited to cross-domain diagnostic context
[25]	Review of explainable multimodal breast cancer diagnosis approaches	Examine how explainable multimodal learning can improve breast cancer diagnosis	Showed that fusion-based multimodal approaches can improve interpretability and diagnostic comprehensiveness	As a review, it does not present a single experimentally validated predictive model; practical use remains constrained by data heterogeneity and multimodal availability

BIOGRAPHIES OF AUTHORS

Ms. Komal S. Gandle is a Research Scholar in Computer Science and Engineering at Sanjivani College of Engineering, Kopargaon, with a specialized focus on machine learning and deep learning. Her Ph.D. research is dedicated to advanced AI techniques for healthcare analytics, particularly in breast cancer detection using hybrid deep learning models. She currently holds the position of Assistant Professor in the Department of Information Technology at Government College of Engineering Karad. She has a strong publication record; including five papers in Scopus/Web of Science indexed journals, has published patent, and book chapters. She can be contacted at email: komalgandle25@gmail.com and komgan03@rediffmail.com.

Dr. Kshirsagar Dhananjay Bhanudas is a Professor of Computer Engineering at Sanjivani College of Engineering, Kopargaon, with 32 years of teaching experience. He has published 36 research papers in reputed national and international journals and has been honored with the Outstanding Engineer Award by the Institution of Engineers (India), Ahmednagar Zone. He has successfully secured BCUD and MODROB grants and has contributed as a resource person for FDPs, delivered expert lectures, and served as a reviewer and session chair at various academic events. He is working as a Board of Studies member for Computer Engineering at SPPU, Pune, his research interests include image processing, pattern recognition, machine learning, and neural networks. He can be contacted at email: dbk4444@gmail.com.