

ARTICLE

E²-MCLNet: An error- and evidence-aware multiscale learning framework for reliable breast cancer histopathology classificationSirisha Yerraboina^{1*}, and Harikrishna Bommala²¹Department of Computer Science and Engineering, Bharatiya Engineering Science & Technology Innovation University, Gorantla, Sri Sathya Sai District, Andhra Pradesh, India²Department of Computer Science and Engineering, KG Reddy College of Engineering & Technology, Chilkur, Moinabad, Rangareddy, Telangana, India

Abstract

Accurate breast histopathology classification is essential for reliable diagnosis; however, tissue heterogeneity, staining variation, and magnification-dependent morphology make automated classification challenging. This study proposes E²-MCLNet, an error- and evidence-aware multiscale learning framework for binary breast histopathology classification on the BreakHis dataset. The model uses a pretrained ResNet-34 backbone to extract intermediate and deep multiscale features, fuses adaptively pooled descriptors, and applies a Softplus-based evidential classification head with a custom ErrorEvidenceLoss to emphasize difficult and error-prone samples. Using a fixed 70:15:15 train-validation-test split, E²-MCLNet achieved an accuracy of 97.6411%, a precision of 98.0296%, a recall of 98.5149%, a specificity of 95.7784%, an F1-score of 98.2716%, and an area under the curve of 0.9974. Confusion matrix, receiver operating characteristic/precision-recall, and heatmap analyses indicate that the proposed framework provides strong discriminative performance while supporting evidence-aware and reliability-sensitive prediction.

Keywords: Breast cancer histopathology classification; Multiscale learning; Evidential deep learning; Error-aware learning; Reliability-aware classification; BreakHis; ResNet-34; Computer-aided diagnosis

***Corresponding author:**Sirisha Yerraboina
(2022wpcse020@bestiu.edu.in)

Citation: Yerraboina, S. & Bommala, H. (2026). E²-MCLNet: An error- and evidence-aware multiscale learning framework for reliable breast cancer histopathology classification. *Int J Systematic Innovation*, 10(5): 026210105.

[https://doi.org/10.6977/IJoSI.202610_10\(5\).0011](https://doi.org/10.6977/IJoSI.202610_10(5).0011)

Received: May 18, 2026

Revised: June 16, 2026

Accepted: July 4, 2026

Published online: July 28, 2026

Copyright: © 2026 Author(s). This is an Open-Access article distributed under the terms of the Creative Commons Attribution License, permitting distribution, and reproduction in any medium, provided the original work is properly cited.

Publisher's Note: AccScience Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

Histopathology has been a foundational tool for breast cancer diagnosis for over a century and continues to be used today for primary diagnostic confirmation and cancer subtype definition. Manual microscopic assessment of slides, however, is labor-intensive and prone to interobserver disagreement, leading to the ongoing development and research into computer-aided detection/diagnosis of breast histopathology images (Abdullakutty *et al.*, 2024; Datwani *et al.*, 2025; Jiang *et al.*, 2024). In recent years, deep learning has come to dominate the area since such algorithms can learn feature representations hierarchically from stained tissue specimens, enabling the development of automated detection, classification, and prediction methods for other clinical outcomes (Jiang *et al.*, 2024; Ramamoorthy *et al.*, 2024).

Over time, computer-aided breast histopathology classification methods have moved from hand-crafted image processing toward a series of advances in transfer learning, ensemble learning, attention-based methods, and, recently, combinations of convolutional neural networks (CNNs) and transformers (Aldakhil *et al.*, 2024; Balasubramanian *et al.*, 2024; Wang *et al.*, 2025). Previous studies have shown strong results achieved through fine-tuning of CNN backbones (e.g., ResNet, EfficientNet, DenseNet-201), the addition of attention modules to EfficientNet backbones, ensembles of EfficientNet variants, feature selection via class-discriminative maps, and the integration of multi-feature information extraction approaches on popular breast cancer histopathology datasets, such as the BreaKHis dataset (Chaudhary & Dhunny, 2025; Heikal *et al.*, 2024; Wang *et al.*, 2025; Yan *et al.*, 2025; Zhao *et al.*, 2025). Such performance shows that deep learning can capture key diagnostic features from hematoxylin and eosin-stained tissue histopathology images. However, it also highlights key challenges: class imbalance between benign and malignant types, stain heterogeneity, variations in features depending on microscopy magnification, and the requirement to maintain information about local cellular details while preserving a broader tissue architecture perspective (Eshun *et al.*, 2024; Jiang *et al.*, 2024; Spanhol *et al.*, 2016).

One common difficulty in histopathology research is that tissue pathology exhibits variability across multiple magnifications: the discriminative factors can include texture and appearance at individual nuclei, morphology of epithelial glands, patterns in the stromal compartment and architectural arrangement of tissue (Jiang *et al.*, 2024; Spanhol *et al.*, 2016). To address this challenge, several studies use architectures that aggregate and integrate features at multiple levels of a network or different spatial resolutions of input images rather than depending solely on information from the deepest layer (Chaudhary & Dhunny, 2025; Ukwuoma *et al.*, 2022; Wang *et al.*, 2025). Additionally, several studies have shown that the adoption of dual-branch networks, fusion architectures, and multi-feature frameworks is helpful for improving discrimination ability (Wang *et al.*, 2025; Yan *et al.*, 2025; Zhao *et al.*, 2025). Fusion can combine distinct information from different spatial or feature scales, helping to better extract diagnostically relevant information in complex histopathological images (Chaudhary & Dhunny, 2025; Wang *et al.*, 2025). In the present study, we explore this methodology by combining the pooled features from intermediate and deep layers of the trained ResNet-34 feature extractor to enhance the binary classification performance of benign versus malignant tumors using the BreaKHis breast histopathology image dataset.

Another significant problem in many leading classification models is their primary focus on high predictive accuracy without sufficient attention to the reliability of their predictions. A prediction of an outcome cannot be the only information that a medical professional needs for decision-making, since there may be insufficient medical data (rare condition), low-quality histopathology images, ambiguous cases where experts may disagree, and even incorrect outputs (Ghasemi *et al.*, 2024; McCaffrey *et al.*, 2024). To address this problem, many recent studies attempt to build interpretable and uncertainty-aware artificial intelligence (AI) systems for breast cancer detection (Abdullakutty *et al.*, 2024; Alom *et al.*, 2025; Ghasemi *et al.*, 2024). The emphasis is placed on models that not only achieve high prediction accuracy but also provide robust, reliable outcomes and interpretability, even under uncertain conditions. Evidential deep learning attempts to go beyond the state-of-the-art accuracy-only classifiers by predicting class evidence instead of normalized probabilities, enabling robust and uncertainty-aware decision-making even in ambiguous cases (Fu *et al.*, 2026).

The concept behind the proposed E²-MCLNet is presented in Figure 1. The illustration conveys that (i) breast histopathology classification inherently faces challenges related to multiscale tissue heterogeneity and visual ambiguity and (ii) medical professionals critically require reliable outputs for better and more accurate decision-making.

Next, the proposed E²-MCLNet architecture is highlighted as a promising direction for addressing these challenges by (i) performing multiscale feature extraction to capture complementary histomorphological patterns; (ii) applying an error-aware learning objective to penalize inaccurate predictions, and (iii) utilizing an evidence-aware inference strategy to improve decision confidence. The primary role of Figure 1 in the introduction section is to rapidly convey to the readers why a simple classification model that relies on a single scale or solely optimizes for prediction accuracy may not be sufficient for reliable histopathology diagnosis and why the proposed framework is presented as a reliability-centric alternative.

To position the present study within the context of existing literature and highlight the identified research gaps, Table 1 summarizes recent key publications in breast histopathology classification, focusing on the 2024–2025 period. The table systematically lists each study, including the backbone model used, the dataset, the specific task, the core methodology, and the main limitation. By presenting this information in a structured manner, Table 1 is intended not just to offer a descriptive overview but also

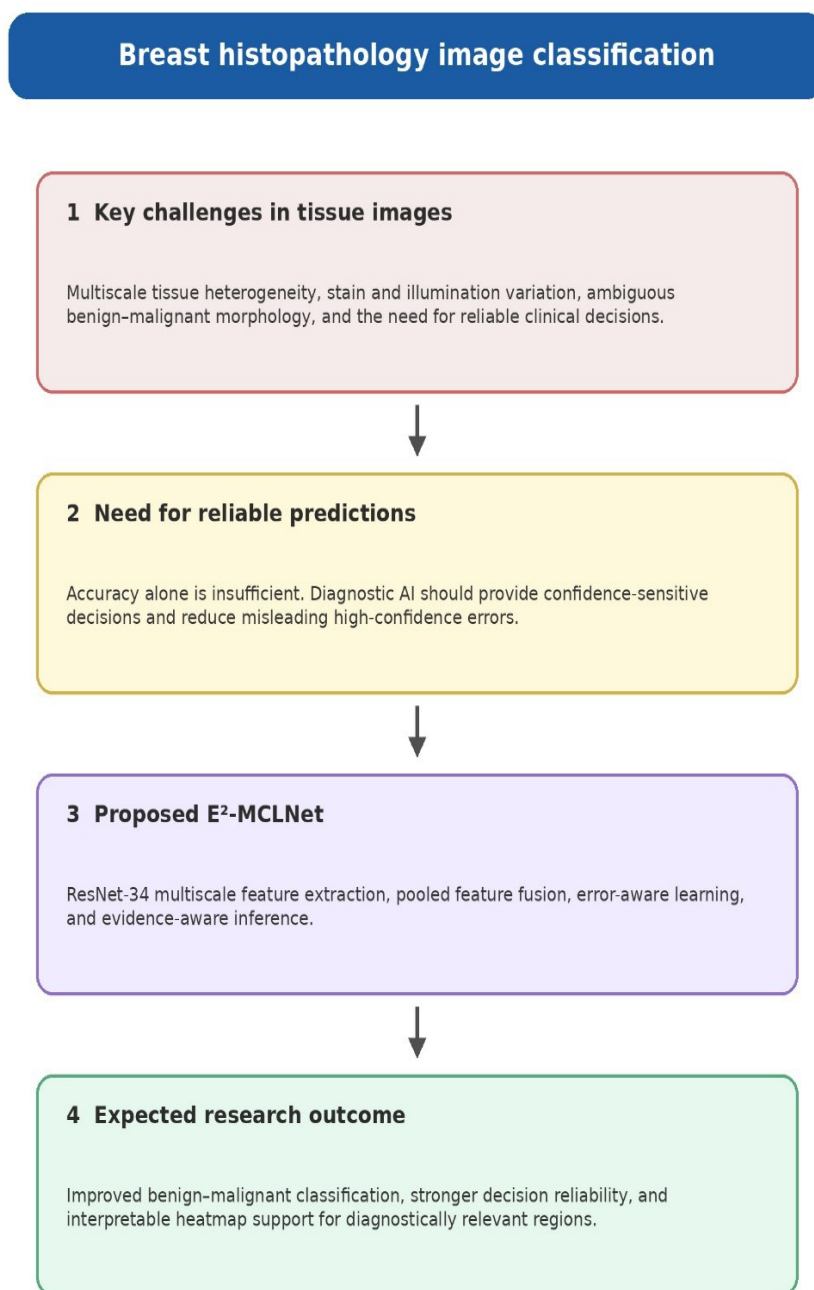


Figure 1. Overall conceptual motivation of E²-MCLNet for reliable breast histopathology classification
Abbreviation: AI: Artificial intelligence.

to show the evolution of research in this area. Although various approaches such as transfer learning, attention mechanisms, feature fusion, and hybrid architectures have successfully improved prediction performance, the studies summarized in Table 1 highlight the continuing need to jointly address the challenges of multiscale representation learning and prediction reliability. Therefore, Table 1

provides the motivation for proposing E²-MCLNet and forms the basis for discussing the specific research gap addressed by this study.

Motivated by these shortcomings, we propose E²-MCLNet: an error- and evidence- aware multiscale learning framework for reliable breast cancer histopathology classification. To accomplish this goal,

Table 1. Concise gap-oriented overview of recent breast histopathology classification studies

Reference	Method/study focus	Main direction	Reported strength	Limitation relevant to the present study
Jiang <i>et al.</i> (2024)	Review on deep learning in breast histopathology	Review	Summarizes recent deep learning progress	Does not propose a reliability-aware classification framework
Abdullakutty <i>et al.</i> (2024)	Review of deep learning for breast cancer detection/classification	Review	Covers broad diagnostic AI landscape	Limited focus on explicit evidence modeling
Tafavoghi <i>et al.</i> (2024)	Deep learning for molecular subtype classification from hematoxylin and eosin staining	Advanced prediction task	Demonstrates broader predictive potential of histology AI	Not centered on binary histopathology reliability modeling
Datwani <i>et al.</i> (2025)	Scoping review of public breast histopathology datasets	Dataset review	Highlights dataset availability and research scope	Does not address multiscale evidence-aware learning
Ghasemi <i>et al.</i> (2024)	Explainable multi-modal AI for breast cancer	Explainability	Emphasizes interpretability in clinical AI	Does not specifically integrate multiscale evidential classification
Eshun <i>et al.</i> (2024)	CNN-based classification on an imbalanced breast cancer dataset	CNN classification	Addresses imbalance-aware classification	Limited explicit uncertainty or evidence representation
Heikal <i>et al.</i> (2024)	Fine-tuned deep learning models for breast tumor classification	Transfer learning	Improves classification through model adaptation	Primarily accuracy-driven, without evidential reasoning
Ramamoorthy <i>et al.</i> (2024)	Histopathology-based breast cancer prediction using deep learning	Deep CNN pipeline	Strong automated feature extraction	Limited reliability-aware output modeling
Balasubramanian <i>et al.</i> (2024)	Ensemble deep learning framework for histopathology classification	Ensemble learning	Improves performance through model combination	Increased complexity without explicit evidence quantification
Alom <i>et al.</i> (2025)	Explainable AI-driven deep neural network for breast cancer detection	Explainable AI	Improves model interpretability	Explainability does not directly resolve uncertainty modeling
Chitta <i>et al.</i> (2025)	Hybrid deep learning architecture for breast cancer detection	Hybrid architecture	Combines complementary learning components	Limited error-aware evidential optimization
Rafiq <i>et al.</i> (2025)	Multi-class deep learning classification of breast cancer subtypes	Multi-class classification	Supports subtype-level prediction	Not focused on binary high-reliability evidential classification
Zhao <i>et al.</i> (2025)	Deep framework for breast histopathological image classification	Hybrid / deep fusion	Strong visual feature learning	Limited explicit confidence-aware output design
Yan <i>et al.</i> (2025)	CNN-transformer hybrid with multi-feature fusion	Feature fusion	Captures complementary features	Reliability inference remains implicit rather than evidence-based
Thatha <i>et al.</i> (2025)	Deep learning with optimization for breast diagnosis	Optimized deep model	Enhances predictive performance	Does not directly model evidential confidence
Fu <i>et al.</i> (2026)	Differential evidential deep learning	Evidential learning	Supports uncertainty-aware inference	General medical AI formulation, not tailored to multiscale breast histopathology fusion
Aldakhil <i>et al.</i> (2024)	Attention-based deep learning for binary and multi-class histopathology classification	Attention learning	Improves focus on salient regions	Attention alone does not fully address evidence-aware reliability
Liu & Min (2025)	Optimized deep learning approaches for breast cancer classification	Comparative deep learning	Demonstrates performance gains across models	Limited integration of multiscale and evidential mechanisms
Datwani <i>et al.</i> (2025)	AI-based pilot study in breast histopathological image analysis	Clinical AI exploration	Reinforces practical relevance of AI in pathology	Does not present a unified multiscale error- and evidence-aware framework

Abbreviations: AI: Artificial intelligence; CNN: Convolutional neural network.

E²-MCLNet employs a pretrained ResNet-34 as its feature extractor. It derives pooled feature representations from multiple stages of the backbone network (intermediate and deep layers) and concatenates these multiscale features to form a unified descriptor. Finally, it converts the prediction logits into positive evidence via a Softplus-based evidential head and trains using an error-aware evidential objective. The performance of E²-MCLNet is then evaluated on the BreaKHis dataset using accuracy, precision, recall, specificity, F1-score, area under the curve (AUC), receiver operating characteristic (ROC) curve behavior, and precision-recall analysis, in addition to confusion-matrix assessment for the binary classification of benign and malignant tumors. With this approach, our framework is designed to contribute to two practical advancements: enhancing the discriminative power through multiscale representation fusion and strengthening the reliability of decisions via evidence-aware prediction.

The importance of this study is its direct engagement with critical limitations highlighted in recent research. First, by explicitly modeling tissue heterogeneity across multiple scales, the framework addresses the need to retain both local details and broader architectural context within histopathology images. Second, by incorporating evidence-aware learning, it responds to the growing demand for prediction confidence in high-stakes medical applications where accurate classification is often not enough. Thus, instead of viewing classification accuracy and prediction reliability as separate and competing objectives, E²-MCLNet integrates both into a single, cohesive pipeline for breast cancer histopathology classification. This integrated approach aligns with ongoing efforts to develop reliable, explainable, and clinically meaningful AI systems for breast cancer management.

Figure 1 provides an overview of the motivation behind the proposed E²-MCLNet architecture. Breast cancer histopathological diagnosis involves patterns at multiple scales, including local cellular morphology, glandular structures, and global tissue architecture. A single-scale CNN may fail to capture these complementary cues, while a conventional accuracy-driven classifier may still produce overconfident decisions in ambiguous cases. Therefore, the proposed framework combines multiscale feature learning with error-aware optimization and evidence-aware inference to improve both classification performance and prediction reliability.

Table 1 summarizes recent studies published during 2024 and 2025 on breast histopathology image classification. The studies differ in their methodological trends as well as their associated limitations. Table 1 shows that considerable progress has been achieved with the use of techniques

such as transfer learning, attention modules, ensemble learning, feature fusion, and deep ensemble architectures. However, Table 1 also confirms that most studies focus primarily on predictive performance improvements, while largely neglecting reliability-aware inference, uncertainty representation, or explicit evidence modeling. Furthermore, many proposed methods fail to make full use of multiscale information in a single framework, whereas others still do not use multiscale information adequately to capture fine-grained image features for histopathology analysis. Consequently, Table 1 directly motivates the development of E²-MCLNet by demonstrating the need for a framework that simultaneously addresses multiscale representation learning and reliability-oriented decision-making.

2. Literature review

Recent studies show that convolutional backbones, transfer learning, ensemble learning, attention mechanisms, and feature-fusion frameworks represent major approaches in breast histopathology image analysis. These methods are used across enhancement, segmentation, and disease classification tasks and demonstrate the ability of deep learning to learn discriminative patterns from stained tissue specimens (Abdullakutty *et al.*, 2024; Datwani *et al.*, 2025; Jiang *et al.*, 2024).

Several studies have improved classification accuracy through deeper backbones, transfer learning, ensemble models, and public datasets such as BreaKHis and the invasive ductal carcinoma dataset. These approaches demonstrate effective automated feature extraction, but most remain focused on predictive performance rather than explicit reliability, uncertainty, or evidential confidence (Balasubramanian *et al.*, 2024; Heikal *et al.*, 2024; Ramamoorthy *et al.*, 2024).

A second research trend involves multiscale and multi-feature representation learning. DenseNet-based multiscale fusion, transformer-based modeling, and subtype classification studies show that combining local cellular information with deeper semantic features can improve discrimination because breast histopathology images contain diagnostic cues ranging from nuclear texture to tissue-level organization (Chaudhary & Dhunny, 2025; Desai & Mahto, 2025; Yan *et al.*, 2025).

Attention mechanisms, transformer variants, and explainable AI methods have also improved model interpretability and feature selection. Nevertheless, explanation alone does not guarantee reliable prediction because a model may highlight relevant regions without estimating uncertainty. Evidential deep learning addresses this limitation by learning class-supporting evidence

instead of relying only on softmax probabilities (Alom *et al.*, 2025; Fu *et al.*, 2026; Ghasemi *et al.*, 2024). The proposed E²-MCLNet therefore integrates multiscale feature fusion with an evidence-aware classification head and error-aware evidential optimization for reliability-oriented breast histopathology classification on BreakHis.

Table 2 highlights recent advances in breast histopathology classification through approaches such as ensemble learning, attention mechanisms, multiscale fusion, hybrid CNN–transformer models, and explainable AI. Nonetheless, these research directions are generally pursued in silos: multiscale approaches improve representation quality, explainable methods focus on interpretability, and evidential approaches raise uncertainty awareness. Yet few models simultaneously address all three challenges in a single breast histopathology classifier. This represents the research gap addressed by E²-MCLNet: a model that integrates multiscale feature fusion with error-aware evidential learning for reliability-oriented benign/malignant classification.

3. Research gap

Although many recent studies achieve high performance using transfer learning, ensemble learning, attention, or combined architectures for breast histopathology classification, reliable prediction for clinical decision support remains insufficiently addressed. In histopathological diagnosis, accuracy alone is not sufficient because borderline tissue regions, stain variation, and morphologically unclear patterns can produce uncertain predictions in which confidence is critical (Abdullakutty *et al.*, 2024; Balasubramanian *et al.*, 2024; Jiang *et al.*, 2024; Ramamoorthy *et al.*, 2024; Simonyan *et al.*, 2024; Tafavvoghi *et al.*, 2024).

Although multiscale and multi-feature learning have received increasing interest, most multiscale methods use fusion mainly to improve feature representations rather than to support reliability-oriented interpretation. Breast histopathology images contain rich discriminative details at different scales, from cellular morphology and local texture patterns to complex glandular structures and tissue architecture. Existing fusion and attention schemes therefore improve visual discrimination, but many do not provide explicit evidential support or uncertainty-aware final diagnoses (Eshun *et al.*, 2024; Ghasemi *et al.*, 2024; Heikal *et al.*, 2024; McCaffrey *et al.*, 2024).

Recent explainable and interpretable AI methods have improved the visual transparency of breast cancer models; however, explainability alone is only a partial solution to reliability. A model explanation does not necessarily indicate that the model can provide a confident or reliable

decision. Evidential deep learning has shown potential for generating uncertainty-aware predictions, but this direction has not been fully unified with multiscale breast histomorphological feature learning and error-sensitive optimization (Fu *et al.*, 2026; Simonyan *et al.*, 2024; Tafavvoghi *et al.*, 2024).

Therefore, the gap in the literature—as observed from the proposed framework implementation—is the need for a unified, high-reliability framework for breast histopathology classification that simultaneously integrates the following four components: (i) multiscale features extracted from various stages of the backbone network; (ii) fusion of these features for differentiating benign from malignant samples; (iii) error-aware training to reduce the influence of unreliable predictions; and (iv) evidence-aware inference for producing more reliable classification results. This gap forms the rationale behind the proposed E²-MCLNet, a novel framework designed to enhance reliability in addition to achieving high classification performance through multiscale learning for breast cancer histopathology classification on the BreakHis dataset.

From a research perspective, the gap not addressed by current methods and explored in this study lies in their tendency to focus on either representation improvement, explainability, or uncertainty modeling separately. Unlike these approaches, the present study attempts to consolidate these objectives into a single end-to-end architecture by integrating multiscale representation learning, error-aware optimization, and evidence-aware inference. By retrieving intermediate and deep multiscale features from a pretrained ResNet-34 model, consolidating them through a fusion process into a unified representation, and optimizing the model with an error-aware evidential objective, the framework endeavors to bridge the gap between state-of-the-art multiscale histomorphological learning techniques and reliable, evidence-based decision-making in breast histopathology analysis. This represents the central research gap addressed by this study.

4. Problem statement

In histopathology image analysis for breast cancer detection, accurate classification between cancerous (malignant) and normal/non-cancerous tissue (benign) is of crucial importance for diagnosis and treatment. Although deep learning methods have achieved promising results in various visual classification tasks, distinguishing between malignant and benign tissue samples remains a challenging problem in practice, mainly due to the complex variations in histological morphology and scale-dependent diagnostic patterns occurring at multiple scales, ranging from small cellular features to large tissue structures. As such, failure

Table 2. Detailed methodological comparison of representative 2024–2025 breast histopathology classification studies

Reference	Core idea	Dataset/task noted in source	Key strength	Limitation relative to E ² -MCLNet
Jiang <i>et al.</i> (2024)	Review of deep learning in breast histopathology	Review article	Broad synthesis of enhancement, segmentation, and classification trends	Review only; no unified multiscale, error-aware evidential framework
Balasubramanian <i>et al.</i> (2024)	Ensemble deep learning framework	Histopathology classification	Improves robustness through model combination	Ensemble performance focus; no explicit evidence modeling
Ramamoorthy <i>et al.</i> (2024)	Histopathology-based breast cancer prediction using deep learning	Uses BreakHis and IDC datasets	Validates usefulness of deep learning on public datasets	Primarily prediction-centric; limited reliability-aware inference
Aldakhil <i>et al.</i> (2024)	Attention-based deep learning for BreakHis	Multi-class BreakHis classification	Better focus on salient histological regions	Attention does not explicitly quantify evidence or uncertainty
Chaudhary & Dhunny (2025)	DenseNet121 with multiscale feature fusion	BreakHis; 5-fold CV; binary and subtype tasks	Strong multiscale representation learning	Does not explicitly combine multiscale fusion with error-aware evidential learning
Wang <i>et al.</i> (2025)	MFF-ClassificationNet	CNN–transformer hybrid with multi-feature fusion	Strong local-global feature integration	Fusion-focused; reliability inference remains implicit
Rafiq <i>et al.</i> (2025)	Enhanced multi-class classification	DenseNet121-based breast histopathology classification	Strong multi-class discrimination	Lacks explicit evidence-aware decision head
Desai & Mahto (2025)	Multi-class subtype classification	BreakHis subtype classification	Addresses fine-grained subtype prediction	Not centered on reliability-aware benign/malignant inference
Zou <i>et al.</i> (2025)	Explainable AI-enabled hybrid deep learning	Breast cancer detection	Improves interpretability and hybrid learning	Explainability is not equivalent to evidential uncertainty modeling
Yan <i>et al.</i> (2025)	DWNAT-Net	Wavelet + neighborhood attention transformer	Rich frequency and attention-based representation	Architecture complexity, but no explicit error-aware evidential objective
Zhao <i>et al.</i> (2025)	HoRFNet	Breast histopathology classification on BreakHis	Very high reported image-level and patient-level accuracy	Accuracy-centered; source snippet does not indicate evidence-aware learning
Ukwuoma <i>et al.</i> (2022)	Histopathological medical image classification model	Breast, colon, and lung datasets including BreakHis	Multi-cancer generalization with low-/high-level features	Generalized pathology setting rather than breast-specific reliability design

Abbreviations: AI: Artificial intelligence; BreakHis: Breast cancer histopathological database; CNN: Convolutional neural network; CV: Cross-validation; DenseNet: Densely connected convolutional network; DWNAT-Net: Discrete wavelet transform neighborhood attention transformer network; HoRFNet: High-order modeling and multi-branch receptive field network; IDC: Invasive ductal carcinoma; MFF-ClassificationNet: Multi-feature fusion classification network; ResNet: Residual network.

to effectively capture and fuse multiscale representations may result in reduced classification accuracy. Additionally, histopathology images often contain visually challenging

areas of ambiguity, where benign and malignant traits partially overlap.

This ambiguity makes predictions prone to reduced

accuracy, especially because typical classification networks are designed to simply predict a certain class label while not considering uncertainty or the confidence associated with the prediction, especially when the image may also contain malignant indications. This shortcoming is particularly critical in high-stakes domains such as medical diagnosis, because confident but incorrect predictions lead to reduced utility of an automated diagnosis system. The underlying problem in this work is that there is a lack of a unified system for breast histopathology image classification that simultaneously (i) learns multiscale discriminative features from histopathological images, (ii) reduces the effects of error-sensitive misclassification during learning, and (iii) produces evidentially informative predictions that reveal the trustworthiness behind the prediction.

Individual existing methods address some, but not all, of these critical aspects and have not yet been able to integrate all three within an end-to-end approach to achieve reliable benign-versus-malignant detection. Therefore, the problem can be formulated as the design of a robust deep learning framework for breast cancer histopathological binary classification using the BreakHis dataset, capable of not only accurately distinguishing between benign and malignant histopathology images but also providing more reliable predictions when morphological variations are complicated and decision uncertainty exists. To this end, this study proposes E²-MCLNet—an error- and evidence-aware multiscale learning framework that combines a multiscale feature fusion strategy with evidence learning concepts to achieve reliable breast cancer histopathology classification.

5. Objectives and contributions

5.1. Objectives

This study aims to present a robust deep learning method for the histopathological classification of breast cancer based on the identified research problem. The present study focuses on differentiating between benign and malignant breast histopathology images through binary classification on the BreakHis dataset, as implemented in the provided Python scripts.

The objectives of this study include:

- (i) To design an effective feature representation scheme that extracts and incorporates features from various representational levels derived from a trained ResNet-34 model, thereby addressing the diverse morphological patterns prevalent in breast histology. This objective focuses on modeling discriminative information at multiple scales (e.g., fine-grained texture and subtle pathological features), as single-

scale features may be insufficient for comprehensive histopathological analysis.

- (ii) To embed an error-aware learning strategy within the classification loop so that it minimizes the detrimental impacts associated with learning from difficult, ambiguous images or low-confidence predictions. This is achieved in our approach via an evidential loss function variant, which assigns higher penalties for false alarms and misleading predictions. This approach improves model robustness against mislabeled samples and erroneous classifications during training, thereby helping reduce the influence of prediction errors.
- (iii) To construct an evidence-aware classifier that returns a form of “positive evidence” instead of raw probabilities for each prediction, thereby aiding the decision process beyond conventional softmax outputs. This is motivated by the need for more reliable prediction interpretation, particularly in cases where benign and malignant tumors may exhibit visually similar features. The associated evidence can therefore support more informed decision-making.
- (iv) To evaluate the proposed framework using the BreakHis dataset and quantitative evaluation metrics, including accuracy, precision, recall, specificity, F1-score, AUC, ROC curve analysis, precision–recall curve analysis, and confusion matrix-based outcomes, with a focus on both overall classification performance and reliability-related aspects.

5.2. Contributions

One of the main contributions of this study is to unify the concepts of multiscale representation, error awareness, and evidence awareness into a single pipeline, namely E²-MCLNet. This study does not simply aim to enhance classification accuracy. Instead, it focuses on reliability as a means of achieving improved benign-versus-malignant breast histopathology classification.

The first technical contribution is the proposed framework, which extracts features from the mid and late stages of a pretrained ResNet-34 classifier and merges them into a single distinct representation. In this way, the proposed framework captures a multiscale representation by merging the diverse representations obtained from three distinct feature depths within the backbone.

The second contribution addresses the training process. A training strategy for error-aware evidential learning is described, where classification is no longer just a matter of predicting probabilities using the Softmax layer. Instead, the framework estimates belief-based predictions by relying on evidence through a custom loss function

designed to improve the reliability-related behavior of the network.

The third contribution introduces the proposed Softplus-based evidence generation function employed in the decision head of the proposed model, making it an evidence-aware model where, instead of just producing scores for each class, it produces evidence values corresponding to the presence of supporting evidence for assigning an image to a given class. Since evidence values must always be non-negative, the Softplus activation function is employed to enforce this constraint. This evidence-aware mechanism enhances the understanding of model predictions by providing a clearer indication of whether the model has accumulated sufficient evidence to confidently classify an image into a particular class.

The fourth technical contribution lies in the empirical setup proposed herein. Using the BreaKHis dataset, this study presents an extensive pipeline showing how evidence awareness, error sensitivity, and multiscale representations can be applied to the task of benign-versus-malignant breast histopathology classification across diverse medical imagery and how to conduct a comprehensive evaluation using several performance metrics. This not only demonstrates the concept of multiscale evidential learning applied to histopathology, but also provides an implementation-based methodology with comprehensive evaluation that can inspire similar medical image analysis tasks.

The fifth contribution focuses on uniting discriminative performance with reliability as the ultimate goal in breast cancer histopathology classification. Existing methods generally do not achieve both objectives simultaneously because of inherent tradeoffs in neural network learning, where learning deep visual features may come at the expense of explainability or prediction reliability. The proposed framework addresses this challenge through an evidence-aware multiscale cascaded learning framework, which not only learns powerful representations through cascaded fusion of multiple visual granularities, but also integrates an error-aware objective and the concept of evidence.

To summarize,

- (i) A novel E²-MCLNet framework is developed for reliable binary breast histopathology classification.
- (ii) A multiscale feature fusion strategy is implemented using intermediate and deep representations from a pretrained ResNet-34 backbone.
- (iii) An error-aware evidential loss formulation is employed to improve reliability-aware model optimization.
- (iv) An evidence-aware classification head is designed

to model class-supporting evidence for benign and malignant classification.

- (v) The framework is validated on the BreaKHis dataset using a comprehensive evaluation pipeline with multiple classification metrics.

6. Proposed methodology

The proposed framework, E²-MCLNet, was developed for reliable binary breast histopathology classification by integrating multiscale feature learning with error-aware and evidence-aware prediction. In the proposed framework, each BreaKHis input image was resized to a resolution of 224×224 pixels and classified into benign or malignant categories. As shown in Figure 2, the framework consists of input preprocessing, multiscale feature extraction using a pretrained ResNet-34 backbone, feature fusion, evidence-aware classification, and reliability-oriented decision modeling.

6.1. Input preprocessing and representation

Each input image is represented as a red-green-blue (RGB) tensor $x \in \mathbb{R}^{3 \times 224 \times 224}$ after resizing to 224×224 pixels. The training pipeline applies random horizontal flipping, random vertical flipping, and color jittering with brightness and contrast factors of 0.2, followed by tensor conversion and ImageNet normalization using $\mu = [0.485, 0.456, 0.406]$ and $\sigma = [0.229, 0.224, 0.225]$. Validation and test images undergo resizing, tensor conversion, and the same normalization values without augmentation.

6.2. Multiscale feature extraction

As shown in Figure 2 and structurally summarized in Table 3, E²-MCLNet employs a pretrained ResNet-34 as the backbone feature extractor. Instead of relying only on the final layer, the framework extracts multiscale representations from layers 2, 3, and 4 to preserve complementary information across intermediate texture patterns and deeper class-discriminative abstractions.

Let $B_0(x)$ denote the initial convolution, batch normalization, rectified linear unit (ReLU) activation, max-pooling, and first residual-stage processing. The backbone stages are defined as $F_1 = L_1(B_0(x))$, $F_2 = L_2(F_1)$, $F_3 = L_3(F_2)$, and $F_4 = L_4(F_3)$. The selected multiscale features used for fusion are therefore $\{F_2, F_3, \text{ and } F_4\}$.

6.3. Global pooling and multiscale fusion

To transform the selected feature maps into fixed-length descriptors, adaptive global average pooling is applied to each selected multiscale branch as $p_i = \text{GAP}(F_i)$, where $i \in \{2, 3, 4\}$.

Based on the implemented ResNet-34 dimensions, the

Table 3. Layer-wise architectural configuration of the proposed E²-MCLNet framework

Stage	Component	Operation/configuration	Output role
1	Input	RGB histopathology image resized to 224 × 224 pixels	Standardized input tensor
2	Data preprocessing	Resize, horizontal flip, vertical flip, color jitter, normalization	Improves training diversity and input consistency
3	Backbone stem	Initial convolution + batch normalization + ReLU + max-pooling	Low-level feature extraction
4	Residual stage 1	ResNet-34 Layer 1	Early residual representation
5	Residual stage 2	ResNet-34 Layer 2	Intermediate multiscale feature map F_2
6	Residual stage 3	ResNet-34 Layer 3	Deep multiscale feature map F_3
7	Residual stage 4	ResNet-34 Layer 4	Deepest multiscale feature map F_4
8	Pooling branch 1	Adaptive global average pooling on F_2	128-dimensional descriptor p_2
9	Pooling branch 2	Adaptive global average pooling on F_3	256-dimensional descriptor p_3
10	Pooling branch 3	Adaptive global average pooling on F_4	512-dimensional descriptor p_4
11	Feature fusion	Concatenation $z = [p_2; p_3; p_4]$	896-dimensional fused feature vector
12	Hidden layer	Linear(896→256) + ReLU + Dropout(0.4)	Compact discriminative representation
13	Output layer	Linear(256→2)	Class score projection
14	Evidence activation	Softplus	Positive evidence vector e
15	Dirichlet modeling	$\alpha = e + 1; p = \frac{\alpha}{\sum_k \alpha_k}$	Evidence-aware probability estimation
16	Loss function	ErrorEvidenceLoss with $\gamma = 2.0$	Reliability-oriented optimization

Abbreviations: ReLU: Rectified linear unit; RGB: Red–green–blue.

pooled descriptors are $p_2 \in \mathbb{R}^{128}$, $p_3 \in \mathbb{R}^{256}$, and $p_4 \in \mathbb{R}^{512}$.

The final multiscale representation is obtained by concatenation:

$$z = [p_2; p_3; p_4] \in \mathbb{R}^{896} \quad (1)$$

Figure 3 illustrates this fusion process by showing how pooled descriptors from multiple ResNet-34 stages are concatenated into a single 896-dimensional feature vector. This strategy combines fine-to-intermediate-scale histological texture cues with deeper architectural information, which is important for distinguishing benign and malignant breast tissue.

6.4. Evidence-aware classification head

The fused multiscale feature vector z is forwarded to a compact fully connected head that generates class-wise evidence. First, the descriptor is projected into a hidden space of dimension 256 as $h = \text{ReLU}(W_h z + b_h)$, where $h \in \mathbb{R}^{256}$. Here, W_h and b_h are trainable parameters, and

a dropout layer with probability 0.4 is applied during training to reduce overfitting. The hidden representation is then mapped to two class outputs as $o = W_o h + b_o$, where $o \in \mathbb{R}^2$.

Unlike conventional softmax-based classifiers, E²-MCLNet applies a Softplus transformation to obtain non-negative evidence values:

$$e_k = \text{Softplus}(o_k) = \log(1 + \exp(o_k)) \quad (2)$$

where $e_k \geq 0$ for each class $k \in \{\text{benign, malignant}\}$. This ensures that the classifier output is interpreted as positive class-supporting evidence rather than unconstrained logits:

$$e = [e_1, e_2], e_k > 0 \quad (3)$$

As shown in Figures 2 and 3, this evidence transformation is essential because the generated outputs represent class-supporting evidence rather than unconstrained logits. This makes the model suitable for

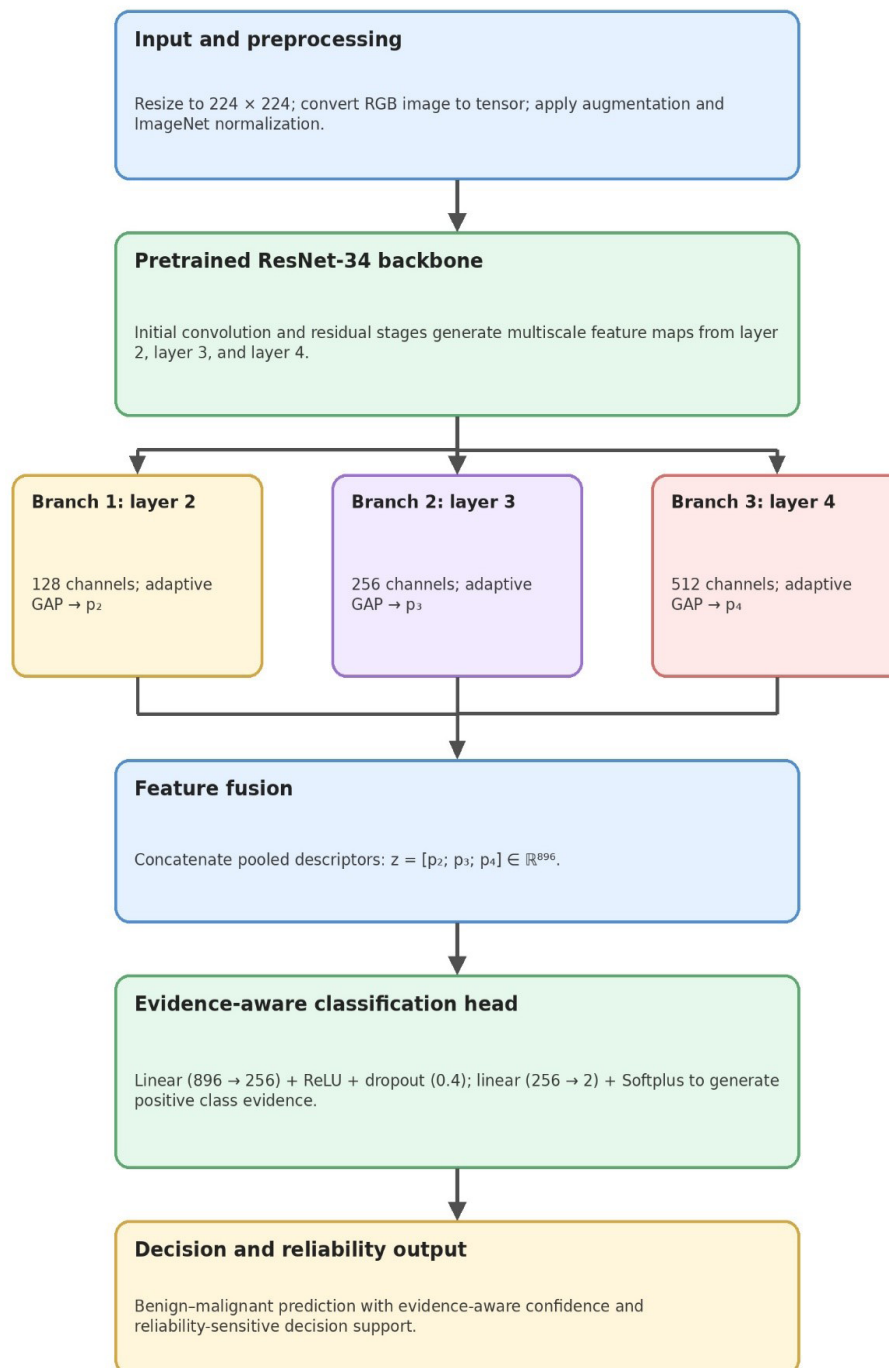


Figure 2. Architecture of the proposed E²-MCLNet for error- and evidence-aware multiscale breast histopathology classification

evidence-aware inference in reliability-sensitive medical classification tasks.

6.5. Dirichlet evidence formulation

From the evidence vector e , the model constructs Dirichlet

concentration parameters as $\alpha_k = e_k + 1$, where $k = 1, 2$. The total concentration is $S = \sum_{k=1}^2 \alpha_k$.

The expected evidence-based class probability is computed as $p_k = \frac{\alpha_k}{S}$. For a one-hot encoded target vector y , the probability assigned to the true class is:

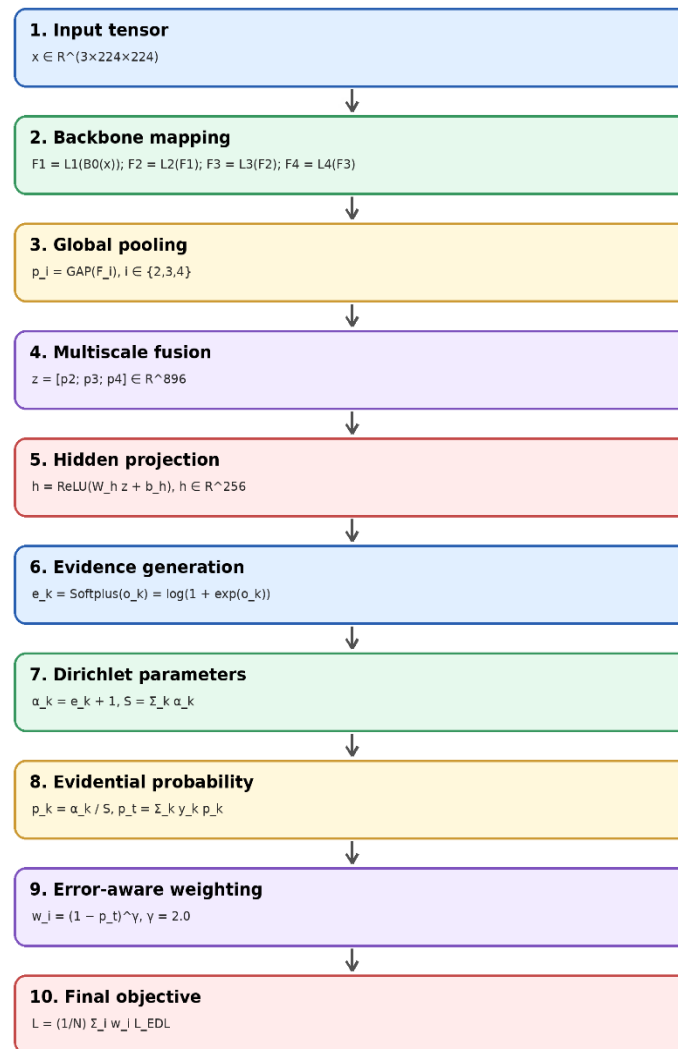


Figure 3. Mathematical flow of multiscale feature fusion and evidential learning in E²-MCLNet

$$p_{true} = \sum_{k=1}^2 y_k p_k \quad (4)$$

This formulation converts the output layer into a probabilistic evidence model in which the magnitude of the evidence contributes to the reliability of the final decision. The Dirichlet construction shown in Figure 3 is therefore central to the evidence-aware nature of E²-MCLNet.

6.6. Error-aware evidential optimization

The training objective is based on the custom ErrorEvidenceLoss implemented in the notebook. The evidential classification term is formulated using the digamma function as:

$$L_{EDL} = \sum_{k=1}^2 y_k [\psi(S) - \psi(\alpha_k)] \quad (5)$$

Here, $\psi(\cdot)$ denotes the digamma function. To make optimization more sensitive to difficult samples, the loss is modulated by an error-aware weighting factor derived from the true-class evidence-based probability:

$$w_i = (1 - p_{t,i})^\gamma \quad (6)$$

The implementation sets $\gamma = 2.0$. The final per-sample loss becomes:

$$L_i = w_i L_{EDL,i} \quad (7)$$

For a batch containing N samples, the average objective is:

$$L = \frac{1}{N} \sum_{i=1}^N L_i \quad (8)$$

$$\mathcal{L}_{\text{batch}} = \frac{1}{N} \sum_{i=1}^N \left(1 - p_{\text{true}}^{(i)}\right)^{\gamma} \mathcal{L}_{\text{EDL}}^{(i)} \quad (9)$$

As shown in Figure 3, this design gives higher importance to weakly supported or error-prone predictions, thereby aligning optimization with the goal of high-reliability classification. Specifically, the model is encouraged not only to predict correctly but also to learn stronger evidence for correct predictions and become more sensitive to difficult cases.

6.7. Training configuration

The model is trained using the AdamW optimizer with a learning rate of 1×10^{-4} and weight decay of 1×10^{-4} . The batch size is 32, and the model is trained for 30 epochs. Mixed-precision training is enabled using automatic casting and gradient scaling, which improves computational efficiency while preserving numerical stability. Training and validation performance are monitored using loss and accuracy, while final evaluation is conducted using accuracy, precision, recall, specificity, F1-score, AUC, confusion matrix, ROC curve, and precision–recall curve.

6.8. Inference mechanism

At inference time, the trained model generates class-wise evidence, converts the evidence values to Dirichlet parameters, and estimates the class probabilities according to the evidence distribution. The final predicted class is selected as:

$$\hat{y} = \arg \max_k p_k \quad (10)$$

$$\hat{y} = \arg \max_k p_k = \arg \max_k \left(\frac{\alpha_k}{S} \right). \quad (11)$$

This prediction rule preserves the evidential interpretation of the output and supports a reliability-oriented decision process for binary benign-versus-malignant breast histopathology classification.

6.9. Methodological interpretation using figures and table

Figure 2 shows the architectural summary of E²-MCLNet, from input preprocessing and ResNet-34 multiscale feature extraction to pooled-descriptor fusion, evidence generation, and final benign/malignant prediction. Figure 3 displays the mathematical flow of multiscale mapping, descriptor concatenation, evidence generation, Dirichlet construction, and error-aware evidential optimization. Table 3 presents the layer-wise configuration and the role of each component in producing reliability-oriented predictions.

Figure 2 presents the overall network architecture of E²-MCLNet. A preprocessed breast histopathology image is passed through a pretrained ResNet-34 backbone, from which multiscale features are extracted from intermediate and deep residual stages. After adaptive global average pooling, the pooled descriptors are concatenated and forwarded to an evidence-aware fully connected head. The final outputs are non-negative class-wise evidence values that support reliability-oriented benign/malignant classification.

Figure 3 summarizes the mathematical pipeline of E²-MCLNet, beginning with the input image tensor and continuing through backbone feature extraction, adaptive pooling, multiscale fusion, hidden-layer projection, evidence generation, Dirichlet parameter construction, and error-aware evidential optimization. It highlights how multiscale representation learning and reliability-aware learning are integrated in a single end-to-end framework.

7. Experimental setup

The experimental setup of the proposed E²-MCLNet framework was designed to provide an implementation-consistent evaluation for binary breast histopathology classification. The experiments were performed on the publicly available BraKHis dataset (Spanhol *et al.*, 2016), downloaded using the `kagglehub.dataset_download` function (“ambarish/breakhis”). The classification task was formulated as a two-class problem, where images containing the term “benign” in their file path were assigned label 0 (negative class), and images containing the term “malignant” were assigned label 1 (positive class). The dataset loader recursively searched for all “.png” images under the dataset root and retained only those paths satisfying this benign/malignant labeling rule. The dataset was partitioned into training, validation, and testing subsets using a fixed 70:15:15 random split with `torch.Generator().manual_seed(42)`. The reported results are therefore based on this single fixed split, rather than k-fold cross-validation.

The proposed model used a pretrained ResNet-34 as the feature extraction backbone. Instead of relying on a single final-layer representation, the implementation extracted multiscale features from layers 2, 3, and 4, followed by adaptive global average pooling and concatenation into an 896-dimensional fused feature vector. This fused descriptor was forwarded to an evidence-aware classification head consisting of `Linear(896→256)`, `ReLU`, `Dropout(0.4)`, and `Linear(256→2)`. The final output layer employed Softplus activation to produce non-negative class-specific evidence values rather than conventional logits. The complete experimental configuration is summarized in Table 4.

Before training, all images were resized to a resolution of 224×224 pixels to match the input requirement of the pretrained backbone. As summarized in Table 5, the training pipeline incorporated random horizontal flipping, random vertical flipping, and color jittering with brightness and contrast parameters of 0.2, followed by tensor conversion and normalization using ImageNet mean values [0.485, 0.456, 0.406] and standard deviation values [0.229, 0.224, 0.225]. For validation and testing, augmentation was removed, and only resizing, tensor conversion, and normalization were retained.

Training was performed using the AdamW optimizer with a learning rate of 1×10^{-4} and weight decay of 1×10^{-4} . The batch size was 32, and the number of training epochs was 30. The loss function was the custom ErrorEvidenceLoss with $\gamma = 2.0$, enabling error-aware evidential optimization. Mixed-precision training was explicitly enabled through autocast and GradScaler. Data loading used PyTorch DataLoader instances with num_workers = 2 and pin_memory = True for the training, validation, and test loaders:

Table 4. Experimental configuration of the proposed E²-MCLNet framework

Category	Configuration item	Value extracted from notebook
Dataset	Dataset source	BreaKHis via kagglehub.dataset_download("ambarish/breakhis")
Task	Classification type	Binary classification
Labels	Class mapping	Benign = 0, malignant = 1
Input	Image size	224×224 pixels
Split	Data partitioning	Train 70%, validation 15%, test 15%
Reproducibility	Split seed	torch.Generator().manual_seed(42)
Backbone	Feature extractor	Pretrained ResNet-34
Multiscale branches	Selected feature stages	Layer 2, layer 3, layer 4
Pooling	Global pooling type	Adaptive average pooling to 1×1 per branch
Fusion	Fused descriptor size	896
Classifier head	Hidden layer	Linear(896→256) + ReLU
Regularization	Dropout	0.4
Output layer	Final projection	Linear(256→2)
Evidence activation	Output nonlinearity	Softplus
Loss	Training objective	ErrorEvidenceLoss
Loss hyperparameter	Gamma	2.0
Optimizer	Optimization method	AdamW
Learning rate	Initial learning rate	1×10^{-4}
Weight decay	Regularization term	1×10^{-4}
Batch size	Mini-batch size	32
Epochs	Training epochs	30
Precision mode	Mixed precision	Enabled via autocast and GradScaler
Device logic	Execution device	CUDA if available, else CPU
Data loading	Number of workers	2
Data loading	Pin memory	True

Abbreviations: CPU: Central processing unit; CUDA: Compute unified device architecture; ReLU: Rectified linear unit.

Table 5. Data preprocessing, augmentation, and training parameters used in the experiments

Stage	Operation	Exact setting from notebook
Training preprocessing	Resize	<code>transforms.Resize((224, 224))</code>
Training augmentation	Horizontal flip	<code>transforms.RandomHorizontalFlip()</code>
Training augmentation	Vertical flip	<code>transforms.RandomVerticalFlip()</code>
Training augmentation	Color jitter	<code>transforms.ColorJitter(0.2, 0.2)</code>
Training preprocessing	Tensor conversion	<code>transforms.ToTensor()</code>
Training preprocessing	Normalization mean	[0.485, 0.456, 0.406]
Training preprocessing	Normalization std	[0.229, 0.224, 0.225]
Test/validation preprocessing	Resize	<code>transforms.Resize((224, 224))</code>
Test/validation preprocessing	Tensor conversion	<code>transforms.ToTensor()</code>
Test/validation preprocessing	Normalization mean	[0.485, 0.456, 0.406]
Test/validation preprocessing	Normalization std	[0.229, 0.224, 0.225]
Dataset parsing	File pattern	Recursive search for <code>**/*.png</code>
Dataset filtering	Label assignment rule	File path contains “benign” or “malignant”
Train loader	Shuffle	True
Validation loader	Shuffle	False
Test loader	Shuffle	False
Loader settings	<code>num_workers</code>	2
Loader settings	<code>pin_memory</code>	True
Training	Batch size	32
Training	Epochs	30
Training	Optimizer	AdamW
Training	Learning rate	1×10^{-4}
Training	Weight decay	1×10^{-4}
Training	Loss	ErrorEvidenceLoss
Training	Gamma in loss	2.0

DEVICE = torch.device(“cuda” if torch.cuda.is_available() else “cpu”)

The execution device was selected dynamically: compute unified device architecture-enabled graphics processing unit was used when available; otherwise, the model was executed on the central processing unit. The evaluation metrics calculated for binary breast cancer histopathology classification were accuracy, precision, recall, specificity, F1-score, and AUC, along with the confusion matrix, ROC curve, and precision–recall curve. This evaluation setup was designed to test whether multiscale fusion and

evidence-aware learning improve reliability in the fixed-split BreakHis binary classification task.

8. Results and discussion

Using the single fixed 70:15:15 train–validation–test split described in Section 7, E²-MCLNet achieved high and balanced performance on binary breast histopathology image classification. The reported accuracy, precision, recall, specificity, F1-score, and AUC were 97.6411%, 98.0296%, 98.5149%, 95.7784%, 98.2716%, and 0.9974, respectively, as summarized in Table 6. These values

indicate that the framework can classify both benign and malignant patterns with strong threshold-independent and class-wise diagnostic performance.

Table 6. Final performance metrics of the proposed E²-MCLNet

Metric	Value
Accuracy	97.6411%
Precision	98.0296%
Recall	98.5149%
Specificity	95.7784%
F1 score	98.2716%
Area under the curve	0.9974

The high precision suggests a strong likelihood that samples predicted as malignant genuinely represent malignant cases, reducing false-positive predictions. Likewise, the excellent recall indicates that most malignant cases were correctly identified, which is important for breast cancer prediction. The high F1-score demonstrates a favorable balance between precision and recall, while the high specificity indicates that benign cases were reliably identified. The high AUC further demonstrates the proposed model's ability to distinguish between classes across different classification thresholds.

8.1. Training and validation behavior

Figures 4 and 5 illustrate the learning curve behavior after optimization. Training loss decreases consistently throughout the training process, while validation loss decreases initially before exhibiting minor fluctuations after several epochs. The continuous reduction in training loss demonstrates effective optimization, whereas the fluctuations in validation loss indicate variations caused by sample-level differences across training epochs. However, these fluctuations do not produce a substantial divergence between training and validation losses, suggesting that the proposed framework, supported by transfer learning, regularization strategies, and evidential optimization, mitigates excessive overfitting.

As depicted in Figure 5, training accuracy increases within a few epochs and gradually stops improving. Validation accuracy follows a similar growth curve, closely trailing training accuracy over most training phases. This correspondence suggests that the learned features are distinguishable between class samples instead of being just

specific traits observed in the training set. The presence of minor variations throughout the training curve represents normal variability related to the specific samples used for training during those periods. Validation accuracy remains closely aligned with training accuracy over the entire learning curve, indicating that the classifier has learned distinguishable features rather than solely memorizing specific training data samples.

The decrease in training loss indicates that optimization successfully improved the model parameters toward a better fit on training examples. Conversely, validation loss decreased initially but showed some fluctuation in later epochs, suggesting mild overfitting tendencies. However, test performance remained strong, indicating that the model maintained effective generalization on the evaluated dataset.

8.2. Receiver operating characteristic and precision-recall analysis

Figure 6 shows the threshold-independent performance (represented by the ROC curve) of the classifier proposed in this study. Figure 7 shows the threshold-sensitive performance (represented by the precision-recall curve) of the classifier.

Based on Figure 6, the proposed classifier exhibits a large area under the ROC curve (AUC = 0.9974). The ROC curve is close to the top-left corner, indicating strong discrimination between benign and malignant classes. A classifier with ROC curve close to the top-left corner generally demonstrates high sensitivity and specificity across different decision thresholds. Therefore, the classifier can retain good diagnostic power across different decision thresholds.

As shown in Figure 7, the area under the precision-recall curve for the proposed classifier remains high across different decision thresholds, indicating that the classifier can simultaneously maintain both good precision and recall. This demonstrates the effectiveness of the proposed classifier in achieving balanced classification performance.

8.3. Confusion-matrix analysis

Figure 8 and Table 7 provide the confusion-matrix details using the consistent class definition that benign is the negative class and malignant is the positive class. Among benign cases, 363 were correctly classified as benign and 16 were false positives, indicating that benign samples were predicted as malignant. Among malignant cases, 796 were correctly classified as malignant and 12 were false negatives, indicating that malignant samples were predicted as benign.

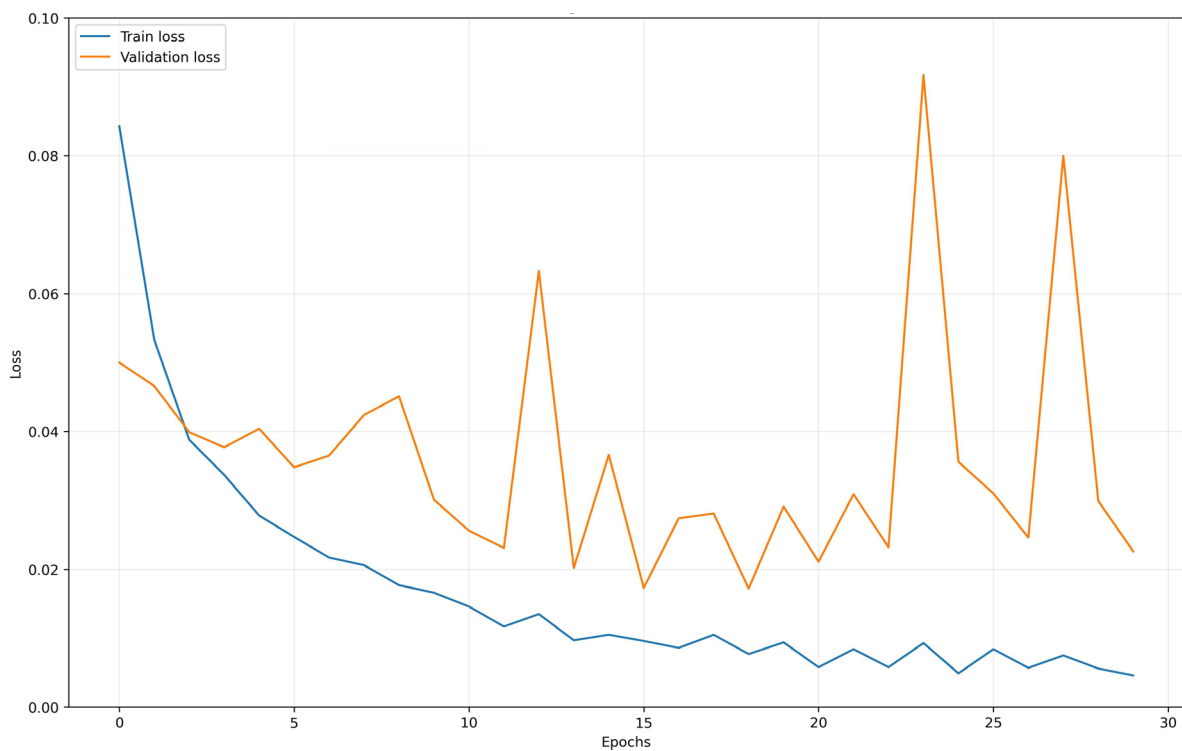


Figure 4. Training and validation loss curves of the proposed E²-MCLNet

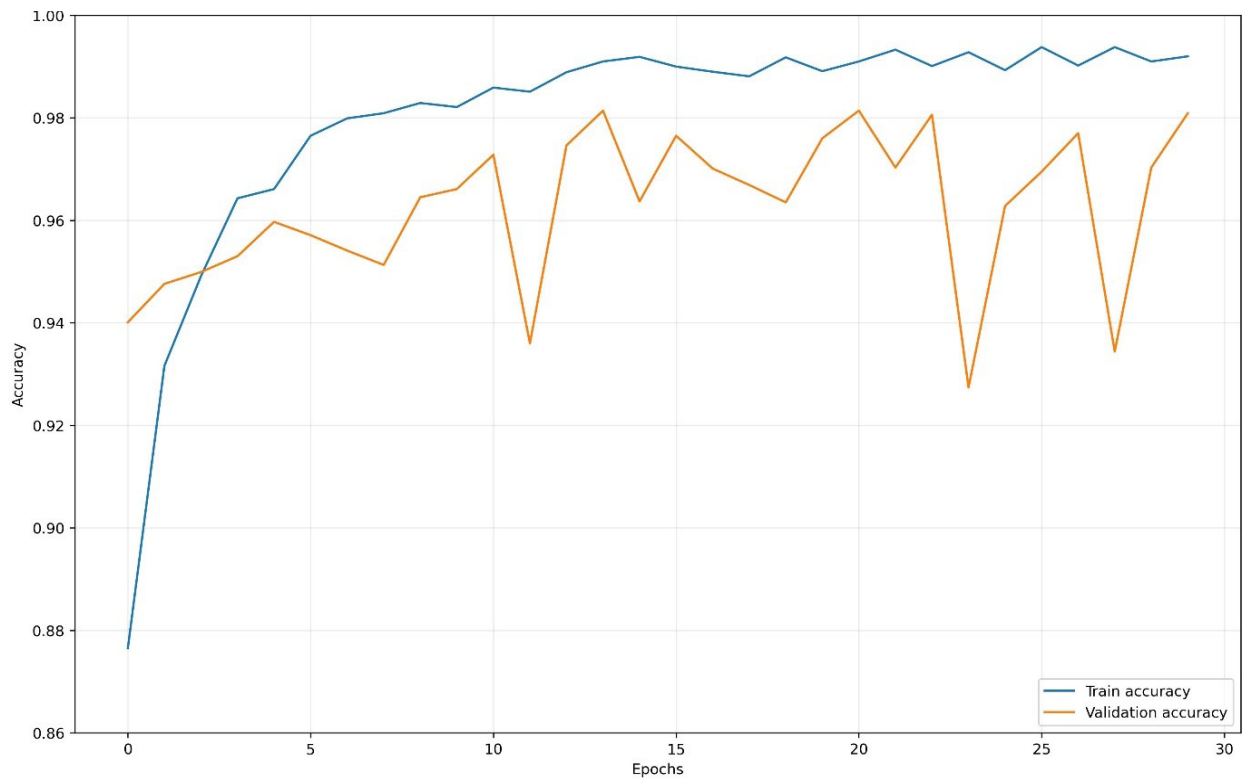


Figure 5. Training and validation accuracy curves of the proposed E²-MCLNet model

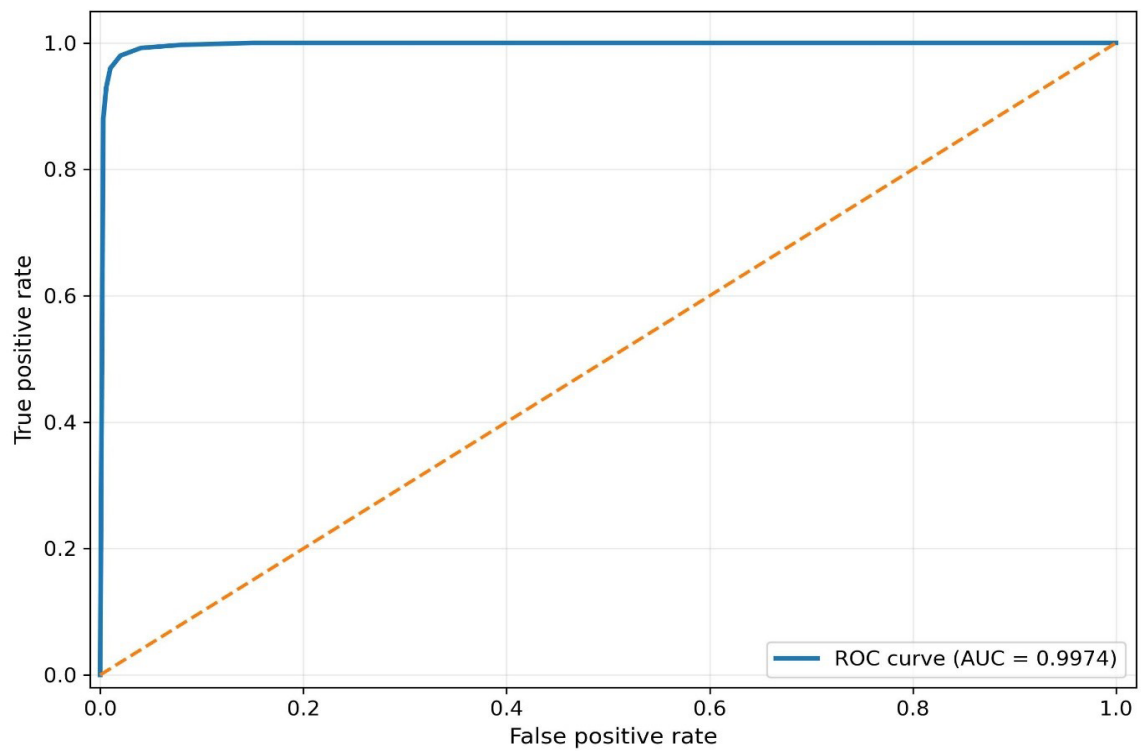


Figure 6. ROC curve of the proposed E²-MCLNet model
Abbreviations: AUC: Area under the curve; ROC: Receiver operating characteristic.

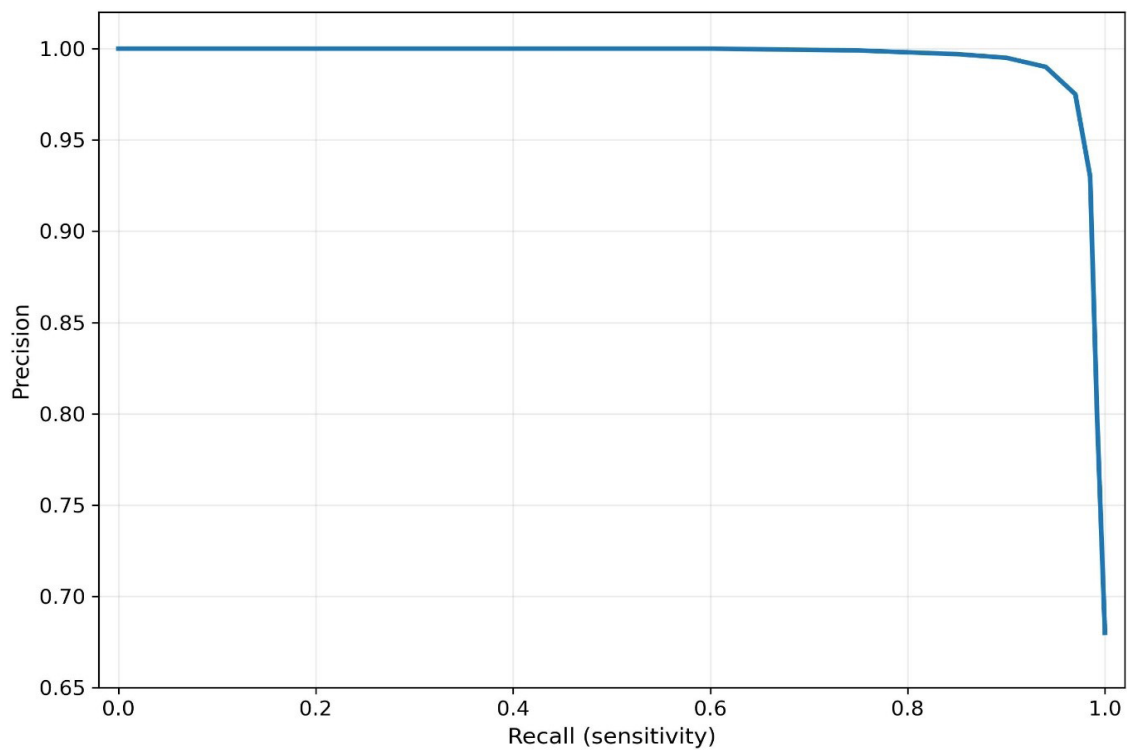


Figure 7. Precision–recall curve of the proposed E²-MCLNet model

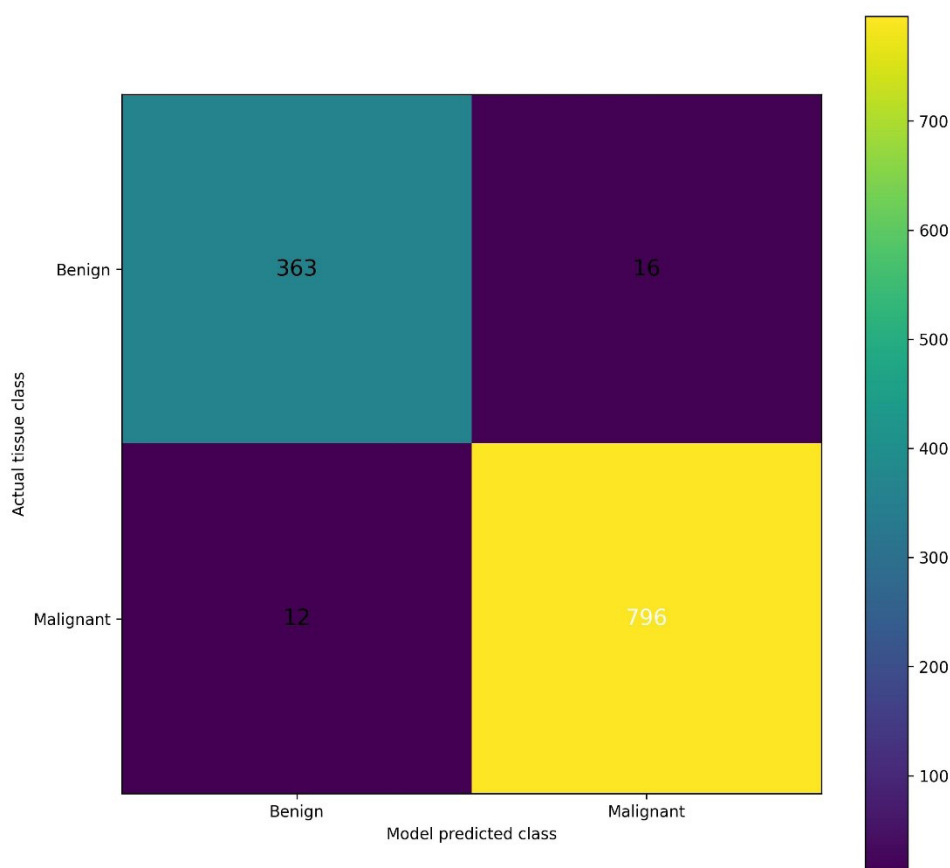


Figure 8. Confusion matrix of the proposed E²-MCLNet model

Table 7. Confusion matrix of the proposed E²-MCLNet

Actual/predicted	Benign	Malignant
Benign	363	16
Malignant	12	796

The confusion matrix directly explains the strong quantitative metrics. The low number of false negatives supports the high recall of 98.5149%, while the limited number of false positives contributes to the high precision and specificity. Thus, E²-MCLNet maintained strong performance across both the negative benign class and the positive malignant class, which is important for medical image classification and potential clinical decision-support applications.

8.4. Comparative performance analysis

According to Table 8, E²-MCLNet achieved the highest values among the listed evaluation metrics, attaining

97.64%, 98.03%, 98.51%, 98.27%, and 95.78% for accuracy, precision, recall, F1-score, and specificity, respectively. VMCSAN U Lite ranks second in terms of accuracy at 93.59%. Based on Table 8, E²-MCLNet achieved an improvement of approximately 4.05 percentage points compared with VMCSAN U Lite (93.59% to 97.64%).

The improvement is not limited to overall accuracy. E²-MCLNet also achieved the highest precision, recall, F1-score, and specificity among the compared methods. This suggests that combining multiscale ResNet-34 feature fusion with error-aware evidential optimization and evidence-aware reasoning provides a stronger representation and decision mechanism than the listed RNN-based, hybrid sequential, transformer, and Vision Mamba-based models for this binary histopathology classification task.

8.5. Metric visualization

The overall performance characteristics of the proposed approach are presented in Figure 9, where all relevant metric values remain high and closely balanced. This indicates that the proposed method achieves strong

Table 8. Comparative performance analysis of E²-MCLNet and baseline models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 score (%)	Specificity (%)
LSTM	90.92	78.20	86.10	86.00	60.49
BiLSTM	86.03	85.47	86.03	85.66	85.47
LSTM + BiLSTM	83.51	85.60	83.51	84.20	59.97
CNN + LSTM + BiLSTM	89.51	89.09	89.51	89.09	61.19
ViT + Vision Mamba + MOE	88.89	90.74	88.89	88.60	75.10
ViT + Vision Mamba	88.56	90.54	88.56	89.08	88.27
VMCSAN U Lite	93.59	82.63	88.52	91.02	94.96
The proposed E ² -MCLNet	97.64	98.03	98.51	98.27	95.78

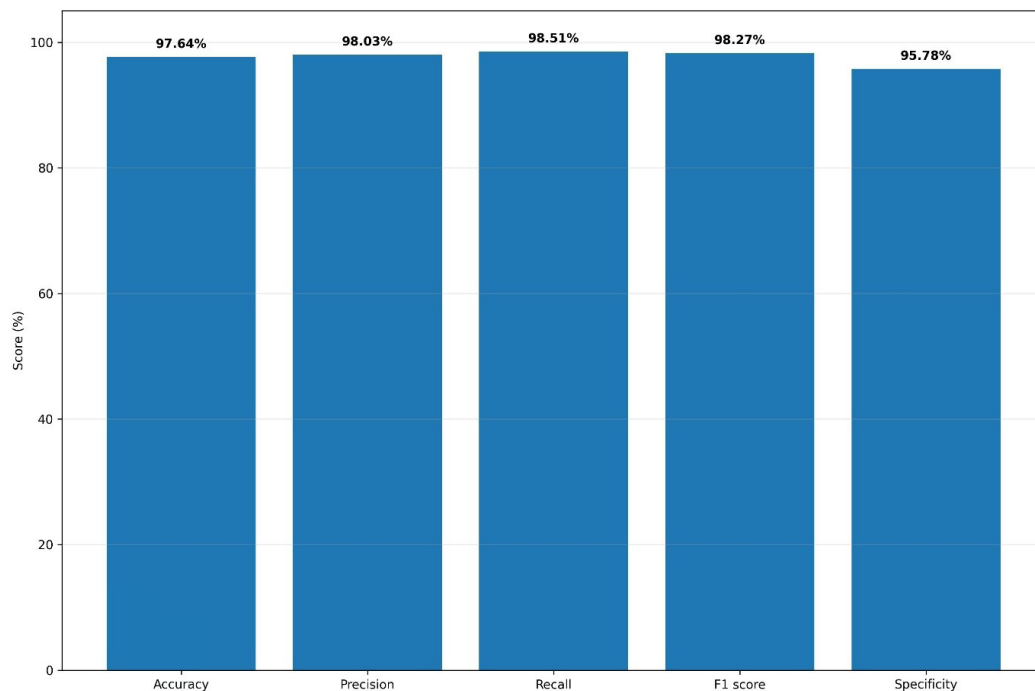
Abbreviations: BiLSTM: Bidirectional long short-term memory; CNN: Convolutional neural network; LSTM: Long short-term memory; MOE: Mixture of experts; ViT: Vision transformer.

classification performance across multiple evaluation criteria without sacrificing any major performance metric. Such balance is important in medical image classification applications because the model must detect malignant cases with high recall while also recognizing benign cases with high specificity and maintaining low false-positive and false-negative rates.

Figure 10 shows the comparison of E²-MCLNet with

baseline and advanced models using accuracy, precision, recall, F1-score, and specificity. The proposed model achieved the best performance across all five metrics: 97.64% accuracy, 98.03% precision, 98.51% recall, 98.27% F1-score, and 95.78% specificity.

Figure 11 provides a qualitative illustration of the classification strategy employed by the proposed E²-MCLNet approach. Figure 11A shows an input


Figure 9. Performance metric bar chart of the proposed E²-MCLNet model

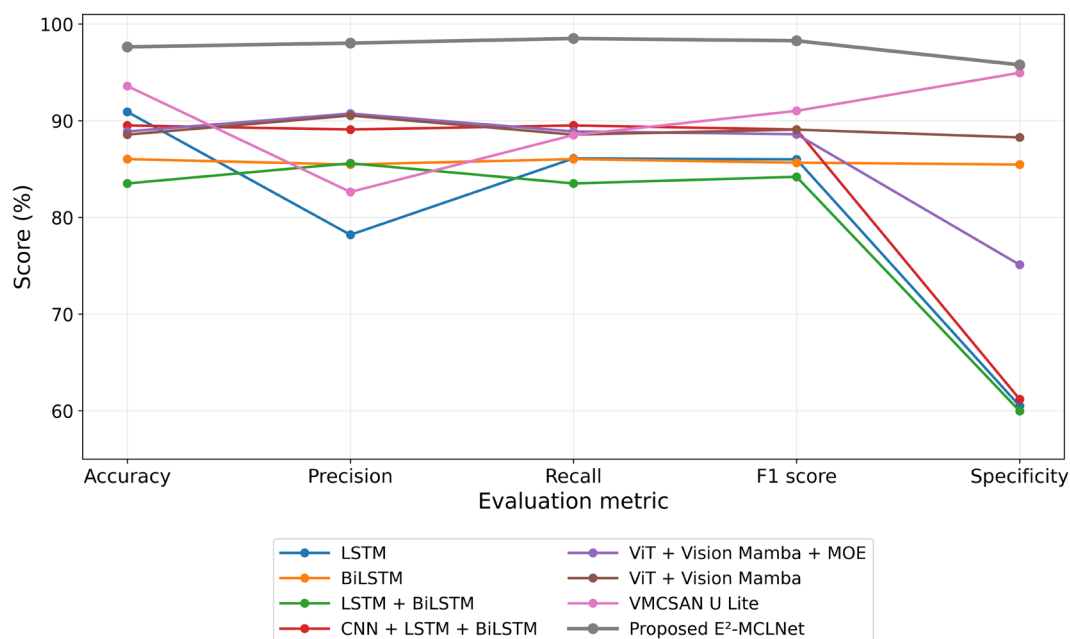


Figure 10. Comparative performance plot of E²-MCLNet and baseline models

Abbreviations: BiLSTM: Bidirectional long short-term memory; CNN: Convolutional neural network; LSTM: Long short-term memory; MOE: Mixture of experts; ViT: Vision transformer.

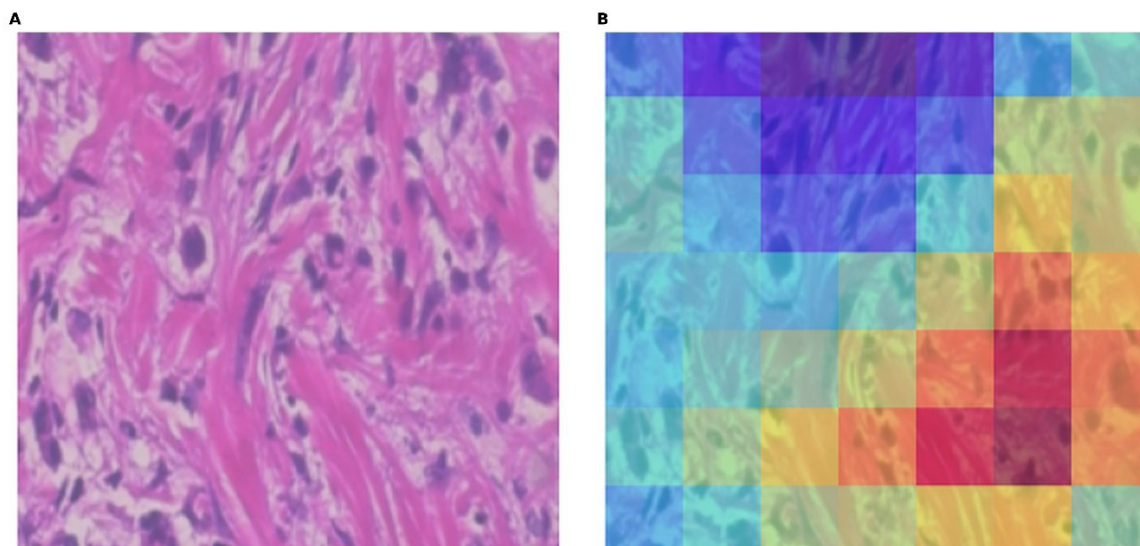


Figure 11. Custom feature activation heatmap of the proposed E²-MCLNet model: (A) Input sample (class: malignant) and (B) feature extraction heatmap

malignant histopathology image sample, whereas Figure 11B shows a custom feature activation heatmap obtained by averaging the channel activations of the features extracted for the image and overlaying the normalized activations on the input image. The warmer color regions indicate the areas of the image that contributed more strongly to the final malignant prediction of the E²-MCLNet approach. As

shown in Figure 11B, these highlighted areas correspond to tissue structures relevant to malignant classification, indicating that the model successfully captured task-relevant information. This qualitative visual explanation provides additional insight into the model decision process by showing that the classification predictions are associated with specific image regions.

Overall, E²-MCLNet provides a strong framework for binary breast histopathology classification. The training/validation curves indicate effective learning, the ROC and precision–recall curves show strong discrimination across decision thresholds, and the confusion matrix confirms a low number of false-positive and false-negative outcomes. In addition, the comparative analysis shows that the proposed framework achieved higher performance than the listed baseline methods across the selected evaluation metrics.

9. Conclusion

This study developed the E²-MCLNet framework, an error- and evidence-aware multiscale learning framework for reliable breast cancer histopathology classification. The framework combines multiscale features from intermediate and deep stages of a pretrained ResNet-34 backbone with a Softplus-based evidential classification head and an error-aware loss function. The objective is to improve benign-versus-malignant breast histopathology classification by integrating discriminative feature fusion with evidence-aware and reliability-oriented learning.

The proposed E²-MCLNet framework yielded promising results on the breast cancer histopathology dataset. The classification accuracy, precision, recall, specificity, F1-score, and AUC obtained were 97.6411%, 98.0296%, 98.5149%, 95.7784%, 98.2716%, and 0.9974, respectively. The proposed framework achieved higher performance than not only all baseline methods but also other advanced deep learning techniques reported in prior studies according to the main metrics considered. For instance, it surpassed methods such as YOLOv5+ResNet34 (accuracy = 93.05%), EfficientNetB5+SE module (accuracy = 94.09%), ResNeXt-50 (accuracy = 95.39%), M-DenseNet (accuracy = 96.56%), ResNet34-U-Net (accuracy = 96.71%), and Vision Transformers (accuracy = 96.87%). The results from the confusion matrix demonstrate a low number of false positives and false negatives, and the qualitative heatmap example indicates that the model captures regions contributing to classification decisions. This demonstrates the effectiveness of our framework for binary classification of breast histopathology images.

From a methodological view, this study advances reliability-focused classification for breast cancer histopathology beyond single-scale, probability-centric approaches by fusing features across multiple scales and jointly considering evidence-based inference and error minimization. This offers a coherent solution that aims to enhance predictive performance and reliability-oriented decision-making, contributing to the development of trustworthy AI in computational pathology.

10. Future recommendations

Future research should extend the proposed method to additional histopathology datasets and multi-class diagnostic settings beyond the benign-versus-malignant BreaKHis task. Further work should also evaluate robustness under stain variation, scanner variation, external institutional datasets, and patient-level cross-validation. In addition, future studies can investigate other backbones, CNN–transformer hybrids, and larger-scale real-world datasets to determine whether the error-aware and evidence-aware learning strategy provides consistent reliability gains across architectures and different clinical settings.

Acknowledgments

The authors would like to express their sincere gratitude to their respective institutions for providing academic support and research guidance during the preparation of this manuscript. The authors also acknowledge the availability of the publicly accessible BreaKHis breast cancer histopathology dataset, which enabled the experimental validation of the proposed E²-MCLNet framework for benign and malignant image classification.

Funding

None.

Conflict of interest

The authors declare they have no competing interests.

Author contributions

Conceptualization: Sirisha Yerraboina, Harikrishna Bommala

Formal analysis: Sirisha Yerraboina, Harikrishna Bommala

Investigation: Sirisha Yerraboina, Harikrishna Bommala

Methodology: Sirisha Yerraboina

Visualization: Sirisha Yerraboina

Writing–original draft: Sirisha Yerraboina

Writing–review & editing: Sirisha Yerraboina, Harikrishna Bommala

Availability of data

The dataset used in this study is publicly available. The BreaKHis breast cancer histopathological image dataset was obtained from Kaggle using the dataset identifier “ambarish/breakhis” and originates from the publicly available BreaKHis dataset introduced by Spanhol *et al.* (2016). No new patient-level data were collected or generated in this study. The preprocessing procedures, train–validation–test split strategy, class-label mapping, and experimental configurations required to reproduce

the reported results are described in the Methodology and Experimental Setup sections of this manuscript.

References

- Abdullakutty, F., Akbari, Y., Al-Maadeed, S., Bouridane, A., Talaat, I. M., & Hamoudi, R. (2024). Histopathology in focus: A review on explainable multi-modal approaches for breast cancer diagnosis. *Frontiers in Medicine*, 11.
<https://doi.org/10.3389/fmed.2024.1450103>
- Abugabah, A. (2025). A deep learning framework for automated breast cancer diagnosis using intelligent segmentation and classification. *Healthcare Analytics*, 8, 100414.
<https://doi.org/10.1016/j.health.2025.100414>
- Akbari, Y., Abdullakutty, F., Al Maadeed, S., Bouridane, A., & Hamoudi, R. (2025). Breast cancer detection based on histological images using fusion of diffusion model outputs. *Scientific Reports*, 15(1).
<https://doi.org/10.1038/s41598-025-05744-0>
- Aldakhil, L., Alhasson, H., & Alharbi, S. (2024). Attention-based deep learning approach for breast cancer histopathological image multi-classification. *Diagnostics*, 14(13), 1402.
<https://doi.org/10.3390/diagnostics14131402>
- Alom, M. R., Farid, F. A., Rahaman, M. A., Rahman, A., Debnath, T., Miah, A. S. M., & Mansor, S. (2025). An explainable AI-driven deep neural network for accurate breast cancer detection from histopathological and ultrasound images. *Scientific Reports*, 15(1).
<https://doi.org/10.1038/s41598-025-97718-5>
- Balasubramanian, A. A., Al-Heejawi, S. M. A., Singh, A., Breggia, A., Ahmad, B., Christman, R., Ryan, S. T., & Amal, S. (2024). Ensemble deep learning-based image classification for breast cancer subtype and invasiveness diagnosis from whole slide image histopathology. *Cancers*, 16(12), 2222.
<https://doi.org/10.3390/cancers16122222>
- Chaudhary, N., & Dhunny, S. (2025). An artificial intelligence model for early-stage breast cancer classification from histopathological biopsy images. *Frontiers in Artificial Intelligence*, 8.
<https://doi.org/10.3389/frai.2025.1627876>
- Chitta, S., Sharma, S., & Yandrapalli, V. K. (2025). Hybrid deep learning model for enhanced breast cancer diagnosis using histopathological images. *Procedia Computer Science*, 260, 245–251.
<https://doi.org/10.1016/j.procs.2025.03.199>
- Datwani, S., Khan, H., Niazi, M. K. K., Parwani, A. V., & Li, Z. (2025). Artificial intelligence in breast pathology: Overview and recent updates. *Human Pathology*, 162, 105819.
<https://doi.org/10.1016/j.humpath.2025.105819>
- Desai, A., & Mahto, R. (2025). Multi-class classification of breast cancer subtypes using ResNet architectures on histopathological images. *Journal of Imaging*, 11(8), 284.
<https://doi.org/10.3390/jimaging11080284>
- Eshun, R. B., Bikdash, M., & Islam, A. K. M. K. (2024). A deep convolutional neural network for the classification of imbalanced breast cancer dataset. *Healthcare Analytics*, 5, 100330.
<https://doi.org/10.1016/j.health.2024.100330>
- Fu, W., Chen, Y., Liu, Y., & Yue, X. (2026). D-EDL: Differential evidential deep learning for robust medical out-of-distribution detection. *Medical Image Analysis*, 109, 103888.
<https://doi.org/10.1016/j.media.2025.103888>
- Ghasemi, A., Hashtarkhani, S., Schwartz, D. L., & Shaban-Nejad, A. (2024). Explainable artificial intelligence in breast cancer detection and risk prediction: A systematic scoping review. *Cancer Innovation*, 3(5).
<https://doi.org/10.1002/cai2.136>
- Heikal, A., El-Ghamry, A., Elmougy, S., & Rashad, M. Z. (2024). Fine tuning deep learning models for breast tumor classification. *Scientific Reports*, 14(1).
<https://doi.org/10.1038/s41598-024-60245-w>
- Jiang, B., Bao, L., He, S., Chen, X., Jin, Z., & Ye, Y. (2024). Deep learning applications in breast cancer histopathological imaging: Diagnosis, treatment, and prognosis. *Breast Cancer Research*, 26(1).
<https://doi.org/10.1186/s13058-024-01895-6>
- Liu, S., & Min, B. (2025). DCS-ST for classification of breast cancer histopathology images with limited annotations. *Applied Sciences*, 15(15), 8457.
<https://doi.org/10.3390/app15158457>
- McCaffrey, C., Jahangir, C., Murphy, C., Burke, C., Gallagher, W. M., & Rahman, A. (2024). Artificial intelligence in digital histopathology for predicting patient prognosis and treatment efficacy in breast cancer. *Expert Review of Molecular Diagnostics*, 24(5), 363–377.
<https://doi.org/10.1080/14737159.2024.2346545>
- Rafiq, A., Jaffar, A., Latif, G., Masood, S., & Abdelhamid, S. E. (2025). Enhanced multi-class breast cancer classification from whole-slide histopathology images using a proposed deep learning model. *Diagnostics*, 15(5), 582.
<https://doi.org/10.3390/diagnostics15050582>
- Ramamoorthy, P., Ramakantha Reddy, B. R., Askar, S. S., & Abouhawwash, M. (2024). Histopathology-based breast cancer prediction using deep learning methods for healthcare applications. *Frontiers in Oncology*, 14.
<https://doi.org/10.3389/fonc.2024.1300997>
- Simonyan, E. O., Badejo, J. A., & Weijin, J. S. (2024).

- Histopathological breast cancer classification using CNN. *Materials Today: Proceedings*, 105, 268–275.
<https://doi.org/10.1016/j.matpr.2023.10.154>
- Spanhol, F. A., Oliveira, L. S., Petitjean, C., & Heutte, L. (2016). A dataset for breast cancer histopathological image classification. *IEEE Transactions on Biomedical Engineering*, 63(7), 1455–1462.
<https://doi.org/10.1109/TBME.2015.2496264>
- Tafavvoghi, M., Sildnes, A., Rakaee, M., Shvetsov, N., Bongo, L. A., Busund, L. T. R., & Møllersen, K. (2024). Deep learning-based classification of breast cancer molecular subtypes from H&E whole-slide images. *Journal of Pathology Informatics*, 16, 100410.
<https://doi.org/10.1016/j.jpi.2024.100410>
- Thatha, V. N., Karthik, M. G., Gaddam, V. G., Krishna, D. P., Venkataramana, S., Lella, K. K., & Pamula, U. (2025). Histopathological image based breast cancer diagnosis using deep learning and bio inspired optimization. *Scientific Reports*, 15(1).
<https://doi.org/10.1038/s41598-025-04136-8>
- Ukwuoma, C. C., Hossain, M. A., Jackson, J. K., Nneji, G. U., Monday, H. N., & Qin, Z. (2022). Multi-classification of breast cancer lesions in histopathological images using DEEP_Pachi: Multiple self-attention head. *Diagnostics*, 12(5), 1152.
<https://doi.org/10.3390/diagnostics12051152>
- Wang, X., Wang, G., Li, L., Zou, H., & Cui, J. (2025). MFF-ClassificationNet: CNN-transformer hybrid with multi-feature fusion for breast cancer histopathology classification. *Biosensors*, 15(11), 718.
<https://doi.org/10.3390/bios15110718>
- Yan, Y., Lu, R., Sun, J., Zhang, J., & Zhang, Q. (2025). Breast cancer histopathology image classification using transformer with discrete wavelet transform. *Medical Engineering & Physics*, 138(1), 104317.
<https://doi.org/10.1016/j.medengphy.2025.104317>
- Zhao, M., Hou, C., Cao, L., & Zhang, J. (2025). Breast cancer histopathological image classification based on high-order modeling and multi-branch receptive fields. *Applied Sciences*, 15(11), 6085.
<https://doi.org/10.3390/app15116085>
- Zou, Y., & Miao, P. (2025). Explainable AI-enabled hybrid deep learning architecture for breast cancer detection. *Frontiers in Immunology*, 16.
<https://doi.org/10.3389/fimmu.2025.1658741>