Journal of Information Systems Engineering and Management
2025, 10(59s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Enhancing Breast Cancer Detection with HNet: A Hybrid Deep Learning Framework

Khaldi Brahim[1], Debakla Mohammed[1] and Djemal Khalifa[2]
[1]*Laboratory of Computer Science and Intelligent Systems (LISYS), Computer Sciences Department, Exact Sciences Faculty, University of Mustapha Stambouli, Mascara, 29000, Algeria*
[2]*Institut Universitaire de Technologie d'Evry, University of Evry Paris-Saclay UEVES, Evry, 91020, France*

ARTICLE INFO

Received: 24 Dec 2024

Revised: 12 Feb 2025

Accepted: 26 Feb 2025

ABSTRACT

Breast cancer remains one of the leading causes of cancer-related mortality among women globally, making early and accurate diagnosis critical for effective treatment. While traditional convolutional neural networks (CNNs) are proficient in extracting local texture features, they often struggle to capture global contextual dependencies and spatial hierarchies inherent in histopathological images. To overcome these limitations, we propose HNet, a novel hybrid deep learning architecture designed to leverage the complementary strengths of multiple techniques. HNet combines EfficientNet for scalable and efficient local feature extraction, Advanced Vision Transformers (AVT) for global context modeling, and Capsule Networks for relational reasoning and spatial hierarchy preservation. This fusion of architectures aims to enhance diagnostic performance and improve interpretability. Evaluated on the BreakHis dataset across multiple image resolutions and data split configurations, HNet demonstrated an accuracy up to 97.52%, showcasing enhanced classification accuracy and generalization. Ablation studies further validated the contribution of each module, highlighting the potential of hybrid deep learning frameworks in enabling robust, real-world breast cancer diagnosis.

Keywords: Hybrid Deep Learning, Breast Cancer Classification, Feature extraction, CNN, Capsule Networks, Transformer Attention Mechanisms.

INTRODUCTION

Cancer is a group of diseases in which cells continue to divide and spread into nearby tissues, forming a lump called a tumor or malignancy [1]. According to the World Health Organization (WHO), breast cancer accounts for nearly 25% of all cancer cases in women worldwide and is the second leading cause of death due to malignant tumors. Early and accurate diagnosis is critical for improving survival rates and ensuring effective treatment. However, this remains a challenging task due to the heterogeneous nature of breast tissue and the subtle morphological differences between benign and malignant tumors.

Traditional diagnostic approaches, such as mammography, ultrasound, and histopathological examination, remain the standard in clinical practice. Nevertheless, these methods are not without limitations. They can be invasive, time-consuming, and often subject to inter-observer variability. In particular, histopathological analysis, while considered the gold standard for definitive diagnosis, requires expert interpretation of complex visual patterns making it susceptible to diagnostic errors, especially under high workload conditions [2].

In recent years, computer-aided diagnosis (CAD) systems have emerged as promising tools to support radiologists and pathologists in the interpretation of medical images [3]. These systems aim to enhance diagnostic efficiency, reduce human error, and provide consistent decision-making, especially in resource-constrained or high-volume clinical settings. The integration of artificial intelligence (AI), particularly in the field of medical imaging, has enabled the automation of complex tasks such as lesion detection, segmentation, and classification. This is particularly relevant in histopathological image analysis, where visual inspection is intricate and prone to variability among

Journal of Information Systems Engineering and Management
2025, 10(59s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

experts. As a result, AI-driven systems have gained considerable attention for their potential to improve the objectivity and accuracy of cancer diagnosis.

Early efforts in automated breast cancer diagnosis relied on classical machine learning (ML) techniques such as support vector machines (SVMs), decision trees, random forests (RF), and k-nearest neighbors (KNN). These models typically require handcrafted features such as texture descriptors, shape metrics, or statistical moments, extracted through domain-specific knowledge. While these approaches have demonstrated reasonable performance in certain scenarios, they often struggle to generalize due to the limited expressiveness of manually designed features. Moreover, handcrafted descriptors may fail to capture the high intra-class variability and subtle morphological differences that are characteristic of histopathological images [4], [5].

The advent of deep learning, particularly convolutional neural networks (CNNs), has significantly advanced the state-of-the-art in medical image analysis[6], [7]. CNNs are capable of learning hierarchical feature representations directly from raw pixel data, thereby eliminating the need for manual feature engineering. In the context of breast cancer histopathology, CNNs have been shown to outperform traditional ML methods in classification tasks by effectively capturing local texture patterns and spatial structures [8], [9]. However, despite their success, CNNs are inherently limited in their ability to model long-range dependencies and global contextual information due to their localized receptive fields. Furthermore, they often fail to preserve spatial hierarchies and pose relationships, which are essential for accurately interpreting tissue organization and diagnosing malignancies.

To address these limitations, recent research has explored alternative deep learning paradigms that go beyond conventional CNN architectures. Vision Transformers (ViTs) have introduced self-attention mechanisms capable of modeling global contextual dependencies across spatially distant regions in an image [10]. Meanwhile, Capsule Networks (CapsNets) aim to preserve spatial relationships and encode part-whole hierarchies using vector-based representations and dynamic routing [11]. While both architectures offer distinct advantages, namely contextual awareness and spatial interpretability, they also face practical challenges when used independently, such as high data requirements for transformers and computational overhead in capsule routing. These observations highlight the need for a unified framework that can jointly leverage local detail, global context, and spatial structure to improve the reliability and interpretability of histopathological image classification.

In this context, we propose HNet, a hybrid deep learning architecture that combines the complementary strengths of three powerful components: EfficientNet for scalable and efficient local feature extraction, Advanced Vision Transformers (AVT) for modeling global contextual dependencies, and Capsule Networks for preserving spatial hierarchies and encoding part-whole relationships. By integrating these modules in a staged and unified framework, HNet addresses the individual shortcomings of each architecture and enhances both diagnostic accuracy and interpretability, which are two key requirements for clinical deployment. To evaluate the effectiveness and robustness of HNet, we conduct extensive experiments on the BreakHis dataset, exploring multiple image resolutions and training/validation/test split strategies. The experimental results demonstrate the effectiveness of HNet in classifying breast cancer histopathological images, with improved performance and generalization compared to standalone models—achieving an F1-score of up to 97.18%.

The remainder of this paper is structured as follows: Section 2 reviews recent work in breast cancer image classification using deep learning; Section 3 describes the proposed methodology and hybrid architecture in detail; Section 4 presents experimental results and discussion; and Section 5 concludes the paper and outlines directions for future research.

RELATED WORKS

Recent advances in breast cancer diagnosis using histopathological images have leveraged deep learning models to address challenges in accuracy and limited annotated data.

More recent work has focused on using deep learning algorithms, such as CNNs, to automate the feature extraction process and improve the accuracy of the classification. Others focused on combining multiple models for processing, or multiple types of medical images, such as histopathological and mammograms to improve classification accuracy.

Journal of Information Systems Engineering and Management
2025, 10(59s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

This has been achieved by using multi-modal CNNs that combine the features extracted from different types of images, or by using ensemble methods that combine the predictions of multiple models.

Raha et al. [12] employ ensemble learning techniques to predict breast cancer from the Wisconsin Breast Cancer Dataset (WBCD). They experimented with multiple machine learning models including Random Forest, XGBoost, SVM, MLP, and Gradient Boosting, with Random Forest yielding the highest performance, achieving an accuracy of 99.46%, precision of 100%, recall of 98.21%, and F1-score of 99.09%. The study further integrated explainable AI methods, SHAP and LIME, to provide global and local interpretability for the Random Forest model. This approach not only delivered high classification accuracy but also enhanced the model's transparency, making it suitable for clinical use in breast cancer prediction.

Singh et al [13] proposed a framework that combines deep learning models and machine learning classifiers for features extraction from histopathological images, aiming for early and cost-effective diagnosis. The study combined deep neural networks with machine learning classifiers. Specifically, a comprehensive super hybrid model combining DenseNet + Logistic Regression an F-score of 0.81, outperforming VGG + Logistic Regression (0.73), VGG + Random Forest (0.74) and DenseNet + Random Forest (0.79), and VGG, DenseNet, and Logistic Regression where, outperforming VGG + Logistic Regression (0.73), and DenseNet + Random Forest (0.79), validated across histopathological datasets.

Aldakhil et al. [14] integrated attention-based deep learning with traditional machine learning by employing ECSAnet (Efficient Channel Spatial Attention Network). Evaluated on the BreakHis dataset across multiple magnifications, ECSAnet combined with classifiers like Decision Trees and Logistic Regression improved classification accuracy.

Priyadarshni et al. [15] proposed a hybrid framework combining deep learning and traditional machine learning for breast cancer detection using the CBIS-DDSM dataset. Features were extracted using VGG-16 and ResNet-101 via transfer learning, then fused and refined through a deep instinctive stacked autoencoder for dimensionality reduction. The resulting 64-dimensional feature vector was classified using enhanced models—EDT, ERF, and ELR—integrating normalized neural weights from DNNs. The proposed EDT model achieved the highest performance with 99.02% accuracy, 99% F1-score, and 98% AUC, followed by ERF (95.75% accuracy, 97% F1-score, 100% AUC), and ELR (86.82% accuracy, 93% F1-score, 78% AUC). The framework outperformed state-of-the-art methods in classification tasks while maintaining computational efficiency and robustness, making it suitable for clinical deployment

Ben Atitallah et al. [9] proposed a CNN-based model that achieved consistently high classification accuracy across varying magnification levels of the BreakHis dataset, recording accuracies of 97.50% at 40×, 97.61% at 100×, 99.06% at 200×, and 97.25% at 400×. The highest performance was attained at 200× magnification, with a precision of 98.43%, recall of 100%, and an F1-score of 99.21%. The model maintained stable results regardless of image resolution, confirming its robustness to magnification variation. Unlike prior approaches such as DenseNet121-AnoGAN and DRDA-Net—which suffered from performance degradation at certain scales or required more complex architectures—the proposed CNN delivered superior accuracy while preserving architectural simplicity and computational efficiency, making it a practical candidate for clinical deployment.

Rafiq et al. [16] proposed a DenseNet121-based architecture for multi-class classification of breast cancer subtypes, achieving state-of-the-art binary classification accuracy of 98.50% and multi-class accuracy of 92.50% on the BreakHis dataset, underscoring the power of deep residual networks in complex cancer subtype differentiation.

Wang et al. [17] addressed data scarcity and feature monotony by introducing an auto-encoder reconstructed semi-supervised domain adaptation model. Their method leveraged feature reconstruction and domain adaptation to achieve superior classification metrics (accuracy 95.24%, F1-score 93.40%) across BreakHis and SNL datasets, illustrating the potential of semi-supervised learning and domain adaptation in histopathology.

Çetin-Kaya [18] trained 20 state-of-the-art models on BreakHis with fine-tuning and proposed MultiHisNet, which achieved 94.69% multi-class accuracy. An ensemble model combining transfer learning and custom architectures reached 96.71% accuracy, emphasizing the efficacy of ensemble learning and advanced attention modules for robust multi-class breast cancer classification.

Journal of Information Systems Engineering and Management
2025, 10(59s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Balasubramanian et al. [19] employed ensemble deep learning techniques across BACH and BreakHis datasets, integrating VGG16 and ResNet variants with novel patching strategies. Their models achieved 98.43% classification accuracy on BreakHis, illustrating the effectiveness of ensemble methods and high-resolution image analysis in improving diagnostic precision.

Cao et al. [20] introduced MbsCANet, an advanced multi-branch spectral channel attention network that functions within the frequency domain. By combining discrete cosine transform features, their model outperformed spatial-domain CNNs with image-level accuracy of 99.01% on BreakHis, highlighting the benefit of frequency domain analysis and attention mechanisms in histopathological image classification.

Kashif and al [21] employed a CNN for classifying mammograms from the DDSM dataset into normal, benign, and malignant categories, achieving an accuracy of 94% through diligent data augmentation and preprocessing techniques to enhance image quality.

Preeti katiyar [22] introduced a deep learning model utilizing transfer learning with pretrained CNN architectures like ResNet50 and VGG-16 to classify mammograms from the MIAS dataset, enhancing breast cancer detection and classification efficiency and accuracy, the model under consideration attained an accuracy rate of 96.00%, an area under the curve (AUC) measurement of 0.95, a sensitivity rate of 94.40%, a specificity rate of 95.65%, a precision rate of 96.01%, and an F score of 96.99%..

Meher et al. [23] proposed a deep learning approach for software bug classification to improve accuracy and automation. The study employs four attention-based deep learning models, achieving a mean F1-Score of 84.78% and a macro-average ROC of 98.25%, significantly outperforming existing methods by 16.88% in F1-Score. This work demonstrates the effectiveness of attention mechanisms in handling large-scale bug classification tasks with high precision.

Feng and Wang [24] developed a deep learning system integrating MobileNetV2, attention mechanisms, and feature pyramid networks for breast cancer classification on X-ray images. The model classifies images into normal, benign, and malignant categories using extensive preprocessing techniques—random cropping, horizontal flipping, Gaussian noise addition, and color variation—to enhance feature extraction. Trained and validated on a balanced Breast Cancer Dataset, the system achieved rapid convergence, 98.56% accuracy, and a minimal loss of 0.075, demonstrating its effectiveness for precise breast cancer diagnosis.

METHODOLOGY

The proposed HNet framework is designed to improve the classification of histopathological breast cancer images by combining the strengths of multiple deep learning paradigms. As depicted in Figure 1, the architecture consists of three main stages. First, input images undergo preprocessing and data augmentation to enhance robustness. Next, the image is simultaneously processed by two parallel branches: EfficientNet, which efficiently captures local and fine-grained visual details, and the Advanced Vision Transformer, which extracts global contextual information through self-attention mechanisms. The feature maps generated by these two branches are then combined and forwarded to a Capsule Network module, which is responsible for preserving spatial hierarchies and modeling part-whons le relationships. Finally, the output capsules are flattened and passed through fully connected layers for final classification. This hybrid and staged design allow HNet to leverage the complementary strengths of local texture, global context, and spatial structure within a unified framework.

Journal of Information Systems Engineering and Management
2025, 10(59s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

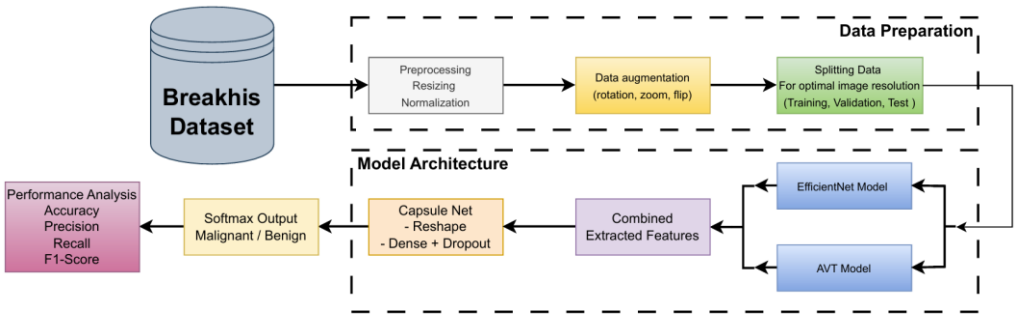

Figure 1: Proposed HNet flowchart

1 Data preparation

The proposed model is designed to perform binary classification (benign vs. malignant) of breast cancer histopathology images. We utilized the BreakHis dataset [25], which contains 7,916 annotated images, 2,487 benign and 5,429 malignant. To mitigate overfitting and improve generalization given the limited dataset size, we applied real-time data augmentation using TensorFlow's ImageDataGenerator. The augmentation techniques included random rotations (90°), zooming (up to 1.2×), horizontal and vertical flips, and pixel intensity shifts. These transformations simulate natural variations in histopathological imaging and enhance the model's ability to learn robust features across classes, thereby addressing potential class imbalance.

To evaluate the model's adaptability to different levels of visual detail, we resized the images to three different resolutions: 96×96, 128×128, and 256×256. All image pixel values were normalized to the range [0, 1] prior to training.

In addition, we explored several data split configurations to assess the model's performance under various training conditions. The dataset was divided into training, validation, and test sets using the following ratios: 70/20/10, 70/15/15, 75/15/10, 80/10/10, 85/10/5, and 90/5/5. This analysis allowed us to study the model's robustness with respect to data availability and balance across subsets.

2 EfficientNet Backbone

The initial segment of the framework utilizes EfficientNet, which has been pretrained on the ImageNet dataset, serving as the foundation for transfer learning. We systematically truncated this architecture subsequent to its convolutional layers, integrating global average pooling followed by a dense layer with 256 units activated by the ReLU function. This model demonstrates superior performance in feature extraction owing to its highly optimized architecture that mirrors the processing capabilities of the human visual system.

Within the preliminary layers, it identifies fundamental features such as edges, gradients, and textures, which are vital for the construction of higher-level representations pertaining to tissue architectures. The subsequent layers adeptly capture intricate patterns, encompassing nuclei forms, cellular boundaries, and subcellular configurations, which are imperative for distinguishing benign from malignant tissues in histopathological imagery. As shown in Figure 2, the structure of EfficientNet highlights its layers and the flow of information within the model.

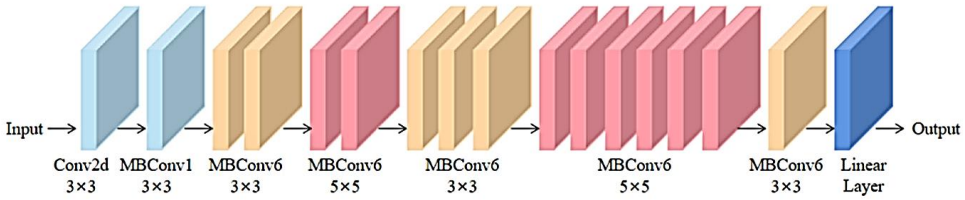

Figure 2: EfficientNet model flowchart

3 Advanced Vision Transformer (AVT)

Journal of Information Systems Engineering and Management
2025, 10(59s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

To capture long-range dependencies and global contextual information in histopathological images, we incorporate an AVT branch in parallel with the EfficientNet module. Vision Transformers have demonstrated strong performance in image classification tasks by modeling relationships between distant regions using self-attention mechanisms, which are particularly valuable in medical imaging where tissue structures exhibit complex spatial arrangements.

The proposed AVT architecture, shown in Figure 3., combines three synergistic components to achieve optimal local-global feature integration for medical image analysis. The framework begins with a modified Inception-Residual hybrid block that employs parallel convolutional pathways (1×1, 3×3, and 5×5 kernels) with residual skip connections, enabling simultaneous multi-scale feature extraction while preserving gradient flow through deep networks. This is followed by a patch-based multi-head attention mechanism that decomposes the feature maps into overlapping patches, where each attention head independently computes depth-wise convolutional projections for queries, keys, and values before applying scaled dot-product attention. This design captures long-range spatial dependencies while maintaining computational efficiency through patch-wise processing rather than global attention.

The attention heads' outputs are dynamically weighted and fused through a learnable 1×1 convolution, producing an attention-refined feature representation. The final output of the AVT branch is a feature vector that complements the local texture descriptors extracted by EfficientNet. This vector is then concatenated with the EfficientNet output and forwarded to the Capsule Network module for further spatial reasoning and classification.

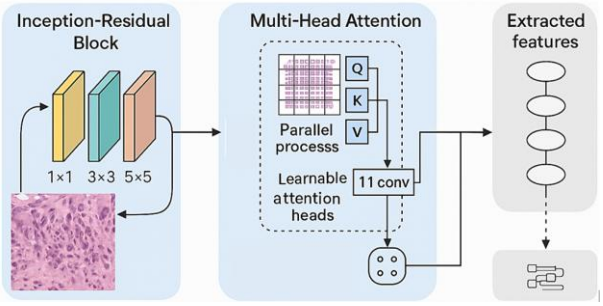

Figure 3: Advanced Vision Transformer flowchart

4 Feature Combination Strategy

After the input images are processed in parallel by the EfficientNet and Advanced Vision Transformer branches, the resulting feature vectors are combined to form a unified representation. This combination strategy is designed to integrate the local features captured by EfficientNet with the global contextual information extracted by the transformer.

To ensure compatibility, each feature vector is projected to a matching dimensional space using dense layers before combination. The vectors are then concatenated, forming a rich and complementary feature embedding that encapsulates both fine-grained textures and high-level semantic information.

This combined feature representation is forwarded to the Capsule Network module, which performs spatial reasoning and part-whole modeling to support the final classification task.

5 Capsule Network

The final stage of the HNet architecture involves a Capsule Network (CapsNet) module shown in Figure 4., which operates on the fused feature vector obtained from the EfficientNet and AVT branches. Unlike traditional neural layers that output scalar activations, capsules produce vector outputs that encode both the probability of the presence of a class and its instantiation parameters, such as pose, orientation, and texture variation.

This property is particularly valuable in histopathological image analysis, where maintaining spatial hierarchies and part-whole relationships is crucial for distinguishing subtle morphological differences between benign and malignant tissues. The capsule network comprises two 4D capsules and three dynamic routing iterations. The output is passed through a softmax classification layer to predict the class label (benign or malignant).

Journal of Information Systems Engineering and Management
2025, 10(59s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

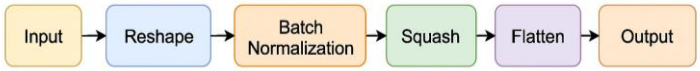

Figure 4: Capsule Network flowchart

6 Training Setup and Optimization

The model was trained using the Adam optimizer, with the following hyperparameters: initial learning rate (LR) = 0.001, $\beta_1 = 0.9$, and $\beta_2 = 0.999$. Binary cross-entropy loss was used for binary classification. To prevent overfitting, we employed an early stopping criterion, which monitored the validation loss over 100 epochs. If no improvement was observed, training was terminated early. Additionally, a reduce-on-plateau scheduler was implemented to halve the learning rate after five epochs of stagnation.

To further improve robustness, the model was trained across multiple data split configurations as detailed in Section 3.1. At each split ratio, the network was retrained from scratch to assess its sensitivity to training set size and ensure consistent performance. Model checkpoints were saved based on validation accuracy, and the final results were reported on the independent test set.

7 Evaluation Metrics and Performance Analysis

The model's performance was evaluated using several metrics, including accuracy, precision, recall and a custom F1-score (harmonic mean of precision and recall). Confusion matrices were generated during testing, providing insights into class-specific performance. Matplotlib was used for visualizing these metrics. These metrics help quantify how well the model distinguishes between benign and malignant tissue in histopathological images. All experiments were conducted on a laptop with an Intel i7-10850H CPU, 32GB of RAM, and a 6GB RTX A3000 GPU, utilizing TensorFlow 2.x.

RESULTS AND DISCUSSION

This section presents the experimental results of the proposed HNet model on the BreakHis dataset under various conditions, including multiple image resolutions and different training/validation/test split configurations. We evaluate the model using several classification metrics such as Accuracy, Precision, Recall and F1-score, to provide a comprehensive analysis of its diagnostic performance. The results are analyzed to assess the impact of input resolution, data availability, and the overall effectiveness of the hybrid architecture compared to standalone models and several state-of-the-art CNN architectures.

1 Impact of Data Splitting Strategies

To assess the robustness and generalization ability of the proposed HNet model under different data availability conditions, we experimented with various training/validation/test split ratios. Specifically, we evaluated six configurations: 70/20/10, 70/15/15, 75/15/10, 80/10/10, 85/10/5, and 90/5/5.

All experiments in this section were conducted using input images resized to 128×128, which represents a balanced compromise between computational efficiency and classification performance (see section 4.7). A detailed analysis of the impact of input resolution is provided in Section 4.2. The results are presented in Table 1, with performance measured by Accuracy, Precision, Recall and F1-score.

As expected, the model's performance improves slightly as the proportion of training data increases. The best results were achieved with the 70/20/10 split, reaching an accuracy of 97.15%. However, even at higher training ratios (e.g., 90%), HNet maintained high and stable performance, demonstrating its strong generalization capability. This consistency reflects the strength of the hybrid design, data augmentation strategy, and the model's ability to learn discriminative features from limited data.

Journal of Information Systems Engineering and Management
2025, 10(59s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Table 1: Classification performance of HNet under different train/validation/test data splits for 128x128 image resolution

Data split Train/Test/Val	Accuracy	Precision	Recall	F1-Score
70/20/10	97.15%	99.04%	95.38%	97.18%
70/15/15	94.98%	97.22%	92.92%	95.02%
75/15/10	95.40%	96.07%	94.76%	95.41%
80/10/10	93.65%	96.41%	91.28%	93.78%
85/10/5	95.76%	94.75%	96.82%	95.77%
90/5/5	96.13%	94.75%	96.15%	95.45%

2 Influence of Input Image Resolution

The quality and resolution of input images are crucial factors in medical image classification, particularly in histopathology, where fine-grained structures such as nuclei, glands, and tissue patterns play a vital diagnostic role. To investigate the effect of image resolution on the performance of HNet, we conducted experiments using three different input sizes: 96×96, 128×128, and 224×224. All experiments in this section were conducted using the 70/20/10 data split, which provided consistently strong results during previous evaluations.

The classification performance at each resolution, measured in terms of Accuracy, Precision, Recall, and F1-score, is presented in Table 2. These experiments aim to highlight how spatial resolution impacts the model's ability to capture fine-grained histopathological features crucial for accurate diagnosis. The highest performance was obtained with 224×224, achieving an accuracy of 97.52% This can be attributed to the preservation of high-frequency details and spatial structures that are essential for accurate discrimination between benign and malignant tissues.

In contrast, the 96×96 resolution, while computationally efficient, led to a decline in classification performance. This suggests that overly compressed images may omit critical morphological information required for reliable diagnosis. The 128×128 resolution offered a strong balance between computational efficiency and classification accuracy, and was thus selected as the default resolution for all subsequent experiments.

Table 2: Classification performance of HNet at different image resolutions using a 70/20/10 data split.

Resolution	Accuracy	Precision	Recall	F1-Score
96x96	94.70%	91.40%	95.50%	93.40%
128×128	97.15%	99.04%	95.38%	97.18%
224×224	97.52%	98.52%	95.43%	96.95%

3 Ablation Study and Component-Level Comparison

To understand the individual contributions of each component within the proposed HNet architecture, we conducted ablation experiments using a fixed resolution of 128×128 and the optimal 70/20/10 data split.

The performance comparison, presented in Table 3, reveals that the EfficientNet-based configuration achieved a strong F1-score of 96.40%, reflecting its ability to extract rich local textures and morphological features. In contrast, the AVT-based configuration showed weaker performance (F1-score: 92.67%), indicating its limitations in capturing fine-grained histological details, despite its strength in modeling global dependencies.

Journal of Information Systems Engineering and Management
2025, 10(59s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

The full HNet model outperformed both individual configurations, achieving the highest f1-score of 96.95%, demonstrating the benefit of hybridizing local and global feature extraction with spatial hierarchy preservation via Capsule Networks. This synergy enhances both the discriminative power and generalization capacity of the model.

Table 3: Component-Level Comparison using F1-Scores

Configuration	Components	F1-score	Model impact
EfficientNet	CNN only	96.40%	EfficientNet excels in feature extraction, capturing local textures well, but struggles with long-range dependencies, resulting in a slightly lower F1-score.
AVT	Transformer only	92.67%	AVT captures global dependencies effectively but lacks in extracting detailed local features like cellular structures, leading to a lower F1-score.
HNet	CNN + AVT + Capsule	96.95%	The hybrid model combines EfficientNet, AVT, and Capsule Networks, enhancing local feature extraction, global context modeling, and spatial hierarchy preservation.

Confusion Matrix and Class-wise Performance

To further evaluate the model's class-wise predictive capabilities, confusion matrices were analyzed across three model configurations: AVT, EfficientNet, and the proposed HNet. This analysis helps in understanding the balance between sensitivity (recall) and specificity, particularly in distinguishing benign from malignant cases.

As shown in Table 4, HNet achieved a strong balance with 1042 true positives (TP) and 1063 true negatives (TN), while maintaining relatively low false positives (FP = 16) and false negatives (FN = 51). Compared to the baseline models, EfficientNet exhibited slightly fewer false positives (FP = 26) and higher true positives (TP = 1035), while AVT showed increased misclassifications (FP = 46, FN = 118), indicating comparatively lower precision and recall.

This comparative analysis confirms that the hybrid HNet model achieves better class-wise discrimination, offering both high sensitivity for malignant detection and low false alarms for benign cases, making it a reliable tool for clinical decision support.

Table 4: Confusion Matrix Results Across Model Variants

Model	True Negatives	False Positives	False Negatives	True Positives
AVT	1033	46	118	975
EfficientNet	1060	26	51	1035
HNet	1063	16	51	1042

5 Training Convergence and Learning Behavior

The proposed HNet model was trained for 100 epochs with real-time monitoring of accuracy and loss. At 224x224 resolution with a 70/20/10 data split, the training curves demonstrated smooth convergence, with validation loss stabilizing after approximately 60 epochs. The application of early stopping and learning rate scheduling effectively mitigated overfitting.

As shown in Figure 5, the learning curves reflect stable optimization dynamics and strong generalization to unseen data, further reinforcing the model's reliability for breast cancer histopathological image classification.

Journal of Information Systems Engineering and Management
2025, 10(59s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

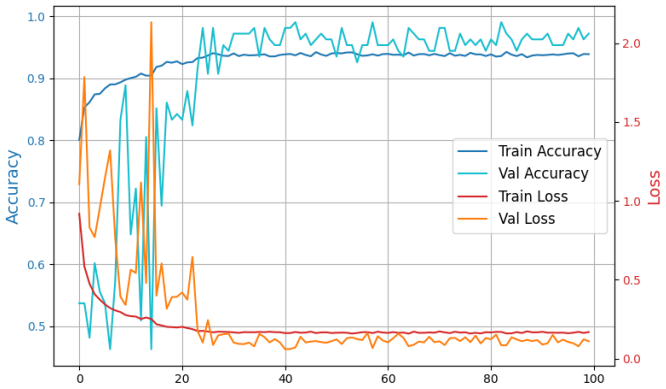

Figure 5: Accuracy and loss over 100 training epochs

6 Comparison with State-of-the-Art

To further validate the effectiveness of the proposed HNet model, we conducted a comparative analysis with several state-of-the-art methods previously applied to the BreakHis dataset. These methods include traditional machine learning approaches, CNN-based models, and recent transformer-enhanced architectures reported in the literature.

Table 5 provides a comparative analysis of several state-of-the-art models and configurations for breast cancer detection, including the proposed HNet architecture. It summarizes the reported classification accuracies and the core architectural components of each approach. The proposed HNet framework achieves an impressive accuracy of 97.52%, leveraging a synergistic integration of EfficientNet, an AVT module, and Capsule Networks. This result either surpasses or closely rivals the performance of existing models that incorporate attention mechanisms, ensemble strategies, or advanced convolutional neural networks, highlighting the effectiveness of the proposed hybrid design.

Table 5. Comparison of the suggested hybrid model with contemporary leading methodologies.

Model / Study	**Accuracy**	**Key Features**
CBAM-EfficientNetV2, [26]	99.01%	EfficientNetV2-XL with Convolutional Block Attention Module (CBAM)
DeepBraestCancerNet [27]	99.35%	Ensemble of ResNet18, ShuffleNet, and Inception-V3Net with transfer learning
CBAM-VGGNet [28]	98.96%	VGG16 and VGG19 fusion with Convolutional Block Attention Module (CBAM)
Inception-ResNet-v2 with Gradient Boosting [29]	96.82%	Inception-ResNet-v2 features with ensemble of CatBoost, XGBoost, and LightGBM classifiers
EfficientNet with Hybrid Attention Mechanisms [30]	91.3%	EfficientNet combined with hybrid attention mechanisms
Proposed HNet	97.52%	Hybrid model EfficientNet + AVT for features extraction, Capsule Networks for classification

7 Computational Cost and Training Time

As detailed in Table 6, the training time varied considerably depending on the image resolution and data split strategy. For the 128×128 resolution using a 70/20/10 split, EfficientNet completed training in approximately 45 minutes, whereas the AVT model required around 1 hour. The hybrid HNet model took approximately 1 hour and 18 minutes under the same configuration. However, increasing the resolution to 224×224 resulted in a substantial rise

in training time. Specifically, the HNet model at 224×224 took nearly 27 hours to complete training and evaluation, emphasizing the significant computational demands of high-resolution image processing in deep learning workflows.

Table 6: Model Execution Time Across Configurations

Model	Resolution	Split	Execution time
EfficientNet	128×128	70/20/10	45 min
AVT	128×128	70/20/10	1 h
HNet	128×128	70/20/10	1 h 18 min
HNet	224×224	70/20/10	27 h

CONCLUSION

In this study, we introduced HNet, a novel hybrid deep learning architecture designed for the classification of breast cancer histopathological images. HNet integrates three powerful components: EfficientNet, Advanced Vision Transformer (AVT), and Capsule Network to exploit local texture patterns, global contextual information, and spatial hierarchies within a unified and scalable framework. Extensive experiments conducted on the BreakHis dataset demonstrated that HNet consistently outperforms standalone models across various image resolutions and data split configurations, achieving an F1-score of up to 97.18%. The model also showed strong generalization capabilities and resilience to data limitations, highlighting its potential for reliable and interpretable cancer diagnosis. Ablation studies and confusion matrix analysis further validated the model's design, showing that hybridizing convolutional and transformer-based representations significantly improves classification performance.

Compared to existing methods including CBAM-EfficientNetV2, DeepBreastCancerNet, and ensemble CNNs, our model HNet offers a compelling balance between performance and interpretability. It avoids the excessive parameter load of deep ensembles while enhancing transparency through capsule-based relational modeling. These attributes make it particularly suitable for clinical decision-support systems, where model accuracy, efficiency, and explainability are equally critical.

Future work will explore the model's adaptability across varying magnification levels (40×, 100×, 200×, and 400×) to assess robustness under different diagnostic conditions. We also plan to integrate multi-scale fusion techniques for better generalization across resolutions, and to extend the framework to multi-class classification of breast cancer subtypes, enabling a more granular and clinically relevant diagnosis beyond binary classification. Additional efforts will focus on explainability via Grad-CAM and capsule activation analysis, as well as cross-dataset generalization using external benchmarks such as BACH and Camelyon16

REFRENCES

[1] M. Sibbering and C.-A. Courtney, 'Management of breast cancer: basic principles', *Surg. Oxf.*, vol. 37, no. 3, pp. 157–163, Mar. 2019, doi: 10.1016/j.mpsur.2019.01.004.
[2] A. Kerhet, M. Raffetto, A. Boni, and A. Massa, 'A SVM-based approach to microwave breast cancer detection', *Eng. Appl. Artif. Intell.*, vol. 19, no. 7, pp. 807–818, Oct. 2006, doi: 10.1016/j.engappai.2006.05.010.
[3] K. Atrey, B. K. Singh, N. K. Bodhey, and R. Bilas Pachori, 'Mammography and ultrasound based dual modality classification of breast cancer using a hybrid deep learning approach', *Biomed. Signal Process. Control*, vol. 86, p. 104919, Sept. 2023, doi: 10.1016/j.bspc.2023.104919.
[4] Z. Jiarui, 'Evaluation of Machine Learning Algorithms in Comparison for Early Breast Cancer Identification', *Highlights Sci. Eng. Technol.*, vol. 120, pp. 35–42, Dec. 2024, doi: 10.54097/gtbfz055.
[5] Ö. Tuğçe and E. Neyhan, 'A Machine Learning Approach to Early Detection and Malignancy Prediction in Breast Cancer', *Int. J. Comput. Exp. Sci. Eng.*, Nov. 2024, doi: 10.22399/ijcesen.516.
[6] K. Atrey, B. K. Singh, and N. K. Bodhey, 'Multimodal classification of breast cancer using feature level fusion of mammogram and ultrasound images in machine learning paradigm', *Multimed. Tools Appl.*, vol. 83, no. 7, pp. 21347–21368, Feb. 2024, doi: 10.1007/s11042-023-16414-6.

[7] M. J. Yaffe, 'Mammographic density. Measurement of mammographic density', *Breast Cancer Res.*, vol. 10, no. 3, p. 209, June 2008, doi: 10.1186/bcr2102.
[8] S. L. Heller and L. Moy, 'Breast MRI Screening: Benefits and Limitations', *Curr. Breast Cancer Rep.*, vol. 8, no. 4, pp. 248–257, Dec. 2016, doi: 10.1007/s12609-016-0230-7.
[9] A. B. Atitallah, J. Kamoun, M. D. Alanazi, T. M. Alanazi, M. Albekairi, and K. Kaaniche, 'An Advanced Medical Diagnosis of Breast Cancer Histopathology Using Convolutional Neural Networks', *Comput. Mater. Contin.*, vol. 83, no. 3, pp. 5761–5779, 2025, doi: 10.32604/cmc.2025.063634.
[10] O. Alqahtani, M. Ghouse, A. Sabahath, O. Hussain, and A. Begum, 'Multi-Scale Vision Transformer with Dynamic Multi-Loss Function for Medical Image Retrieval and Classification', *Comput. Mater. Contin.*, vol. 83, no. 2, pp. 2221–2244, 2025, doi: 10.32604/cmc.2025.061977.
[11] M. Haq, M. Athar, N. Aoun, A. Alluhaidan, S. Ahmad, and Z. Farid, 'CapsNet-FR: Capsule Networks for Improved Recognition of Facial Features', *Comput. Mater. Contin.*, vol. 79, no. 2, pp. 2169–2186, 2024, doi: 10.32604/cmc.2024.049645.
[12] A. D. Raha *et al.*, 'Modeling and Predictive Analytics of Breast Cancer Using Ensemble Learning Techniques: An Explainable Artificial Intelligence Approach', *Comput. Mater. Contin.*, vol. 81, no. 3, pp. 4033–4048, 2024, doi: 10.32604/cmc.2024.057415.
[13] S. Singh, 'Hybrid Models for Breast Cancer Detection via Transfer Learning Technique', *Comput. Mater. Contin.*, vol. 74, no. 2, pp. 3063–3083, 2023, doi: 10.32604/cmc.2023.032363.
[14] L. A. Aldakhil, H. F. Alhasson, and S. S. Alharbi, 'Attention-Based Deep Learning Approach for Breast Cancer Histopathological Image Multi-Classification', *Diagnostics*, vol. 14, no. 13, Art. no. 13, Jan. 2024, doi: 10.3390/diagnostics14131402.
[15] V. Priyadarshni, S. K. Sharma, M. K. I. Rahmani, B. Kaushik, and R. Almajalid, 'Machine Learning Techniques Using Deep Instinctive Encoder-Based Feature Extraction for Optimized Breast Cancer Detection', *Comput. Mater. Contin.*, vol. 78, no. 2, pp. 2441–2468, 2024, doi: 10.32604/cmc.2024.044963.
[16] A. Rafiq, A. Jaffar, G. Latif, S. Masood, and S. E. Abdelhamid, 'Enhanced Multi-Class Breast Cancer Classification from Whole-Slide Histopathology Images Using a Proposed Deep Learning Model', *Diagnostics*, vol. 15, no. 5, Art. no. 5, Jan. 2025, doi: 10.3390/diagnostics15050582.
[17] P. Wang, J. Zhang, Y. Li, Y. Guo, and P. Li, 'Breast Histopathological Image Classification Based on Auto-Encoder Reconstructed Domain Adaptation', *Appl. Sci.*, vol. 14, no. 24, Art. no. 24, Jan. 2024, doi: 10.3390/app142411802.
[18] Y. Çetin-Kaya, 'Equilibrium Optimization-Based Ensemble CNN Framework for Breast Cancer Multiclass Classification Using Histopathological Image', *Diagnostics*, vol. 14, no. 19, Art. no. 19, Jan. 2024, doi: 10.3390/diagnostics14192253.
[19] A. A. Balasubramanian *et al.*, 'Ensemble Deep Learning-Based Image Classification for Breast Cancer Subtype and Invasiveness Diagnosis from Whole Slide Image Histopathology', *Cancers*, vol. 16, no. 12, Art. no. 12, Jan. 2024, doi: 10.3390/cancers16122222.
[20] L. Cao, K. Pan, Y. Ren, R. Lu, and J. Zhang, 'Multi-Branch Spectral Channel Attention Network for Breast Cancer Histopathology Image Classification', *Electronics*, vol. 13, no. 2, Art. no. 2, Jan. 2024, doi: 10.3390/electronics13020459.
[21] K. Ishaq and M. Mustagis, 'Computer Aided Detection and Classification of mammograms using Convolutional Neural Network', Sept. 04, 2024, *arXiv*: arXiv:2409.16290. doi: 10.48550/arXiv.2409.16290.
[22] preeti katiyar, 'A Transfer Learning-Based Deep Learning Model for Automated Breast Cancer Identification in Mammograms', Mar. 29, 2024, *Research Square*. doi: 10.21203/rs.3.rs-3749398/v4.
[23] J. P. Meher, S. Biswas, and R. Mall, 'Deep learning-based software bug classification', *Inf. Softw. Technol.*, vol. 166, p. 107350, Feb. 2024, doi: 10.1016/j.infsof.2023.107350.
[24] H. Feng and C. Wang, 'Research on Classification Methods of Breast Cancer Based on Deep Learning', in *2024 4th International Conference on Electronic Information Engineering and Computer Science (EIECS)*, Sept. 2024, pp. 983–986. doi: 10.1109/EIECS63941.2024.10800722.
[25] F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, 'A Dataset for Breast Cancer Histopathological Image Classification', *IEEE Trans. Biomed. Eng.*, vol. 63, no. 7, pp. 1455–1462, July 2016, doi: 10.1109/TBME.2015.2496264.

Journal of Information Systems Engineering and Management
2025, 10(59s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

[26] N. Sengodan, 'Breast Cancer Histopathology Classification using CBAM-EfficientNetV2 with Transfer Learning', May 13, 2025, *arXiv*: arXiv:2410.22392. doi: 10.48550/arXiv.2410.22392.
[27] G. Boumediene Ghaouti and B. Meftah, 'An Optimized Clustering Approach using Tree Seed Algorithm for the Brain MRI Images Segmentation', *Intel. Artif.*, vol. 26, no. 72, pp. 44–59, June 2023, doi: 10.4114/intartif.vol26iss72pp44-59.
[28] A. Raza *et al.*, 'A Hybrid Deep Learning-Based Approach for Brain Tumor Classification', *Electronics*, vol. 11, no. 7, p. 1146, Apr. 2022, doi: 10.3390/electronics11071146.
[29] M. R. Abbasniya, S. A. Sheikholeslamzadeh, H. Nasiri, and S. Emami, 'Classification of Breast Tumours Based on Histopathology Images Using Deep Features and Ensemble of Gradient Boosting Methods', Sept. 03, 2022, *arXiv*: arXiv:2209.01380. doi: 10.48550/arXiv.2209.01380.
[30] W. Xie *et al.*, 'Patch-based deep learning models for breast mammographic mass classification', in *Proceedings of the 2023 15th International Conference on Bioinformatics and Biomedical Technology*, in ICBBT '23. New York, NY, USA: Association for Computing Machinery, Nov. 2023, pp. 13–22. doi: 10.1145/3608164.3608167.