An Explainable Multi-Model Deep Learning Framework for Breast Cancer Diagnosis from Histopathological Images

Muhammad Nabeel Mehmood[1], Muhammad Hassaan Ashraf[1,2*]
[1]Faculty of Computing, Riphah International University, Islamabad, 45200, Pakistan
[2]Department of Computer Science, COMSATS University Islamabad, Islamabad, 45550, Pakistan
E-mail: mhassaan.scholar@gmail.com
[*]Corresponding author

Keywords: Computer-aided diagnosis, biomedical image processing, histopathological image analysis, explainable artificial intelligence, attention mechanism, scale adaptive dense connection, deep learning, transfer learning, ensemble learning, convolutional neural networks, DenseNet201, HFF-Net, feature fusion techniques, contrast limited adaptive histogram equalization (CLAHE), breast cancer diagnosis

Received: September 20, 2025

Breast cancer is a significant cause of cancer-related deaths among women worldwide. Its early identification and screening are essential for improved patient outcomes and reduced mortality rates. Histopathological image analysis is considered as the gold standard for the diagnosis and prognosis of breast cancer. Nevertheless, the complexity of Whole-Slide Images (WSI) and their manual examination make this task time consuming, and prone to pathologist subjectivity. Recently, Deep Learning (DL) technology has achieved remarkable success in computer vision. However, their application still faces critical challenges in pathology analysis, including Region-of-Interest (RoI) scale variations, inter- and intra-class heterogeneity, diverse staining protocols, and the scarcity of annotated datasets. Furthermore, DL model's findings are opaque and lack decision-level transparency. This study proposes a novel explainable multi-model DL framework for breast cancer classification leveraging histopathological images. The framework integrates Contrast Limited Adaptive Histogram Equalization (CLAHE) for image contrast enhancement, and diverse data augmentation to mitigate class imbalance and overfitting. Proposed architecture ensembles two branches, one employes DenseNet201 benefiting from Transfer Learning (TL) via ImageNet weights, while other utilizes a custom light weight attention based Hierarchal Feature Fusion (HFF) Network. DenseNet201 utilizes multilevel features to effectively tackle gradient vanishing issues and capture intricate feature representations, while HFF-Net, designed specifically for biomedical imaging, leverages HFF stem and multiscale feature extraction with Swish activation to enhance learning stability. Attention mechanism introduced within HFF-Net further refines the output features. Final feature vectors from the two branches are fused at the Global Average Pooling (GAP) layer, consolidating discriminative information. Experimental results on BRACS dataset demonstrate the proposed framework achieves 97.15% accuracy, 92.59% precision, 93.81% recall, and a 93.19% F1-score in screening tasks, while for grading tasks, it attains 84.08% accuracy, 83.39% precision, 83.64% recall, and 83.44% F1-score on 4391 test samples. Additionally, Gradient-Class Activation Mapping (Grad-CAM) saliency heatmap are generated for visual representation of proposed model's choices, thereby increased transparency. The integration of these advanced techniques significantly enhances diagnostic reliability, addressing the challenges in histopathological image analysis.

Povzetek: Študija predstavlja razložljiv večmodelni pristop globokega učenja za natančnejše in preglednejše odkrivanje ter razvrščanje raka dojke na histopatoloških slikah.

1 Introduction

Breast cancer is a common and life-threatening malignancy worldwide, accounting to cancer-related fatalities among women [1], [2]. Risk factors like genetic predispositions, hormonal imbalances, lifestyle and environmental exposures are the wide spectrum causes of breast cancer. According to World Cancer Research Fund (WCRF), breast cancer as second most common cancer worldwide caused 2,296,840 new cases of breast cancer, and 666,103 deaths among women in 2022 [3]. These alarming figures underscore the urgent need for effective diagnostic and treatment strategies. Early detection plays a key role in improving survival rates by favoring patient's prognosis, resulting in enhanced treatment outcomes [4]. However, it requires reliable diagnostic tools and methodologies that can cope with the inherent challenges of the disease's presentation and progression.

Over the years, various strategies developed for breast cancer detection are generally classified into two broad categories: non-invasive and invasive approaches [5]. Non-invasive imaging techniques such as mammography,

ultrasound, Computed Tomography (CT), and Magnetic Resonance Imaging (MRI) are frequently employed for screening and preliminary diagnosis. Mammography, in particular, remains a cornerstone for population-wide screening due to its wide availability, but these approaches may cause incorrect False-Positive (FP) and False-Negative (FN) readings, that can lead to either unnecessary biopsies or missed diagnoses, delaying treatment. Moreover, such non-invasive techniques do not provide tissue-level information necessary for determining specific tumor subtypes.

In contrast, invasive techniques such as biopsies offer a gold standard for precise and accurate diagnosis by extracting tissue-level samples and examining them microscopically as shown in the Figure 1. Histological staining procedures enrich histopathological slides with cellular structures allowing pathologists to examine with considerable details. By assessing parameters such as nuclear morphology, cellular arrangement, and tissue architecture, histopathologists determine whether a tumor is malignant, classify its subtype, and measure its severeness. Pathologists are often required to examine very large, complex images in which staining, contrast, and illumination can vary significantly across tissue sections, even within slides. These insights are critical for designing targeted treatment regimens. However, Manual inspection is time-consuming, subject to inter-observer variability, and prone to human error, making it essential to optimize this process for both efficiency and accuracy.

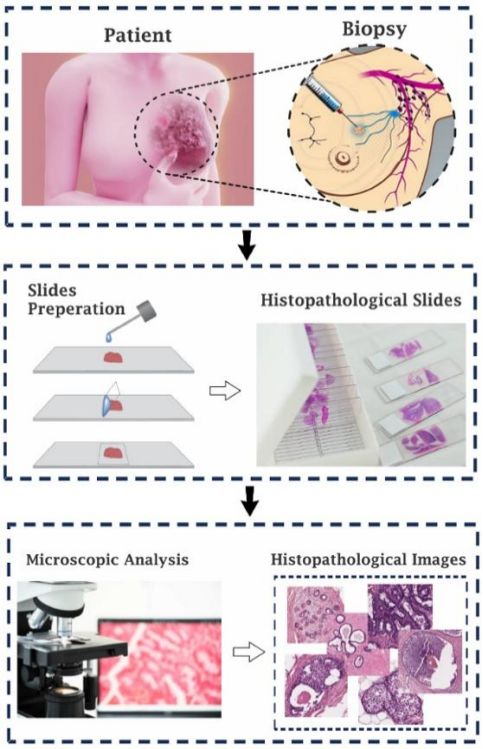

Figure 1: Breast biopsy and histopathological images preparation.

Machine learning and DL methods, especially those with CNNs, have shown over the past decade a remarkable ability to learn useful aspects of visual data, and have been successfully applied to complex tasks like tumor detection and classification in biomedical imaging [6], [7], [8], [9]. In contrast to conventional machine learning techniques based on hand-written features, CNNs can automatically learn discriminative features useful to the classification problem, and have advantages in different aspects of feature-learning. For instance, deeper networks such as ResNet and DenseNets better approximate multi-level representations by using residual and dense connections, whereas multi-scale networks such as Inception-based networks best approximate fine and coarse features by using different filter sizes within a single layer.

One of the key points that affect the performance of CNN in histopathology is the quality and variety of training datasets. Although there are a few publicly available datasets like BreaKHis, BACH and more recently BRACS dataset, each benchmark has images cropped at different magnification levels, staining protocols, and annotation criteria. This heterogeneity may hinder the generalization of model in different clinical settings. Besides, big and highly annotated datasets are rare because every picture takes a lot of time and knowledge of several skilled pathologists. To overcome such shortcomings, TL has become a critical method by pre-training CNNs with weights learned on massive datasets such as ImageNet, it allows the model to inherit generic feature extraction properties that can in turn be refined to smaller and more domain specific histopathology datasets. The strategy is not only efficient in accelerating convergence but also reducing the risk of overfitting when the data is scarce [10].

Nevertheless, no alone CNN architecture universally outperforms all other state-of-the-art (SOTA) CNNs in breast histopathological images classification. Models vary in how effectively they capture multiscale information, mitigate vanishing gradients, or handle image artifacts. Consequently, Ensemble Learning (EL) has made possible to harness the complementary strengths of multiple models by fusing features from various CNN architectures. Ensemble-based approaches achieve more robust and accurate classification compared to single networks, particularly advantageous differentiating complex subtypes of breast cancer.

In this context, this study proposes a multi-model architecture, that combines the strengths of two CNN models designed for solely for multilevel and multiscale features extraction, for breast cancer classification from histopathological images. First branch takes advantage from dense connectivity of DenseNet201 to mitigate gradient vanishing problem and maximize feature reuse. To maximize the generalizability, Densnet201 is assigned the pre-trained ImageNet weights. Second branch uses the light weight HFF-Net proposed by Ashraf et al. [8], that utilizes HFF-stem to hierarchically extract features at varying scales and fuse them to generate enriched feature representations. HFF-Net is proven to capture fine and coarse features in biomedical imaging as it draws inspiration from Inception-like modules for multiscale processing, Xception modules for depth-wise separable

convolutions, but fails to generalize across disease specific patterns. To address this, we introduced attention blocks in HFF-Net's architecture that refines the features before passing them to next block. DensNet201 and HFF-Net originally utilize ReLU and Swish activations, and cross-entropy and sparse_categorical_cross-entropy loss functions, respectively. The proposed architecture merges the output feature vectors from both branches at GAP layer to form a richer feature space promising high-precision classification.

We have utilized the BRACS dataset, introduced in 2022 [11], which encompasses complexities like diverse staining protocols and numerous cancer subtypes, thereby providing divergent test set for evaluation. Our proposed framework employs CLAHE for contrast enhancement to reduce noise and address uneven illumination across the tissue sections. Various geometric transformations (i.e., rotations at 90° & 270°, and horizontal & vertical flips) were applied to perform data augmentation to increase the dataset's variability and mitigate overfitting. Proposed architecture's decisions are interpreted by generating Grad-CAM heatmaps. It is an XAI based technique employed for effective lesion localization and improved diagnostic transparency. The proposed framework consistently outperforms the baseline SOTA CNNs over the same test sets.

Overall, this study demonstrates the usefulness of multi-level and multi-scale feature extracting methods combined in strategic DL framework to enhance automated breast cancer grading. Furthermore, by utilizing Grad-CAM saliency maps, we have made diagnostics more transparent by providing visual explanations of localized lesions. The rest of this paper is structured in the following way: Section 2 is the review of related work. Section 3 explains the constituents of the proposed framework, such as the data preprocessing, architecture of the proposed model, and evaluation metrics. Section 4 presents the experimental design and the results of our empirical study, and Section 5 comments on the implications of these results. Lastly, Section 6 concludes the paper, summing up our most important contributions and findings, along with the future work.

1.1 Contribution

This work offers a novel DL framework for histopathological image-based breast cancer classification. Following is the summary of main contributions:

- The BRACS dataset is utilized in the study to increase classification robustness, as it covers a significant variety of breast cancer subtypes.
- CLAHE enhancement is done to improve contrast, reduce noise, and fix the illumination irregularities. Moreover, to enhance variability and model's generalizability, geometric augmentation is employed on the training and validation set.
- A multi-model architecture is proposed combining the strengths of fine-tuned DenseNet201 using the pre-trained ImageNet weights, and the light weight HFF-Net taking advantage from attention blocks to refine features.
- For improved post-hoc interpretability, Grad-CAM maps are generated to emphasize the decision-making regions and increase transparency.
- A detailed ablation study demonstrating the effect of each component employed in proposed architecture and the comparative analysis with baseline SOTA CNNs is presented.

These developments aid in the creation of a clinically interpretable AI-powered breast cancer histopathology diagnostic system.

2 Literature insights

This section describes previous studies that have used the publicly accessible BRACS dataset for building DL systems to diagnose breast cancers from histopathological images. BRACS dataset is distinguished by its thorough classification of 4391 breast tumor samples form 547 WSIs across seven subtypes [11]. Granular annotations in the dataset facilitate DL research, but introduce problems like inconsistent metadata and tumor subtype variability that affect the generalizability of the models.

Wang et al. [12] used 4,391 Hematoxylin and Eosin (H&E)-stained tumor samples from seven classes present in BRACS dataset for classifying breast cancer utilizing their proposed Heterogeneous Knowledge Distillation (HKD) framework. To address inter- and intra-observer variability, annotations were validated through consensus among pathologists. The HKD framework incorporates a transformer-based Graph Neural Network (GNN) to capture detailed tissue-level representations, and Multi-Teacher GNN (MP-GNNs) for fine-grained cell-level analysis. In order to align biological entities across scales, preprocessing steps included tissue graph construction and cell segmentation using HoVer-Net. Using cross-entropy and biological affiliation losses for optimization, the study focused on enhancing classification consistency across the heterogeneous data from BRACS.

Tafavvoghi et al. [6] classified molecular subtypes (LumA, LumB, HER2, Basal) from H&E-stained whole-slide images (WSIs) using BRACS in conjunction with TCGA-BRCA and CPTAC-BRCA. The extraction of 512×512 tiles for tumor vs. non-tumor classification was made possible by BRACS's annotated tumor regions. Tile augmentation and HER2-Warwick data integration helped to address issues like HER2 subtype imbalance. To minimize inter-slide variability, preprocessing included tile generation using QuPath and Macenko color normalization. The effectiveness of BRACS in addressing tumor heterogeneity and scanner variability was demonstrated by XGBoost aggregation in conjunction with ResNet-18-based one-vs-rest classifiers.

Li et al. [13] incorporated Global Attention Pooling and Bayesian Collaborative Learning (BCL) into a GNN model for breast cancer classification. The BRACS dataset was reclassified as malignant, atypical, and benign. Using SLIC to extract local regions for superpixel segmentation, a ResNet34 backbone pre-trained on ImageNet processed the data. They enhanced the model's robustness by augmentation techniques which included rotation and flipping. Semantic ambiguity between subtypes was addressed by a soft classification head. The framework demonstrated the practicality of BRACS in addressing complex diagnostic problems by iteratively optimizing graph- and patch-level classifiers.

Zheng et al. [14] proposed a TL framework that uses six pre-trained CNNs (VGG16, InceptionV3, ResNet50, etc.) for histopathological image classification from the BreakHis dataset. Class imbalance was addressed through augmentation via horizontal and vertical flipping, and performance was enhanced through EL across accuracy-based weighted voting. Their methodology shows the potential of hybrid architectures for BRACS's grading task's challenges.

Kumari et al. [15] and Potsangbam et al. [16] recommended Magnification-independent classification with the help of ResNet50 and DenseNet121, respectively. TL made the extraction of features possible in sparse data and preprocessing methods such as resizing and augmentation dealt with inter-class variability. These experiments show that the TL is flexible to histopathological data.

Maurya et al. [17] presented FCCS-Net to classify images for breast cancer diagnosis. It is a multi-level attention-based TL model, developed on the foundation of the ResNet18 architecture that enhances channel-wise and spatial features discrimination using a fully convolutional attention mechanism. To give the robustness of varying optical zoom levels, this design focus on hierarchical feature extraction by adding more residual connections and multi-granular analysis. The framework was tested on histopathological datasets such as BreakHis, IDC and BACH and demonstrated flexibility with a variety of image resolutions.

Singh et al. [10] explored hybrid TL frameworks in detecting breast cancer with the help of such architectures as DenseNet, VGG, ResNet, and EfficientNet. Their analysis was concerned with class imbalance of Kaggle IDC dataset by using augmentation methods such as rescaling, zooming, and shearing and preprocessing methods including normalization and undersampling. ReLU activation and cross-entropy loss were used to optimize the models, and hybrid combinations that included DenseNet and Logistic Regression were focused on.

Karthik et al. [18] combined Dual Attention Multi-scale CNN (DAMCNN) with Channel-Spatial Attention ResNet (CSAResNet). The design had batch normalization and channel-spatial attention modules to minimize overfitting. EL is able to balance out the class distributions with augmentation methods such as rotation and zooming hence tackling FPs and magnification sensitivity.

Majumdar et al. [19] used TL to create a Gamma function-based ensemble of GoogleNet, VGG11, and MobileNetV3_Small for the classification of breast cancer. The rank-based confidence scoring made the framework useful in a variety of datasets, including BRACS, as it improved robustness with magnification.

Ashraf et al. [9] propose a hybrid MSHFF-Net, an explainable Multi-Scale Hierarchical Feature Fusion (MSHFF) architecture designed for breast cancer grading task using histopathological images. The model integrates adaptive multi-scale blocks with a feature-fusion stem, to capture of both tissue-level patterns. MSHFF-Net processes input images of size 320×320 and is trained on BRACS dataset. The architecture employs Leaky ReLU activation and utilizes sparse categorical cross-entropy as the loss function. To enhance interpretability and clinical trust, the study incorporates the Grad-CAM technique to visualize the decision-making process of MSHFF-Net.

Although BRACS has been used for molecular classification [6] and tumor subtyping [12], [20], few studies have directly utilized its seven-class hierarchy for cancer classification [9]. The effectiveness of hybrid architectures (such as attention mechanisms and gamma-based weighting) in addressing issues like inter-class ambiguity and data imbalance is demonstrated by existing frameworks [14], [15], [16], [18], [19]. The proposed framework, which combines EL and TL to handle the particular complexities of histopathological images, is driven by these advances.

3 Proposed methodology

The proposed framework for screening and grading of breast cancer via histopathological image analysis comprises a dual branch CNN architecture, enhanced by integrating TL and attention mechanisms. Figure 2 presents the complete pipeline of our proposed framework, which include data acquisition, preprocessing pipeline detailing both enhancement and augmentation strategy, and proposed architecture's training and evaluation. Local contract enhancement emphasizes on subtle nuclear and architectural patterns to distinguish between atypical (like ADH, FEA) and malignant (like DCIS, IC) lesions. Moreover, the dataset is subjected to geometric augmentation in order to eliminate class imbalance and increase generalizability. The following sections describes thorough dissection of these components.

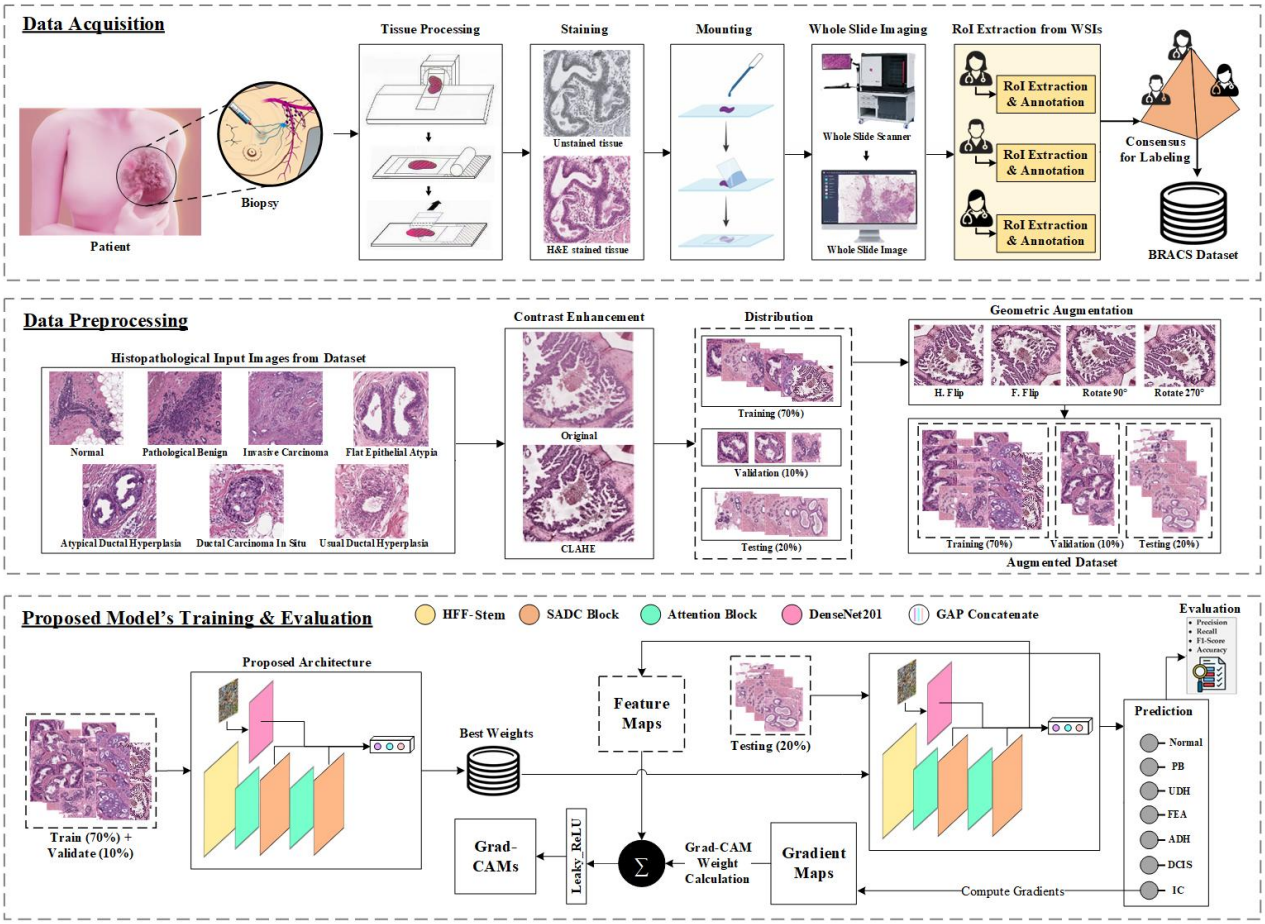

Figure 2: Proposed framework designed for breast tumor diagnosis using histopathological images analysis

3.1 Data acquisition

The BRACS (BReAst Carcinoma Subtyping) dataset is a detailed and robust benchmark aimed at advancing research in breast cancer histopathology image analysis. The dataset contains total of 4,391 images extracted from the 547 WSIs of 189 different patients, which were utilized in this study. A WSI is a high-resolution, panoramic image of a tissue sample, obtained from entire microscope slides. An annotated set of WSIs were divided and distributed to three pathologists, and each one of them extracted RoI image samples, which were later annotated with the consensus of all three pathologists [11]. The classification of lesions into seven subtypes portraits BRACS as a valuable resource for breast cancer screening and grading. The dataset is scanned at a resolution of 0.25 µm/pixel with 40× magnification using an Aperio AT2 scanner, and includes diagnostically challenging lesions like Flat Epithelial Atypia (FEA) and Atypical Ductal Hyperplasia (ADH), which are critical as precursors to Invasive Carcinoma (IC). These subtypes along with other Normal (N), Pathological Benign (PB), Usual Ductal Hyperplasia (UDH) and Ductal Carcinoma In Situ (DCIS) classes, makes BRACS a valuable tool for tackling real-world diagnostic complexities. Figure 3 shows lesion spots of all 6 diseases from BRACS dataset.

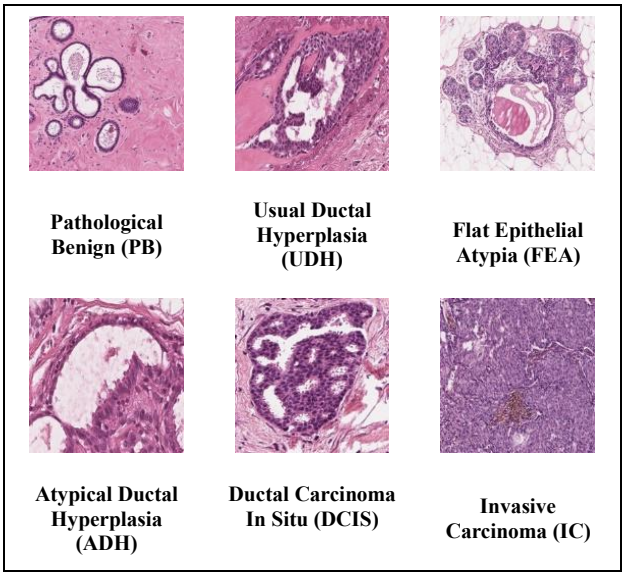

Figure 3: Visual representation of lesion classes from BRACS dataset

Compared to other datasets like BreakHis and BACH, BRACS excels in diversity [11], [20] and richness. Unlike datasets focused on binary classification, BRACS supports multi-class subtyping, offering a realistic and nuanced benchmark for breast cancer diagnosis. Its inclusion of artifact-prone and heterogeneous images better represents clinical scenarios, allowing researchers to develop algorithms that generalize effectively to complex diagnostic tasks. This make BRACS particularly valuable for benchmarking AI systems, improving automated subtyping, and advancing computational pathology.

3.2 Data preprocessing

The BRACS dataset comes with a predetermined set of training, testing, and validation splits, but they did not align with the desired 70%-20%-10% distribution, used in our study. To address this, all samples from each class in the original splits were pooled together and redistributed into training (70%), testing (20%), and validation (10%) subsets. This reorganization ensured that each subset is in 70%-20%-10% distribution, as mentioned in Table 1, providing a consistent and equitable foundation for model training and evaluation.

Table 1: BRACS screening (2-class) and grading (7-Class) datasets summary

Dataset	Classes	Train (70%)	Validate (10%)	Test (20%)
BRACS Screening	Cancerous	2715	387	777
	Normal	358	51	103
BRACS Grading	N	358	51	103
	PB	530	75	153
	UDH	329	47	95
	ADH	397	56	115
	FEA	548	78	157
	DCIS	524	74	151
	IC	385	55	110

Before training, all images were resized to a standardized resolution of 224×224 pixels using bicubic interpolation. This step ensured uniform input dimensions for CNN models, optimizing computational efficiency while preserving critical morphological details.

3.2.1 Contrast limited adaptive histogram equalization (CLAHE)

The contrast enhancement technique employed was CLAHE to enhance the image quality and model performance [21]. Noise, uneven staining and illumination are present in histopathological images, and they may obscure critical cellular characteristics. CLAHE solve these issues by enhancing local contrast, and by preventing over-amplification of noise, retaining fine morphological and textual features required to carry out classification.

To apply CLAHE, the images were broken into uniform, non-overlapping blocks so that contrast could be adjusted on a local basis. Each image was divided into 64 tiles using a 8 × 8 grid that offers a trade-off between the local contrast adjustments and the computational requirements. The Corner Regions (CR), Border Regions (BR), and Inner Regions (IR) were done separately in order to preserve structural integrity. To manage contrast enhancement and avoid noise over-amplification, a clip limit of 2.0 was used to restrict contrast enhancement to a controlled amount, ensuring that the enhancement does not overly exaggerate. The Cumulative Distribution Function (CDF) was calculated using the adjusted histograms which fine-tuned the pixel intensity values within each tile. CLAHE decreased variability brought on by uneven staining and improved the model's capacity to distinguish between various tumor subtypes by making subtle histopathological structures more visible that can be seen in the Figure 4.

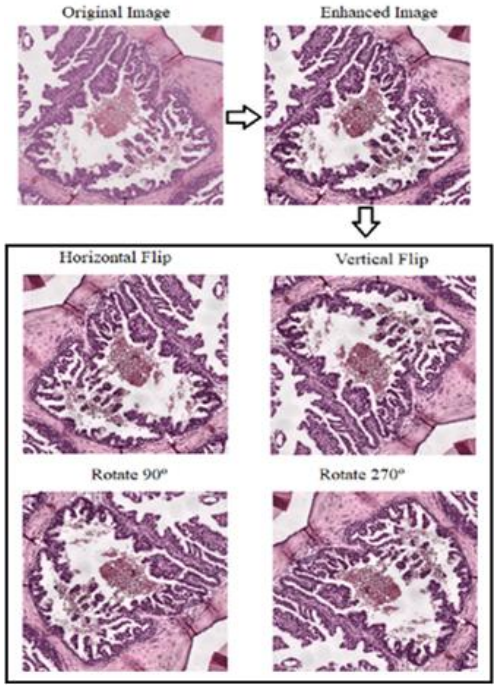

Figure 4: Impact of CLAHE enhancement and geometric augmentation on input image

3.2.2 Geometric augmentation

Geometric augmentation was applied on the BRACS dataset in order to increase variability and enhance the proposed framework's capacity for generalization. By augmenting all the subset, biases in model evaluation are avoided, and the testing accurately represent actual clinical situations. The dataset size was essentially increased by employing geometric transformations to create four augmented variants for every contrast-

enhanced image that can be visualized in the Figure 4. The augmented images were added to their enhanced variants increases the dataset size by a factor of five. Pre- and post-augmentation samples are mentioned in Table 2. The augmentation transformations used are as follows:

- **Horizontal Flip:** Reflecting along vertical axis.
- **Vertical Flip:** Flip along the horizontal axis.
- **90° Rotation:** Rotating the image clockwise.
- **270° Rotation:** Counterclockwise rotation.

Table 2: pre-and post-augmentation dataset samples

Dataset	Classes	Before Aug.	After Aug.
BRACS Screening	Cancerous	3879	19395
	Normal	512	2560
BRACS Grading	N	512	2560
	PB	758	3790
	UDH	471	2355
	ADH	568	2840
	FEA	783	3915
	DCIS	749	3745
	IC	550	2750

By assisting the CNN models in creating feature representations that are both rotation- and orientation-invariant, these transformations decreased the possibility of overfitting to particular spatial arrangements. This augmentation method was critical in enhancing the strength and adaptability of the classification framework, as it increased the variability of the training data.

3.3 CNN architecture

CNNs that implement simple feed-forward architectures such as AlexNet, process the data sequentially using multiple layers, and take advantage of the fixed kernel sizes, such as 3x3 filters with a stride of 1, to extract fine-grained features. With 61 million training parameters, AlexNet uses dropout to minimize overfitting and non-saturating neurons to accelerate the training process. When used on histopathology imaging, it commonly fails at complex tasks in grading due to its fixed kernel sizes [22].

Multi-level CNNs tackle the limitations of simple feed-forward models by reusing features through skip or dense connections. Models like ResNet18 and ReNet50 incorporates residual blocks with skip connections to address the gradient vanishing issue, that enables efficient training and features learning [23]. These models accept the input image of size 224 × 224 × 3, and have 44 and 23 Million (M) trainable parameters, respectively. These networks enhance the learning of both low- and high-level features, however, their computational demand and risk of overfitting is high for datasets like BRACS.

DenseNet101 and DenseNet201 with their innovative design that connects each layer to every other layer, optimizing feature reuse across the layers with dense connections, they use 10 and 20 M trainable parameters, respectively. They process the input image of 224 × 224 × 3 and integrate a series of convolutional layers, batch normalization, ReLU activation, max and average pooling and depth concatenation layers. All these architectural components collaborate in promoting performance and efficiency in feature extraction.

Multi-scale CNNs tackle the scale variation issues by incorporating multiple filters of different kernel sizes within a single layer, allowing the network to detain both fine and broad patterns, making it suitable for datasets like BRACS, having varying patterns and lesion sizes. InceptionV3 employs Inception modules which combine parallel convolutional and pooling operations with varying kernel sizes, and Xception-Net utilizing the depth-wise separable convolutions, both take the input image of 299 × 299 × 3, with 315 and 71 layers, and 23.9 and 22.8 M trainable parameters, respectively. Despite the advantages, multis-scale CNNs may encounter challenges, such as feature loss in deeper layers and inefficiencies in fusing multi-level features when skip connections are absent.

In DL, researchers propose hybrid CNN models, that use the advantages of many SOTA CNN architectures to produce a classification framework that is more resilient and flexible, and capture relevant features in their respective domains like agriculture [24], speed estimation [25] and biomedical imaging [7], [9]. Hybrid models leverage deep networks' high-level feature abstraction capabilities, combined in a single CNN architecture. Ashraf et al. [8] proposed HFF-Net to overcome the drawbacks of traditional CNNs, by incorporating hierarchical multiscale feature extraction minimizing data loss for biomedical imaging. The network is organized into two primary parts: an HFF-stem module that captures low-level data hierarchically using four parallel convolutional layers which concatenates at various depths, and a branch module that extracts high-level features by incorporating two Scale Adaptive Dense Connection (SADC) blocks. Two GAP layers reduce overfitting while use of swish activation and sparse_categorical_cross-entropy loss helps in mitigating dead neuron and overfitting issue. The model maintains the spatial information necessary for accurate lesion detection by using only 1.18 M trainable parameters. However, when applied to complex datasets such as BRACS, which exhibit significant heterogeneity in lesion types, these structures may not distinguish subtle features and be prone to focusing on irrelevant or noisy information, which emphasizes on the need for additional refinement in the architecture.

3.4 Ensemble learning

EL combines multiple models to leverage the strengths of both to enhance classification accuracy, improve robustness, and achieve greater generalizability compared to individual models. This approach has shown immense potential in capturing complex histopathological patterns in breast cancer diagnosis, particularly for analyzing H&E-stained histological images [14], [26], which often present challenges for single architecture. The EL process is typically divided into three key stages:

- **Data Sampling/Selection:** Ensuring diversity in training data is essential for effective learning. In this study, we divided the dataset into training (70%), testing (20%), and validation (10%) subsets to maintain a structured learning process. Moreover, the training and validation subset underwent augmentation to increase generalizability to more comprehensive feature representation.
- **Feature Extraction and Concatenation:** Instead of training multiple base models independently, we leveraged the strengths of different CNN architectures by extracting their learned features and concatenating them into a unified feature space. The extracted discriminative features serve as inputs to a final classification layer, enhancing the robustness of the predictions.
- **Combining Results:** Rather than employing traditional ensemble strategies like majority or weighted voting, our approach fuses the feature representations from the two branches before classification. This feature-level integration enables the model to learn a richer and more informative data, ultimately improving classification performance.

3.5 Transfer learning

TL moves over the knowledge that is obtained through solving one problem into another related problems. This approach has been useful in breast histopathology, where issues like small, annotated datasets and the low performance of DL models can be encountered. DenseNet121 has been used as a feature extractor due to its dense connection, which facilitates a smooth flow of information across layers [16]. Potsangbam et al. [27] showcased the power of this model by resizing histopathological images to 224 × 224 pixels and processing them through DenseNet121. The extracted deep features were then classified into benign or malignant categories using fully connected layers, leading to significant improvements in classification accuracy. This approach also addressed the issue of dataset imbalance through the use of augmentation techniques. TL is implemented using two primary strategies:

- **Feature Extraction**: In this method, pre-trained models are used as feature extractors. The convolutional layers of these models, trained on large datasets like ImageNet, are frozen, and the extracted features are then fed into task-specific classifiers. For instance, Maurya et al. utilized ResNet18 combined with multi-level attention mechanisms to extract fine-grained features from histopathological images [17]. This approach achieved high classification accuracy on datasets such as IDC and BreakHis.
- **Fine Tuning**: Fine-tuning involves selectively unfreezing specific layers of a pre-trained model to adapt it to the target domain. An example can be seen in the FCCS-Net model, where attention mechanisms were added to ResNet18 [17], that delivered state-of-the-art performance across multiple datasets. These mechanisms focused on both spatial and channel-specific features, allowing the model to capture intricate cellular patterns.

By addressing key challenges such as limited data availability, class imbalance, and computational resource constraints, TL has become an essential tool in histopathological image analysis for breast cancer diagnosis.

3.6 Overview of proposed architecture

The proposed architecture adapts an enhanced features flow within architectural blocks for precise and accurate breast cancer classification. The architecture is designed to capture tissue level patterns from H&E-stained images, to address the morphological heterogeneity, and diagnostic complexity inherent in histopathological images. Figure 5 illustrates the proposed architectural diagram.

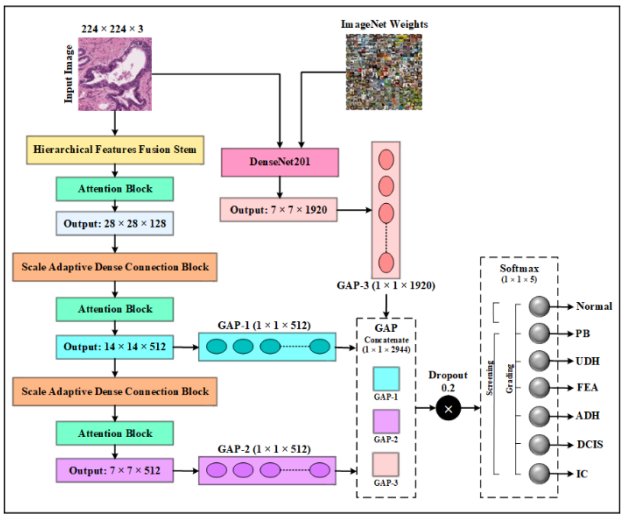

Figure 5: Proposed architecture's diagram

The model accepts the input image with dimensions 224 × 224 × 3 pixels, as this resolution maintains essential cellular details and maximizes computational efficiency. In the first branch, input is processed via HFF-stem [8], which processes the input image using four parallel convolutional layers to obtain hierarchical features at various scales as demonstrated in Figure 6. The initial layer (Conv.1) consists of a 3 × 3 convolution with Stride (S) value 1 to extract Feature Maps (FM) of size 224×224×64, whereas Conv.2 is downsamples the input to 112×112×64 using 3×3 convolution and S=2. Third and

fourth convolutions utilizes 5×5 and 7×7 kernels to extract 56×56×32 FMs with S=4, respectively. These multiscale features are integrated at various levels in the stem. Conv.1's FMs are then downsampled with Max Pooling (MP) layer having 2×2 kernel and S=2. In this way, the FMs from Conv.1 and Conv.2 are concatenated, forming a 112×112×128 tensor. These concatenated FMs are downsampled to 112×112×64 with a 1×1 convolution, known as the Channel Pooling Layer (CPL). Two parallel convolution layers consisting of 3×3 filters with S=1 and S=2 are used to process the CPL output. The S=1 convolution output is then subjected to MP, and the output is joined with the S=2 convolution output and the Conv.3 (5×5, S-4) FMs. This concatenated set of 56×56×160 features is then fed into another CPL that downsizes the features to 56×56×64 for the final stem's output.

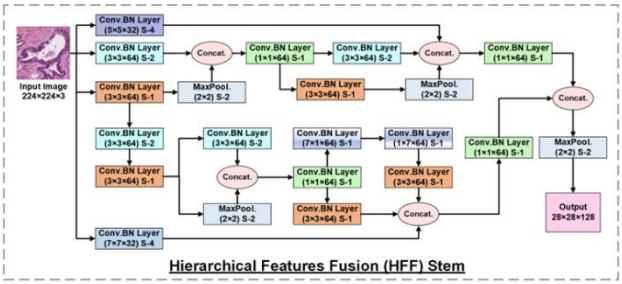

Figure 6: HFF-stem used in HFF-Net [8]

Moreover, the FMs of Conv.1 are sent through a S=2 convolution layer to generate 112×112×64 FMs which are followed by a standard convolution layer with 64 filters. The resultant features are then subjected to parallel S=2 convolution and MP layers and then concatenated to create 56×56×128 FMs. These FMs are subjected to CPL, and then 3×3, 1×7, 7×1, and 3×3 filters all having S=1 are applied. Lastly, the Conv.4 output is finally added to the output of these layers to produce a 56×56×160 tensor. CPL is used to cut the 160 channels into 64 channels which are in turn combined with the 64 channels of other parallel layers. The ultimate output of the HFF-stem is a collection of 128 FMs of size 56×56, optimized to produce low-level features at the lowest possible information loss. Finally, a standard MP layer is applied to the HFF-Stem's output reducing the features dimensions to 28×28.

The refinement of output FMs from HFF-stem is done through an attention block, as demonstrated in Figure 7. The attention block adaptively recalibrates channel-wise responses, improving the features before passing to the next layers [7]. The GAP is applied to the spatial dimensions, and the input-tensor is reduced into a 1×1×128 descriptor with global contextual information of each channel, and then directed to a Fully Connected (FC) layer that reduces the channel dimension to 1×1×16 with a reduction factor of r = 8. A Leaky_ReLU activation then comes after this reduction and this is non-linear but does not break the gradient flow so that the attention mechanism is allowed to focus on more useful features [24]. These representations are then upsampled to the original channel size using a second FC layer, yielding a 1×1×128 attention vector. A sigmoid activation is used to normalize the attention weights. The resulting channel-wise attention weights are finally multiplied element-wise with the original 28×28×128 FMs to obtain recalibrated features, which are sent to the subsequent block.

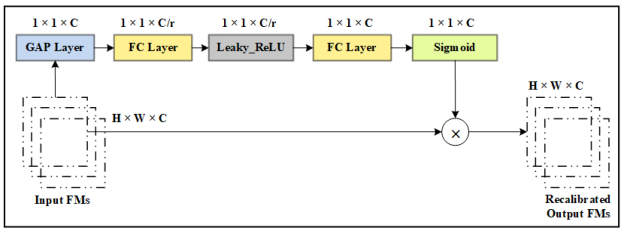

Figure 7: Attention block employed in Enhanced HFF-Net

The polished FMs are fed into the first SADC block, which is illustrated in Figure 8. It processes the input in four parallel branches each serving different purposes. A 3×3 MP with S=1 is used to maximize the information passing through the last unit, and then 1x1 convolution to reduce the FMs to 128 channels. The second, third, and fourth branches start with a 1×1 convolution to reduce the FMs to 64 channels, thus reducing computational complexity. The second branch then applies three closely spaced 3×3 convolutional layers, each of which extracts 64 FMs, making the total 192 FMs. The third branch uses 32 closely interspersed 5×5 convolutional kernels, which produce 96 FMs. The bigger 5×5 kernel captures a wider scope of spatial features, and thus it can detect larger patterns that smaller kernels might fail. The fourth branch employs 32 closely spaced 3×3 convolutional kernels with dilation rate of 2, which produce additional 96 FMs. The rate of dilation expands the receptive field, without any additional parameters, and improves the capacity of the network to encode contextual information with computational efficiency. The FMs of all of these branches are concatenated and subjected to a MP layer to reduce the dimensionality, resulting in 14×14×12 FMs. This representation is subsequently subjected to an attention block and afterwards sent to the first GAP layer and the second SADC block. The second SADC block works similar to the first one and generates 7×7×512 FMs that are refined by a third attention block, before finally sending them to the second GAP layer.

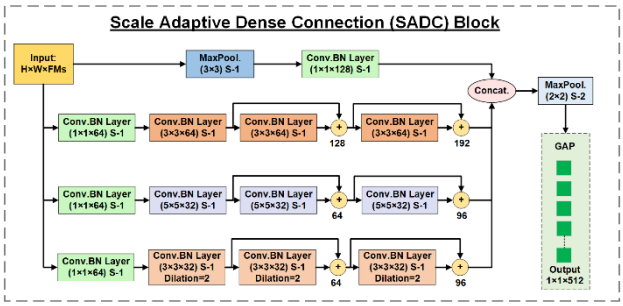

Figure 8: SADC block used in HFF-Net [8]

The second branch obtains multi-level features via the densely connected DenseNet201's architecture. It uses pre-trained ImageNet weights, and thus generalizes to the

heterogeneous patterns in the histopathological RoIs. DenseNet201 generates output FMs of 7×7×1920, which are then fed into the GAP-3 layer, creating a 1×1×1920 tensor. All three GAP layers, having discriminative features at different spatial depths, are concatenated in a GAP concatenation unit. This design guarantees that the classification layer is provided with a rich hierarchical representation of multi-level and multi-scale features, which include low-level and high-level cues. In order to avoid overfitting, dropout rate of 0.2 is employed prior to final classification. A Softmax layer is then used to produce class probability distributions of histopathological images, in this case, breast cancer diagnosis.

3.7 The Activation and loss function

In this research, we have employed the swish activation function, which is given in the Equation (1):

$$f(x) = x \cdot \sigma(x), \tag{1}$$

where $\sigma(x) = \frac{1}{1+e^{-x}}$ denotes the sigmoid function.

Swish activation maintains high convergence rates while effectively addressing the gradient vanishing problem, which significantly enhances the training efficiency, leading to notable improvements in classification performances.

To further optimize the training over the diverse BRACS dataset, we employed the "sparse_categorical_cross-entropy" loss function, derived from the traditional categorical cross-entropy, specifically to work with integer-encoded target labels instead of one-hot encoded labels. Its formulation provides a suitable mechanism for managing grading tasks in this context. The loss function is formally defined in Equation (2).

$$L = -\Sigma_{i=1}^{N}\ log\left(P_{i,y_i}\right) \tag{2}$$

Here, L represents the loss function, while N denotes the total number of samples in the dataset. The variable y_i corresponds to the integer-encoded label for the $i-th$ sample, and P_{i,y_i} is the predicted probability that the $i-th$ sample belongs to its true class, as indicated by y_i.

3.8 Evaluation metrics

Evaluation metrics which include recall, precision, F1-score, and overall accuracy, are utilized to assessed the proposed framework's efficiency. The Confusion Matrix (CM) is used to calculate these metrics, which are reported for each class label. Additionally, $CM[i][j]$ is the number of instances that are predicted to be in class j but are actually in class i. Following are the performance metrics for each class:

$$\boldsymbol{Precision}_i = \frac{TP_i}{TP_i+FP_i} = \frac{CM[i][i]}{CM[i][i]+\sum_{j\neq i} CM[j][i]} \tag{3}$$

$$\boldsymbol{Recall}_i = \frac{\boldsymbol{TP}_i}{\boldsymbol{TP}_i+\boldsymbol{FN}_i} = \frac{CM[i][i]}{CM[i][i]+\sum_{j\neq i} CM[i][j]} \tag{4}$$

$$\boldsymbol{F1-Score}_i = \frac{2\times \text{Precision}_i \times Recall_i}{\text{Precision}_i + Recall_i} \tag{5}$$

$$\boldsymbol{Accuracy} = \frac{\sum_{i=1}^{n} CM[i][i]}{\sum_{i=1}^{n}\sum_{j=1}^{n} CM[i][j]} \tag{6}$$

where

- n is the number of classes.
- $CM[i][i]$ represents the True Positive (TP) samples for class i.
- $\sum_{j=1,j\neq i}^{n} CM[j][i]$ represents the False Positive (FP) samples for class i, which are instances predicted as class i but belong to other classes.
- $\sum_{j=1,j\neq i}^{n} CM[i][j]$ represents the False Negative (FN) samples for class i, these instances are actually in class i, but are predicted as other classes.

3.9 Grad-CAM interpretability analysis

The proposed framework applies Grad-CAM as an Explainable Artificial Intelligence (XAI) to enhance model transparency, and permit reliable decisions, by producing localized maps of the per class instances, that highlight the significant spatial areas contributing in class prediction. Heatmaps are extracted from the last convolutional block prior to the GAP layer, as it retains higher semantic information and spatial structure. For an input image and target class c, the gradient of the class score y_c with respect to the FMs FM_k from the selected convolutional layer is computed. The weight a_k^c of each channel of the FM_k is then calculated as an average of the gradients with respect to the spatial dimensions:

$$\alpha_k^c = \frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W}\frac{\partial\, y_c}{\partial FM_k(i,j)} \tag{7}$$

H and W are the width and height of FMs, that are measures of the contribution that each channel has to the class prediction. The weighted FMs are then summed to obtain the Grad-CAM heatmap, which is then followed by a Leaky ReLU activation:

$$L_{Grad-CAM}^{c} = \text{Leaky\_}ReLU\sum_{k=1}^{k}\alpha_k^c\, FM^k \tag{8}$$

The conventional ReLU is replaced with leaky ReLU [24], that process small negative gradients, mitigating dead neurons, and improve gradient flow. The heatmap is then resampled to the size of the input image and overlaid onto the original image, indicating areas of tumor that contributed in the model decision-making.

4 Experiments and evaluation

Substantial experiments on the BRACS dataset were conducted to thoroughly assess the performance of the proposed framework, with particular focus on both screening (binary classification) and grading (multi-class classification) tasks. To evaluate the effectiveness of integrated components in proposed framework, an

ablation study was conducted using DenseNet201, InceptionV3 and HFF-Net [8]. Our framework was analyzed comparatively against baseline SOTA CNNs (which include AlexNet, Mobile-Net, ResNet18/50, DenseNet101/201, Xception and Inception), HFF-Net and a recently published work by Ashraf et al. [9], where they conducted experiments on the BRACS dataset using their proposed MSHFF-Net. Same training and evaluation environments were carried out, which makes it a relevant choice for comparison.

All of these experiments were carried out by using the NVIDIA Tesla T4, with 40 streaming multiprocessors, 16 GB of GDDR6 memory and 6 MB shared L2 cache, which is exceptionally powerful and capable of up to 8.1 TFLOPS of FP32 and 65 TFLOPS of FP16 precision. The GPU with clock speed of 1.59 GHz and a memory bandwidth of 300 GB/s, provided the required performance to process the augmented sets of datasets and complex computations needed by DL frameworks. A batch size of 32, which offered a balance between memory and gradient stability. AdamW helped in achieving rapid convergence and stable training dynamics, due to its adaptive learning rate. The model was trained over 60 epochs providing sufficient time to capture complicated patterns. Moreover, best weights were saved automatically, to maintain best performance until the last evaluation stage. During the testing phase, these weights were used to evaluate the generalization ability of the models and their classification accuracy. This methodology made sure that the reported results are the best performances for histopathological images classification for breast tumor detection.

4.1 Ablation study for binary class classification

A rigorous ablation study presented in Table 3 was carried out to observe that how the components integrated in proposed framework contributes in accurate diagnoses of breast cancer.

Table 3: Ablation study for binary classification

Model	TL	EL	Att.	Enh.	Aug.	Binary Class Classification			
						Pr. %	Rec.%	F1 Score %	Acc.%
DenseNet201	✗	✗	✗	✗	✗	81.04	75.93	78.16	91.70
Inception V3	✗	✗	✗	✗	✗	83.80	69.49	74.07	91.47
HFF-Net [8]	✗	✗	✗	✗	✗	91.96	72.98	78.90	93.18
Enhanced HFF-Net [V1]	✗	✗	✓	✗	✗	91.81	79.88	84.53	94.54
DenseNet201 [V2]	✓	✗	✗	✗	✗	87.14	77.68	81.47	93.29
Inception V3 [V3]	✓	✗	✗	✗	✗	91.02	77.16	82.23	93.86
Dense+Incp [V4]	✓	✓	✗	✗	✗	87.40	86.45	86.91	94.65
Dense+E.HFF [V5]	✓	✓	✓	✗	✗	89.25	88.58	88.91	95.45
Dense+E.HFF [V6]	✓	✓	✓	✓	✗	92.50	88.13	90.15	96.13
Dense+E.HFF [V7]	✓	✓	✓	✓	✓	92.59	93.81	93.19	97.15

✗**:** Component not included, ✓**:** Component included, **TL:** Transfer Learning, **EL:** Ensemble Learning, **Enh.:** Enhancement, **Aug.:** Augmentation, **Att.:** Attention Block

To assess the influence of individual elements on, we started by analyzing a number of baseline architectures, trained without improvements for binary classification task. DenseNet201, Inception V3, and HFF-Net became top performers by obtaining accuracy of 91.70%, 91.47%, and 93.18%, respectively. HFF-Net showed a promising edge compared to other baselines. In the next step, we incorporated the attention block in HFF-Net architecture, resulting in enhanced HFF-net [V1] that achieved 94.54% accuracy along with a stable precision of 91.81%, demonstrating the effectiveness of attention mechanism. We fine-tuned DenseNet201 [V2] and InceptionV3 [V3] on ImageNet weights to evaluate the impact of TL instead of learning from scratch. With accuracies of 93.29% and 93.86%, respectively, both versions showed gains over standalone baselines. Moving forward, the ensemble of DenseNet201 and Inception V3 (Dense+Incp) [V4] stated 94.65% accuracy, while combining DenseNet201 with Enhanced HFF-Net (Dense+E.HFF) [V5] gained remarkable performance of 95.45%, with E.HFF-Net being attention-based hierarchically adaptive and DenseNet201 leveraging advantage from ImageNet weights.

Introducing contrast enhancement via CLAHE increased the visibility of essential features, resulting in improved model's (Dense+E.HFF) [V6] performance of 96.13% accuracy, with stable improvements across all evaluation metrics. Augmentation training and validation set further enhanced model's generalizability and improved the accuracy. The final variant (Dense+E.HFF) [V7] achieved an overall accuracy of 97.15%, F1-score of 93.19, recall of 93.81 and precision of 92.59, showing the effectiveness of all the integrated components of proposed framework. The CM in Figure 9 give an overview of the screening performance of proposed model on the BRACS's test set. The diagonal cells represent correct

predictions, while off-diagonal cells tells where the model confuses the labels.

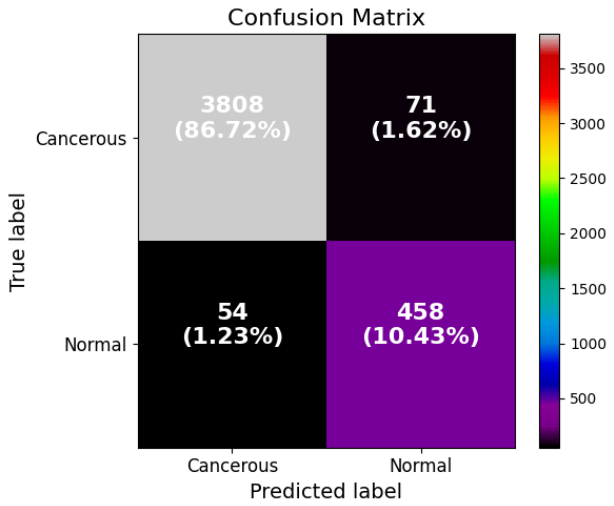

Figure 9: The proposed framework's CM for screening task.

The CM shows strong separation between Cancerous and Normal tissue, with 3808 samples correctly identified as Cancerous (TP), and 458 samples correctly identified as Normal (TN), resulting in 4266/4391 correct predictions and the overall accuracy of 97.15%. The error rates are limited to 54 FPs, and 71 FNs. Interpreting Cancerous as the positive class, the counts TP = 3808, TN = 458, FP = 54, and FN = 71, imply a high sensitivity for Cancerous detection and specificity for Normal cases recognition. Notably, the small FN (71 cases) suggests that the proposed framework hardly misses very few cancerous patterns, which make it suitable for breast cancer histopathological image analysis under screening environment.

4.2 Ablation study for multi-class classification

The impact each component was also evaluated under multi-class classification task. The goal was to assess the performance across a broader range of tumor categories. Training from scratch produced accuracies of 67.42%, 64.36%, and 70.92% for DenseNet201, Inception V3 HFF-Net, respectively, as shown in Table 4. In the baseline comparisons, HFF-Net distinguished itself with high accuracy of 70.92%, appearing as a good fit for difficult and complex classification tasks. Furthermore, incorporating attention block increased the accuracy to 76.02% with stable precision and F1-score, indicating the efficacy of design choice for enhanced HFF-Net (E.HFF-Net) [V1]. Fine-tuned versions of baselines produced substantial improvements, with Inception V3 [V3] reaching 73.75% and DenseNet201 [V2] reaching 76.69% accuracy. The Dense+Incp [V4] ensemble fell short of DenseNet201's [V2] performance by achieving only 74.88%, indicating that unstructured increase of features for prediction can confuse the model. On the other hand, the Dense+HFF [V5] demonstrated an accuracy of 78.61% with balanced precision and recall, suggesting that the distinct features capturing ability of HFF-Net effectively supplemented DenseNet201 in multi-class classification setting.

Table 4: Ablation study for multi-class classification

Model	TL	EL	Att.	Enh.	Aug.	Binary Class Classification			
						Pr. %	Rec.%	F1 Score %	Acc.%
DenseNet201	✗	✗	✗	✗	✗	67.31	65.02	64.81	67.42
Inception V3	✗	✗	✗	✗	✗	62.61	62.71	61.60	64.36
HFF-Net [8]	✗	✗	✗	✗	✗	69.50	69.32	68.04	70.92
Enhanced HFF-Net [V1]	✗	✗	✓	✗	✗	75.91	76.11	75.64	76.02
DenseNet201 [V2]	✓	✗	✗	✗	✗	76.21	75.90	75.87	76.69
Inception V3 [V3]	✓	✗	✗	✗	✗	74.37	73.19	72.26	73.75
Dense+Incp [V4]	✓	✓	✗	✗	✗	74.73	74.89	74.40	74.88
Dense+E.HFF [V5]	✓	✓	✓	✗	✗	77.65	77.74	77.57	78.61
Dense+E.HFF [V6]	✓	✓	✓	✓	✗	79.20	79.28	79.12	80.31
Dense+E.HFF [V7]	✓	✓	✓	✓	✓	83.39	83.64	83.44	84.08

✗: Component not included, ✓: Component included, **TL:** Transfer Learning, **EL:** Ensemble Learning, **Enh.:** Enhancement, **Aug.:** Augmentation, **Att.:** Attention Block

The accuracy was further raised to 80.31% by adding CLAHE-based enhancement (Dense+HFF) [V6], suggesting that enhancement provides additional benefits in identifying classes that might differ slightly and are difficult in distinguishing otherwise. Data augmentation (Dense+E.HFF) [V7] yielded the top performance with an accuracy of 84.08%, and stable precision, recall and F1-score of 83.39, 83.64 and 83.44, respectively. The multi-class classification performance on the BRACS's test set is summarized by the CM in Figure 10, where true labels are on the y-axis and predicted labels are on the x-axis. In a multi-class setting, TP/FN/FP/TN are interpreted per

class using a one-vs-rest view. This lens is essential as errors often occur among morphologically adjacent entities like hyperplasia, atypia and in situ carcinoma, and overall accuracy alone can mask these clinically meaningful results.

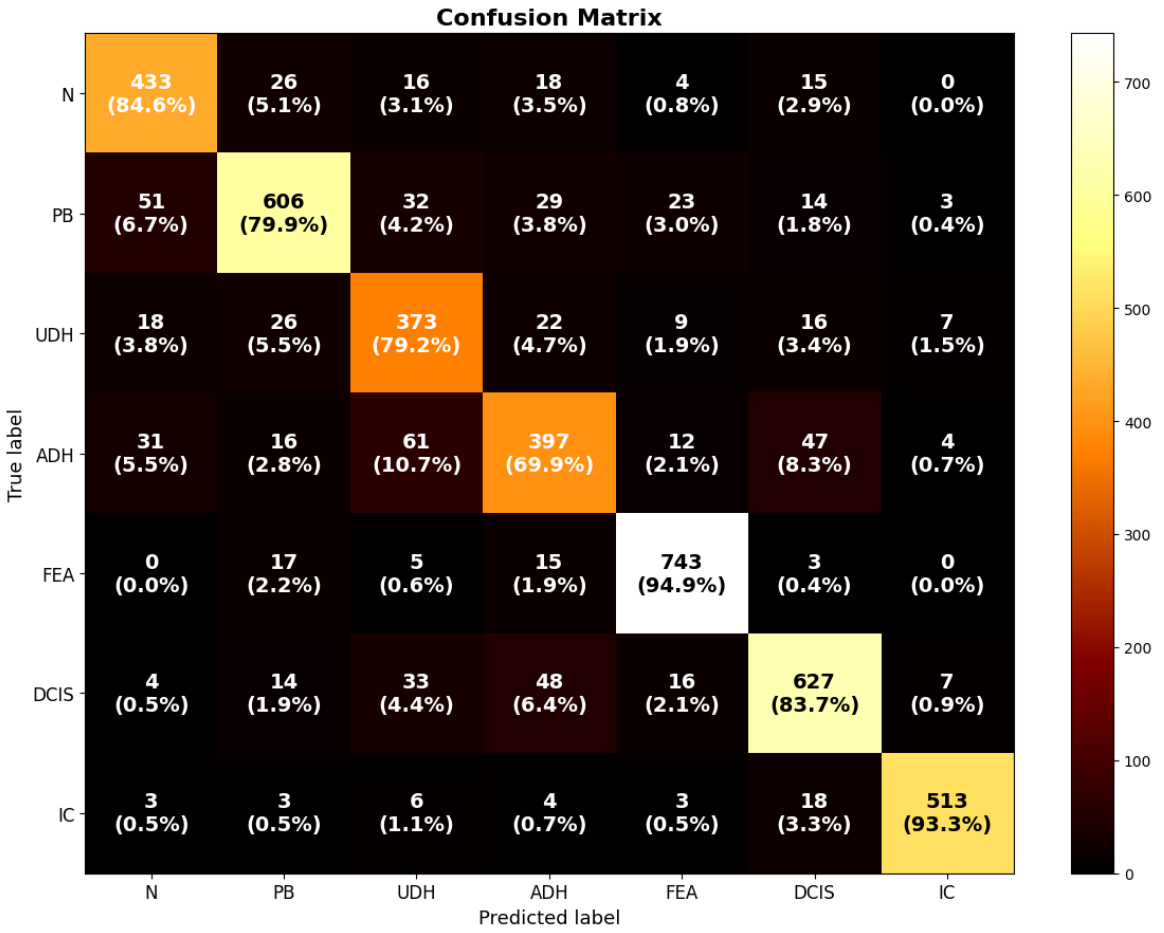

Figure 10: The proposed framework's CM for grading task

The diagonal of CM is predominant, which confirms that the proposed framework (Dense+HFF) has achieved the greatest overall grading accuracy, with 3692 correctly classified samples of 4391 (84.08% accuracy). On class level, the diagonal wise entities are the class-wise TP, the other cells in that row are the FN, the cells in that column are the FP, and all other cells are treated as TN. The model demonstrates best recognition of FEA (743/783, 94.9% recall) and IC (513/550, 93.3% recall), which suggests that these tumor types are well discriminated against in H&E RoIs. The performance for (DCIS 627/749, 83.7%) and Normal (433/512, 84.6%) are also high, which proves that the separation of malignancy tissue and normal tissue is performed with high reliability. Most residual errors are found in the moderate ADH, having highest FN load and lowest recall (397/568, 69.9%), with errors being due to confusion with severe (DCIS:48, FEA:15, IC:4) and mild (PD:29, UDH:22, N:18) cases, while DCIS is most frequently confused with ADH (47 cases) and other mild (45) tumors. These misclassifications are consistent with the BRACS label space, and includes tough atypical lesions that contribute to the difficulty of diagnosis, and they explain why gains such as contrast enhancement and augmentation are quantified by reinforcing subtle morphology needed to resolve borderline categories. Overall, this study shows that substantial performance improvements can be achieved by progressively integrating components that helps in refining and focusing on the disease relevant features.

5 Discussion

The diagnosis of breast histopathology is based on the identification of epithelial proliferations in ducts and lobules typically on H&E-stained RoIs. The diagnostic continuum of this study is normal tissue, benign and proliferative intraductal lesions and malignant entities that incorporate in situ and invasive disease. The documented findings suggest that there is a two-phase performance pattern. In the lesion detection task that contrasts cancerous tissue against normal tissue, the CM yields 3808 TPs and 458 TNs out of 4391 test samples, which corresponds to 97.15% overall accuracy, indicating that the proposed framework has a high sensitivity, and minimum abnormal missed cases. While grading into seven classes, the CM indicates 3692 correct predictions out of 4391 samples, which corresponds to 84.08% overall accuracy. The overall result is consistent with multi class histopathology benchmarks where morphological

heterogeneity and staining variability constrain ceiling performance.

Class wise recall is highest for FEA at 743 of 783 cases (94.9%) and IC at 513 of 550 cases (93.3%). Only a few non-invasive types are predicted to be IC, resulting in low overcalled rate. The robust structural signatures are plausibly decreasing the rate of spurious invasive predictions. Strong performance is also found in DCIS and normal tissue with recall of 83.7% and 84.6%, respectively. The trend suggests that lesions having unique epithelial lining or stromal invasion phenotypes are separated efficiently. The CM demonstrates that the most residual errors are between intraductal categories that are biologically adjacent to each other. ADH has least recall (397 of 568 cases 69.9%), and the highest FN load, with the majority of the false identifications made with UDH (61 cases) and to DCIS (47 cases). DCIS has been most commonly misdiagnosed with 48 DCIS cases.

UDH and PB tissue show intermediate recall of 79%, bidirectionally confused with normal tissue and ADH, due to the overlapping heterogeneous populations of epithelium, and low-grade neoplastic patterns in small foci. There is a partially operational and extent-based boundary separating ADH and low-grade DCIS, which is why the adjacent groups are challenging to detect by even experts. Grad-CAM offers class discriminative localization through activation map weighting with gradients of the target class, resulting in a heatmap indicating areas most helpful to prediction. Such visual explanations come in handy especially in histopathology as they can be matched with known diagnostic cues of ductal and stromal. The strongest claims of interpretability are those in which the heatmaps are consistent with pathology priors.

The analysis by Grad-CAM shows that the proposed framework is weighted towards architectural integrity and ductal organization in the prediction of PB tissue as shown in Figure 11. The attention-based refinement focuses on well-structured glands and activation on features of proliferative or malignant progression. This pattern indicates that the decision process is based on structure as opposed to unrelated visual correlations. This interpretability finding is consistent with the levels of moderate recall of PB tissue at 606 of 758 cases (79.9%), since benign patterns may be stromal and ductal varying.

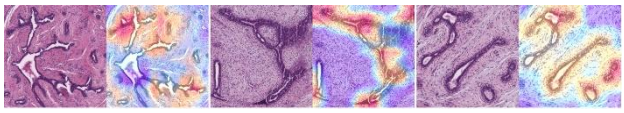

Figure 11: Grad-CAM heatmaps for PB-class.

Heatmap analysis of UDH suggests that it is sensitive to intraductal proliferation of epithelial cells and cellularity of ductal spaces. Activations move towards boundary centered patterns observed in benign tissue to density centered patterns observed in ducts, consistent with the histopathological description of UDH as a heterogeneous intraductal proliferation. This reinforcement of the interpretation shows that the model relies on structural and density related cues, and is not based on spurious texture, as illustrated in Figure 12.

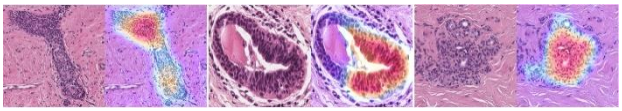

Figure 12: Grad-CAM heatmaps for UDH-class

Figure 13 shows Grad-CAM outputs of ADH, which focused its attention on localized localities in ducts where the epithelial cells proliferated abnormally with minimal attention being paid to the stroma. The maps indicate the presence of a shift in diffuse patterns that is typical of benign proliferations to tight high intensity focal areas, and which is congruent with architectural irregularity and the emergence of cytologic monotony. The interpretability trend promotes diagnostic plausibility, but the low recall of 69.9% shows focal atypia continues to pose a challenge in minute conditions.

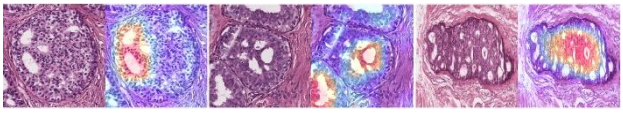

Figure 13: Grad-CAM heatmaps for ADH-class.

The FEA Grad-CAM analysis reveals that the focus was on the ductal linings, lumen facing epithelial layers and not on the intraductal masses. Figure 14 demonstrates that the activation geometry is aligned edge and this is consistent with the dependence on the epithelial thickness and slight atypia as opposed to overall hypercellularity. This action is congruent with the microscopic description that FEA has enlarged ductules or acini surrounded with a thin layer of low-grade atypical cells. The recall of FEA was extremely high with 94.9% indicating that the attention-based E.HFF-net reliably detect fine structures.

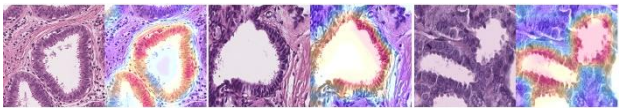

Figure 14: Grad-CAM heatmaps for FEA-class.

The heatmaps analysis for DCIS shows broad activation across entire ductal regions filled by densely packed malignant cells, and the maps appear large and continuous within ducts, as illustrated in Figure 15. This activation pattern supports the claim that the network distinguishes diffuse malignant proliferation from localized atypical change, which aligns with the 83.7% recall level.

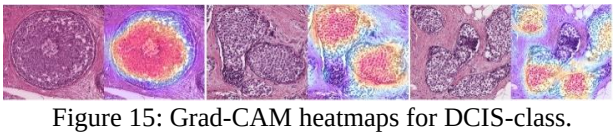

Figure 15: Grad-CAM heatmaps for DCIS-class.

The Grad-CAM analysis in Figure 16 for IC shows attention distributed across disorganized stromal regions rather than confined ductal spaces, and it highlights patterns compatible with tissue infiltration. Invasive breast carcinoma is fundamentally defined by basement membrane breach with stromal infiltration, and reporting protocols emphasize stromal invasion and associated desmoplastic response. The high recall for IC at 93.3% therefore coheres with an interpretable focus on invasion anchored morphology.

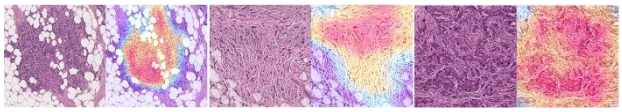

Figure 16: Grad-CAM heatmaps for IC-class.

Table 5 summarizes the performance of SOTA CNNs for breast cancer screening using histopathological images. MobileNet, being lightweight (4 million parameters) and with 87.04% accuracy, also faces trade-offs in precision and recall, which is reflected by its lower F1-score (62.49%). AlexNet (17 million parameters) achieved an accuracy of 88.64%, yet has poor precision (72.87 percent) and recall (75.88 percent), indicating that it is not able to identify small features of the lesions. ResNet-18 and ResNet-50 demonstrate a slight improvement, where ResNet-50 attains 87.61% accuracy, which is higher than 85.68% of ResNet-18, highlighting the advantage of a deeper architecture to capture more intricate features. DenseNet101 and DenseNet201 proved more effective than ResNets, with DenseNet201 reaching an accuracy of 89.43%, which shows the benefit of densely connected layers in reusing features. Xception Net (91.70% accuracy) and Inception Net (91.47% accuracy) demonstrate that multi-scale and depthwise separable convolutions improve the performance, especially in irregular tissue areas.

Table 5: Comparative analysis for screening.

CNN Backbone	Pr. (%)	Rec. (%)	F1-Score (%)	Acc. (%)
AlexNet [30]	72.87	75.88	74.23	88.63
Mobile-Net[31]	66.11	60.66	62.49	87.04
ResNet-18 [32]	65.03	64.52	64.77	85.68
ResNet-50 [32]	66.57	57.61	59.55	87.61
DenseNet101[28]	68.31	70.87	69.45	86.47
Xception Net [33]	81.04	75.93	78.16	91.70
DenseNet201 [28]	75.46	63.70	67.10	89.43
Inception Net [29]	83.80	69.49	74.07	91.47
HFF-Net [8]	91.96	72.98	78.90	93.18
Proposed Framework	92.59	93.81	93.19	97.15

The HFF-Net, which is inspired by architectural patterns of Inception, Xception, and DenseNet, achieves high accuracy of 93.18% and F1-score of 78.90, indicating that it is able to detect cancerous lesions with a low number of FPs. The proposed framework retains the best performance with the accuracy of 97.15%, precision of 92.59%, recall of 94.81%, and F1-score of 93.19%.

Grading tasks are inherently more challenging due to the increased number of classes, ranging from normal tissue to invasive carcinoma. The simpler architectures such as AlexNet (17 million parameters) retained only 34.04% accuracy, which is not good enough to distinguish subtle features between the seven tumor classes. MobileNet has a lightweight design and slightly higher accuracy (39.70%), although the precision (42.56%) and recall (39.18%) are low, indicating that it is challenging to use on multi-class tasks. ResNet-18 and ResNet-50 achieved 54.29% and 42.87% accuracies, respectively, with ResNet-50 performing poorly, potentially because of its complicated architecture and the difficulty of balancing class distributions. DenseNet101 (60.06% accuracy) has the advantage of densified connectivity, which enhances the preservation of features, yet it remains unable to detect subtle differences between classes. Xception Net retained an accuracy of 44.68%, which indicates its inability to detect subtle morphological differences. Inception V3 (64.36% accuracy) converges well at multi-scale processing but did not outperform DenseNet201 (67.42%), which uses densely connected layers to extract features more efficiently.

Table 6: Comparative analysis for grading.

CNN Backbone	Pr. (%)	Rec. (%)	F1-Score (%)	Acc. (%)
AlexNet [30]	42.21	33.34	27.31	34.04
Mobile-Net[31]	42.56	39.18	39.39	39.70
ResNet-18 [32]	50.61	52.41	50.72	54.29
ResNet-50 [32]	55.01	40.66	40.50	42.87
Xception Net [33]	47.74	43.06	40.70	44.68
DenseNet101[28]	61.99	60.37	58.24	60.06
DenseNet201 [28]	62.61	62.71	61.60	64.36
Inception Net [29]	67.31	65.02	64.81	67.42
HFF-Net [8]	69.50	69.32	68.04	70.92
MSHFF-Net [9]	73.20	72.99	72.32	73.94
Proposed Framework	83.39	83.64	83.44	84.08

With a slightly higher F1-Score (68.04%) and an accuracy of 70.927%, HFF-Net outperforms all SOTA CNNs by leveraging hierarchical feature extraction, demonstrating superior adaptability to varying lesion sizes and shapes. HFF-Net's comparatively low parameter count (1.26 million) demonstrates an effective architecture design that maintains a significant amount of discriminative power. MSHFF-Net achieves an impressive accuracy of 73.94%, with significant improvements in precision (73.20%), recall (72.99%), and F1-score (72.32%), demonstrating superior performance through its multi-scale hierarchical feature fusion approach. Our proposed framework improves the input

images with CLAHE and performs augmentation for generalizability, and refines the output features by utilizing attention mechanisms, all of which contributes to achieving the highest overall accuracy of 84.08%, precision of 83.39%, recall of 83.64%, and F1-score of 83.44%, as listed in Table 6. Comparing this to the SOTA and other proposed hybrid CNNs, there is a noticeable improvement. The resulting performance demonstrates how this approach can capture highly localized patterns, which are essential for differentiating the most challenging BRACS subtypes (like DCIS, IC, FEA and ADH).

All things considered, the proposed framework shows clinically significant grading accuracy, and emphasizes the enhancement and stain robust preprocessing, while supporting the claim that augmentation plays a vital role in reducing generalization error. This helps explain that the attention mechanisms and hierarchical features fusions can improve lesion grading by enabling models to prioritize discriminative structures and suppress irrelevant background. The present performance is strong, but domain shift remains a central barrier for clinical translation. Intraductal lesions are defined partly by extent and distribution across ducts, and small regions can underrepresent lesion scales, which is a known source of discordance for ADH versus low grade DCIS. Moreover, interpretability evidence should also be treated as supportive rather than definitive. Clinical deployment benefits from pairing Grad-CAM with complementary checks such as occlusion-based sensitivity, counterfactual testing, and reader studies aligned to user needs.

6 Conclusions and future work

This study proposed a novel explainable multi-model DL framework to screen and grade breast cancer using histopathological analysis. The proposed algorithm achieved strong results on the BRACS data in terms of binary (screening) and multi-class (grading) classification task, and outperformed all the standalone SOTA CNN architectures, such as AlexNet, MobileNet, and ResNet variants, DenseNet variants, Inception Net Xception Net, MSHFF-Net and HFF-Net, through the use of ImageNet weights, feature level concatenation, CLAHE-based contrast enhancement, and geometric data augmentation. The results indicate the importance of dense feature propagation and attention based hierarchical feature extraction in overcoming the challenges of H&E-stained images.

Although these developments have been made, there is still room for more improvement. Future research will be directed towards coming up with more advanced hybrid models that will combine multi-level and multi-scale feature extraction methods to minimize both FPs and FNs, especially on complex subtypes of breast lesions. The quantitative assessment with the use of perturbation-based measures will also be performed in the future. This strategy will evaluate the changes in the confidence of the model with the removal or addition of the most relevant pixels in the input image. The deletion measure will be used to measure the decrease in model confidence when most significant pixels are substituted with a baseline value (e.g., blurred image), and the insertion measure will estimate the gain in model confidence as the most significant pixels are reintroduced into the image. We will also consider more sophisticated data augmentation methods, including DCGANs or other generative models, to further reduce the effects of class imbalances and increase model robustness to changes in tissue morphology and staining protocols. Additional testing of the generalizability of the model in various datasets, such as BreakHis and BACH, will contribute to proving the applicability of the framework in the clinical environment. Lastly, although the proposed model is good at grading tasks, it can be improved by increasing feature fusion and layer connectivity to improve the accuracy of classification, particularly in difficult and subtle lesion types. The solution of these areas will help in the creation of a more reliable, scalable and clinically viable breast cancer histopathology diagnostic system.

References

[1] M. Xie *et al.*, “Multi-resolution consistency semi-supervised active learning framework for histopathology image classification,” *Expert Syst. Appl.*, vol. 259, Jan. 2025, doi: 10.1016/j.eswa.2024.125266.

[2] C. C. Ukwuoma *et al.*, “Enhancing histopathological medical image classification for Early cancer diagnosis using deep learning and explainable AI – LIME & SHAP,” *Biomed. Signal Process. Control*, vol. 100, p. 107014, Feb. 2025, doi: 10.1016/J.BSPC.2024.107014.

[3] World Cancer Research Fund, “Breast cancer statistics.”

[4] F. Yu *et al.*, “BiaCanDet: Bioelectrical impedance analysis for breast cancer detection with space–time attention neural network,” *Expert Syst. Appl.*, vol. 269, p. 126223, Apr. 2025, doi: 10.1016/J.ESWA.2024.126223.

[5] M. Zubair, S. Wang, and N. Ali, “Advanced Approaches to Breast Cancer Classification and Diagnosis,” *Front. Pharmacol.*, vol. 11, p. 632079, Feb. 2021, doi: 10.3389/FPHAR.2020.632079/BIBTEX.

[6] M. Tafavvoghi *et al.*, “Deep learning-based classification of breast cancer molecular subtypes from H&E whole-slide images,” *J. Pathol. Inform.*, vol. 16, Jan. 2025, doi: 10.1016/j.jpi.2024.100410.

[7] M. H. Ashraf *et al.*, “HIRD-Net: An Explainable CNN-Based Framework with Attention Mechanism for Diabetic Retinopathy Diagnosis Using CLAHE-D-DoG Enhanced Fundus Images,” 2025, doi: 10.3390/xxxxx.

[8] M. H. Ashraf and H. Alghamdi, “HFF-Net: A hybrid convolutional neural network for diabetic retinopathy screening and grading,” *Biomedical*

Technology, vol. 8, pp. 50–64, Dec. 2024, doi: 10.1016/J.BMT.2024.09.004.

[9] M. H. Ashraf, M. Ahmed, M. N. Mehmood, M. E. Qureshi, and H. Alghamdi, "MSHFF-Net: An Explainable Multi-Scale Hierarchical Feature Fusion CNN for Breast Cancer Diagnosis from Histopathological Images," in *2025 27th International Multitopic Conference (INMIC)*, Islamabad: IEEE, Jan. 2026, pp. 1–6.

[10] S. Singh *et al.*, "Hybrid Models for Breast Cancer Detection via Transfer Learning Technique," *Computers, Materials and Continua*, vol. 74, no. 2, pp. 3063–3083, 2023, doi: 10.32604/cmc.2023.032363.

[11] N. Brancati *et al.*, "BRACS: A Dataset for BReAst Carcinoma Subtyping in H&E Histology Images," *Database*, vol. 2022, 2022, doi: 10.1093/database/baac093.

[12] K. Wang, F. Zheng, D. Guan, J. Liu, and J. Qin, "Distilling heterogeneous knowledge with aligned biological entities for histological image classification," *Pattern Recognit.*, vol. 160, Apr. 2025, doi: 10.1016/j.patcog.2024.111173.

[13] Q. He *et al.*, "Global attention based GNN with Bayesian collaborative learning for glomerular lesion recognition," *Comput. Biol. Med.*, vol. 173, May 2024, doi: 10.1016/j.compbiomed.2024.108369.

[14] Y. Zheng *et al.*, "Application of transfer learning and ensemble learning in image-level classification for breast histopathology," *Intelligent Medicine*, vol. 3, no. 2, pp. 115–128, May 2023, doi: 10.1016/j.imed.2022.05.004.

[15] V. Kumari and R. Ghosh, "A magnification-independent method for breast cancer classification using transfer learning," *Healthcare Analytics*, vol. 3, Nov. 2023, doi: 10.1016/j.health.2023.100207.

[16] J. Potsangbam and S. Shuleenda Devi, "Classification of Breast Cancer Histopathological Images Using Transfer Learning with DenseNet121," in *Procedia Computer Science*, Elsevier B.V., 2024, pp. 1990–1997. doi: 10.1016/j.procs.2024.04.188.

[17] R. Maurya, N. N. Pandey, M. K. Dutta, and M. Karnati, "FCCS-Net: Breast cancer classification using Multi-Level fully Convolutional-Channel and spatial attention-based transfer learning approach," *Biomed. Signal Process. Control*, vol. 94, Aug. 2024, doi: 10.1016/j.bspc.2024.106258.

[18] R. Karthik, R. Menaka, and M. V. Siddharth, "Classification of breast cancer from histopathology images using an ensemble of deep multiscale networks," *Biocybern. Biomed. Eng.*, vol. 42, no. 3, pp. 963–976, 2022, doi: 10.1016/j.bbe.2022.07.006.

[19] S. Majumdar, P. Pramanik, and R. Sarkar, "Gamma function based ensemble of CNN models for breast cancer detection in histopathology images," *Expert Syst. Appl.*, vol. 213, p. 119022, Dec. 2023, doi: 10.1016/J.ESWA.2022.119022.

[20] M. Tafavvoghi, L. A. Bongo, N. Shvetsov, L. T. R. Busund, and K. Møllersen, "Publicly available datasets of breast histopathology H&E whole-slide images: A scoping review," Dec. 01, 2024, *Elsevier B.V.* doi: 10.1016/j.jpi.2024.100363.

[21] Pramod K.B. Rangaiah, B.P. Pradeep Kumarb, and Robin Augustinea, "Histopathology-driven prostate cancer identification: A VBIR approach with CLAHE and GLCM insights," *Elsevier* , vol. 182, 2024.

[22] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Adv. Neural Inf. Process. Syst.*, vol. 25, 2012, [Online]. Available: http://code.google.com/p/cuda-convnet/

[23] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904–1916, Dec. 2015, doi: 10.1109/TPAMI.2015.2389824.

[24] M. H. Ashraf, F. Jabeen, M. Waqar, and A. Kim, "HFA-Net: Explainable Multi-Scale Deep Learning Framework for Illumination-Invariant Plant Disease Diagnosis in Precision Agriculture," *Sensors*, vol. 26, no. 7, p. 2067, Mar. 2026, doi: 10.3390/s26072067.

[25] M. H. Ashraf, F. Jabeen, H. Alghamdi, M. S. Zia, and M. S. Almutairi, "HVD-Net: A Hybrid Vehicle Detection Network for Vision-Based Vehicle Tracking and Speed Estimation," *Journal of King Saud University - Computer and Information Sciences*, vol. 35, no. 8, p. 101657, Dec. 2023, doi: 10.1016/J.JKSUCI.2023.101657.

[26] L. Solorzano, S. Robertson, B. Acs, J. Hartman, and M. Rantalainen, "Ensemble-based deep learning improves detection of invasive breast cancer in routine histopathology images," *Heliyon*, vol. 10, no. 12, Jun. 2024, doi: 10.1016/j.heliyon.2024.e32892.

[27] J. Potsangbam and S. S. Devi, "Classification of Breast Cancer Histopathological Images Using Transfer Learning with DenseNet121," in *Procedia Computer Science*, Elsevier B.V., 2024, pp. 1990–1997. doi: 10.1016/j.procs.2024.04.188.

[28] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," 2017. [Online]. Available: https://github.com/liuzhuang13/DenseNet.

[29] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," 2016.

[30] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Adv. Neural Inf. Process. Syst.*, vol. 25, 2012, [Online]. Available: http://code.google.com/p/cuda-convnet/

[31] A. Howard *et al.*, “Searching for MobileNetV3.”
[32] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” 2016. Accessed: Jun. 27, 2023. [Online]. Available: http://image-net.org/challenges/LSVRC/2015/
[33] F. Chollet, “Xception: Deep Learning With Depthwise Separable Convolutions,” 2017.