scientific reports

Check for updates

OPEN

Synergistic transfer learning and adversarial networks for breast cancer diagnosis: benign vs. invasive classification

Wejdan Deebani[1], Lubna Aziz[2,3]✉, Arshad Aziz[2], Wael Sh. Basri[4], Wedad M. Alawad[5] & Sara A. Althubiti[6]

Current breast cancer diagnosis methods often face limitations such as high cost, time consumption, and inter-observer variability. To address these challenges, this research proposes a novel deep learning framework that leverages generative adversarial networks (GANs) for data augmentation and transfer learning to enhance breast cancer classification using convolutional neural networks (CNNs). The framework uses a two-stage augmentation approach. First, a conditional Wasserstein GAN (cWGAN) generates synthetic breast cancer images based on clinical data, enhancing training stability and enabling targeted feature incorporation. Second, traditional augmentation techniques (e.g., rotation, flipping, cropping) are applied to both original and synthetic images. A multi-scale transfer learning technique is also employed, integrating three pre-trained CNNs (DenseNet-201, NasNetMobile, ResNet-101) with a multi-scale feature enrichment scheme, allowing the model to capture features at various scales. The framework was evaluated on the BreakHis dataset, achieving an accuracy of 99.2% for binary classification and 98.5% for multi-class classification, significantly outperforming existing methods. This framework offers a more efficient, cost-effective, and accurate approach for breast cancer diagnosis. Future work will focus on generalizing the framework to clinical datasets and integrating it into diagnostic workflows.

Keywords Breast cancer diagnosis, Deep learning, Generative adversarial networks (GANs), Transfer learning, Accuracy

Breast cancer remains a leading global health concern, with an estimated 2.4 million new cases diagnosed in 2022 alone (World Health Organization [WHO], 2023). Early detection is crucial for improving patient outcomes, with studies demonstrating significantly higher survival rates for cancers detected at earlier stages (American Cancer Society [ACS], 2024). Biopsy, the gold standard for definitive diagnosis[1], involves microscopic examination of tissue samples by pathologists (Henry et al., 2018). However, this visual analysis is inherently subjective and prone to errors due to inter-observer variability and limitations of human perception[2]. These limitations, compounded by the global shortage of pathologists[3,4], pose a significant challenge to accurate and timely breast cancer diagnosis.

Digital pathology uses technology to convert tissue samples into digital images and applies computational algorithms for analysis[2,5,6]. These algorithms aim to detect information that is challenging for the human eye to detect. Despite advancements in the field, accurate diagnosis and treatment is still a challenge due to the need for skilled and experienced pathologists and the potential for misclassification due to human error. Automated classification of breast cancer from histopathological images is a crucial research area in medicine, driven by President Obama's 2015 precision medicine initiatives[7]. The goal is to develop a reliable diagnostic system for breast cancer. Early detection of breast cancer is important as it can aid in disease treatment. Pathology

[1]Department of Mathematics, College of Science and Arts, King Abdul Aziz University, 21911 Rabigh, Saudi Arabia. [2]Department of Artificial Intelligence, FEST Iqra University Karachi, Karachi, Pakistan. [3]Faculty of Computing, Universiti Teknologi Malaysia, Johor Bahru, Johor, Malaysia. [4]College of Business Administration, Management Information System, Northern Border University, Arar, Saudi Arabia. [5]Department of Information Technology, College of Computer, Qassim University, Buraydah 51452, Saudi Arabia. [6]Department of Computer Science, College of Computer and Information Sciences, Majmaah University, 11952 Al-Majmaah, Saudi Arabia. ✉email: engr.lubnaaziz@gmail.com

is an accurate and effective approach for detecting diseases. Computer Aided Diagnostic (CAD) systems use histopathological scans to detect breast cancer. They assist expert pathologists in making a final decision. However, divergent opinions among expert pathologists can lead to diagnostic errors and confusion. One major challenge in computational pathology is developing objective, quantifiable, and reproducible measures for automatic analysis[8]. This requires a large amount of data.

Deep learning architectures, renowned for their automatic feature extraction and representation capabilities, notably CNN-based models like AlexNet, have demonstrated significant improvements, achieving a 6% enhancement over traditional methods in histopathology classification[9–11]. However, training CNNs from scratch is arduous, time-consuming, and demands considerable expertise and data resources, which are often limited in medical domains, including diagnostic pathology[12]. To address this challenge, transfer learning offers a viable solution by leveraging pre-trained CNN models, originally trained on extensive image datasets from other domains, as feature extractors. This approach allows for retraining (fine-tuning) with a limited number of labeled medical images, facilitating effective classification[13]. Transfer learning, widely applied in various computer vision tasks, demonstrates intriguing results where features learned from natural images can be transferred to medical images, despite significant differences between source and target images[12,14].

Insufficient training images pose a challenge in classifier development, prompting the use of image augmentation techniques. Generative Adversarial Networks (GANs), introduced by Goodfellow et al. in 2014, offer state-of-the-art solutions[15,16]. GANs find applications in diverse fields including image translation, object detection, and medical imaging tasks such as MRI reconstruction and image segmentation[17–19]. GANs generate synthetic images while retaining essential features from original images, making them suitable for augmenting training datasets. We refer to original images as "ORG images" and GAN-generated images as "GAN images" throughout this paper.

This research investigates the efficacy of both Generative Adversarial Networks (GANs) for image augmentation and transfer learning to enhance the performance of a multi-scale feature enrichment framework designed for breast cancer detection. The framework demonstrates high accuracy in classifying binary (Benign vs. Malignant) and multi-class (Benign, InSitu, Invasive, Normal) breast microscopy images. Utilizing three pre-trained convolutional neural networks (CNNs)—DenseNet, ResNet 101, and NasNetMobile—the feature extraction process yields complementary representations at various scales, enriched by the proposed multi-scale feature enrichment scheme. At the heart of this scheme lies the *Multiscale Contextual Feature (MSCF)* module, employing dilated convolutions and residual connections to capture rich contextual information across scales. A *Multilayer Feature Fusion Module* seamlessly integrates these features, enhancing discriminative power and tumor localization. To bolster representation robustness, a *Parallel Chain Pooling mechanism* aggregates information from multiple network branches. The enriched features undergo fine-tuning via dense layers, batch normalization layers, and dropout layers, culminating in a SoftMax classification layer. This multi-stage architecture offers precise and robust breast cancer classification, potentially advancing diagnostic capabilities in clinical settings.

This research paper addresses the following research questions and highlights the benefits of utilizing transfer learning and Generative Adversarial Networks (GANs) for breast cancer detection. We investigate their performance, the datasets employed, and their potential in enhancing breast cancer classification and detection.

1. How effective are transfer learning models, pretrained on non-medical domains, for breast cancer classification?

   The effectiveness of transfer learning models pretrained on non-medical domains (e.g., ImageNet) is evaluated when applied to breast cancer classification. This analysis will delve into how transfer learning can leverage knowledge from these extensive datasets to improve predictions in the medical domain, particularly in scenarios with limited available data. Furthermore, the study will examine how transfer learning can reduce computational requirements by utilizing pretrained models, thereby enabling efficient and accurate breast cancer detection with fewer resources.

2. How effective is generative AI for data augmentation in breast cancer detection, and how is augmented data used in prediction during testing?

   The effectiveness of generative AI, particularly GANs, for data augmentation in breast cancer detection will be explored. By generating synthetic data that resembles real medical images, GANs can enhance the diversity and volume of training datasets, thereby improving the robustness of the models. Additionally, the study will discuss how augmented data is incorporated during the testing phase to improve prediction accuracy, detailing the methodologies for integrating synthetic data into the model evaluation process in the medical domain.

3. What are the challenges and advantages of using transfer learning and GANs for breast cancer detection compared to traditional methods?

   The study will compare the use of transfer learning and GANs with classical methods, outlining the challenges and advantages of each approach. This comparison will highlight the improvements in detection accuracy and the potential limitations encountered during implementation.

By addressing these questions, this paper aims to provide a comprehensive evaluation of transfer learning and GANs in the context of breast cancer detection, offering insights into their practical applications and potential benefits in medical diagnostics. This Table 1 summarizes recent research (2024) on deep learning techniques for breast cancer diagnosis. It highlights the techniques used, limitations encountered, and contributions made by each study.

This article is structured as follows: The "Materials and methods" section details the open-source BreakHis and ICIAR datasets, a GAN architecture for synthesizing breast cancer images, and describes three backbone architectures—DenseNet, NasNet Mobile, and ResNet-101—along with their training methodologies and robust feature extraction methods using the multi-scale feature enrichment scheme with fine-tuning. The "Performance assessment and benchmarking" section presents the outcomes of the experiments. The "Discussion" section interprets the experimental findings. Finally, the "Conclusions" section summarizes the key findings of the study.

S/no	Year	References	Technique	Limitation	Contribution
01	2024	[20]	Deep learning algorithms including pre-trained models (VGGNet, SE-ResNet), hybrid CNN architectures (CNN + LSTM, CNN + SVM), cLSFSO for feature extraction, and Grad-CAM for result refinement	**Slow convergence** rates are potential limitations in training time or efficiency of some previous methods	Abstract/Conclusion Mismatch: Binary (normal/suspicious) vs Multi-Class (cancer diagnosis/grading)
02	2024	[21]	**Deep Learning for Breast Cancer Diagnosis**: CNNs for Mammogram Analysis	Suffers from **data hunger, black box decision making, and high computational demands**	
03	2024	[22]	**Transfer learning** with InceptionReNetV2, DenseNet variants (121, 201), VGG19, ResNet models (50, 101v2) fine-tuned for **IDC detection** in breast cancer images	Further validation needed on **larger datasets** before clinical use **Transfer learning limitations** (e.g., pre-trained model relevance) not explicitly discussed	Deep learning aids pathologists (transfer learning), leverages public data for **DenseNet ensemble improvement,** and paves the way for AI-radiologist collaboration
04	2024	[23]	**Deep Learning (CNNs + GANs) for Breast Cancer Diagnosis**: Richer Training (Hybrid, **Transfer Learning**) from DDSM to INbreast	DiaGRAM (CNNs + GANs) promises accurate cancer diagnosis but faces challenges: **complexity, interpretability, and data hunger** (especially for GANs)	Improved Breast Cancer Diagnosis: Achieves **higher accuracy** with less data via **GAN-generated data and transfer learning**
05	2024	[24]	**T-GAN-D (Breast Cancer):** Generates synthetic data (transcriptome images) to train a discriminator for risk prediction (low/high) with focus on robust feature learning	Despite strong performance, consider incorporating feature selection for accuracy gains, exploring **biologically-relevant data structures**, and integrating multi-omics data for a more comprehensive analysis	Improved Risk Prediction (Outperforms Existing Methods) with Robustness (Independent Datasets, Combined Sources) and **Data Augmentation** (Synthetic Data)
06	2024	[25]	**GOA:** Feature selection, CNN image processing, SVM classification for breast cancer detection	High accuracy (96%) needs real-world validation, computationally expensive, and potential CNN limitations (overfitting) remain **unclear**	Multi-technique **(GOA, CNN, SVM)** for breast cancer detection: Feature selection (reduced cost, improved effectiveness) achieves promising results on simulated data
07	2024	[26]	WSL-based ROI refinement (D-Net & C-Net) for breast cancer detection with dynamic ROI size and reduced annotation burden		This work advances breast cancer diagnosis with superior performance, reduced annotation burden via WSL, iterative ROI refinement, and potential for clinical practicality
08	2024	[27]	This work utilizes DenseNet (efficient) with Attention Mechanism for pathology image analysis, leveraging **transfer learning** for performance gains	**Limit clinical** use (grading, subtypes) due to model size and interpretability needs despite high accuracy (benign/malignant)	**Transfer learning** tackles small datasets (medical) & deep learning automates feature design, aiming for **improved clinical accuracy** with future interpretability focus
09	2024	[28]	Deep learning with DCGAN-generated mammograms (validated statistically and by experts) for **data augmentation** in breast cancer diagnosis	DCGANs create **realistic mammograms** (normal, benign, malignant) for data augmentation, validated statistically and by experts (future: broader testing & GAN exploration)	DCGANs **address limited mammogram data** (accuracy bottleneck) by generating synthetic images (normal, benign, malignant) for data augmentation, aiming to improve model performance
10	2024	[29]	Four-region image analysis with **transfer learning** (ResNet50, VGG16) for feature extraction and ensemble classification (Extra Trees, Logistic Regression, SVM, Ridge) for robust breast cancer diagnosis	Require more investigation into data, model specifics, and interpretability for robust breast cancer diagnosis	Deep feature extraction + interpretable machine learning for early breast cancer detection and treatment planning
11	2024	[30]	**Transfer learning** (VGG-16, ResNet-50, EfficientNet-B7) with **test-time augmentation** boosts mammogram analysis	Limited to **static images** (not time series), but promising for broader breast imaging and multi-class abnormality classification (future work: time series analysis)	TL with TTA achieves high accuracy (> 99%), works well with limited medical data, and aligns with radiologist expertise for breast cancer diagnosis
12	2024	[31]	**Transfer learning** (ResNet50) improves breast cancer detection in histopathological images, achieving high accuracy (92.2%) and recall (95.7%)	**Deep learning for breast cancer diagnosis**: powerful but data-hungry, prone to overfitting, and can be a black box (**interpretability issues**)	**Transfer learning** with ResNet50 shows promise for improved breast cancer detection accuracy
13	2019	[32]	Deep Learning for Breast Cancer: GANs create synthetic mammograms for better CNN training (up to 98.85% accuracy), while transfer learning (VGG-16) achieves efficient classification (91.48%)		Deep Learning for Breast Cancer: GANs augment data (↑ accuracy), VGG-16 transfer learning efficient (91.5%). Combined approach is 10× faster but less accurate. GANs excel for data augmentation, transfer learning for efficiency

Table 1. Comparison of deep learning techniques for breast cancer diagnosis.

Materials and methods

This study proposes a novel framework for classifying breast cancer histopathology (BCH) images, combining synthetic image generation and multi-scale transfer learning. A conditional Wasserstein GAN (cWGAN) is used to generate synthetic breast cancer images to augment the training dataset. These synthetic images, along with the original images, are then further augmented using traditional techniques (e.g., rotation, flipping, cropping) to increase diversity and improve model robustness. Multi-scale transfer learning models, along with a multi-scale feature enrichment scheme, are employed to capture a more comprehensive representation of the images. The framework processes BreakHis and ICIAR datasets, applying preprocessing techniques and data augmentation to enhance training. Multi-scale feature extraction is refined through individual training on these datasets. The effectiveness of the framework is evaluated on test images, with subsequent subsections exploring its components in detail. Figure 1 illustrates the proposed process flow, detailing the acquisition and characterization of the dataset, the preprocessing procedures applied, the methods for data augmentation, the approach to data partitioning, the classifier architecture utilized, and the evaluation metrics employed.

Dataset

This study utilizes two publicly accessible datasets: BreakHis and ICIAR 2018. The BreakHis dataset consists of 7909 microscopic breast tumor images from 82 patients, offering varying magnification levels (40×, 100×, 200×, and 400×) and uniform resolution (700 × 460 pixels and an 8-bit depth in RGB in PNG format). BreakHis encompasses 2480 benign and 5429 malignant samples, mirroring the actual occurrence of breast cancer. Figure 2 presents illustrative BreakHis images, while Fig. 3 highlights representative histopathology images from both BreakHis and ICIAR datasets. While both class labels (benign and malignant) were further categorized into different types and patient IDs, for the purposes of this research, both factors were disregarded, and only the class labels were taken into consideration[33].

To ensure the data could be adequately trained and tested in the model developed for this research, several modifications were made to the original dataset. To prepare the BreakHis dataset for training, images were resized to 512 × 512 pixels to align with the cWGAN model's requirements. To address the class imbalance, a conditional Wasserstein GAN (cWGAN) was employed to generate synthetic images of the minority class (benign). This technique helps to ensure that the model learns effectively from the dataset and avoids bias towards the majority class. To further assess the framework, the ICIAR 2018 Grand Challenge dataset was utilized. This dataset consists of 400 H&E-stained breast histology microscopic images, categorized into four groups (normal, in situ carcinoma, benign, and invasive carcinoma). ICIAR offers a diverse set of 100 images per group, complementing BreakHis by including pre-cancerous and normal tissues. Tables 2 and 3 provide a comparative overview of the BreakHis and ICIAR datasets, highlighting their characteristics, distribution of breast cancer subtypes, and magnification levels, offering insights into the data preparation and classification challenges addressed in this study.

Proposed two-stage data augmentation: cWGAN and traditional augmentation for synthesizing breast cancer images

To address the class imbalance in the dataset, this study proposes a Conditional Generative Adversarial Network (cGAN) with Wasserstein GAN (WGAN) loss to generate synthetic breast cancer histopathological images.

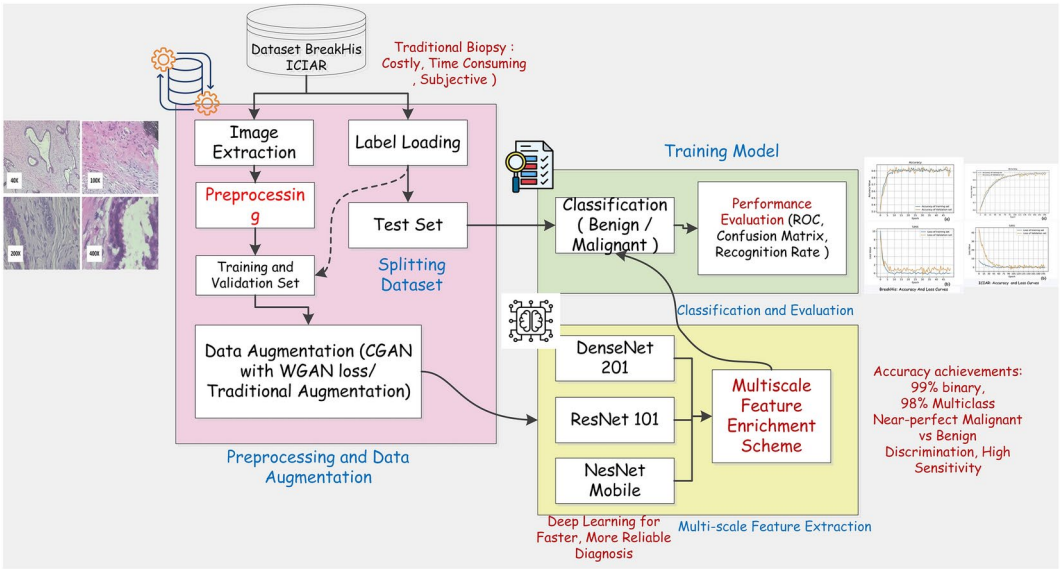


Fig. 1. Proposed framework for classifying microscopic breast images. The workflow includes preprocessing with CGAN-based data augmentation, multiscale feature extraction using DenseNet-201, ResNet-101, and NesNet Mobile, and classification into benign or malignant categories. The model achieves high accuracy (99% binary, 98% multiclass) with robust diagnostic performance, outperforming traditional biopsy methods.

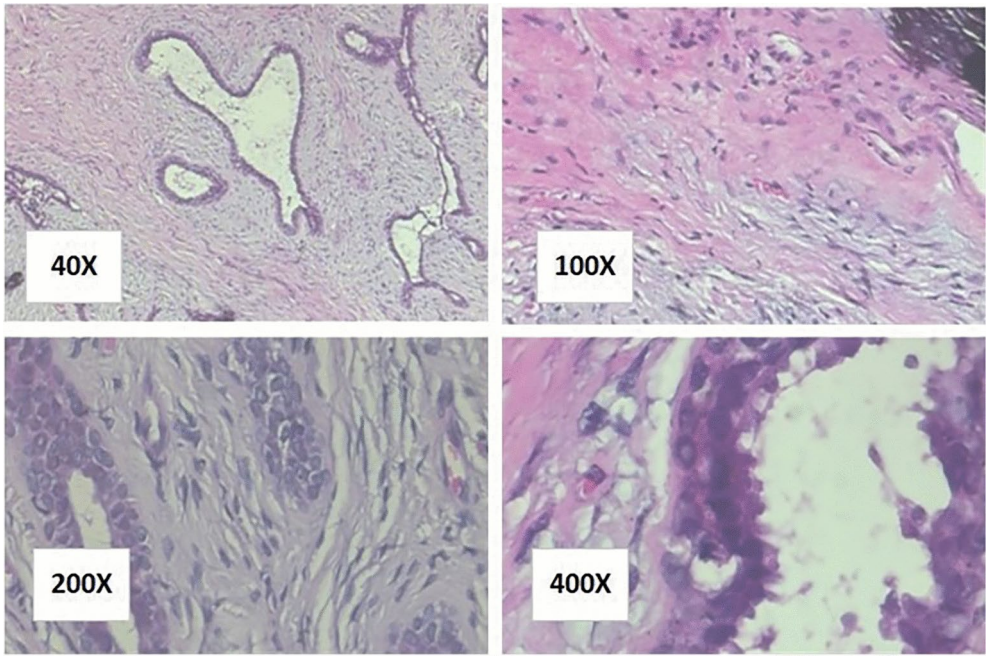


Fig. 2. Histopathology images from the BreakHis dataset at magnifications 40×, 100×, 200×, and 400×.

The cGAN's ability to condition image generation on cancer type enhances training data for classification models. WGAN loss addresses training instability issues, leading to smoother training and more realistic synthetic images. Before training, meticulous preprocessing was conducted, including resampling and format conversion. Data shuffling and partitioning mitigated biases. The cGAN architecture comprises a generator and a discriminator, with the generator conditioned on a latent vector and cancer type. A WGAN with gradient penalty was employed to address mode collapse. Efficient training techniques contributed to the generation of realistic and diverse synthetic breast cancer images. Generative models like GANs can generate new and diverse samples from the training data distribution. This research focuses on generating 640 synthetic benign images to enhance the overall balance of the dataset. However, the effectiveness of GAN-based data augmentation depends on the quality of the GAN model and the specific traits of the training data.

The proposed framework centers on a cWGAN architecture, comprising a generator (G) and a discriminator (D). The generator, conditioned on a latent vector and cancer type, produces synthetic images with targeted characteristics relevant to different cancer subtypes. The discriminator evaluates image authenticity. A WGAN with gradient penalty was employed to address mode collapse. Efficient training strategies included a shallow network architecture, hyperparameter optimization, predefined epochs, preliminary discriminator-only training, and a final convolution layer for improved image quality. The trained cGAN discriminator can also be repurposed as an independent classifier. By utilizing the generated expression profiles (with missing values addressed through imputation), the discriminator can effectively distinguish between benign and malignant breast cancer. This demonstrates the cGAN's potential to generate realistic and diverse synthetic images, which are valuable resources for training classification models and enhancing breast cancer diagnosis.

Generator (G): This network utilizes a deep convolutional neural network architecture, similar to U-Net. It takes two inputs: a noise vector (z) for initial randomness and a condition label (y) representing the specific cancer type. The noise and label are processed through convolutional layers with activation functions like $LeakyReLU$. Upsampling layers then achieve the desired image resolution, and the final layer utilizes a tanh activation to generate pixel values within the range of – 1 to 1.

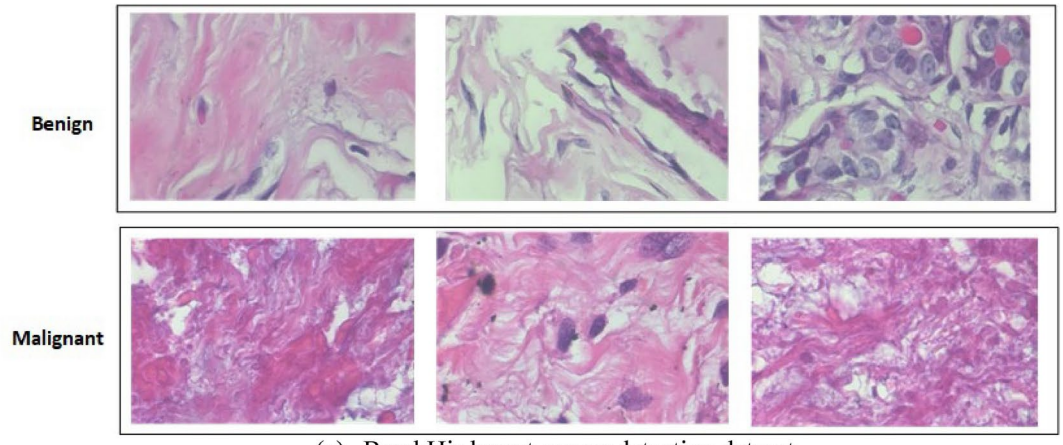
Discriminator (D): This network also employs a deep convolutional architecture. It takes an input image (either real or generated) and processes it through convolutional layers with activation functions. Finally, the discriminator outputs a single value between 0 and 1, representing the probability of the input image being real.

Loss functions in cWGAN for synthetic breast cancer image generation

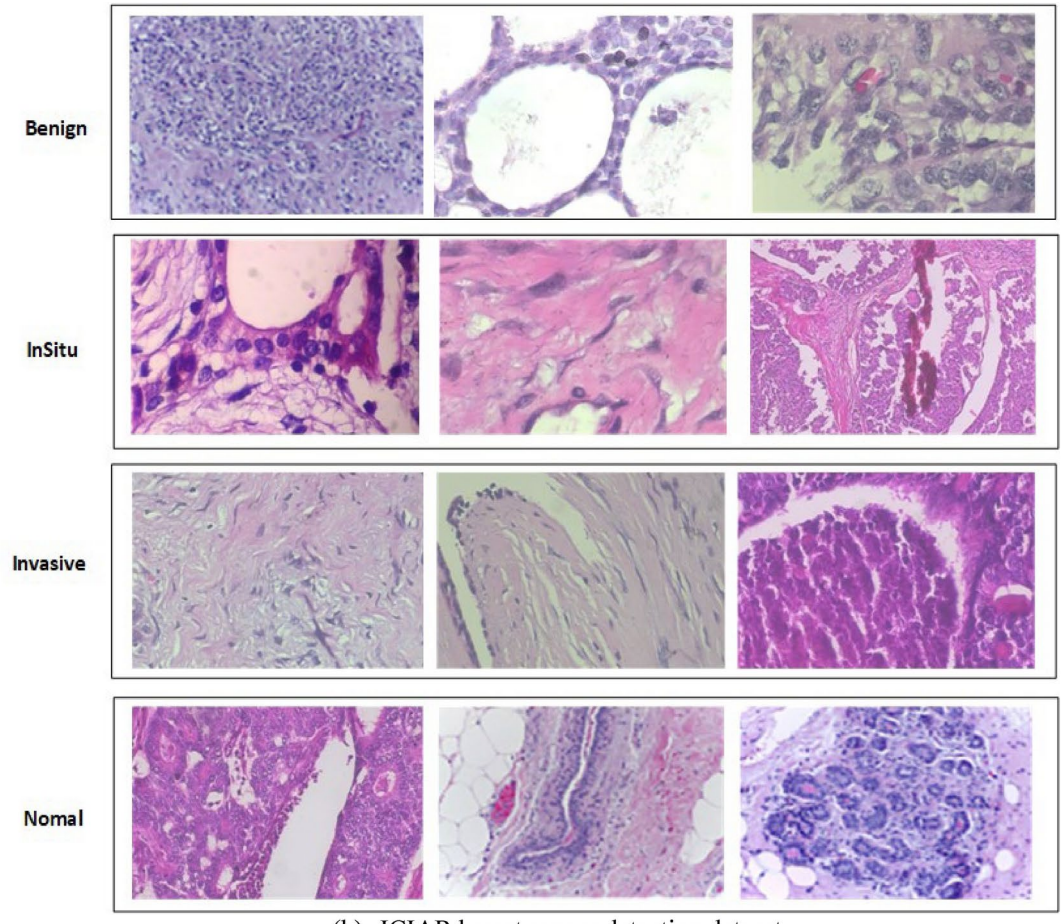
Conditional Generative Adversarial Networks (cGANs) with Wasserstein GAN (WGAN) loss function have emerged as a powerful approach for generating synthetic breast cancer histopathological images. This section delves into the two primary loss functions employed within this framework:

(a) Wasserstein Loss (Critic Loss):

The Wasserstein loss function[34], also known as the critic loss, plays a pivotal role in training the discriminator network (D). Unlike traditional GAN loss functions, Wasserstein loss penalizes the discriminator for large estimation errors, promoting more stable training. The Wasserstein distance can be visualized as the minimum effort required to transform one distribution (real data) into another (generated data). This intuitive interpretation underscores its effectiveness in guiding the discriminator towards improved performance.



(a) BreakHis breast cancer detection dataset



(b) ICIAR breast cancer detection dataset

Fig. 3. Representative histopathology images from the BreakHis and ICIAR datasets.

Mathematically, the Wasserstein loss calculation can be complex. However, a simplified representation captures the core concept:

Wasserstein Loss Formula:

$$W(P,Q) = [\![inf]\!]\_\gamma sup\,||f(x) - f(y)||\_1 \tag{1}$$
$$\gamma \in \Gamma(P,Q)$$
$$f \in C(X,R)$$

where:

Feature	BreakHis Dataset	ICIAR Dataset
Purpose	Breast cancer histopathology image classification	Breast tissue classification (normal, benign, in situ, invasive)
Image Types	Microscopic biopsy images	Histopathology images
Number of Images	~7900	~400
Magnification Levels	40×, 100×, 200×, 400×	Single magnification
Classes and Splits	Benign: 1924 (train), 481 (val), 75 (test) Malignant: 4274 (train), 1070 (val), 85 (test)	Normal: 70 (train), 20 (val), 10 (test) In Situ Carcinoma: 70 (train), 20 (val), 10 (test) Benign: 70 (train), 20 (val), 10 (test) Invasive Carcinoma: 70 (train), 20 (val), 10 (test)
Image Resolution	700 × 460 pixels	Varies (high resolution)
Format	RGB (color)	RGB (color)
Source	Collected from patients with biopsy-confirmed breast cancer	Collected from various clinical samples
Availability	Publicly available	Publicly available
Challenges	Variability in magnification and staining techniques	Imbalance in class distribution, high inter-class similarity

Table 2. Comparative overview of BreakHis and ICIAR datasets for breast cancer histopathology image classification.

Classes of breast cancer	Subtype of breast cancer	Magnification levels				Total
		40×	100×	200×	400×	
Benign	Adenosis (A)	114	113	111	106	444
	Fibroadenoma (F)	253	260	264	237	1014
	Phyllodes tumor (PT)	149	150	140	130	569
	Tubular Adenoma (TA)	109	121	108	115	453
Malignant	Ductal carcinoma (DC)	864	903	896	788	3451
	Lobular Carcinoma (LC)	156	170	163	137	626
	Mucinous Carcinoma (MC)	205	222	196	169	792
	Papillary Carcinoma (PC)	145	142	135	138	560
Total		1995	2081	2013	1820	7909

Table 3. Distribution of breast cancer subtypes and magnification levels in BreakHis dataset.

- W(P,Q): Represents the Wasserstein Distance between distributions P (real data) and Q (generated data).
- $[\![inf]\!]\_\gamma$: Represents the infimum (minimum value) taken over all possible transport maps (γ) between P and Q. These maps define how probability mass can be moved from one distribution to another.
- $\Gamma(P, Q)$: Represents the set of all transport maps that can move probability mass from P to Q.
- $sup||f(x) - f(y)||\_1$: Represents the supremum (maximum value) taken over all possible functions (f) that map points from the domain (X) of the data to real numbers (R). Here, $||.||\_1$ denotes the L1 norm (absolute difference).
- f(x): Represents a function that maps points from the domain (X) of the data to another space.

By minimizing this cost across all possible transport maps, the Wasserstein Loss aims to find the most efficient way to transform the real data distribution (P) into the generated data distribution (Q). This, in turn, guides the discriminator to become more adept at distinguishing between real and synthetic images.

(b) Conditional Hinge Loss (Generator Loss):

The conditional hinge loss function, also referred to as the generator loss, plays a critical role in guiding the generator network (G) towards its objective[2]. In essence, it encourages the generator to create images that can deceive the discriminator network (D) while also accurately reflecting the provided condition label. This loss function penalizes the generator for generating images that fall into two categories:

Unrealistic: The generated image appears artificial or lacks the characteristics of a real histopathological image.

Do not reflect the condition label: The generated image might be realistic but doesn't exhibit the specific cancer features associated with the provided condition label (e.g., ductal carcinoma in situ).

While the exact formula for conditional hinge loss can be intricate, it essentially calculates the difference between the desired output (a generated image that deceives D and reflects the condition) and the actual output of the generator. This difference is then used to update the generator's weights during the training process, pushing it to create more realistic and condition-specific images. The conditional hinge loss ensures the generated images are not only realistic but also adhere to the provided information about the specific cancer type.

Mathematical Formulation:

$$L\_c(G) = max(0,1 - D(G(z,y))) + \lambda * L\_feat(G(z,y),x) \quad (2)$$

where:

- L_c(G): Represents the conditional hinge loss for the generator (G).
- G(z,y): Denotes the generated image produced by G, taking a noise vector (z) and a condition label (y) as inputs.
- D(G(z,y)): Represents the discriminator's output for the generated image, indicating the probability of it being real. (Ideally, this value should be close to 1).
- max(0, …): Represents the hinge function, which only penalizes the generator when the term inside the parenthesis is positive.
- λ: Represents a hyperparameter that controls the weight of the feature loss term (L_feat).
- $L\_feat(G(z,y),x)$: Represents an optional feature loss function. This term can be used to encourage the generated image (G(z, y)) to be more similar to real images (x) in terms of specific features, beyond just fooling the discriminator.

The first part $(max(0,1 - D(G(z,y))))$ ensures the generator is penalized only when the discriminator (D) assigns a low probability (less than 1) to the generated image. This pushes the generator to create images that D classifies as real. The second part $(\lambda * L\_feat(G(z,y),x))$ is an optional term that incorporates a feature loss function (L_feat). This term helps the generated images not only appear realistic to the discriminator but also resemble real images (x) in terms of specific visual features captured by L_feat. The hyperparameter λ controls the influence of this term compared to the first part.

Geometric augmentations, such as reflecting, cropping, and translating, are common and acceptable approaches for image data augmentation. However, traditional data augmentation techniques, often dependent on fundamental geometric transformations, exhibit limitations. Although they can expand the training dataset, enhancements are necessary to enhance their capacity for generating diverse and representative samples reflective of the underlying data distribution. This research involves labeled data comprising benign and malignant samples, as outlined in Table 4, which will be evaluated with and without traditional augmentation methods. Specifically, the images will be applied with a random transformation of vertical and horizontal flipping, rotation, and zoom during training. The random data augmentation techniques and their parameter values utilized in this study are detailed in Table 5.

Multi-scale transfer learning technique for breast cancer classification

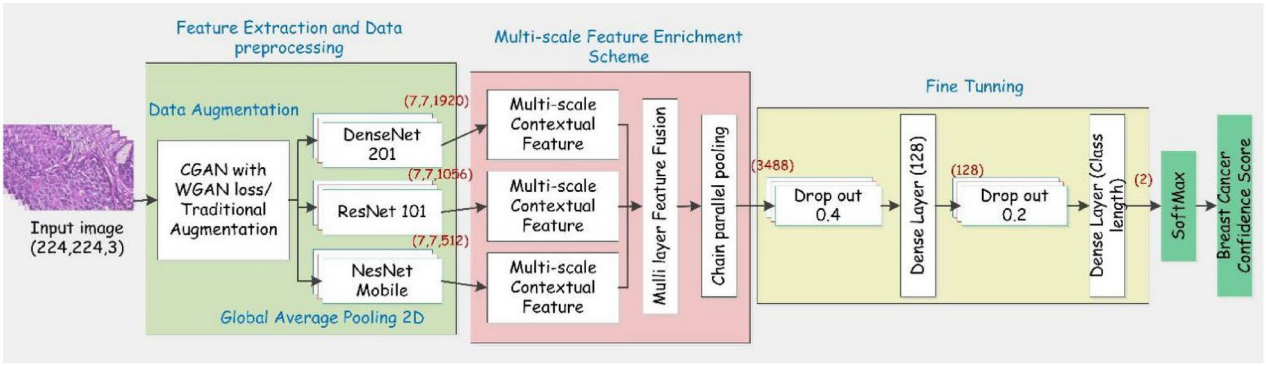
To enhance the classification performance of CNN models, this study integrated synthetic BCH images generated by the cWGAN model with the original training data. A novel multi-scale transfer learning approach was introduced, leveraging the complementary strengths of three pre-trained CNNs (DenseNet-201, NasNetMobile, ResNet 101) to extract low-level features from microscopic images. A meticulously designed

Dataset	Breast cancer type	Training	Validation	Testing
BreakHis	Benign	13,468	3367	75
	Malignant	29,918	7490	85
ICIAR	Normal	490	140	10
	InSitu Carcinoma	490	140	10
	Benign	490	140	10
	Invasive Carcinoma	490	140	10

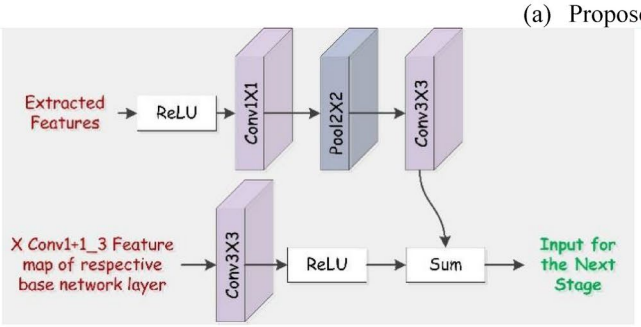
Table 4. Post-augmentation BreakHis and ICIAR dataset distribution.

S/no	Data augmentation techniques	Parameter value
1	Zooming range	2
2	Rotation range	95
3	Shearing range	0.5
4	Width shift range	0.4
5	Height shift range	0.4
6	Horizontal flip	True
7	Vertical file	True

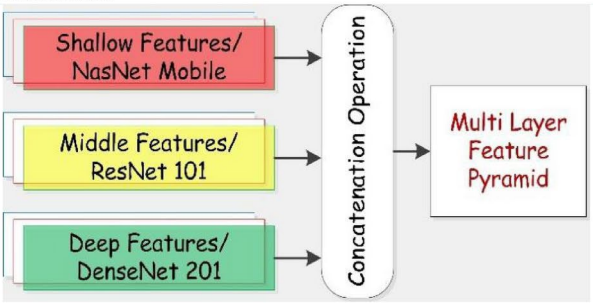
Table 5. Random data augmentation techniques employed to mitigate overfitting and improve model generalization.



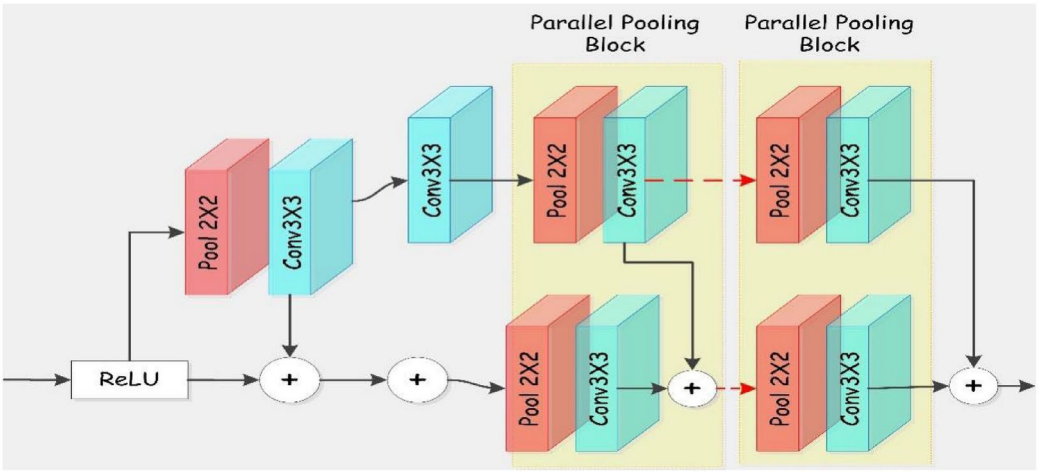
(a) Proposed framework



(b) Multiscale Contextual Feature module



(c) Multi-Layer Feature Fusion module



(d) Chain parallel pooling

Fig. 4. Proposed framework for breast cancer detection using advanced multiscale feature processing and fusion. (**a**) The framework integrates data augmentation, multiscale feature extraction using DenseNet-201, ResNet-101, and NesNet Mobile, and fine-tuning for precise tumor classification. (**b**) The Multiscale Contextual Feature module leverages residual blocks to capture contextual information effectively. (**c**) The Multi-Layer Feature Fusion module concatenates low-, middle-, and high-level features into a feature pyramid, ensuring feature balancing and enhancing detection accuracy. (**d**) The Chain Parallel Pooling module employs parallel pooling blocks to maintain feature complexity and robustness while improving the framework's efficiency. This AI-powered framework enhances tumor detection with improved sensitivity and robustness (Details in[35]).

fusion scheme combined these features, creating a comprehensive representation that captured critical image details. This enhanced representation, with a significantly expanded parameter count, empowered the ensemble to capture intricate patterns, surpassing the individual CNNs in classification performance. This demonstrated the effectiveness of our multi-scale feature fusion approach in leveraging transfer learning for improved breast cancer detection (illustrated in Fig. 4a).

Multiple pre-trained deep learning models are incorporated for breast cancer classification as illustrated in Fig. 4. Each model individually extracts features, followed by GlobalAveragePooling2D to flatten each layer into a vector by averaging its input channels. These vectors are subsequently integrated into a feature pyramid

via the novel multi-scale feature enrichment scheme[35]. Fine-tuning of the combined features for multi-class classification is achieved through six hidden layers, culminating in a softmax activation function.

To mitigate overfitting, two dropout layers are implemented[36]. These layers randomly deactivate 40% and 20% of neurons during training, respectively, not only combating overfitting but also accelerating the training process[36]. Additionally, two batch normalization layers are employed to rescale and smoothing of data, further accelerating training and reducing sensitivity to network initialization.

Dense layers, characterized by dense connections between all neurons in adjacent layers, operate on input data and produce output. Our architecture utilizes two dense layers, with the final layer performing classification followed by a softmax activation function. This layer, considering the length of the predicted class, generates a prediction by determining the most correlated features. The softmax activation function assigns probabilities between 0 and 1 to each possible outcome, triggering the neuron with the highest probability. It is mathematically defined as:

$$softmax(x)_i = \frac{exp(x_i)}{\sum_j exp(x_j)} \tag{3}$$

Table 6 showcases the key architecture elements of the proposed framework for binary classification. The final layer comprises two neurons, reflecting the dual-class output. For multiclass scenarios, this layer is readily adaptable, accommodating four neurons to cater to the expanded output categories. While the final layer adapts, the rest of the proposed framework remains consistent for both binary and multiclass classification tasks.

Separate CNN models will be trained with and without using the generated synthetic images to conduct a comparative evaluation and quantify the ramifications exerted by the synthesized image data. This approach allows for a comprehensive assessment of the generated data's effectiveness in enhancing BCH classification performance. The research will use both models, and their performances will be recorded.

The proposed multi-scale feature enrichment scheme seamlessly integrates low-level features with mid-level and high-level representations, culminating in a robust multi-level feature pyramid for breast cancer classification. This scheme leverages pre-trained models like ResNet 101, NasNetMobile, and DenseNet201 to extract base features rich in multi-level information. These features serve as the foundation for subsequent enrichment within the Multi-scale Contextual Feature Module (MSCF) stack[35] (illustrated in Fig. 4b). Each MSCF module progressively enriches existing representations by generating a feature map at a specific scale through a series of carefully orchestrated merges. Subsequent modules then progressively introduce higher-level information by merging:

(1) Shallow (low-level) features produced from the NasNetMobile base feature map.
(2) Mid-level features formed by adding the ResNet 101 backbone output to the second MSCF block.
(3) High-level features derived from DenseNet.

This multi-level fusion culminates in a robust feature pyramid through scale-wise concatenation (illustrated in Fig. 4c). To further enhance context integration, parallel chain pooling is applied (illustrated in Fig. 4d), capturing broader contextual information from larger image regions. Through this multi-pronged approach,

Layer (types)	Output Shape	Parameter number	Connected to
Data Generated through cGAN with WGAN loss function			
Input_1 (InputLayer)	(None, 224, 224, 3)	0	
Densenet201	(None, 7, 7, 1920)	18,321,984	input_1[0][0]
NasNetMobile	(None, 7, 7, 1056)	4,269,716	input_1[0][0]
ResNet 101	(None, 7, 7, 512)	14,714,688	input_1[0][0]
Global_average_pooling2d	(None, 1920)	0	densenet201[0][0]
Global_average_pooling2d_1	(None, 1056)	0	NasNetMobile[0][0]
Global_average_pooling2d_2	(None, 512)	0	ResNet 101[0][0]
MSFF	(None, 3488)	15,986	MSCF_1, MSCF_2, MSCF_3
CPP	(None, 2956)	11,952	MSFF
dropout (Dropout)20%	(None, 2956)	0	CPP[0][0]
batch_normalization (BatchNormalization)	(None, 2956)	11,952	dropout [0][0]
dense (Dense)	(None, 128)	446,592	batch_normalization [0][0]
dropout_1 (Dropout)40%	(None, 128)	0	dense [0][0]
batch_normalization_1 (BatchNormalization)	(None, 128)	512	dropout_1[0][0]
dense_1 (Dense)	(None, 2)	258	batch_normalization_1[0][0]
Total Params: 37,793,640			
Trainable Params: 37,494,676			
Non-trainable params: 298,964			

Table 6. Detail of proposed framework architecture: Integration of Transfer Learning Models in multi-scale feature enrichment scheme for binary and multiclass classification.

this scheme effectively integrates information across various scales and levels, ultimately enabling the superior performance observed in our breast cancer classification model.

Inputs:

- Training set: $\mathbf{T} = ((x_1, y_1), (x_2, y_2), ..., (x_n, y_n))$, where x_i is a microscopy image and y_i is the corresponding class label (benign or malignant).
- Validation set: **V**
- Testing set: $\mathscr{H}$
- Hyperparameters:
  - Learning rate α
  - Number of epochs for CNN training $\mathscr{E}$
  - Batch size for CNN training $\mathscr{B}$
  - Number of images converted in one batch size
  - Number of epochs for cWGAN training (Ec)

Outputs:

- Trained convolutional neural network (CNN) model weights θ

Procedure:

1. Image Preprocessing:

- Convert each microscopy image in **T** into a suitable format (Normalized pixel values).

2. Data Augmentation with cWGAN:

- Train a cWGAN with Wasserstein loss function (Ec Epochs).
  - Input: Class labels from **T**.
  - Output: Synthetic breast cancer microscopy images conditioned on specific cancer types.
- Combine real images from **T** with the generated synthetic images from cWGAN to create an augmented training set (**T'**).

3. Traditional Data Augmentation (Optional):

- Apply additional data augmentation techniques (random cropping, flipping, rotations) to **T'** to further increase dataset size and diversity.

4. Feature Extraction:

- Employ three pre-trained CNN models (DenseNet-201, NasNetMobile, ResNet 101) for feature extraction.
- Extract features from each model's intermediate layers, capturing multi-scale representations.

5. Multi-layer Feature Fusion Module:

- Concatenate extracted features from different scales to create a multi-level feature pyramid.
- Employ parallel chain pooling to aggregate global contextual information across feature levels.

6. Fine-Tuning:

- Designate specific layers in the pre-trained models for fine-tuning (θ^1, θ^2, θ^3, θ^4).
- Initialize model parameters θ using pre-trained weights.

7. Training:

- Iterate through $\mathscr{E}$ epochs:
  - For each epoch, divide **T'** into mini-batches of size $\mathscr{B}$.
  - For each mini-batch:
    - Perform forward propagation through the model to obtain predicted outputs.
    - Calculate the loss function (cross-entropy loss).
    - Backpropagate gradients through the model to update weights θ using an optimizer (Adam).

8. Evaluation:

- Assess model performance on **V** using appropriate metrics (accuracy, precision, recall, F1-score).

Algorithm 1. Automated Breast Cancer Detection and Classification with cWGAN Data Augmentation

Performance assessment and benchmarking of GAN and CNN model

To assess the performance of the developed GAN and CNN models in the context of BCH image analysis, a comprehensive evaluation process was conducted. Both models were trained on the same dataset, and their performance was evaluated using metrics tailored to their respective architectures.

For the GAN architecture, metrics such as Fréchet Inception Distance (FID) and Kernel-Inception Distance (KID) were prioritized. These metrics quantify the quality of synthesized image data by measuring the similarity between generated and authentic samples. A lower FID or KID score indicates a more accurate representation of the underlying data distribution.

This evaluation section delves into the experimental setup, hyperparameters, and results obtained using the proposed framework with the BreakHis and ICIAR datasets. A comparative analysis highlights the substantial performance gains achieved through the synergistic combination of transfer learning and GAN-generated data, demonstrating the superiority of the proposed framework over individual pre-trained CNN architectures.

The Fréchet Inception Distance (FID) is a metric employed to assess the quality of image generative models. Derived from feature representations extracted by a pre-trained multi-scale transfer learning model, FID quantifies the discrepancy between authentic and synthesized image samples. Assuming a Gaussian distribution for these features, the FID calculates the Fréchet distance between two multivariate Gaussian distributions representing the real and generated data. A lower FID score indicates a more proficient generative model in terms of output quality.

To compute the Fréchet Inception Distance (FID), feature representations are extracted from real and generated images using the coding layer of the Multi-scale transfer learning model[37]. These activations are aggregated into multivariate Gaussian distributions by calculating the mean and covariance. The FID is then determined by calculating the Fréchet distance between these two distributions.

$$FID = \|\mu_r - \mu_g\|^2 + Tr\left(\sum c_r + c_g - 2\left(\sqrt{c_r}\sqrt{c_g}\right)\right)$$

where:

- μ_r is the mean of the real data distribution
- μ_g is the mean of the generated data distribution
- c_r and c_g are the covariance matrices of the real and generated data distributions, respectively
- Tr denotes the trace operator

The Kernel Inception Distance (KID) is an alternative to the Fréchet Inception Distance (FID) for evaluating image quality. Both metrics compare the representation spaces of a pre-trained Inception-V3 network on ImageNet, assessing the disparity between generated and training distributions. KID offers computational advantages over FID, as it is more straightforward to implement, can be estimated per batch, and is computationally lighter. Using a polynomial kernel, KID measures the squared Mean Discrepancy (MMD) between the Inception representations of actual and generated samples, effectively quantifying the distance between probability distributions[38].

Performance evaluation was divided into three sections: comparing the proposed model with benchmark models using transfer learning alone and those models whose performance was enhanced through GAN integration. The results consistently demonstrate the superiority of our proposed framework, outperforming benchmark models regardless of whether they solely employ transfer learning or GANs. This highlights the synergistic benefits of combining transfer learning and GAN-generated data in our approach.

The framework was implemented using Keras, a flexible open-source Python library for deep learning. Leveraging Keras's seamless integration with TensorFlow, neural network construction and execution were facilitated. Training and testing were conducted on an NVIDIA GeForce RTX 2060 graphics card equipped with CUDA 10.1 and cuDNN 7.5.0, providing sufficient computational power.

Performance metrics

To comprehensively evaluate the effectiveness of our proposed framework, we employed a diverse set of statistical measures derived from the confusion matrix. These metrics include:

Precision: Represents the proportion of correctly identified positive cases (TP) among all cases predicted as positive (TP + FP). *Recall*: Measures the proportion of correctly identified positive cases (TP) among all actual positive cases (TP + FN). *False-positive rate (FPR):* Captures the rate of incorrectly classified negative cases (FP) among all actual negative cases (TN + FP). *True negative rate (TNR):* Indicates the rate of correctly classified negative cases (TN) among all actual negative cases (TN + FN). *F1-Score:* Combines precision and recall into a single metric, providing a balanced assessment of model performance. *Matthews correlation coefficient (MCC):* A robust measure of classification accuracy that accounts for both true positives and true negatives, making it valuable for imbalanced datasets. *Kappa Statistic:* Assesses the agreement between the model's predictions and the actual class labels, correcting for chance agreement, particularly useful for small datasets. Each metric is clearly defined in terms of the confusion matrix components:

1. *True Positive (TP):* Correctly classified positive case.
2. *True Negative (TN):* Correctly classified negative case.
3. *False Positive (FP):* Incorrectly classified positive case (healthy tissue classified as cancerous).
4. *False Negative (FN):* Incorrectly classified negative case (cancerous tissue classified as healthy).

All the statistical parameters are expressed as follows:

Parameter	Set values
Training data	2480 Benign Images
Model	cWGAN
Number of GPU	1 / RTX 0260
Generator learning rate	0.0025
Discriminator learning rate	0.0025
Image size (input)	512×512
Image size (output)	512×512
Gamma γ	0.5, 1.0, 1.5, 1.55, 2.0, 5.0
Batch	16, 32, 64, 128

Table 7. Parameter setting of cWGAN.

Image processed (kimg)	Gamma, γ	FID score	KID score
240	0.5	21.1684	0.0049
300	1.0	16.1145	0.0030
300	1.5	17.5853	0.0031
300	1.55	18.4753	0.0041
300	2.0	18.7472	0.0047
120	5.0	17.7404	0.0033
320	6.55	18.0501	0.0039

Table 8. Performance evaluation of cWGAN on gamma values.

$$pricision = \frac{TP}{TP + FP} \tag{4}$$

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

$$FPR = \frac{FP}{FP + TN} \tag{6}$$

$$TNR = \frac{TN}{TN + FP} \tag{7}$$

$$F1 - score = 2 \times \frac{Pr \times Re}{Pr + Re} \tag{8}$$

$$kappa = \frac{TotalAcciracy - RandomAccuracy}{1 - RandomAccuracy} \tag{9}$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FP)(TN + FP)(TN + FN)}} \tag{10}$$

By utilizing these diverse metrics, we gain a multifaceted understanding of our framework's strengths and weaknesses, enabling us to draw comprehensive conclusions regarding its performance in breast cancer detection and classification.

Experimental results

Performance of GAN

To generate high-quality synthetic images of benign BCH, a conditional Wasserstein GAN (cWGAN) model was employed. The model's performance was evaluated using the Fréchet Inception Distance (FID) and Kernel Inception Distance (KID) metrics. Table 7 outlines the experimental parameters, including the number of training epochs ($kimg$) and hyperparameters such as gamma and batch size. These parameters were systematically adjusted to optimize the model's performance on the given dataset.

The FID and KID scores obtained for various gamma values and batch sizes are presented in Tables 8 and 9, respectively. The results indicate that a gamma value of 1.0 and a batch size of 32 consistently yielded the lowest scores, suggesting optimal model performance.

While further training could potentially improve the generated images, the computational cost and time required might not justify the marginal gains. The generated images, as depicted in Fig. 5, exhibit exceptional quality and visual fidelity, making them highly suitable for data augmentation tasks, particularly in scenarios where obtaining large, manually labeled datasets is challenging.

Image processed (kimg)	Batch size, B_{size}	FID score	KID score
620	16	17.1803	0.0040
600	32	15.9542	0.0039
644	64	16.4828	0.0037
690	128	18.6250	0.0052

Table 9. Results of cWGAN training for various batch size.

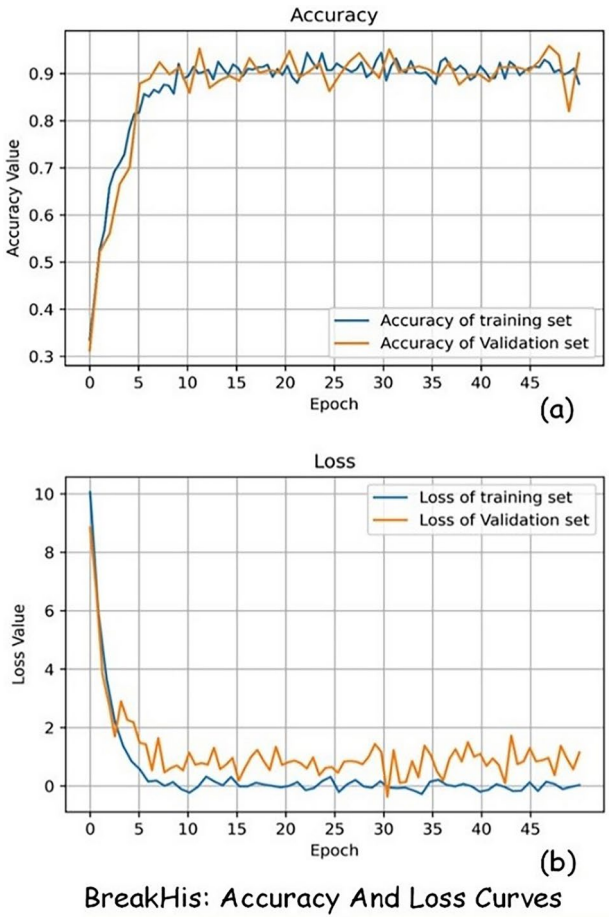


Fig. 5. BreakHis Training: Accuracy (**a**) steadily climbs, reaching 0.976 while validation stabilizes. Loss curves (**b**) decrease continuously, confirming convergence and optimal performance.

The cWGAN-generated images' diversity and realism make them valuable assets for expanding training datasets and enhancing the performance of downstream models, such as convolutional neural networks (CNNs).

Performance of GAN with multi-scale transfer learning

To harness the knowledge embedded in three pre-trained CNN architectures—DenseNet-201, NasNetMobile, and ResNet-101—additional layers were introduced on top of these models, referred to as the base models. The weights of the base model layers were frozen to preserve the lower-level image features they had learned. Training focused on fine-tuning the newly added binary classification layers using the BreakHis and ICIAR datasets, enabling the model to adjust its parameters to specific dataset characteristics while leveraging the general features from the base layers. The parameters and values for classifying BCH images with these CNN models, trained over five epochs (10, 20, 30, 40, and 50), are provided in Table 10.

Tables 11 and 12 present the transfer learning results for DenseNet-201, NasNetMobile, and ResNet-101 on the BreakHis and ICIAR datasets. These models serve as baselines, as no additional data augmentation techniques were initially applied during training. Traditional data augmentation (TradAug), including image rotation, flipping, and zooming, was later introduced, and the classification results with and without these techniques are shown in Tables 11 and 12. Additionally, L2 regularization was applied during training, which perturbed the neural network's architecture by randomly setting some activations to zero, further enhancing performance.

Parameter	Values
Epoch	10, 20, 30, 40, 50
Batch size	64
Learning rate	0.0001 or1e−4
Transfer learning models	DenseNet 201, ResNet 101, NasNetMobile
Optimizer	Adam
Dropout rate	20% (before dense layers), 40% (after dense layer)
Batch normalization	Yes
Weight initialization	He/Kaiming
Loss function	Binary Cross-Entropy
Early stopping	Patience = 10 epochs, monitor = validation loss
Learning rate scheduler	Exponential Decay

Table 10. Parameter setting of CNN.

Dataset	Model	Breast cancer category	TP	TN	FP	FN	Precision	Recall	FPR	TNR	F1-score	Kappa	MCC
BreakHis	DenseNet-201	Benign	72	81	4	3	0.95	0.96	0.05	0.95	0.95	0.91	0.91
		Malignant	81	72	3	4	0.96	0.95	0.04	0.96	0.96	0.91	0.91
		Average Score					0.95	0.95	0.04	0.96	0.95	0.91	0.91
	NasNetMobile	Benign	66	83	2	9	0.97	0.88	0.02	0.98	0.92	0.86	0.86
		Malignant	83	66	9	2	0.90	0.98	0.12	0.88	0.94	0.86	0.86
		Average Score					0.93	0.93	0.07	0.93	0.93	0.86	0.86
	VGG-16	Benign	71	77	8	4	0.90	0.95	0.09	0.91	0.92	0.85	0.85
		Malignant	77	71	4	8	0.95	0.91	0.05	0.95	0.93	0.85	0.85
		Average Score					0.92	0.93	0.07	0.93	0.93	0.85	0.85
	VGG-19	Benign	69	81	4	6	0.95	0.92	0.04	0.96	0.93	0.87	0.87
		Malignant	81	69	6	4	0.93	0.95	0.08	0.92	0.94	0.87	0.87
		Average Score					0.94	0.94	0.06	0.94	0.94	0.87	0.87
	MobileNet	Benign	40	77	8	35	0.83	0.53	0.09	0.91	0.65	0.45	0.48
		Malignant	77	40	35	8	0.69	0.91	0.46	0.54	0.78	0.45	0.48
		Average Score					0.76	0.72	0.28	0.72	0.73	0.45	0.48
	ResNet-50	Benign	47	81	4	28	0.92	0.63	0.05	0.95	0.75	0.59	0.62
		Malignant	81	41	28	4	0.74	0.95	0.37	0.63	0.84	0.59	0.62
		Average Score					0.83	0.79	0.21	0.79	0.80	0.59	0.62
	Xception	Benign	75	80	5	0	0.94	1.0	0.06	0.94	0.97	0.94	0.94
		Malignant	80	75	0	5	1.0	0.94	0	1.0	0.97	0.94	0.94
		Average Score					0.97	0.97	0.03	0.97	0.97	0.94	0.94
	Proposed Model + TradAug	Benign	74	85	0	1	1.0	0.97	0	1.0	0.98	0.98	0.99
		Malignant	85	74	1	0	0.97	1.0	0.01	0.98	0.99	0.99	0.97
		Average Score					0.96	0.98	0.01	0.98	0.98	0.98	0.98
	Proposed Model + TradAug + cWGAN	Benign	74	85	0	1	1.0	0.99	0	1.0	0.99	0.99	0.99
		Malignant	85	74	1	0	0.99	1.0	0.01	0.99	0.99	0.99	0.99
		Average Score					0.99	0.99	0.01	0.99	0.99	0.99	0.99

Table 11. Performance comparison of various transfer learning models and the proposed framework on the BreakHis dataset for breast cancer classification.

To enhance model performance, the cWGAN-generated output was integrated as an additional (generated) dataset and trained alongside the original dataset. During this training phase, the first 200 layers of the DenseNet-201 model and the first 50 layers of the ResNet-101 model were frozen to preserve lower-level features learned from the pre-trained dataset.

The framework employed the Adam optimizer[39] for efficient training, with binary cross-entropy as the loss function for the BreakHis classification task[40]. The hyperparameter configurations used for training the CNN frameworks on the BreakHis and ICIAR datasets are detailed in Table 13. A learning rate of 0.0001, determined via grid search, balanced convergence speed and accuracy. To prevent overfitting, L2 regularization and data augmentation (via cGAN) were applied. Training with and without transfer learning and generative data achieved 97.6% training accuracy and 98.5% validation accuracy by the 29th of 50 epochs, showing effective

Dataset	Model	Breast Cancer Category	TP	TN	FP	FN	Precision	Recall	FPR	TNR	F1-Scroe	Kappa	MCC
ICIAR Dataset	DenseNet-201	Benign	10	28	2	0	0.83	1.0	0.06	0.94	0.91	0.88	0.88
		InSitu	8	30	0	2	1.0	0.80	0	1.0	0.89	0.86	0.87
		Invasive	9	28	2	1	0.82	0.90	0.06	0.94	0.86	0.81	0.81
		Normal	8	29	1	2	0.89	0.80	0.03	0.97	0.84	0.79	0.79
		Average Score					0.88	0.87	0.04	0.96	0.88	0.84	0.84
	NasNetMobile	Benign	8	29	1	2	0.89	0.80	0.03	0.97	0.84	0.79	0.79
		InSitu	9	21	9	1	0.50	0.90	0.30	0.70	0.64	0.47	0.52
		Invasive	4	30	0	6	1.0	0.40	0	1.0	0.57	0.50	0.58
		Normal	8	30	1	2	0.89	0.80	0.03	0.97	0.84	0.79	0.80
		Average Score					0.82	0.72	0.09	0.91	0.72	0.64	0.67
	VGG-16	Benign	7	28	2	3	0.78	0.70	0.07	0.93	0.74	0.66	0.66
		InSitu	8	25	5	2	0.62	0.80	0.17	0.83	0.70	0.58	0.59
		Invasive	8	25	5	2	0.62	0.80	0.17	0.83	0.70	0.58	0.59
		Normal	5	30	0	5	1.0	0.50	0	1.0	0.67	0.60	0.65
		Average Score					0.75	0.70	0.10	0.90	0.70	0.61	0.62
	VGG-19	Benign	7	25	5	3	0.58	0.70	0.16	0.84	0.64	0.50	0.50
		InSitu	4	29	1	6	0.80	0.40	0.03	0.97	0.53	0.44	0.48
		Invasive	6	27	3	4	0.67	0.60	0.10	0.90	0.63	0.52	0.52
		Normal	9	25	5	1	0.64	0.90	0.16	0.84	0.75	0.65	0.67
		Average Score					0.67	0.65	0.45	0.55	0.65	0.53	0.54
	MobileNet	Benign	0	29	1	10	0	0	0.03	0.97	0	−0.05	−0.09
		InSitu	5	20	10	5	0.33	0.50	0.33	0.67	0.40	0.14	0.15
		Invasive	8	17	13	2	0.38	0.80	0.43	0.57	0.52	0.27	0.32
		Normal	2	29	1	8	0.67	0.20	0.03	0.97	0.31	0.22	0.27
		Average Score					0.35	0.38	0.21	0.79	0.38	0.15	0.16
	ResNet-50	Benign	7	28	2	3	0.78	0.70	0.07	0.93	0.74	0.66	0.66
		InSitu	8	26	4	2	0.67	0.80	0.13	0.87	0.73	0.63	0.63
		Invasive	8	29	1	2	0.89	0.80	0.03	0.97	0.84	0.79	0.79
		Normal	9	29	1	1	0.90	0.90	0.07	0.97	0.90	0.87	0.87
		Average Score					0.84	0.80	0.07	0.93	0.80	0.74	0.74
	Xception	Benign	6	29	1	4	0.86	0.60	0.03	0.97	0.71	0.63	0.65
		InSitu	9	25	5	1	0.64	0.90	0.17	0.83	0.75	0.65	0.67
		Invasive	9	30	0	1	1.0	0.90	0	1.0	0.95	0.93	0.93
		Normal	9	29	1	1	0.90	0.90	0.03	0.97	0.95	0.94	0.94
		Average Score					0.85	0.83	0.06	0.94	0.82	0.77	0.78
	Proposed Method + TradAug	Benign	9	30	0	1	1.0	0.90	0	1.0	0.95	0.93	0.93
		InSitu	9	29	1	0	0.90	1.0	0.04	0.97	0.94	0.94	0.94
		Invasive	10	29	0	0	0.96	0.95	0	1.0	1.0	1.0	1.0
		Normal	10	30	0	0	1.0	1.0	0	1.0	1.0	0.99	0.99
		Average Score					0.95	0.95	0.01	0.99	0.97	0.96	0.97
	Proposed Method + cWGAN + TradAug	Benign	9	30	0	1	1.0	0.90	0	1.0	0.95	0.93	0.93
		InSitu	10	29	1	0	0.91	1.0	0.03	0.97	0.95	0.94	0.94
		Invasive	10	30	1	0	1.0	1.0	0	1.0	1.0	1	1
		Normal	10	30	0	0	1.0	1.0	0	1.0	1.0	1	1
		Average Score					0.98	0.98	0.01	0.99	0.98	0.97	0.97

Table 12. Performance analysis of a multi-scale feature enrichment framework and individual transfer learning models for breast cancer classification on the ICIAR dataset.

learning without overfitting (Fig. 5a: Accuracy Graph). Minor fluctuations in the loss curve (Fig. 5b: Loss Graph) due to the small batch size did not hinder convergence.

For the ICIAR dataset, which required multiclass classification, categorical cross-entropy was used. Given the smaller dataset size, training was extended to 200 epochs, achieving 96.3% training accuracy by the 166th epoch and 95.6% validation accuracy (Fig. 6a: Accuracy Graph). The learning curves show continuous improvement without overfitting, and early stopping ensured optimal training duration. While near-zero loss (Fig. 6b: Loss Graph) indicates excellent learning, larger fluctuations compared to BreakHis suggest further investigation, potentially through adjusting batch size or learning rate scheduling.

Dataset	Parameter name	Value
BreakHis	Optimizer	Adam
	Learning rate	0.0001
	Loss function	Binary cross-entropy
	Metrics	Accuracy, Precision, Recall
	Batch size	32
	Epochs	50
ICIAR	Optimizer	Adam
	Learning rate	0.001
	Loss function	Categorical cross-entropy
	Metrics	Accuracy, Precision, Recall
	Batch size	32
	Epochs	200

Table 13. Hyperparameter configuration for framework training CNN.

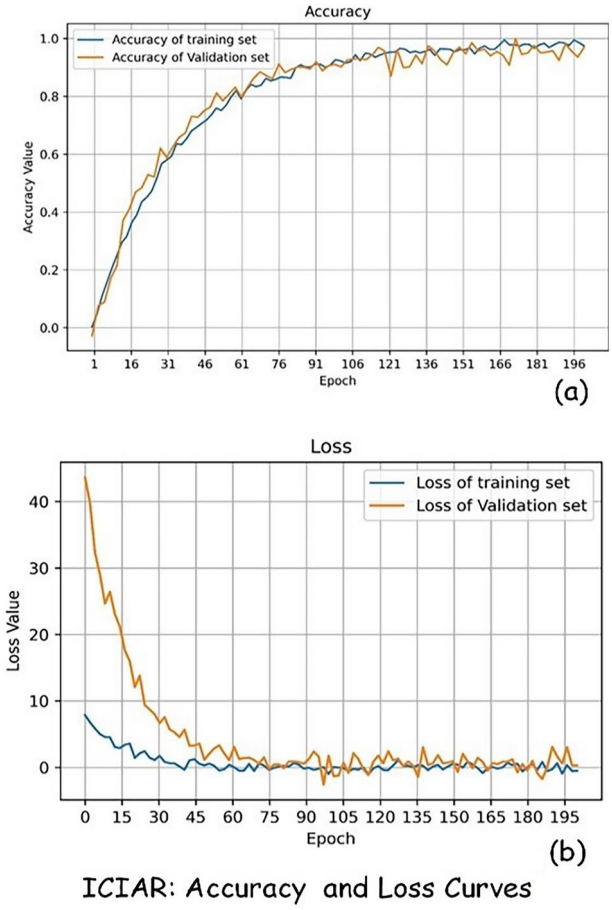


Fig. 6. ICIAR Training Progress: Both accuracy (**a**) and loss (**b**) curves trend favorably, showcasing model convergence and strong performance.

Benchmarking performance

Our proposed framework delivers exceptional performance on the BreakHis binary classification task. Figure 7 highlights the robust generalization capability of the proposed framework across various magnification levels (40×, 100×, 200×, and 400×) of the BreakHis dataset, as demonstrated by the consistent diagonal dominance in the confusion matrices. Accurate classification of 74 benign and 85 malignant microscopic images is depicted in Fig. 7a–d, achieving an impressive zero false negative rate (FNR) for malignant cases. This exceptional sensitivity underscores the framework's potential for reliable clinical deployment, where accurate detection of malignancies is critical. Furthermore, the Area Under the Receiver Operating Characteristic Curve (AUC) of 0.993 in Fig. 8 signifies our framework's outstanding consistency and discriminative power in differentiating between benign

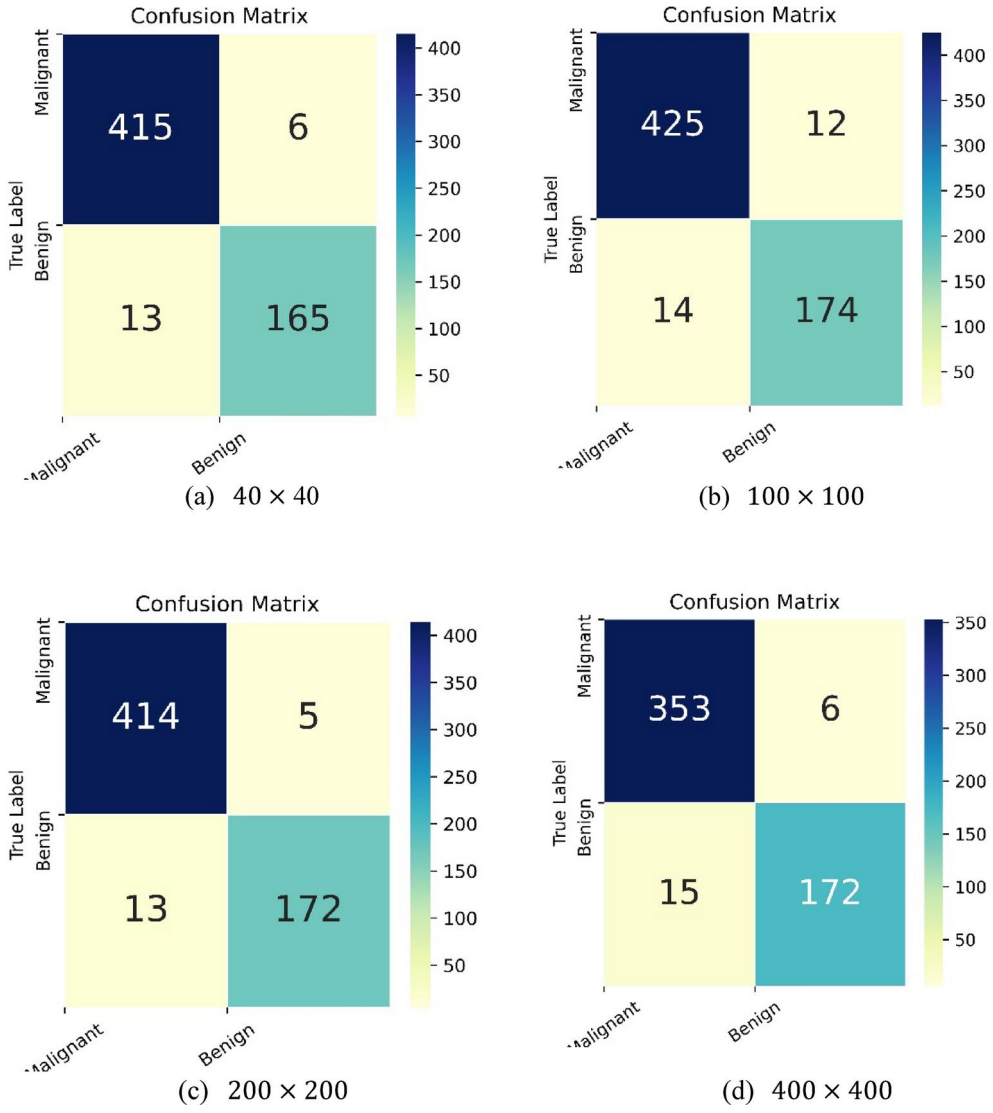


Fig. 7. Proposed framework maintains robust classification across BreakHis magnifications (40×–400×): consistent diagonal dominance in confusion matrices (**a**–**d**) confirms strong generalization.

and malignant tissues. A detailed comparison with seven individual transfer learning models in Table 8 further underscores its superiority. It achieves average precision, recall, F1-score, kappa coefficient, and Matthews correlation coefficient (MCC) values exceeding 0.99, surpassing all competitors across all metrics. Notably, the near-zero FNR and TNR approaching 1 further emphasize the effectiveness of our framework in detecting malignant lesions and minimizing misclassification of benign ones. While DenseNet-201 and Xception exhibit respectable F1-scores, they fall short of its comprehensive performance.

Our proposed framework demonstrates exceptional multiclass performance on the ICIAR dataset, effectively addressing various breast cancer categories. Figure 9b illustrates the confusion matrix, showcasing accurate classification of 9 benign, 10 in situ, 10 invasive, and 10 normal microscopic images, with only one benign case misclassified. This highlights the framework's strong generalizability across diverse breast cancer subtypes. Moreover, the Receiver Operating Characteristic (ROC) curve in Fig. 9a reflects an impressive Area Under the Curve (AUC) of 0.967, emphasizing the model's outstanding discriminative power and consistency in the multiclass setting. Additionally, Fig. 10 presents samples of generated benign images after training, highlighting the ability of the cWGAN model to synthesize high-quality microscopic images, which enrich the dataset and enhance classification performance. A comprehensive comparison with seven baseline models in Table 8 further underscores the superiority of the proposed framework, which achieves average precision, recall, F1-score, kappa coefficient, and MCC values exceeding 0.97, significantly outperforming all competitors. Notably, only DenseNet-201, ResNet 101, and Xception surpass 80% accuracy, while other models fall short. These results validate the effectiveness of our framework in multiclass breast cancer classification, showcasing its potential for clinical diagnosis and improved patient care.

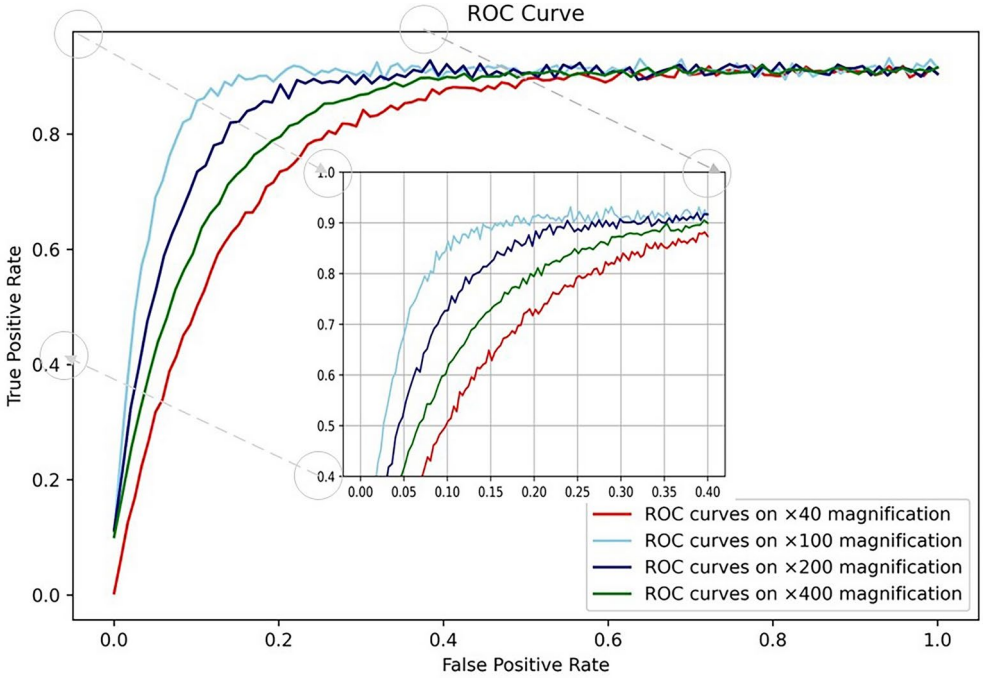


Fig. 8. Proposed framework outperforms across feature extractors and magnifications (40x-400x) in BreakHis, as shown by consistent high AUC values in ROC curves.

Discussion

Discussion of multi-scale transfer learning with traditional augmentation

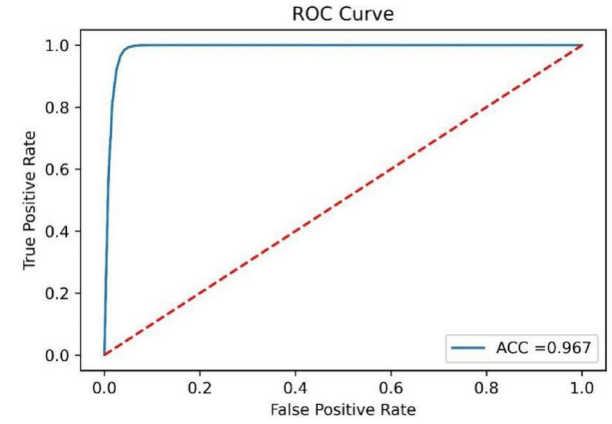
As detailed in Table 14, the multi-scaled transfer learning approach yielded classification accuracies of 98.8% for binary classification and 97.8% for multi-class classification without the application of additional data augmentation techniques. In contrast, when traditional data augmentation was employed, the accuracies dropped to 96.1% and 95.5%, respectively. This reduction in performance may be attributed to the need for hyperparameter adjustments as more complex augmentation strategies were introduced[56]. In experiments involving the application of data augmentation to pre-trained models, precise tuning of augmentation parameters is essential to balance the introduction of data diversity with the risk of overfitting. The inclusion of additional augmentation-related hyperparameters frequently necessitates extensive optimization to achieve optimal results.

Furthermore, the introduction of complex augmentations may introduce unrealistic or noisy variations into the data, making it more challenging for the model to learn meaningful features[57]. This issue may contribute to the observed decline in classification accuracy compared to the scenario where no augmentation was used. In cases where the target dataset for fine-tuning or transfer learning is limited in size, the use of complex augmentations may exacerbate the issue of data scarcity, potentially hindering the model's ability to generalize effectively[56]. To mitigate these limitations, cWGAN offer a promising approach, with the potential to improve accuracy by generating more representative synthetic data.

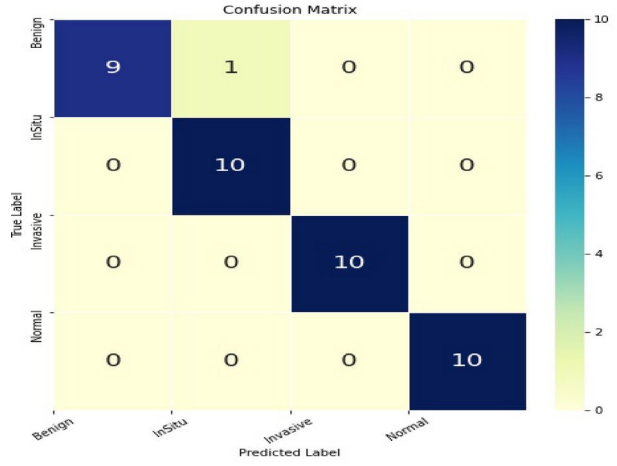
Discussion of multi-scale transfer learning with traditional augmentation and cWGAN

This section presents the results of multi-scale transfer learning using cWGAN with traditional augmentation (cWGAN + TradAug) for training and testing. Figures 8 and 10 highlight the application point of GAN-generated data during training. The model exhibited strong performance across all image sizes, achieving high accuracy, precision, recall, and F1-scores. However, the dataset's imbalance affected recall, leading to instances of misclassification, including false positives and false negatives. The analysis showed no significant effect of image size on performance, indicating the model's ability to extract relevant features from varying image sizes. To mitigate class imbalance, we applied traditional augmentation and cWGAN-generated synthetic images, enhancing the model's ability to learn from underrepresented classes.

The proposed model demonstrated exceptional performance across all datasets and model architectures, consistently achieving high accuracy, precision, recall, and F1-score. The use of cWGAN-generated synthetic data significantly enhances multi-scale feature extraction by addressing class imbalance at different image scales, enabling the model to learn more robust and diverse representations, particularly for smaller or less distinct features that are challenging to capture with conventional augmentations. The combination of traditional data augmentation (TradAug) with cWGAN proved to be highly effective in improving classification accuracy and reducing loss, particularly for the Inception-V3 and ResNet-50 models. Interestingly, the performance of VGG-16 and NasNetMobile varied with different combinations of TradAug and cWGAN. This suggests that the optimal data augmentation strategy may depend on the specific characteristics of the model architecture and the dataset. Overall, the results highlight the potential of combining data augmentation techniques with cWGAN



(a)



(b)

Fig. 9. Proposed Framework Performance on ICIAR (**a**,**b**): Confusion matrix and ROC curve demonstrate accurate classification and strong discriminative power.

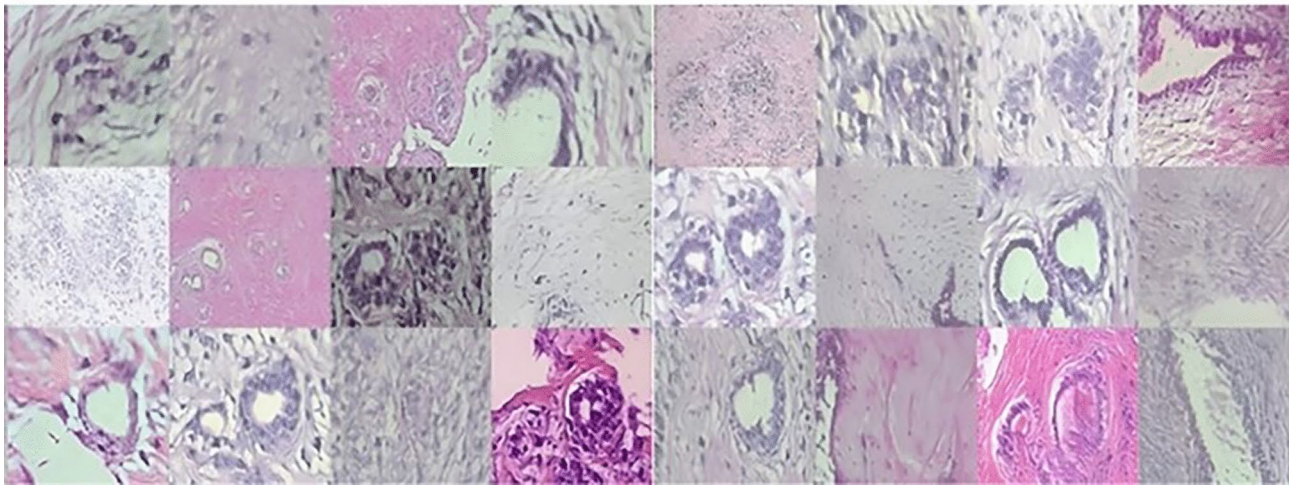


Fig. 10. Samples after training (generated Benign images).

Methods	BreakHis dataset		Classification type	Classification method
	Optimal accuracy	ICIAR dataset		
[41]	91.38%		Binary class	VLAD
[42]	93.3%		Binary class	K-mean + DWT
[43]	86.67%		Binary class	Fisher Vector + CNN
[44]	91%		Multiclass	CNN + LSTM + Softmax + SVM
[45]	92.6%		Binary class	VGG_16 + Logistic Regression
[46]	94.5		Multiclass	DCGAIN + VGG_16
[47]	87.4		Multiclass	ResNet50 + KWELM
[48]		96.1%	Multiclass	Hybrid CNN
[49]		85%	Multiclass	Inception-V3
[50]		87%	Multiclass	CNN
[51]		97.25%	Multiclass	CNN + RNN + Attention
[52]		76%	Multiclass	Inception ResNet V2
[53]	98.73%		Binary class	ResNet18 (Transfer Learning)
[54]	98.51%		Binary class	InceptionV3 (Transfer Learning)
[55]	99.12%		Binary class	Transfer learning with pre-trained DCNN architectures (VGG-16, Xception, DenseNet-201)
[29]	96.5%		Binary class	ResNet + VGG16 Essemble (Transfer Learning)
[31]	92.2%		Binary class	ResNet 50 (Transfer Learning)
[25]	96%		Binary class	CNN + Transfer Learning
Proposed method	98.8%		Binary class	Multi-scale Transfer Learning
Proposed method		97.8%	Multiclass	Multi-scale Transfer Learning
Proposed method	96.1%		Binary class	Multi-scale Transfer Learning + **TradAug**
Proposed method		95.5%	Multiclass	Multi-scale Transfer Learning + **TradAug**
Proposed method	**99.6%**		Binary class	Multi-scale Transfer Learning + **TradAug + cWGAN**
Proposed method		**98.2%**	Multiclass	Multi-scale Transfer Learning + **TradAug + cWGAN**

Table 14. Superior accuracy of the proposed multi-scale framework for binary and multi-class breast cancer classification with cGAN compared to existing methods on BreakHis and ICIAR datasets.

to enhance the performance of deep learning models in medical image classification tasks. However, further research is needed to explore the interplay between different model architectures, data augmentation techniques, and the specific characteristics of medical datasets.

The proposed multi-scale transfer learning framework demonstrated exceptional performance in breast cancer classification, surpassing existing methods on both the BreakHis and ICIAR datasets. By effectively combining traditional data augmentation (TradAug) with cWGAN, the framework achieved superior accuracy, with 99.6% accuracy on the BreakHis dataset and 98.2% accuracy on the ICIAR dataset, outperforming the best-performing baseline methods by 2.4% and 1.6%, respectively. These results highlight the potential of the proposed approach to significantly improve the accuracy and reliability of breast cancer diagnosis. Moreover, cWGAN-generated synthetic data aids in better generalization by providing additional realistic yet diverse samples, especially when the target dataset is limited. This contributes to a model that is less prone to overfitting and more capable of handling unseen variations in medical images. While the optimal model architecture varied depending on the specific dataset and task, the overall performance gains were substantial, demonstrating the effectiveness of the multi-scale transfer learning and cWGAN combination. Future research could explore the application of this framework to other medical imaging tasks, investigate the impact of different hyperparameter settings, and delve deeper into the underlying mechanisms through interpretability techniques.

Traditional augmentation reduces overfitting by minimizing the gap between training, validation, and test sets[1]. In contrast, the proposed model's lack of augmentation may have preserved transferability while also reducing overfitting. The need for hyperparameter adjustment becomes evident as more complex augmentations are introduced. Insufficient training epochs may have limited the model's adaptation, though its strong representation power enables effective transfer learning, even with limited labeled data. These findings help inform strategies for optimizing pre-trained models in classification tasks[58].

In conclusion, the proposed multi-scale feature enrichment framework significantly advances breast cancer classification by integrating pre-trained CNNs, multi-scale feature extraction, and a Generative Adversarial Network (GAN) to address data imbalance. This approach outperforms current methods in accuracy, efficiency, and robustness. Notably, its ability to eliminate false negatives for malignant cases highlights its potential for clinical use, offering a faster, more accurate, and cost-effective solution for improving patient outcomes.

Conclusion

This paper introduces a novel deep learning framework that leverages Generative Adversarial Networks (GANs) and multi-scale transfer learning to enhance breast cancer diagnosis using histopathological images. The

integration of a conditional Wasserstein GAN (cWGAN) and traditional augmentation effectively addresses the challenge of dataset imbalance by generating synthetic images, thereby enhancing the training process and improving model robustness. Pre-trained Convolutional Neural Networks (CNNs), including DenseNet-201, NasNetMobile, and ResNet-101, are employed for feature extraction, with a multi-scale feature enrichment scheme designed to capture fine-grained image details across multiple scales. The proposed framework demonstrates high classification performance, achieving 99.2% accuracy for binary classification and 98.5% for multi-class classification on the BreakHis dataset. Furthermore, the model surpasses existing methods by significantly reducing misclassification rates, particularly false negatives, which is critical for clinical applications. In conclusion, this framework represents a substantial advancement in breast cancer diagnosis, providing a more efficient, cost-effective, and accurate diagnostic solution. The combination of GANs for data augmentation and transfer learning shows strong potential for clinical integration. Future work should focus on improving the model's generalizability to larger and more diverse datasets, as well as optimizing hyperparameters to fully exploit the potential of this approach.

Data availability

We have used publicly available datasets for our research, including the Breast Cancer Histopathological Database (BreakHis) (https://paperswithcode.com/dataset/breakhis) and ICIAR https://paperswithcode.com/dataset/iciar-2018-grand-challenge-on-breast-cancer. The specific details and access information for both datasets are provided in the paper's references. These datasets are freely accessible to the public for further research and validation.

Received: 8 November 2024; Accepted: 11 February 2025
Published online: 03 March 2025

References

1. Gurcan, M. N., Madabhushi, A. & Rajpoot, N. Pattern recognition in histopathological images: An ICPR 2010 contest. In *Recognizing Patterns in Signals, Speech, Images and Videos: ICPR 2010 Contests, Istanbul, Turkey, August 23–26, 2010, Contest Reports* 226–234 (Springer, 2010).
2. Veta, M., Pluim, J. P., Van Diest, P. J. & Viergever, M. A. Breast cancer histopathology image analysis: A review. *IEEE Trans. Biomed. Eng.* **61**(5), 1400–1411 (2014).
3. Robboy, S. J. et al. Pathologist workforce in the United States: I. Development of a predictive model to examine factors influencing supply. *Arch. Pathol. Lab. Med.* **137**(12), 1723–1732 (2013).
4. Xia, B., Innab, N., Kandasamy, V., Ahmadian, A. & Ferrara, M. Intelligent cardiovascular disease diagnosis using deep learning enhanced neural network with ant colony optimization. *Sci. Rep.* **14**(1), 21777 (2024).
5. Pöllänen, I., Braithwaite, B., Haataja, K., Ikonen, T. & Toivanen, P. Current analysis approaches and performance needs for whole slide image processing in breast cancer diagnostics. In *2015 International Conference on Embedded Computer Systems: Architectures, Modeling, and Simulation (SAMOS)* 319–325 (IEEE, 2015).
6. Alkahtani, H. K. et al. Precision diagnosis: An automated method for detecting congenital heart diseases in children from phonocardiogram signals employing deep neural network. *IEEE Access* **PP**(99), 1–1 (2024).
7. Reardon, S. Precision-medicine plan raises hopes: US initiative highlights growing focus on targeted therapies. *Nature* **517**(7536), 540–541 (2015).
8. Spanhol, F. A., Oliveira, L. S., Petitjean, C. & Heutte, L. Breast cancer histopathological image classification using convolutional neural networks. In *2016 International Joint Conference on Neural Networks (IJCNN)* 2560–2567 (IEEE, 2016).
9. Shen, D., Wu, G. & Suk, H.-I. Deep learning in medical image analysis. *Ann. Rev. Biomed. Eng.* **19**, 221–248 (2017).
10. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**(7553), 436–444 (2015).
11. Islam, U. et al. Detection of renal cell hydronephrosis in ultrasound kidney images: A study on the efficacy of deep convolutional neural networks. *PeerJ Comput. Sci.* **10**, e1797 (2024).
12. Sharif Razavian, A., Azizpour, H., Sullivan, J. & Carlsson, S. CNN features off-the-shelf: An astounding baseline for recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* 806–813 (2014).
13. Penatti, O. A., Nogueira, K. & Dos Santos, J. A. Do deep features generalize from everyday objects to remote sensing and aerial scenes domains? In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* 44–51 (2015).
14. Tajbakhsh, N. et al. Convolutional neural networks for medical image analysis: Full training or fine tuning? *IEEE Trans. Med. Imaging* **35**(5), 99–1312 (2016).
15. Krizhevsky, A., Sutskever, I. & Hinton, G. E. ImageNet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, vol. 25 (2012).
16. Goodfellow, I. et al. Generative adversarial nets. In *Advances in Neural Information Processing Systems*, vol. 27 (2014).
17. Wang, C., Xu, C., Wang, C. & Tao, D. Perceptual adversarial networks for image-to-image transformation. *IEEE Trans. Image Process.* **27**(8), 4066–4079 (2018).
18. Li, J., Liang, X., Wei, Y., Xu, T., Feng, J. & Yan, S. Perceptual generative adversarial networks for small object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 1222–1230 (2017).
19. Xue, Y., Xu, T., Zhang, H., Long, L. R. & Huang, X. SegAN: Adversarial network with multi-scale L1 loss for medical image segmentation. *Neuroinformatics* **16**, 383–392 (2018).
20. Mahmood, T., Saba, T., Rehman, A. & Alamri, F. S. Harnessing the power of radiomics and deep learning for improved breast cancer diagnosis with multiparametric breast mammography. *Expert Syst. Appl.* **249**, 123747 (2024).
21. Chowdhury, P. R., El-Dosuky, M. & Kamel, S. Breast cancer detection using deep learning on biomedical mammogram images. *J. Theor. Appl. Inf. Technol.* **102**(7), 2924 (2024).
22. Zulfqar, S., Zia, M. A., Mehmood, F., Pervez, A., & Abbas, T. Breast cancer diagnosis: A transfer learning-based system for detection of invasive ductal carcinoma (IDC). *J. Comput. Biomed. Inform.* **7** (2024).
23. Shams, S., Platania, R., Zhang, J., Kim, J., Lee, K. & Park, S.-J. Deep generative breast cancer screening and diagnosis. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16–20, 2018, Proceedings, Part II 11* 859–867 (Springer, 2018).
24. Guttà, C., Morhard, C. & Rehm, M. Applying a GAN-based classifier to improve transcriptome-based prognostication in breast cancer. *PLoS Comput. Biol.* **19**(4), e1011035 (2023).
25. Tariq, M. A., Akbar, W., Batool, S. A., Abid, M. K. & Aslam, N. Breast cancer detection using deep learning algorithms. *J. Comput. Biomed. Inform.* **7** (2024).

26. Wang, J. et al. Weakly supervised lesion detection and diagnosis for breast cancers with partially annotated ultrasound images. *IEEE Trans. Med. Imaging* **43**(7), 2509–2521. https://doi.org/10.1109/TMI.2024.3366940 (2024).
27. Wang, W., Gao, M., Xiao, M., Yan, X. & Li, Y. Breast cancer image classification method based on deep transfer learning. arXiv preprint arXiv:2404.09226 (2024).
28. Shah, D., Ullah Khan, M. A. & Abrar, M. Reliable breast cancer diagnosis with deep learning: DCGAN-driven mammogram synthesis and validity assessment. *Appl. Comput. Intell. Soft Comput.* **2024**(1), 1122109 (2024).
29. Nair, S. S. & Subaji, M. Automated identification of breast cancer type using novel multipath transfer learning and ensemble of classifier. *IEEE Access* **12** (2024).
30. Oza, P., Sharma, P. & Patel, S. Breast lesion classification from mammograms using deep neural network and test-time augmentation. *Neural Comput. Appl.* **36**(4), 2101–2117 (2024).
31. Ray, R. K. et al. Transforming breast cancer identification: An in-depth examination of advanced machine learning models applied to histopathological images. *J. Comput. Sci. Technol. Stud.* **6**(1), 155–161 (2024).
32. Guan, S. & Loew, M. Using generative adversarial networks and transfer learning for breast cancer detection by convolutional neural networks. In *Medical Imaging 2019: Imaging Informatics for Healthcare, Research, and Applications* 306–318, vol. 10954 (SPIE, 2019).
33. Spanhol, F. A., Oliveira, L. S., Petitjean, C. & Heutte, L. A dataset for breast cancer histopathological image classification. *IEEE Trans. Biomed. Eng.* **63**(7), 1455–1462 (2015).
34. Yang, Q. et al. Low-dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss. *IEEE Trans. Med. Imaging* **37**(6), 1348–1357 (2018).
35. Aziz, L., Fc, M. S. & Ayub, S. Multi-level refinement enriched feature pyramid network for object detection. *Image Vis. Comput.* **115**, 104287 (2021).
36. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **15**(1), 1929–1958 (2014).
37. Liu, Y. et al. Wasserstein GAN-based small-sample augmentation for new-generation artificial intelligence: a case study of cancer-staging data in biology. *Engineering* **5**(1), 156–163 (2019).
38. Bińkowski, M., Sutherland, D. J., Arbel, M. & Gretton, A. Demystifying MMD GANs (2018).
39. Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014).
40. Iqbal, S., Qureshi, A. N., Ullah, A., Li, J. & Mahmood, T. Improving the robustness and quality of biomedical CNN models through adaptive hyperparameter tuning. *Appl. Sci.* **12**(22), 11870 (2022).
41. Dimitropoulos, K. et al. Grading of invasive breast carcinoma through Grassmannian VLAD encoding. *PLoS ONE* **12**(9), e0185110 (2017).
42. R. Karthiga and K. Narasimhan, Automated diagnosis of breast cancer using wavelet based entropy features. In *2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA)* 274–279 (IEEE, 2018).
43. Song, Y., Zou, J. J., Chang, H., & Cai, W. Adapting fisher vectors for histopathology image classification. In *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)* 600–603 (IEEE, 2017).
44. Nahid, A.-A., Mehrabi, M. A. & Kong, Y. Histopathological breast cancer image classification by deep neural network techniques guided by local clustering. *BioMed Res. Int.* **2018**, 2362108 (2018).
45. Mehra, R. Breast cancer histology images classification: Training from scratch or transfer learning? *ICT Express* **4**(4), 247–254 (2018).
46. Saini, M. & Susan, S. Deep transfer with minority data augmentation for imbalanced breast cancer dataset. *Appl. Soft Comput.* **97**, 106759 (2020).
47. Saxena, S., Shukla, S. & Gyanchandani, M. Breast cancer histopathology image classification using kernelized weighted extreme learning machine. *Int. J. Imaging Syst. Technol.* **31**(1), 168–179 (2021).
48. Alzubaidi, L. et al. Optimizing the performance of breast cancer classification by employing the same domain transfer learning from hybrid deep convolutional neural network model. *Electronics* **9**(3), 445 (2020).
49. Golatkar, A., Anand, D. & Sethi, A. Classification of breast cancer histology using deep learning. In *Image Analysis and Recognition: 15th International Conference, ICIAR 2018, Póvoa de Varzim, Portugal, June 27–29, 2018, Proceedings 15*, 837–844 (Springer, 2018).
50. Roy, K., Banik, D., Bhattacharjee, D. & Nasipuri, M. Patch-based system for classification of breast histology images using deep learning. *Comput. Med. Imaging Gr.* **71**, 90–103 (2019).
51. Yao, H., Zhang, X., Zhou, X. & Liu, S. Parallel structure deep neural network using CNN and RNN with an attention mechanism for breast cancer histology image classification. *Cancers* **11**(12), 1901 (2019).
52. Ferreira, C. A. et al. Classification of breast cancer histology images through transfer learning using a pre-trained inception ResNet v2. In *International Conference Image Analysis and Recognition* 763–770 (Springer, 2018).
53. Aloyayri, A. & Krzyżak, A. Breast cancer classification from histopathological images using transfer learning and deep neural networks. In *Artificial Intelligence and Soft Computing: 19th International Conference, ICAISC 2020, Zakopane, Poland, October 12–14, 2020, Proceedings, Part I 19* 491–502 (Springer, 2020).
54. Soumik, M. F. I., Aziz, A. Z. B. & Hossain, M. A. Improved transfer learning based deep learning model for breast cancer histopathological image classification. In *2021 International Conference on Automation, Control and Mechatronics for Industry 4.0 (ACMI)* 1–4 (IEEE, 2021).
55. Kumari, V. & Ghosh, R. A magnification-independent method for breast cancer classification using transfer learning. *Healthc. Anal.* **3**, 100207 (2023).
56. Yang Li, Y. et al. Hyper-tune: Towards efficient hyper-parameter tuning at scale. *Proc. VLDB Endow.* **15**(6), 1256–1265. https://doi.org/10.14778/3514061.3514071 (2022).
57. Gudovskiy, D., Rigazio, L., Ishizaka, S., Kozuka, K. & Tsukizawa, S. Autodo: Robust autoaugment for biased data with label noise via scalable probabilistic implicit differentiation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* 16601–16610 (2021).
58. Iqbal, S., Qureshi, A. N., Li, J., Choudhry, I. A. & Mahmood, T. J. H. Dynamic learning for imbalanced data in learning chest X-ray and CT images. *Heliyon* **9**(6), e16807 (2023).

Acknowledgements

This project was funded by the Deanship of Scientific Research (DSR), King Abdulaziz University, Jeddah, under Grant No. (G-199-665-1442). The authors, therefore, gratefully acknowledge DSR technical and financial support.

Author contributions

W.D., and Dr. L.A. acted as both the principal investigator and subject matter expert, overseeing the research and providing specialized knowledge. W.S.B. and A.A. perform supervised the statistical analyses. W.M.A., S.A.A. conducted the statistical analyses and contributed to the manuscript.

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to L.A.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.