Received: 11 September 2025 | Revised: 29 April 2026 | Accepted: 14 August 2026 | Published online: 26 October 2026

RESEARCH ARTICLE

Journal of Computational and Cognitive Engineering
2026, Vol. 00(00) 1–14
DOI: 10.47852/bonviewJCCE62027642

BON VIEW PUBLISHING

# AI-Enhanced Histopathological Breast Cancer Diagnosis Using ACO-Tuned L2-Regularized ResNet50: Toward Early Detection and Clinical Integration

**Debabrata Swain[1], Mahi Shah[1], Dev Solanki[1], Harsh Shah[1], Biswaranjan Acharya[2], Davinder Paul Singh[1] and Fernando Moreira[3,*]**

[1] *Department of Computer Science and Engineering, Pandit Deendayal Energy University, India*

[2] *Jagadguru Kripalu Institute of Technology, Jagadguru Krupalu University, India*

[3] *REMIT – Research on Economics, Management and Information Technologies, Portucalense University, Portugal*

**Abstract:** The most common form of cancer, which is found in females in society, is breast cancer. The diagnosis of the disease is very complex and hence often results in the mortality of the patient. Healthcare practitioners generally use different types of medical imaging-based reports for the precise and timely detection of the disease. In this work, a custom ResNet50 model is trained using the histopathological image samples of benign and malignant cases. A hybrid optimization approach has been used to improve the model performance by resizing the image to 224 × 224 normalization and handling the class imbalance by augmenting the benign samples. By limiting the augmentation to the training set, data leakage has been prevented. In this work, searching the hyperparameter space and finding the optimal values of batch size, learning rate, and epsilon is treated as one of the most crucial parts for which ant colony optimization (ACO) has been used. During model training, the amount of regularization is adjusted in such a way that it minimizes overfitting and enhances generalization based on the gradient's behavior, which has been introduced in an adaptive L2 ridge regularization scheme. Other than fine-tuning the model, ACO-based optimization and adaptive L2 regularization are performed to enhance the generalization of the model on the histopathological data. The model has shown a promising performance of 98.86% validation accuracy, which has proven the effectiveness of the customizations made to the model during implementation.

**Keywords:** histopathology image analysis, breast cancer detection, ResNet50, ant colony optimization, adaptive L2 regularization

## 1. Introduction

At present, breast cancer is appearing as one of the most common and life-threatening diseases for females in society, with an overall mortality rate of 10.6% [1]. Survival from this disease is only possible if it is detected at an early stage. But once the disease has gone to its advanced stage, then the treatment generally becomes more complicated, and chances of survival become lower. The disease generally starts in the mammary glands and later affects the milk-producing structures known as lobules or the milk-carrying channels known as ducts. Mostly, the occurrence of the disease starts with the irregular development of any muscle that creates a tumor, which may be cancerous (malignant) or non-cancerous (benign). A malignant tumor usually grows and affects its neighboring organs rapidly. The common symptoms that are found at the initial stage of the disease are a breast lump, nipple discharge, skin dimpling, change in breast shape, and chronic pain. Frequent examinations and imaging diagnostics are essential for the reason being that many warning signals are not always taken into consideration. Breast cancer still remains one of the most common reasons causing cancer-related deaths in women throughout the world. It accounts for approximately 15% of all female cancer deaths.

Today, the traditional methods for detecting breast cancer include various techniques such as mammography, ultrasound, and MRI scanning. For early-stage screening procedures, mammography is considered the most frequently utilized technique. However, it might lead to false negative results as this technique has low sensitivity in dense breasts. Whereas the secondary diagnostic techniques include ultrasound and MRI scanning, they are highly dependent on the operator and cost, ultimately leading to less availability in developing countries. So, in order to conclude, the study and the interpretation of the tissue have been done under a microscope using trained experts as part of the histopathological

***Corresponding author:** Fernando Moreira, REMIT – Research on Economics, Management and Information Technologies, Portucalense University, Portugal. Email: fmoreira@upt.pt

© The Author(s) 2026. Published by BON VIEW PUBLISHING PTE. LTD. This is an open access article under the CC BY License (https://creativecommons.org/licenses/by/4.0/).

examination of the tissue samples. There has been a serious need for more advanced and automated options that has arisen due to a lot of inconsistency in interpretation, false positives, and delayed diagnosis. Convolutional neural networks (CNNs) are one of the deep learning techniques that have come into the picture for detecting breast cancer as more and more information becomes available in the form of medical imaging. CNNs have been found to automatically learn to extract prominent features from histopathology images, which results in the reduction of manual assessments and thereby enhancing diagnostic accuracy. A special technique known as residual learning is being used by a powerful CNN model called ResNet50, which trains very deep neural networks effectively. To allow multiple layers and better visual imagery features in the network and improve the accuracy of classification, the issue of vanishing gradients is addressed. As the performances are heavily dependent on learning rate, batch size, and regularization strength, deep learning models are extremely sensitive to hyperparameter choice. The selection of the optimum value for the hyperparameters using manual observation makes the process tedious and computationally expensive. To avoid this issue, a nature-inspired optimization algorithm known as ant colony optimization (ACO) is used in this case.

The ACO algorithm works on the principle that is followed by the ants during their food collection. The ants generally adopt the optimum path selected through pheromone signals. In a similar fashion, the ACO algorithm explores all possible values in the feature space to discover the best values of the hyperparameters to be used in the proposed deep learning model. L2 regularization is a method that helps to avoid the overfitting problem. In this work, a customized L2 regularization strategy is developed that varies the regularization strength as a function of gradient fine-tuning, instead of a constant regularization strength. This step helps the model to reach the point of convergence. Finally, a transfer learning-based ResNet50 model is used along with the inclusion of ACO-based optimization and customized L2 regularization to improve the generalizability in terms of breast cancer identification using histopathological data. Model training and validation are performed using the BreakHis benchmark dataset. The swarm-based hyperparameter optimization and adaptive regularization have helped the model deliver high classification performance, as shown in the experimental results. The ability of the system to detect breast cancer was improved at an effective and consistent level. Hence, the diagnosis process tends to become more reliable and automated. Most research studies are based on single training conditions with fixed regularization and hyperparameters to be adjusted manually, despite the popularity of ResNet50 in medical image analysis, and it is adaptable to the dynamic training characteristics of histopathological images. ResNet50 is also utilized in this work only as a backbone feature extractor, and the main idea is an adaptive training framework, which integrates ACO-based hyperparameter optimization with gradient-aware L2 regularization to enhance generalization.

## 2. Literature Review

Alam et al. [2] proposed a deep learning framework based on the Unet3+ architecture for the classification of breast cancer from images taken using ultrasound, where the optimization of hyperparameters is done with the Adam optimizer and skip connections with deep supervision. The procedure includes image processing and pixel-wise segmentation, comparing Unet3+ against Fully Convolutional Network (FCN), Unet, SegNet, DeepLabV3+, and PSPNet. The structure performs better than the other structures with 90.99% total accuracy, 82.53% mean accuracy, 52.57% mean Intersection over Union (IoU), and 89.14% weighted IoU. In reference [3], a modified ResNet50 has been used with Gradient-weighted Class Activation Mapping (Grad-CAM) for breast cancer detection. Adam optimizer has been used for the same by fine-tuning the last 10 layers for hyperparameter optimization. The approach that has been used in this paper has included preprocessing using enhanced spectral mammography images, data augmentation, and Grad-CAM heatmaps. The model that has been used in this research work has successfully outperformed VGG16, VGG19, MobileNetV2, EfficientNet B7, and standard ResNet50 by achieving 89.20% Contrast-Enhanced Spectral Mammography (CESM) and 98.30% on Mammographic Image Analysis Society (MIAS). In the research work [4], an optimized CNN-based model for breast cancer classification has been used. ReduceLROnPlateau and Early Stopping for hyperparameter tuning have been the preferred methods utilized in the work that has been presented. Preprocessing of both mammograms and histopathology images, CNN-based feature extraction, and dynamic learning rate optimization have been included in the approach. The model that has been trained used cross-entropy loss and was optimized with Adam and RMSprop, which achieved 95.2% accuracy, overcoming the conventional models. In reference [5], the application of multiple deep learning architectures has been explored, which has included CNN, ResNet50, and Xception models for breast cancer detection. For performance enhancement, data augmentation, contrast correction, and batch normalization have been used. Preprocessing of the mammography images, segmentation, and feature extraction have been used in the approach. Adam optimizer has been used, and it has achieved an accuracy of 98.3% for CNN, 92.2% for ResNet50, and 90.2% for Xception, and it has been found out to be effective. Sureshkumar et al. [6] proposed a hybrid framework combining Convolutional Neural Networks with Pruned Ensembled Extreme Learning Machine (HCPELM) for breast cancer detection, in which they achieved 86% accuracy. This approach included preprocessing mammography images, segmentation, and feature extraction. Various optimization techniques included data augmentation, dropout, and batch normalization, which also enhanced model generalization. Abunasser et al. [7] utilized a deep learning framework based on a Breast Cancer Convolutional Neural Network (BCCNN) and pre-trained models Xception, InceptionV3, VGG16, MobileNet, and ResNet50 in classifying breast cancer. They followed methods such as preprocessing the MRI images of breast cancer and data augmentation using a Generative Adversarial Network (GAN) fine-tuned with deep learning models. The Adam optimizer is employed, with the use of data augmentation, and the hyperparameters were tuned. The proposed BCCNN model shows 98.28% accuracy, which beats the other models. Hai et al. [8] presented a transfer learning-based breast cancer histopathology classification system, in which fine-tuning the pre-trained CNNs (VGG16, ResNet50, and InceptionV3) on the BreakHis dataset was done under binary classification of benign and malignant tissue samples. The training model used feature extraction using frozen base layers and then selective fine-tuning of the additional layers. As compared to other models, ResNet50 performed the best at 95.4% accuracy. The measures of model performance that were used with ResNet50 were accuracy, precision, recall, F1-score, and AUC-ROC. Standard preprocessing and data augmentation methods have been used in the study. After that, the hyperparameters were manually optimized with the Adam optimizer. The research work by Humayun et al. [9] has proposed a deep learning framework based on InceptionResNetV2 as the starting model for

predicting breast cancer risk. The methodology that has been involved includes preprocessing images of breast histopathology, transfer learning, and classification. The optimization strategies that have been used are fine-tuning the final layers, data augmentation, and dropout with the help of the Adam optimizer. Having achieved 91% accuracy in this research work, it has been stated as a better model than the conventional deep learning approaches. In reference [10], the work has introduced an improved Transformer-based detection system, which emphasizes efficient multi-scale feature fusion. There has been an exchange between local and global representations due to the optimization, which has been used to enhance the efficiency and performance of the detection. The suggested architecture, with the help of structural feature fusion, has been able to solve the limitations of traditional transformer detectors. This article has been able to highlight the importance of building innovation in the development of object identification technologies. The paper by Yang et al. [11] has introduced an all-in-one modality fusion framework for hyperspectral and LiDAR classification through a Vision Transformer. By making use of a cross-modal attention mechanism, architecture that is proposed is able to effectively integrate heterogeneous spectral-spatial features, which in turn shows excellent performance on benchmark datasets. In this research work, a modality fusion block is used. Now, this fusion block includes both cross-attention and spectral self-attention mechanisms. These mechanisms help to reduce the dependency on alignment and retain the spatial information. The outstanding classification results are better than those of traditional deep learning methods, as shown in the framework. The introduction of an efficient monocular depth estimation system is done, where a combination of panoptic segmentation and instance-based depth prediction takes place in the paper by Raman et al. [12]. To enhance the practicability, the model has been able to correctly and effectively differentiate the relationships between occluders and oculars that too without the sensors being used. The results have been quite excellent with both benchmark datasets and custom datasets and have been able to show a high degree of robustness with occlusion-aware perception tasks being used. The vision-based transformer models have the capability of continuously growing, which highlights many advancements leading to extract and fuse precise spatial, spectral, and depth information from highly complex environments.

In order to identify the key research gaps, a thorough investigation of the literature and methodologies has been done. The detailed summary of the literature review has been provided in the Table 1.

1) Many times, it has happened that there has been a requirement for models requiring hand-tuning of hyperparameters to identify breast cancer, but that might result in inaccuracy and poor scaling to various other data collections.
2) There have been many cases in which overfitting has been found, along with compromised regularization effectiveness by most commonly used algorithms like Adam that have failed to decouple weight decay from gradient updates.
3) Multiple studies have been found to employ static L2 regularization that has been found to struggle to adapt to the altering training behavior and fail to avoid overfitting in deep networks.
4) Various earlier models fail to address the class imbalance problem, specifically in histopathology datasets where malignant samples tend to dominate as compared to other samples.
5) For the complete model training and output receiving part, one of the most important parts is the integration of metaheuristic optimization techniques like ACO, which often have been omitted for exploring intricate, high-dimensional hyperparameter domains.
6) Even though adaptive regularization strategies that can adapt to the gradient behavior during training have played a crucial role in enhancing model capability, there have been only a few studies that have been carried out in this domain and area to generalize medical image analysis tasks.
7) While some of the models demonstrate strong classification results, the practical insights into applicability in the clinical environment still remain uncertain and fail to satisfy the needs of medical exposure. Many of the problems have been found to be insufficiently addressable like inference speed, medical imaging system compatibility, and robustness across various imaging environments.
8) Although ResNet50 has been widely adopted for breast cancer classification, the proposed work distinguishes itself by optimizing model hyperparameters and adaptive L2 regularization. ACO has been used to enhance both predictive performance and generalization capability for the same.

**Table 1**
**Summary of models, optimization methods, and performance in literature review**

| Paper | Model used | Optimization | Hyper-parameter tuning | Accuracy |
|---|---|---|---|---|
| [2] | UNet3+ | Adam | Skip connections, deep supervision | 90.99% |
| [3] | Modified ResNet50 | Adam | Fine-tuning last 10 layers | 89.20% |
| [4] | Optimized CNN | Adam, RMSprop | ReduceLROnPlateau, Early Stopping | 95.2% |
| [5] | CNN, ResNet50 | Adam | Data augmentation, batch norm | 98.3% |
| [6] | Hybrid CNN | Adam | Dropout, batch normalization | 86% |
| [7] | BCCNN, Xception | Adam | GAN augmentation, fine-tuning | 98.28% |
| [8] | VGG16, ResNet50, InceptionV3 | Adam | Fine-tuning | 95.4% |
| [9] | Inception--ResNetV2 | Adam | Fine-tuning, dropout | 91% |

## 3. Research Methodology

### 3.1. Research design

In this study, a research framework has been developed which is using a supervised deep learning model for binary classification of breast cancer based on histopathology images. Two types of tissue samples, that is, benign and malignant, have been used in this research work. Using the ACO algorithm, a 6-framework global hyperparameter search has been driven with a gradient-aware adaptive L2 regularization technique, which has enhanced the model's generalization. The framework has been

strictly focused on standard performance measures like accuracy, precision, recall, and area under the curve (AUC) so that it can pass all the reliability clinical tests and can also be compared with the current existing literature work.

## 3.2. Participants

The information used in this paper was based on the BreakHis dataset derived from histopathological images of a clinical study on 82 patients. These individuals were subject to surgical open biopsies to provide their tissues to be examined. A total of 7,909 images were produced from these patients and shared among various magnified image levels to record various cell structures.

*3.2.1. Instruments*

Integrated with a digital camera, the Olympus BX 50 microscope system is the main imaging technique used for the acquisition of images in the BreakHis dataset. For high-resolution histopathological images of breast tissue slide samples, this device was used. Through surgical open biopsy, the acquisition of these samples was done, which is also known as partial mastectomy or excisional biopsy. To follow the standard data acquisition process, certain specific technical details, methods, and features were utilized:

1) Staining Technique (Hematoxylin and Eosin): In order to increase contrast for enhancing the visibility of the structures, the staining process was performed on the tissue samples.

2) Magnification Levels: To obtain the microscopic images, four different optical magnifications were applied. The magnifications of those microscopic images were 40×, 100×, 200×, and 400×.

3) Image Resolution: Raw images were acquired from the tissue samples, and they were in the form of 3-channel RGB true color with an initial resolution of 700 × 460 pixels.

4) Digital Format: The data acquisition has been carried out using a bit depth of 8 bits per pixel and PNG format in order to get accurate results by following the digital format properly.

## 4. Proposed System

The proposed methodology is illustrated in Figure 1.

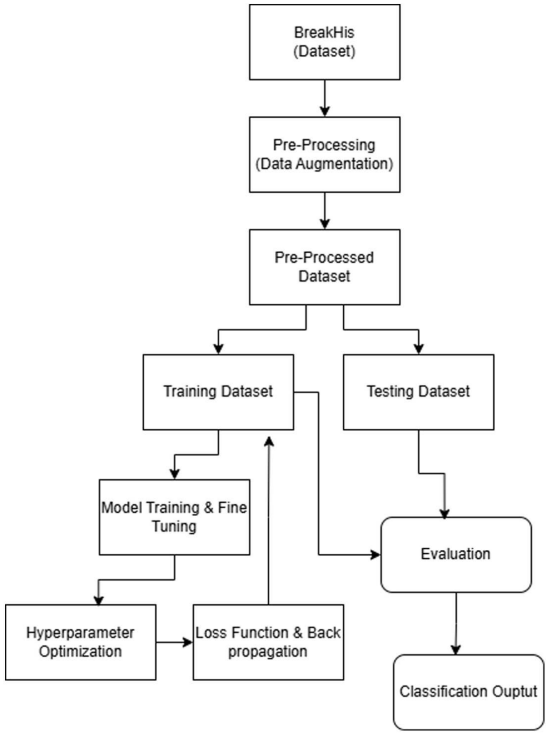
## 4.1. Dataset collection

The BreakHis dataset contained breast tissue's high-resolution cellular-level images, so for precise diagnoses of cancer, it was critical to make use of cellular-level images. Usage of the BreakHis dataset does not require specific licensing as it is available publicly for academic purposes. 7,909 histopathological images gathered from 82 patients are present within the dataset and distributed through different magnifications such as 1,995 images at 40×, 2,081 images at 100×, 2,013 images at 200×, and 1,820 images at 400×.

The images in the dataset have a standard resolution of 700 × 460 pixels and are mainly classified into two categories, namely, benign (2,480 images) and malignant (5,249 images), whose samples are shown in Figure 2. The overall number of samples present in both categories shows a class imbalance.

For better classification of cancerous tissues, it was important to have highly detailed inner and outer layers of tissues at the cellular level, which was made possible with the adoption of histopathological images because of their microscopic nature.

**Figure 1**
**Methodology flow diagram**




Traditional imaging methods such as X-ray and conventional photography fail to detect early stages of malignancies due to their nature of imaging at surface levels, as cancer usually develops and spreads from within. Histopathological imaging allows deep tissue analysis, hence playing a critical role in detecting cancers in their developing stages and identifying their classes accurately, making it a reliable choice for developing AI-based diagnostic systems.

## 4.2. Data preprocessing

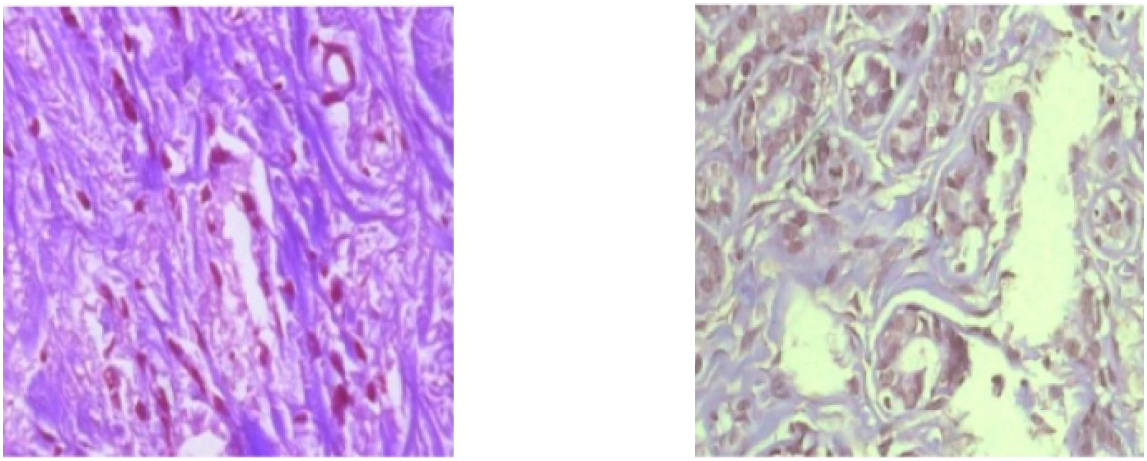
*4.2.1. Data augmentation*

Application of image augmentation as a preprocessing step [2] was carried out to handle the class imbalance present in the dataset and contribute to improved model performance [13]. With the use of this technique, new samples were generated in order to handle the class imbalance in the dataset. By this, the model would be more robust toward identifying the features of both classes and better generalize across different input combinations without biasing toward the dominant class.

Methods involved the use of operations such as random rotations, zooming, cropping, scaling, and horizontal and vertical flipping, which are mainly used for geometric manipulation [4, 7]; alongside this, different positions of the tissue samples were synthesized by applying a specific rotation of ±40°. Furthermore, a zoom factor of 0.2 and a shear range transformation of 0.2 were utilized during preprocessing to emulate physical stretching of tissue and different magnification levels. Moreover, to ensure the focus remains on the histological features, the movement range of the horizontal and vertical flipping was constrained to 20% of the image dimensions. The transformations that are used were a particular combination of specific parameters. This was used for selection in order to produce the maximum variety of the

**Figure 2**
**Histopathological images of breast tissue: (a) malignant sample; (b) benign sample**

(a) (b)



samples while preserving the underlying biological morphology of the benign and malignant samples. Through the transformations, the augmented images derived from the normal images are shown in Figure 3. The model to extract features from images at various positions, angles, and scales was enabled by the transformations that generated variations. Data augmentation techniques resulting in 5,249 samples per class equalized the sample count across both classes as a result. The distribution of the samples in both categories before and after applying the data augmentation approach is presented in Figure 4 accordingly. This modification helped achieve structural consistency in the dataset, which enhanced the training conditions and strengthened the model's recognition capability with regard to the initially underrepresented benign class [2, 4].

Moreover, the use of a special operation known as *fill mode* was employed during the augmentation process. The empty borders that are observed by the rotation of an image at specific angles are filled using this, as depicted in Equation (1). The use of neighboring pixels is made to interpolate and construct empty regions in the fill mode technique to complete the image [4, 7], providing visual continuity and monitoring formations so that no unnatural patterns are formed that the system would have difficulty interpreting. It should be noted that to avoid data leakage, no augmentation was performed on the test set, allowing evaluation metrics to reflect real-world scenario performance without bias [2, 4, 7].

$$(x_{\text{nearest}}, y_{\text{nearest}}) = \arg\min_{(x,y)\in I} \sqrt{(x'-x)^2 + (y'-y)^2} \quad (1)$$

where $I$ represents the original image, $(x, y)$ are the original pixel coordinates, $(x', y')$ is the transformed pixel location, argmin finds the $(x, y)$ in $I$ that is closest to $(x', y')$, and $(x_{\text{nearest}}, y_{\text{nearest}})$ is the selected nearest pixel used to fill the new position.

It was noted that once these techniques were implemented, a significant enhancement was made in the generalization capabilities of the model. Deep architectures like ResNet50, which are prone to overfitting to a biased medical dataset, were greatly reduced. The process of backpropagation became fairer through the trade-off between the benign class, which resulted in greater recall of the undersampled class. Furthermore, the divergence gap between training and validation accuracy was reduced out of the model's necessity toward learning invariant structure rather than memorizing particular orientations of pixels; as a result, a balanced dataset was obtained by applying these techniques.

#### *4.2.2. Normalization considerations*

As a step in the preprocessing pipeline, all dimensions of images from the BreakHis dataset were resized from their original resolution of $700 \times 460$ pixels to $224 \times 224$ pixels to effectively train the ResNet50 model, meeting its input requirements [3, 7]. This standardized size is commonly used across many CNNs and provides consistency while reducing computational load, which is beneficial given the dataset's large size [7, 8].

Image resizing leads to loss of some fine-grained image details, but the positive outcomes that come from using standardized input sizes and increased computational efficiency overcome this limitation. Also, pixel values were scaled down to the range $[0, 1]$ through division by 255 with the use of normalization; to avoid data leakage, this step was only performed on the training set [2].

## 5. Model Training

### 5.1. ResNet50

In terms of deep learning applications related to complex medical images, one of the most difficult challenges is the decline in model accuracy associated with the increasing network depth [8]. In order to overcome this problem, a CNN with a depth of 50 layers (ResNet50) was used based on the approach of residual learning (Figure 5). Instead of the direct input-output mapping, which is the output of the neural network, the network learns the residual function, which represents the difference between the input and output of the neural network. The output according to Equation (2) [3] is shown below.

$$H(x) = F(x) + x \quad (2)$$

Here, $H(x)$ denotes the target function, $x$ represents the input, and $F(x)$ represents the residual that is added to the input to produce the final output. This was made possible through the use of skip connections, which allowed the input to bypass some layers and be added directly to the output (Figure 5).

**Figure 3**
**Visual comparison of normal and augmented tissue samples: (a) normal sample 1; (b) augmented sample 1; (c) normal sample 2; (d) augmented sample 2**

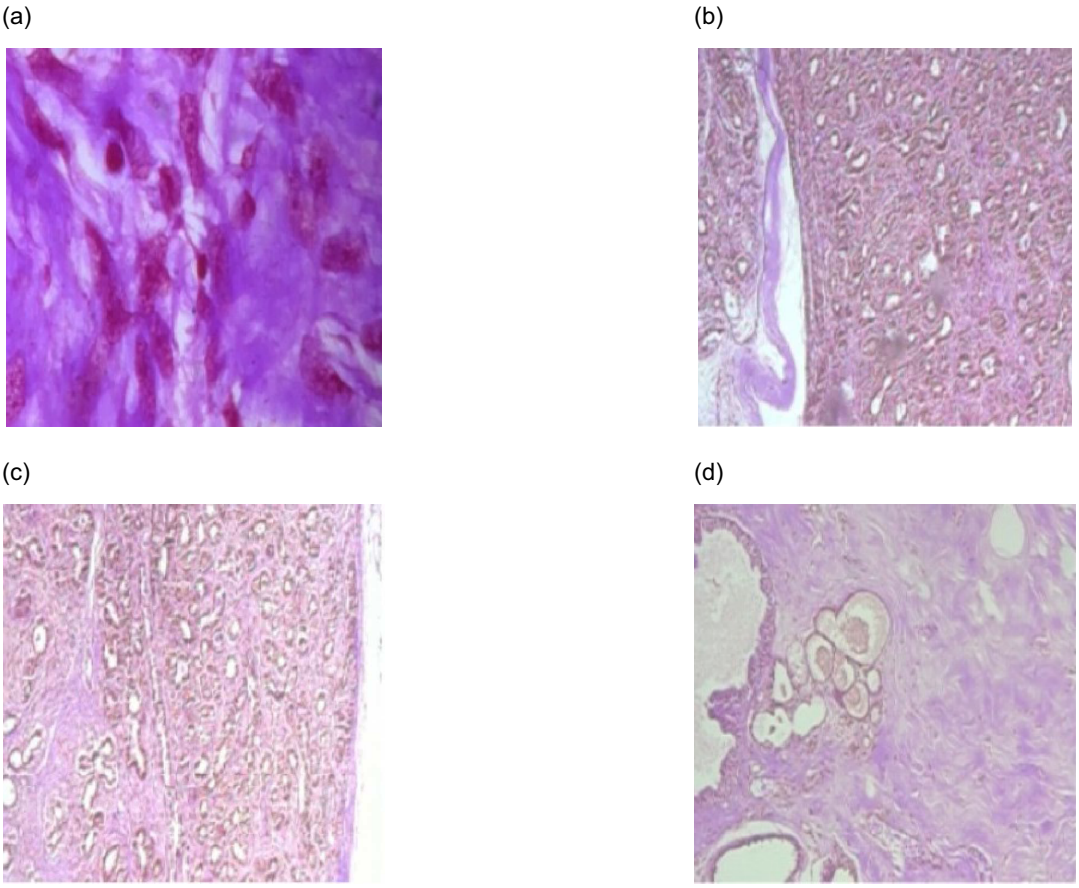
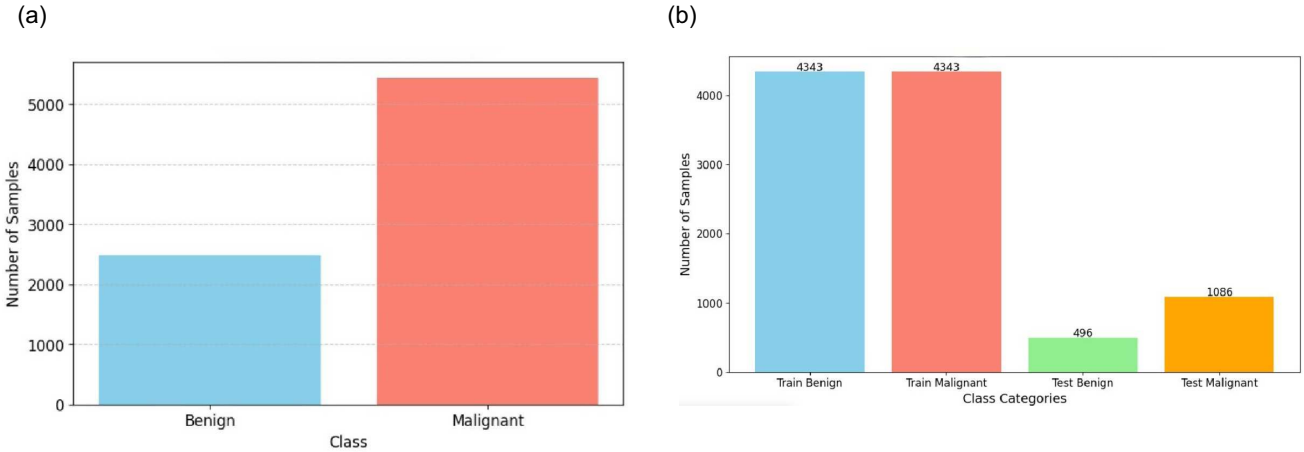



**Figure 4**
**Class distribution analysis: (a) before augmentation; (b) after augmentation**

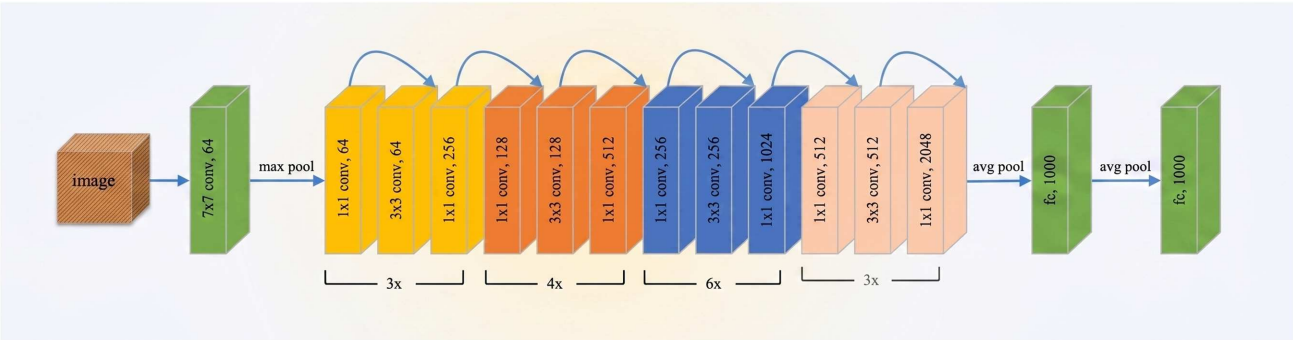



Skip connections prevent vanishing gradients during training to help maintain gradient flow. To enhance the overall learning efficiency.

The ResNet50 architecture starts with a $7 \times 7$ convolutional layer with 64 filters and a stride of 2. It then uses Batch normalization layers to add stability to training and rectified linear unit (ReLU) activation layers for nonlinearity (Figure 5). A $3 \times 3$ max pooling layer follows to decrease spatial size while keeping essential features, which also has a stride of 2. The network then goes through four key stages after this initial stage: Residual blocks [3, 8] are used to generate composite blocks. The design of each residual block is a bottleneck design with a convolution of $1 \times 1$. A $3 \times 3$ convolution is used to capture spatial features, and a final $1 \times 1$ convolution is used to restore the original dimensions. Between these layers, except after the last layer, batch normalization and ReLU are applied in each block.

Batch normalization throughout the network assists in network stability, quick convergence, and better performance.

**Figure 5**
**Architecture of the proposed ResNet50 model**




Training efficiency is improved by using normalization of the input values of the layer. This is for consistency of activation during training improves the values and minimizes the internal covariate shift [7]. Allows quicker convergence rates, and a larger learning rate was used, thus making the network more powerful and less dependent on the initial weights. In the identity blocks, the input is fed directly into the output, and in the convolution blocks, the input is convoluted. A $1 \times 1$ convolution is applied to make the output dimensional. This model allows for successful learning, spanning several scales and depths of features, including detecting fine-grained and complex features, which is essential. The histopathology images contain patterns that can be used to detect them [5].

## 5.2. AdamW optimizer

In this study, AdamW is an optimization algorithm that was derived from the Adam optimizer and was used here. To train the deep neural network successfully and correctly on the BreakHis dataset, there needs to be a better and more efficient technique to do that. It should be noted that the Adam optimizer usually performs poorly when it comes to weight decay regularization due to its too-simple way to put on weight decay during the gradient update, which can incite unwanted interference with the adaptive learning mechanism and reduce the effectiveness of regularization [4]. Models trained using Adam are able to converge fast, but sometimes, they overfit the training data, which is very common in medical image classification tasks [9].

Separating the weight decay from the gradient calculation helps make the optimization approach better. Instead of mixing decay into the gradient step, AdamW applies decay directly to the parameters after the gradient step. The mathematical update formula for AdamW is given in Equation (3), where the parameter value is updated at time step $t + 1$. This small modification ensures that regularization does what it is supposed to do: draw the weights successively and separately from the learning dynamics toward zero [14].

$$\theta_{t+1} = \theta_t - \eta \left( \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \varepsilon} \right) + \lambda \theta_t \tag{3}$$

Here, $\varepsilon$ is a small constant providing numerical stability and preventing division by zero when the denominator becomes very small [4]. For the proposed model, tuning this $\epsilon$ value was a crucial step, especially due to the high-resolution nature of histopathological images, where even minor numerical instabilities could impact gradient updates [12]. To counter the risk of overfitting, AdamW was employed. The aim of AdamW is to enhance the ability of the model to generalize to unseen samples.

## 5.3. Adaptive L2 regularization

L2 regularization is an approach that mitigates the problem of using excessively large weights in the model. However, it helps to prevent the overly complex pattern from being learned when training the model, since this is what it makes sure of in training. This is because penalties are applied to large weight values in this method [8]. It operates by adding a penalty term to the loss function, which is proportional to the sum of the squared weights as in Equation (4) [7, 15]. This is critical in avoiding overfitting, which is a major problem with deep learning models used in medical image analysis so that the model can generalize well to unknown data [5, 16].

$$\text{Loss}_{\text{total}} = \text{Loss}_{\text{original}} + \frac{\lambda}{2} \sum_i w_i^2 \tag{4}$$

In this case, $\lambda$ is a regularization parameter that determines the force of the penalty. The importance of proper tuning of $\lambda$ is very Important: If it is too large, it can lead to underfitting; if it is too small, it will not be effective in preventing overfitting. The optimum balance in the model was determined using cross-validation [12, 15].

For the implementation, the L2 regularization was added as the weight decay parameter in the optimizer. Since the AdamW optimizer was employed, which separates the weight decay from the gradient updating, the L2 penalty was better and more effectively implemented than the one with Adam [4]. This led to smaller and more uniform weight values, thereby making smoother decision boundaries and resulting in better generalization [10]. This is also consistent with Bayesian construal, working similarly to a Gaussian prior on weights [16, 17]. This model was able to be generalized and fused with L2 regularization; it achieves reduced overfitting but maintains the performance of the network during training. With L2, the regularization caused the model to not overfit on the training data and allowed it to generalize, without sacrificing the training performance [18]. The proposed method never uses a fixed $\lambda$ as with traditional $L_2$ regularization and scales the $\lambda$ dynamically, depending on the changes in gradient magnitudes from one epoch to another.

$$\lambda_t = \lambda_0 \cdot (1 + \gamma \cdot ||\nabla W_t||) \tag{5}$$

The update of the penalty term is dependent on the norm of the weight gradients in Equation (5), which ensures that the

strength of regularization is adjusted in accordance with the training dynamics.

## 5.4. Hyperparameter optimization using ant colony algorithm

Due to the huge search space of ResNet50 and its sensitivity to various combinations of hyperparameters, tuning the hyperparameters of ResNet50 was difficult. Hyperparameters such as the learning rate, batch size, and dropout rate ranged widely, resulting in a complex, highly nonlinear, and non-smooth search space [2]. This makes it difficult and tricky to explore the search space through conventional optimization techniques [10].

A nature-inspired metaheuristic optimization algorithm called Ant Colony Optimizer (ACO) was taken into account. This is partly because it can survive and function in very complicated environments [19]. ACO simulates the ants' foraging behavior—they generate a possible solution from a probability distribution that can be determined using pheromone trails and heuristic cues. Such trails become reinforced in areas offering more promising results, making it possible to concentrate efforts toward high-yielding directions without trapping into local optima [20]. ACO allowed for efficient exploration of such hyperparameter space and helped improve model generalization and robustness based on pheromone deposition mechanisms [21]. Instead of applying ACO solely for tuning the values of learning rate or batch size as performed in existing studies, this work tried to integrate ACO into a dual optimization loop that jointly influenced the learning rate, dropout rate, and the adaptive regularization procedure, hence changing the effective training trajectory of ResNet50.

The ants, in this study, symbolize the hyperparameter configuration candidates that are used to help train the model [22]. The trajectory followed by the ants in searching hyperparameters can be described with the help of a discrete vector:

$$S = \{b, \eta, d\}$$

where:

1) $b$ represents the batch size,
2) $\eta$ represents the learning rate, and
3) $d$ represents the dropout rate.

The ants sequentially select hyperparameter values from the predetermined hyperparameter nodes for ACO search graph construction.

### 5.4.1. Pheromone initialization

The optimization process begins by initializing pheromone trails for all hyperparameters. The pheromone value of each hyperparameter option, denoted as $\tau_j$, is initialized at $t = 0$ as follows:

$$\tau_j(0) = \tau_0 \tag{6}$$

whereas $\tau_0$ is a very small positive value for initialization. This will make sure that all hyperparameter configurations are done uniformly, so that exploration can be done in an unbiased fashion during the process in the initial stages.

### 5.4.2. Heuristic information

Heuristic information helps ants make informed decisions when identifying superior hyperparameter configurations. The heuristic evaluation function $\eta_j$ for a hyperparameter option is defined as:

$$\eta_j = \frac{1}{L_{val} + \varepsilon} \tag{7}$$

where $L_{val}$ stands for the validation loss and $\epsilon$ is a small constant used to prevent division by zero. The evaluation function facilitates convergence quickly while favoring hyperparameter settings that improve generalization by minimizing the loss function.

### 5.4.3. Pheromone trail update

$$T_{ij}(t+1) = (1-\rho)T_{ij}(t) + \Delta T_{ij} \tag{8}$$

Here, $T_{ij}(t)$ denotes the pheromone intensity on edge $(i, j)$ at time $t$, $\rho$ is the evaporation rate controlling pheromone decay, and $\Delta T_{ij}$ represents the new pheromone deposited on the edge during the current iteration.

### 5.4.4. Pheromone deposition

$$T_{ij} = \sum_k \Delta T_{ij}^k \tag{9}$$

where $\Delta T_{ij}^k$ represents the pheromone deposited by ant $k$ on edge $(i, j)$.

### 5.4.5. Ant's probabilistic decision

$$P_{ij}^k = \frac{[T_{ij}(t)]^\alpha [\eta_{ij}]^\beta}{\sum_{l \in \text{allowed}_k} [T_{il}(t)]^\alpha [\eta_{il}]^\beta + \varepsilon} \tag{10}$$

In this equation, $P_{ij}^k$ represents the likelihood of ant $k$ choosing edge $(i, j)$ when constructing its path. $T_{ij}(t)$ indicates the amount of pheromone deposited on edge $(i, j)$ at time t, while $\eta_{ij}$ is the heuristic value of edge $(i, j)$, for example, the reciprocal of its distance. $\alpha$ and *beta* regulate the impact of the intensity of pheromones and heuristic value, respectively. $textallowed_k$ refers to the set of edges that ant $k$ is allowed to select from at a given step.

A small constant $\varepsilon$ (typically $10^{-10}$) is added to the denominator for numerical stability and to prevent division by zero [23]. The computational complexity of the proposed ACO method for hyperparameter tuning is defined as follows:

$$O(A \times I \times E)$$

where:

1) $A$ = the number of all ants in a colony.
2) $I$ = number of iterations of the algorithm performed
3) $E$ = number of epochs in evaluation of each hyperparameter.

In this model, $E$ is the number of partial epochs in the final epochs of training during the training phase, to ensure the computation is efficient. The continual updates over the course of the trail by evaporation and condensation are called iterative trail updating. With reinforcement, the algorithm will be able to converge to the optimum configurations quickly [24]. The technique has lower computational complexity than the Grid Search technique, making it much more efficient. It could be adapted for optimal hyperparameter selection in real-time applications [25].

## 6. Results of Using Optimized ResNet50 for Breast Cancer Classification Based on Histopathological Samples

The performance of the proposed model for benign and malignant classes was assessed with standard performance measures such as accuracy, precision, recall, ROC, and confusion matrix. These metrics are a good way of measuring the ability of the model, as both false positives and false negatives have tremendous implications in the medical domain [26].

### 6.1. Performance metrics

*6.1.1. Accuracy*

The accuracy score helps in computing the overall correctness of the system [27]. Regardless of the class of the image being tested, this metric considers the overall correctly classified images out of the total number of test images [6]. The accuracy of the model is determined with the help of the confusion matrix by the following equation: no. of correctly predicted positives by the classifier, no. of positives not predicted by the classifier, no. of negatives predicted by the classifier, and no. of negatives not predicted by the classifier. Accuracy = true positive rate * true negative rate + (1-true positive rate) * (1-true negative rate). That is, we can use Equation (11) to calculate the accuracy of the classifier:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{11}$$

*6.1.2. Precision*

Precision, also called positive predictive value, is the number of correctly predicted malignant cases out of the total number of cases predicted as malignant [5]. For breast cancer diagnosis, if it is more precise, the false positive increases are decreased, which reduces unnecessary treatment in benign patients [28]. Recently, precision was calculated as shown in Equation (12):

$$\text{Precision} = \frac{TP}{TP + FP} \tag{12}$$

*6.1.3. Recall*

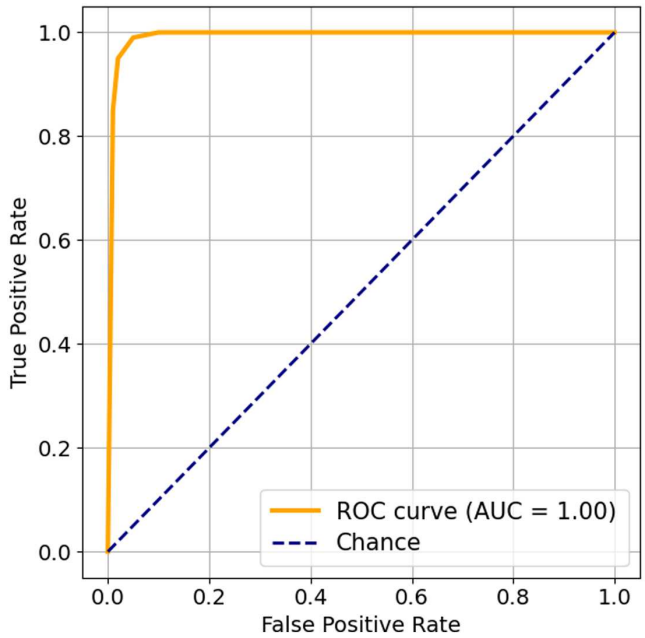
Recall, also referred to as sensitivity, measures the ability of the model to classify malignant tumors accurately [29]. A high recall implies that most of the malignancies have been identified correctly. Recall can be determined using the confusion matrix with the help of Equation (13):

$$\text{Recall} = \frac{TP}{TP + FN} \tag{13}$$

*6.1.4. ROC-AUC curve*

The receiver operating characteristic (ROC) curve, as shown in Figure 6, graphically represents the performance of the model's true positive rate (recall) against the false positive rate varying over different thresholds. The area under the curve (AUC) gives the numerical summary of the entire ROC plot into a single value, whereby a value of 1.0 means perfect classification while a score of 0.5 represents a random guess [8]. A larger value of AUC denotes a good discrimination ability of the model to distinguish between benign and malignant classes [7].

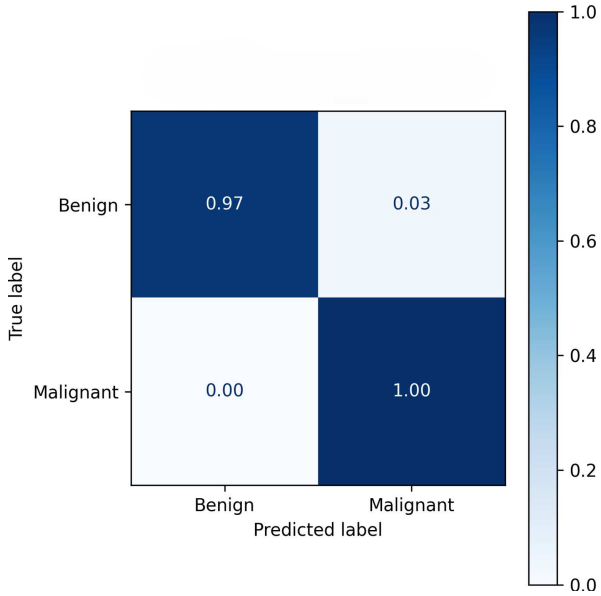
**Figure 6**
**The proposed framework's receiver operating characteristic(ROC) curve**




*6.1.5. Confusion matrix*

The graphical representation of the results obtained by the classification performed by the proposed model is carried out with the use of the confusion matrix as shown in Figure 7 showcasing the numbers in expressions of TP (true positives), TN (true negatives), FP (false positives), and FN (false negatives) based on the model's output as shown in Table 2.

**Figure 7**
**Confusion matrix based on the metrics of the proposed system**




### 6.2. Hyperparameter tuning

The performance of the ResNet50 model was optimized by rigorous experimentation on the major hyperparameters of the

**Table 2**
**Confusion matrix of the proposed framework**

| | Predicted benign | Predicted malignant |
|---|---|---|
| Actual benign | True negative (TN) | False positive (FP) |
| Actual malignant | False negative (FN) | True positive (TP) |

model, which are batch size, dropout of fully connected layers of the CNN, and learning rate. Initially, the behavior of the model during convergence was effectively studied by manually selecting different values of these hyperparameters while keeping the other two constant.

*6.2.1. Batch size*

Learning rate = 0.0001, Dropout = 0.2

**Observations:** The model with batch size 32 shows early convergence and stability but was unable to improve after a certain epoch. Batch sizes 64 and 128 had increased instability initially due to overfitting, with batch size 128 being the most unstable. Batch size 64 achieved the highest accuracy as shown in Table 3.

**Table 3**
**Batch size effect on the model's performance**

| Batch size | Accuracy | Precision | Recall |
|---|---|---|---|
| 32 | 98.29% | 0.9611 | 0.99 |
| 64 | 98.48% | 0.9615 | 1 |
| 128 | 97.91% | 0.952 | 1 |

*6.2.2. Dropout*

Learning rate = 0.0001, Batch size = 64

**Observations:** Dropout follows a linear trend of decreased accuracy performance over increased dropout due to underfitting as shown in Table 4. The graphs also indicate increased instability over epochs.

**Table 4**
**Dropout effect on the model's performance**

| Dropout | Accuracy | Precision | Recall |
|---|---|---|---|
| 0.1 | 98.67% | 0.9615 | 1 |
| 0.15 | 98.48% | 0.9615 | 1 |
| 0.3 | 98.10% | 0.9611 | 0.99 |

*6.2.3. Learning rate*

Batch Size = 64, Dropout = 0.15

**Observations:** Higher learning rates show instability during convergence but higher performance over epochs, and lower learning rates show high stability but delayed convergence, hence lower performance as indicated in Table 5.

*6.2.4. Ant colony optimization*

To obtain the optimal set of hyperparameters, the ACO technique was utilized, rather than relying on manual selection, which might not yield the most effective configuration [30]. ACO, being a metaheuristic algorithm [31], tries to mimic the ants' exploration behavior, which is considered an ideal choice for solving discrete optimization problems [32], such as selecting the best configurations yielding the highest accuracy and searching through the hyperparameter search space [33].

**Table 5**
**Learning rate effect on the model's performance**

| Learning rate | Accuracy | Precision | Recall |
|---|---|---|---|
| 0.0001 | 98.35% | 0.9615 | 1 |
| 0.00001 | 98.16% | 0.9523 | 1 |
| 0.000001 | 94.05% | 0.8461 | 0.99 |

To explore the hyperparameter [34] search space in each iteration, 10 ants were appointed as agents in this study. The different values of batch size, learning rate, and dropout considered in the ACO are presented in Table 6. As a result, a total of 10 combinations of hyperparameters were evaluated in each iteration, while only 5 iterations were performed to carry out the whole process [35]. The distinct points showing the parameter ranges.

**Table 6**
**Different values for batch size, learning rate, and dropout used in ACO**

| Batch size | Learning rate | Dropout |
|---|---|---|
| 32 | 0.0001 | 0.1 |
| 64 | 0.00001 | 0.15 |
| 128 | 0.000001 | 0.2 |
| - | - | 0.3 |

The proposed framework underwent 15 epochs of training and testing as shown in Figure 8 to yield the best accuracy each time with the combination of hyperparameters. Therefore, the best-yielding combination of hyperparameters using the ACO is:

1) Batch size: 32
2) Learning rate: 0.00001
3) Dropout: 0.1

The final model's performance is shown in Table 7.

The model's reliability is demonstrated by achieving a high accuracy of 98.86%. The precision of 0.9708 suggests a very low false positive rate. The recall of 1.0 indicates high sensitivity of the model, successfully identifying all malignant cases. The near 1.0 AUC indicates strong classification and generalization capability.

## 7. Comparative Analysis

This particular model is the best of all the comparative models on all measures. Special attention is given to the high accuracy, in particular, not only due to the model architecture but also because of some strategic improvements that are selected such as the optimization techniques.

Residual convolutional architecture, ResNet50, is used for stable gradient propagation and effective deep feature learning. In this study, richer cellular information is obtained from the histopathological pictures, which has improved the performance compared to other traditional mammographic or ultrasound modalities. The hyperparameters are adjusted dynamically by the suggested ACO-based adaptive regularization scheme and

**Figure 8**
**Accuracy and loss curves that have been observed during the training of the model: (a) accuracy over epochs; (b) loss over epochs**

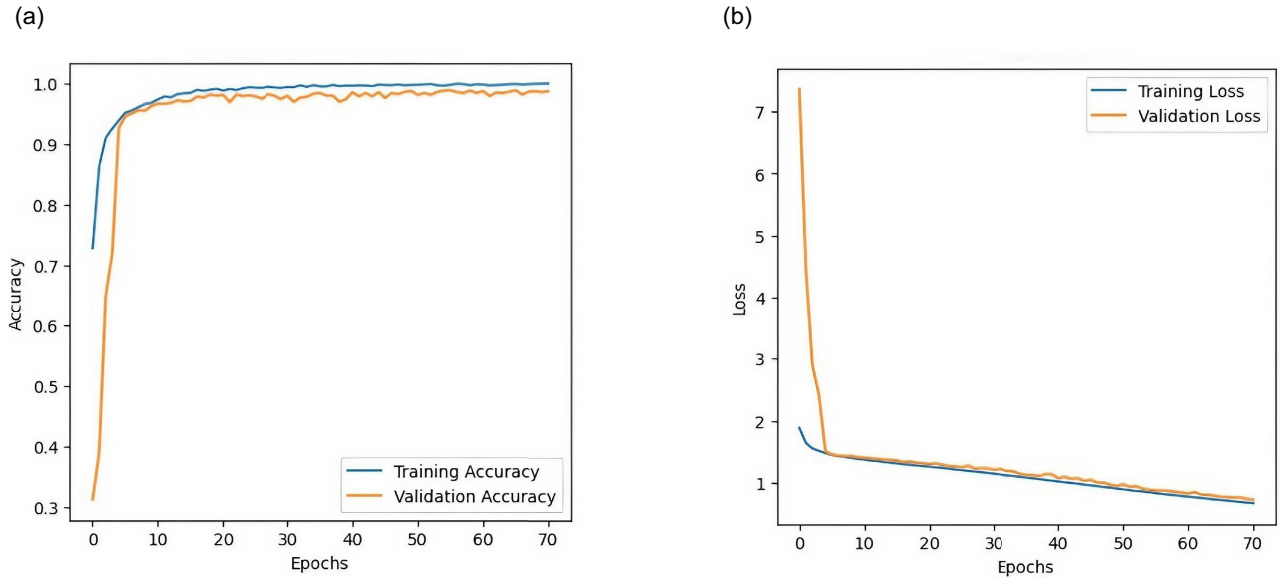



**Table 7**
**Performance of the final model after ACO-based hyperparameter tuning**

| Metric | Value |
|---|---|
| Accuracy | 98.8% |
| Precision | 0.9708 |
| Recall | 1.0 |
| AUC | ~1.0 |

**Table 8**
**The proposed ResNet50 model vs comparative analysis of existing models**

| Reference | Model used | Accuracy |
|---|---|---|
| [2] | UNet3+ | 90.99% |
| [3] | ResNet50 with Grad-CAM | CDD-CESM: 89.20%, MIAS: 98.30%, |
| [4] | Optimized CNN | 95.2% |
| [5] | CNN, ResNet50, Xception | CNN: 98.3%, ResNet50: 92.2%, Xception: 90.2% |
| [6] | Hybrid CNN (HCPELM) | 86% |
| [7] | BCCNN, Xception, ResNet50 | 98.28% |
| [8] | VGG16, ResNet50, InceptionV3 | 95.4% |
| [9] | InceptionResNetV2 | 91% |

regulates the intensity of regularization in training results. One can easily say that this increase in performance is ultimately the result of the combination and mixture of the swarm-based hyperparameter optimization and gradient-aware regularization and not only ust the architecture. The summary of the comparative analysis is as shown in Table 8.

1) From this study, it was analyzed that the ACO method for hyperparameter tuning is more efficient than the traditional Adam optimizer because it is one of the reasons in enabling a global search for the best parameters, such as batch size, learning rate, and dropout. This ensures faster convergence and prevents local minima traps. The class imbalance problem is handled by our model through strategic image augmentation. There is no class bias in the dataset, and the dataset is balanced having 4334 images per class. This is also a limitation in other models that rely on real-world or uneven datasets.
2) In order to reduce overfitting and improve generalization, adaptive ridge regularization is applied to the model, and by applying this, the penalties are adjusted dynamically based on the weight magnitudes. By this, the model is being made robust to noisy or high-dimensional histopathological images.
3) If one compares the ResNet50 model with other models like HCPELM or standard CNNs, the ResNet50 turns out to be enabling the maintainability of the deep feature hierarchies along with superior gradient flow while having less manual pruning as well. It can also retain interpretability if gets to be integrated with methods like Grad-CAM.
4) To differentiate from various other methods (e.g., [2], [4], [5]) that depend on manual tuning and fixed configurations, our framework stands out in such a way that it consists of an automated and optimized training pipeline to reduce the time and increase scalability.
5) Models such as Ensemblers and InceptionResNetV2, which are extremely powerful although they are still lagging in clinical applications. Their structure is complex, and they are so configuration-dependent that they are not adaptable. Compared to these, our model is lightweight, fast convergence, and consistent, which is suitable for clinical deployment.

## 8. Conclusion

In this research work, in order to obtain better classification results for breast cancer diagnosis, an adaptive deep learning model applying swarm-based hyperparameter optimization and gradient-aware regularization to the ResNet50 model is used.

Here, one of the major things kept in mind is ACO, which was used for effective hyperparameter tuning, where many of the model's key parameters were used, such as batch size, dropout rate, and learning rate. Along with this, the model performance has been improved by amalgamating L2 regularization and the AdamW optimizer.

The results of the evaluation metrics of the model on the BreakHis dataset were calculated as an accuracy of 98.8%, precision of 0.9708, and recall of 1.0.

Apart from this, the classification strength and the generalization features of the model led to a nearly perfect AUC score. These results demonstrate that the adaptive deep learning system can be applied effectively in medical image processing and cancer detection.

## 9. Future Work

The proposed model is to be trained with large-sized and diverse images in the near future. The use of a larger and more diverse dataset will achieve greater generalizability across all data samples. The proposed method ensures optimization and also considers all design aspects for obtaining better results across all clinical diagnosis test cases. The investigation has proven the deep learning model to be a useful tool for the analysis and prediction of cancer disease using medical imaging data. The further extension of this work can be made for the multi-stage classification of breast cancer by clearly specifying different statuses of the tumor such as fibroadenoma or carcinoma. To achieve better trust in the prediction made by the model, explainable AI (XAI) can be integrated. Along with XAI, Grad-CAM heat maps can also be helpful to better visualize and understand why the prediction is made by the model. To identify the most frequently used algorithm, various metaheuristic algorithms can be compared to each other using a computationally efficient model. Personalized regularization techniques can be used during the actual "inferencing."

## Recommendations

Achieving 98.86% classification accuracy with the use of the ResNet50 model framework optimized using the ACO technique demonstrates its potential to serve as an assistance tool for radiologists to classify cancerous (malignant) cells of human organs with high precision. Thus, it was theorized that the system can be applied in medical screening procedures to give real-time clinical decision-making capabilities along with reducing manual human inspection, especially in case of intricate diagnostic cases. Moreover, the system's work efficiency in different types of high-dimensional clinical image-based modalities was improved with the utilization of adaptive L2 regularization.

## Funding Support

This research is completely supported by national funds through FCT – Fundação para a Ciência e a Tecnologia, I.P., under the Programme Contract UID/05105/2025. https://doi.org/10.54499/UID/05105/2025

## Ethical Statement

The data used in this study were obtained from a publicly available and fully de-identified dataset, the Breast Cancer Histopathological Image (BreakHis) dataset licensed under CC BY 4.0. Therefore, ethical approval and informed consent were not required.

## Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

## Data Availability Statement

The BreaKHis datasets that support the findings of this study are openly available at: https://web.inf.ufpr.br/vri/databases/breast-cancer-histopathological-database-breakhis/.

## Author Contribution Statement

**Debabrata Swain:** Conceptualization, Validation, Resources, Visualization, Supervision, Project administration. **Mahi Shah:** Conceptualization, Software, Formal analysis, Resources, Writing – original draft, Writing – review & editing, Visualization. **Dev Solanki:** Conceptualization, Methodology, Software, Formal analysis, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Harsh Shah:** Conceptualization, Methodology, Investigation, Visualization. **Biswaranjan Acharya:** Validation, Data curation, Supervision, Project administration. **Davinder Paul Singh:** Validation, Investigation, Supervision, Project administration. **Fernando Moreira:** Validation, Writing – review & editing, Supervision, Project administration, Funding acquisition.

## References

[1] Rahman, M., Deb, K., Dhar, P. K., & Shimamura, T. (2024). ADBNet: An attention-guided deep broad convolutional neural network for the classification of breast cancer histopathology images. *IEEE Access*, *12*, 133784–133809. https://doi.org/10.1109/ACCESS.2024.3419004

[2] Alam, T., Shia, W.-C., Hsu, F.-R., & Hassan, T. (2023). Improving breast cancer detection and diagnosis through semantic segmentation using the Unet3+ deep learning framework. *Biomedicines*, *11*(6), 1536. https://doi.org/10.3390/biomedicines11061536

[3] Gharaibeh, H., Almaiah, M., Nasayreh, A., Al Mamlook, R. E., Al-Na'amneh, Q., Nahar, K. M. O., & Alqaseem, M. A. (2025). Improved breast cancer detection using modified ResNet50-based on gradient-weighted class activation mapping. In *Bridging Digital Innovation and Technology for Business Transformation—ICTIM*, 43-53. https://doi.org/10.1007/978-3-031-90558-2_7

[4] Mahesh, T. R., Thakur, A., Gupta, M., Sinha, D. K., Mishra, K. K., Venkatesan, V. K., & Guluwadi, S. (2024). Transformative breast cancer diagnosis using CNNs with optimized ReduceLROnPlateau and early stopping enhancements. *International Journal of Computational Intelligence Systems*, *17*(1), 14. https://doi.org/10.1007/s44196-023-00397-1

[5] Hariharaganesh, M., & Themburane, S. P. (2024). Harnessing deep learning techniques for identification of breast tumors. *Tuijin Jishu/Journal of Propulsion Technology*, *45*(4), 311–323. https://propulsiontechjournal.com/index.php/journal/article/view/8053/5122

[6] Sureshkumar, V., Prasad, R. S. N., Balasubramaniam, S., Jagannathan, D., Daniel, J., & Dhanasekaran, S. (2024). Breast cancer detection and analytics using hybrid CNN and extreme learning machine. *Journal of Personalized Medicine*, *14*(8), 792. https://doi.org/10.3390/jpm14080792

[7] Abunasser, B. S., Al-Hiealy, M. R. J., Zaqout, I. S., & Abu-Naser, S. S. (2023). Convolution neural network for breast cancer detection and classification using deep learning. *Asian Pacific Journal of Cancer Prevention*, *24*(2), 531–544. https://doi.org/10.31557/APJCP.2023.24.2.531

[8] Hai, H. H., Manoharan, S., M, S. D., & Raja, K. (2025). An explainable transfer learning framework for breast cancer histopathology classification using CNNs. In *2025 IEEE 7th International Conference on Computing, Communication and Automation*, 1–7. https://doi.org/10.1109/ICCCA66364.2025.11325662

[9] Humayun, M., Khalil, M. I., Almuayqil, S. N., & Jhanjhi, N. Z. (2023). Framework for detecting breast cancer risk presence using deep learning. *Electronics*, *12*(2), 403. https://doi.org/10.3390/electronics12020403

[10] Liu, F., Zheng, Q., Tian, X., Shu, F., Jiang, W., Wang, M., . . . , & Saponara, S. (2025). Rethinking the multi-scale feature hierarchy in object detection transformer (DETR). *Applied Soft Computing*, *175*, 113081. https://doi.org/10.1016/j.asoc.2025.113081

[11] Yang, B., Wang, X., Xing, Y., Cheng, C., Jiang, W., & Feng, Q. (2024). Modality fusion vision transformer for hyperspectral and LiDAR data collaborative classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *17*, 17052–17065. https://doi.org/10.1109/JSTARS.2024.3415729

[12] Raman, R., Jiang, W., & Bakshi, S. (2025). PODE: Panoptic-aware object-wise depth estimation for occlusion prediction. *IEEE Transactions on Consumer Electronics*, *71*(4), 10906–10916. https://doi.org/10.1109/TCE.2025.3610083

[13] Alzoubi, H., Amin, M., Aljehani, A. M., Alhatamleh, S., Madain, R., Madae'en, S., . . . , & Alomani, A. A. (2026). Enhancing breast cancer diagnosis with ResNet50 and salp swarm-based feature reduction on BreakHis dataset. *Biomedical Signal Processing and Control*, *114*, 109319. https://doi.org/10.1016/j.bspc.2025.109319

[14] Rahman, R., Sharif, S. T., & Sikder, M. H. (2025). An ensemble approach: Stacked pre-trained deep learning based and logistic regression meta-learner for breast cancer histopathology analysis. In *2025 International Conference on Quantum Photonics, Artificial Intelligence, and Networking*, 1–5. https://doi.org/10.1109/QPAIN66474.2025.11172189

[15] Venkatesh, Nayana, N., & Nagaraju, Y. (2025). A transfer learning approach for breast cancer grading. In *2025 IEEE International Conference for Women in Innovation, Technology & Entrepreneurship*, 1–6. https://doi.org/10.1109/ICWITE64848.2025.11306949

[16] Pulvera, E. V. J., & Lao, D. M. (2024). Enhancing deep learning-based breast cancer classification in mammograms: A multi-convolutional neural network with feature concatenation, and an applied comparison of best-worst multi-attribute decision-making and mutual information feature selections. In *2024 9th International Conference on Intelligent Informatics and Biomedical Sciences*, 1–8. https://doi.org/10.1109/ICIIBMS62405.2024.10792816

[17] Sharmin, S., Ahammad, T., Talukder, M. A., & Ghose, P. (2023). A hybrid dependable deep feature extraction and ensemble-based machine learning approach for breast cancer detection. *IEEE Access*, *11*, 87694–87708. https://doi.org/10.1109/ACCESS.2023.3304628

[18] Hiregoudar, V., & Das, S. (2024). Automated detection of breast cancer in histopathology images using convolutional neural networks. In *2024 4th International Conference on Innovative Sustainable Computational Technologies*, 1–6. https://doi.org/10.1109/CISCT62494.2024.11134214

[19] Elkorany, A. S., Marey, M., Almustafa, K. M., & Elsharkawy, Z. F. (2022). Breast cancer diagnosis using support vector machines optimized by whale optimization and dragonfly algorithms. *IEEE Access*, *10*, 69688–69699. https://doi.org/10.1109/ACCESS.2022.3186021

[20] Ben Ammar, M., Ayechi, F. L., Ksantini, R., & Mahjoubi, H. (2023). Pre-trained deep convolutional neural network architectures for breast cancer diagnosis in mammography: Current state-of-the-art. In *2023 International Conference on Innovations in Intelligent Systems and Applications*, 1–7. https://doi.org/10.1109/INISTA59065.2023.10310612

[21] Aguerchi, K., Jabrane, Y., Habba, M., & El Hassani, A. H. (2024). A CNN hyperparameters optimization based on particle swarm optimization for mammography breast cancer classification. *Journal of Imaging*, *10*(2), 30. https://doi.org/10.3390/jimaging10020030

[22] Alzahrani, R. M., Sikkandar, M. Y., Begum, S. S., Babetat, A. F. S., Alhashim, M., Alduraywish, A., . . . , & Ng, E. Y. K. (2025). Early breast cancer detection via infrared thermography using a CNN enhanced with particle swarm optimization. *Scientific Reports*, *15*(1), 25290. https://doi.org/10.1038/s41598-025-11218-0

[23] Nasser, M., & Yusof, U. K. (2023). Deep learning based methods for breast cancer diagnosis: A systematic review and future direction. *Diagnostics*, *13*(1), 161. https://doi.org/10.3390/diagnostics13010161

[24] Obayya, M., Maashi, M. S., Nemri, N., Mohsen, H., Motwakel, A., Osman, A. E., . . . , & Alsaid, M. I. (2023). Hyperparameter optimizer with deep learning-based decision-support systems for histopathological breast cancer diagnosis. *Cancers*, *15*(3), 885. https://doi.org/10.3390/cancers15030885

[25] Li, H., Govindarajan, V., Ang, T. F., Shaikh, Z. A., Ksibi, A., Chen, Y.-L., . . . , & Por, L. Y. (2025). MSPO: A machine learning hyperparameter optimization method for enhanced breast cancer image classification. *Digital Health*. *11*, 20552076251361603. https://doi.org/10.1177/20552076251361603

[26] Mu, J. (2023). Breast cancer detection using ResNet with hyperparameter tuning. In *2023 IEEE 3rd International Conference on Power, Electronics and Computer Applications*, 1245–1249. https://doi.org/10.1109/ICPECA56706.2023.10076197

[27] Kshirsagar, J. P., Phulpagar, B., & Patil, P. (2024). Enhancing breast cancer image classification: Minimizing features with optimized feature extraction and transfer learning techniques. In *2024 Parul International Conference on Engineering and Technology*, 1–8. https://doi.org/10.1109/PICET60765.2024.10716184

[28] Poongodai, A., Pavani, S., Khalil, S. M., Prakash, P. J., & Ujwal, C. R. (2025). Improvised breast cancer detection using CNN with particle swarm optimization (PSO). In *2025 6th International Conference on Inventive Research in Computing Applications*, 1–8. https://doi.org/10.1109/ICIRCA65293.2025.11089807

[29] Zhang, Y., Liu, Y.-L., Nie, K., Zhou, J., Chen, Z., Chen, J.-H., . . . , & Su, M.-Y. (2023). Deep learning-based automatic diagnosis of breast cancer on MRI using Mask R-CNN for detection followed by ResNet50 for classification. *Academic Radiology*, *30*(Suppl. 2), S161–S171. https://doi.org/10.1016/j.acra.2022.12.038

[30] Khan, N. K., & Islam, M. N. (2025). Breast cancer detection using deep learning and quantum-inspired hyperparameter optimization. In *2025 International Conference on Quantum Photonics, Artificial Intelligence, and Networking*, 1–6. https://doi.org/10.1109/QPAIN66474.2025.11171695

[31] Bilal, A., Imran, A., Baig, T. I., Liu, X., Nasr, E. A., & Long, H. (2024). Breast cancer diagnosis using support vector machine optimized by improved quantum inspired grey wolf optimization. *Scientific Reports*, *14*(1), 10714. https://doi.org/10.1038/s41598-024-61322-w

[32] Satyanarayana, U., & U, M. (2025). Enhanced ant colony optimization-driven hyperparameter tuning for transfer learning-based breast cancer classification using ResNet101. In *2025 International Conference on Sustainable Communication Networks and Application*, 1390–1396. https://doi.org/10.1109/ICSCN67106.2025.11308387

[33] Jiang, X., Hu, Z., & Xu, Z. (2024). CSO-CNN: Cat swarm optimization-guided convolutional neural network for mobile detection of breast cancer. *Mobile Networks and Applications*, *29*(6), 1886–1896. https://doi.org/10.1007/s11036-024-02298-9

[34] Manimurugan, S., Karthikeyan, P., Aborokbah, M., Narmatha, C., & Ganesan, S. (2024). Breast cancer diagnosis model using stacked autoencoder with particle swarm optimization. *Ain Shams Engineering Journal*, *15*(6), 102734. https://doi.org/10.1016/j.asej.2024.102734

[35] Kulothungan, V., Ramamoorthy, T., Sathishkumar, K., Mohan, R., Tomy, N., Miller, G. J., & Mathur, P. (2024). Burden of female breast cancer in India: Estimates of YLDs, YLLs, and DALYs at national and subnational levels based on the National Cancer Registry Programme. *Breast Cancer Research and Treatment*, *205*(2), 323–332. https://doi.org/10.1007/s10549-024-07264-3

**How to Cite:** Swain, D., Shah, M., Solanki, D., Shah, H., Acharya, B., Singh, D. P., & Moreira, F. (2026). AI-Enhanced Histopathological Breast Cancer Diagnosis Using ACO-Tuned L2-Regularized ResNet50: Toward Early Detection and Clinical Integration. *Journal of Computational and Cognitive Engineering*. https://doi.org/10.47852/bonviewJCCE62027642