

Original Article

A Logarithmic Mean Optimized EfficientNet with Joint Contrastive Learning for Improving Breast Cancer Detection from Mammogram Images

Sujitha Priya Duraipandi¹, Radha Venkatachalam²

^{1,2}Department of Computer Science, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, Tamil Nadu, India.

¹Corresponding Author : 21phcsf004@avinuty.ac.in

Received: 18 April 2026

Revised: 17 August 2026

Accepted: 26 August 2026

Published: 30 September 2026

Abstract - Early prediction of breast cancer plays a critical role in improving patient survival rates and effective treatment planning. Mammography is the most often utilized imaging method for early breast cancer monitoring because of its ability to detect microcalcifications and subtle tissue abnormalities at an early stage. However, Mammography suffers from low sensitivity in dense breast tissue, overlapping structures and inter-observer variability, leading to high false positive rates and potential misdiagnosis. Recently, Deep Learning (DL) techniques have been utilised for mammography-based breast cancer detection, as they can automatically learn features from images. However, some models struggle with the conflation of malignant and benign areas, irrelevant feature extraction, and suboptimal training strategies. Their reliance on labelled data and classification-focused learning limits discriminative feature representation, resulting in poor differentiation of visually similar benign and malignant tissues. Hence, in this paper, an advanced DL approach is developed for efficient breast carcinoma prediction using mammography images. Initially, collected mammography images are augmented to enhance data diversity and reduce overfitting. Then, EfficientNet B7 with Joint Contrastive Learning (ENB7-JCL) is developed to learn discriminative visual representations. In ENB7-JCL, EfficientNetB7 is employed for deep feature extraction from mammography images, while JCL jointly optimizes contrastive and classification objectives by modelling similarity and dissimilarity relationships between samples. This improves the class separability and reduces misclassification of visually similar benign and malignant tissues. Additionally, the Logarithmic Mean Optimization Algorithm (LMOA) is applied to optimize the hyperparameters of EfficientNet B7 for effective training. Finally, a softmax classifier is used to perform breast cancer classification. The complete work is termed Optimized ENB7-JCL (OENB7-JCL). The experimental states that the proposed model achieves 94.85% and 95.12% of accuracy on the MIAS and INBreast dataset outperforming other existing models.

Keywords - Breast cancer, Mammography, Deep Learning, Feature extraction, Hyperparameter tuning.

1. Introduction

Breast tumour is a cancerous condition that develops within breast tissues, most commonly the milk-producing lobules and ducts [1]. It is the leading cancer diagnosis among women and continues to be a major contributor to fatalities from malignancy cancer across the world [2]. According to the World Health Organization, more than 3 million women are treated for new breast carcinoma each year and account for nearly one-fourth of all instances of female cancer in the world [3]. The stage at which the disease is detected has a big impact on the success of therapy. Cancers, when detected in the early stages, have a much higher survival rate as well as a better treatment outcome [4]. Unfortunately, breast cancer is hard to detect in its early stages due to the presence of the disease, which often develops without any clear clinical symptoms or symptoms without clinically significant [5]. Magnetic

Resonance Imaging (MRI) and Positron Emission Mammography (PEM) are important imaging modalities to diagnose breast tumours [6]. Mammography is the most widely utilized screening technology because of its ability to obtain high-specific structural information of the breast tissue through the use of a low-energy X-ray [7].

Mammography has a high spatial resolution, and as such, it is very good at finding microcalcifications and small tumours, which are usually an early sign of breast cancer [8]. These characteristics are not easily detected early on by ultrasonography or MRI. In addition, it is relatively inexpensive and widely available in the therapeutic setting. The amount of radiation used is extremely low, and one can safely be screened repeatedly [9]. Digital mammography enhances the detection performance of mammograms over the



Screen Film Mammography (SFM) due to increased contrast resolution and picture augmentation [10, 11]. This is particularly a benefit to women who have a lot more breast tissue. As a consequence, mammography has long been the most popular way to detect early breast cancer.

Despite these benefits, mammography has major limitations. The interpretation of mammography is subject to some variation among radiologists, with an accompanying interobserver variability. Highly fibro-glandular breast tissue and structural features may obscure abnormalities, and therefore, there is the potential for undiagnosed or incorrectly diagnosed lesions. Variations in tumour form, size, and surrounding tissue density make precise diagnosis difficult and may lead to unnecessary diagnostic examinations and/or follow-up treatments.

Deep Learning (DL) models have recently started being an effective approach to predict and diagnose breast cancer [12]. DL approaches are efficient in units with medical techniques to take useful insights from the complex visual datasets [13]. Convolutional Neural Networks (CNNs) are a specific technique in DL methods that has the ability of automatic learning of complex and discriminative features from medical images that detect the existence of subtle patterns and anomalies that can appear in dense breast tissue [14]. DL-based systems help doctors to provide quicker, error-free and accurate breast cancer diagnosis by reducing differences in image interpretations and the number of false positive diagnoses [15].

Various DL models were established recently for breast cancer prediction using mammography. A DL approach [16] was developed for efficient breast carcinoma detection, inspired by GoogLeNet and residual learning. This model leveraged multi-scale feature extraction, shortcut connections and learnable activation functions to improve model performance. A Transfer Learning (TL) and advanced Al-Biruni Earth Radius Optimization (ABERO) [17] were developed for breast cancer classification. This model utilizes data augmentation, removing features by using AlexNet via TL and optimum identification through CNN. A Breast cancer prediction model [18] was suggested using gated attentive multimodal DL. This model produces layered characteristics via a sigmoid-gated attention CNN and employs flattening, thick layers, and loss techniques for bi-modal attention.

1.1. Problem Statement

Although recent DL models have significantly improved breast cancer detection, several important limitations still remain. Mammogram images often contain dense breast tissue, overlapping structures, and subtle differences between benign and malignant lesions, making accurate diagnosis difficult. Existing models often struggle to distinguish visually similar benign and malignant tissues. In other words, most methods focus only on improving classification accuracy and

do not effectively learn the similarity and dissimilarity between different samples, which may reduce feature discrimination. In addition, hyperparameters like learning rate, optimizer, batch size, and dropout are selected manually, resulting in unstable convergence, suboptimal model performance, and reduced generalization across different datasets.

1.2. Main Contributions

To solve the above-mentioned issues, this article proposes an Optimized EfficientNet B7 with Joint Contrastive Learning (OENB7-JCL) model for accurate and efficient breast cancer prediction. First, augmentation techniques like rotation, horizontal flipping, and random cropping are applied to amend the diversity of the dataset and reduce overfitting. Next, EfficientNet-B7 is employed to obtain deep and meaningful features from mammogram images.

To improve the discrimination between visually similar benign and malignant tissues, Joint Contrastive Learning (JCL) is incorporated, which effectively learns similarity and dissimilarity relationships among image samples, thereby enhancing feature representation. Furthermore, a Logarithmic Mean Optimization Algorithm (LMOA) is employed to automatically optimize important hyperparameters of EfficientNet-B7, leading to better convergence and improved classification performance. Finally, a Softmax classifier predicts whether the mammogram image is benign or malignant. Overall, the suggested OENB7-JCL improves the functionality of the early diagnosis and still retains the computational efficiency in clinical environments. The residual sections are planned as the following: literature survey covers earlier research based on DL based breast cancer predictions. The third part delineates the proposed technique, whereas the fourth part illustrates its evaluation of efficacy. The fifth part completes the manuscript and recommends avenues for further enhancement.

2. Literature Survey

A Multi-Feature Attention Network (MFAN) [19] was developed for breast cancer categorization. In this model, two important modules, like the Multi-Scale Spatial Channel Attention Module (McSCAM) and the Global-Local Attention Module (GLAM), were used. In stream combination, McSCAM preserves spatial properties and extracts high-order data from statistics, whereas GLAM assist to mitigating size discrepancies across attributes.

The regional and worldwide focus nodes assist the system in accurately identifying tiny disease areas via acquiring both globally and locally relevant data. However, this model results in high overfitting issues on small data and lowers generalizability. A preliminary mixed deep neural network to forecast malignancies in mammography utilizing Fourier characteristics [20] was suggested. In this model, CLAHE and a Wiener filter were applied to reduce the noise problem

through restricting contrasted enhancement for every surrounding pixel. Then, the 2D Discrete Wavelet Transform (DWT) was used for feature extraction. Finally, a CNN model was applied for breast cancer prediction. However, lower accuracy as the model struggles to learn the complex patterns from the images.

An Ensemble-based Channel and Spatial Attention Network (ECS-A-Net) [21] was devised for automated categorization of Breast Cancer images. In this method, various augmentation methodologies were applied to increase the volume of input. Then, a group method was utilized to parallel boost customized SE-ResNet50 as well as InceptionV3 provide a backbone for obtaining traits, subsequently using Channel Attention (CA) and Spatial Attention (SA) components for feature selection. Finally, Softmax was applied for breast cancer prediction. However, lower accuracy resulted due to increased model complexity and overfitting from ensemble backbones and extensive data augmentation. A Patch-based mammogram malignancy identification employing Attention-Driven CNNs [22] was created. Using a patch-based methodology, mammograms have been divided into 50×50 pixel patches to capture regional irregularities. A circular attention strategy was used to retrieve the important areas, enhancing sensitivity and reducing incorrect positives and fake negatives. Furthermore, histopathological data and microscopic cellular information were integrated to enhance the diagnostic performance. However, the computational complexity of this model was high.

A Dual-View DL model [23] was created for accurate breast cancer detection in mammograms. In this method, EfficientViewNet was applied to analyse both craniocaudal view and mediolateral oblique view mammograms to capture complementary breast tissue information that single-view models often miss. Two separate EfficientNet-based pathways are used to extract view-specific features, which are later fused for final classification. An optimized learning rate scheduling strategy is applied to enhance convergence and reduce false positives and false negatives. However, a missing or poor image view might result in decreased performance.

A DL model for breast carcinoma recognition with mixed removing attribute methods [24] was presented. In this method, separation utilizing Modified Otsu thresholding and identification of Canny edges was applied for the pre-processing task. Then, Gray Level Co-occurrence Matrix (GLCM), Gray Level Run Length Matrix (GLRLM), and initial-stage empirical descriptions were utilized for feature extraction. These features were input into a 2D BiLSTM-CNN network developed to recognize spatial and sequential structures in mammography images. However, dependence on handcrafted features and extensive preprocessing reduces scalability across diverse mammogram images. A DL based evaluation of mammographic images [25] was devised to

identify breast tumours via TL. In this model, CLAHE was applied for an effective pre-processing task. Then, previously trained approaches like Squeeze Net, VGG-16, AlexNet, Inception v3, and ResNet-50 were applied for obtaining features and recognition of malignancy phases in mammographic images. However, the hyperparameters of these models were not fine-tuned adequately, resulting in inefficient results.

A multi-autoencoder and attention-based structure [26] was presented for the diagnosis of breast carcinoma utilizing mammographic images. Multiple Autoencoders (MAE) were used to extract diverse and complex features to capture variations and detailed patterns in breast tissue. Attention Mechanisms (AM) were applied to emphasize the most important features and regions in the images. Finally, XGBoost (XGB) was used to classify images into malignant, benign, and normal cases. But lower performance due to higher computational cost and potential overfitting on limited datasets. An accurate classification technology was developed using mammography images to detect breast cancer [27].

The first process was to pre-process the mammography images using a filter to remove unwanted noise. A Level Set was then utilised on the image to isolate the significant portions of the image using a Thresholding-Based Level Set. It was then used to obtain the features using the Seagull Optimization Algorithm and the Improved Grey Wolf Optimization. Finally, the category of the samples (benign or malignant, non-cancerous or cancerous) was assessed with the help of the CatBoost. Nevertheless, it had unsatisfactory outcomes shown on larger datasets.

2.1. Research Gaps

Although existing DL models have shown promising performance in breast cancer detection, they still suffer from several limitations. Many models exhibit high computational complexity, overfitting, and limited generalization, particularly on small mammography datasets. In addition, the use of extensive preprocessing, complex attention mechanisms, and handcrafted feature extraction increases model complexity and reduces training efficiency.

Patch-based and multi-view approaches are also highly sensitive to image quality, which affects their robustness. Furthermore, inadequate hyperparameter optimization limits training stability and overall classification performance. Hence, there is a need for an efficient DL framework to enhance the precision, robustness, and generalization of breast cancer detection.

3. Proposed Methodology

3.1. Image Augmentation

This part provides a comprehensive illustration of the proposed framework. Figure 1 explains the structure of the suggested approach. The description of each module is

explained in detail for the exact prediction of breast cancer. DL models require large and diverse datasets to be reliable and widely applicable, however the amount of mammography datasets accessible for free use in breast cancer diagnosis is frequently limited.

This deficit reduces model generalization and increases the risk of overfitting. To address this issue, data augmentation is used to artificially extend the training dataset while keeping the diagnostic features of mammography images. This study uses geometric augmentation techniques to increase data diversity.

Each mammography patch is rotated at four distinct angles: 0, 90, 180, and 270 degrees. Horizontal flipping is then used to improve the final photos. Furthermore, random cropping is used to force the model to concentrate on different regions of the local breast tissue. This multimodal augmentation strategy helps the model learn features with rotation-invariance-, reflection-, and translation-invariance, improves the model's robustness to orientation fatigue and spatial fluctuations, and provides eight more samples from each of the original patches. As a result, overall classification performance improves while overfitting is decreased.

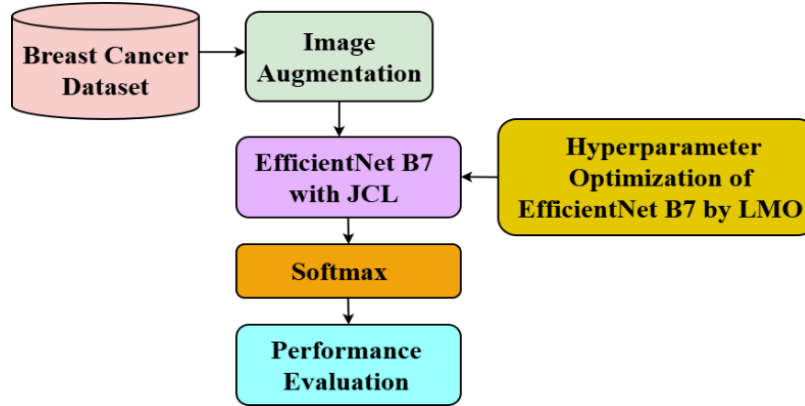


Fig. 1 Pipeline of the proposed work

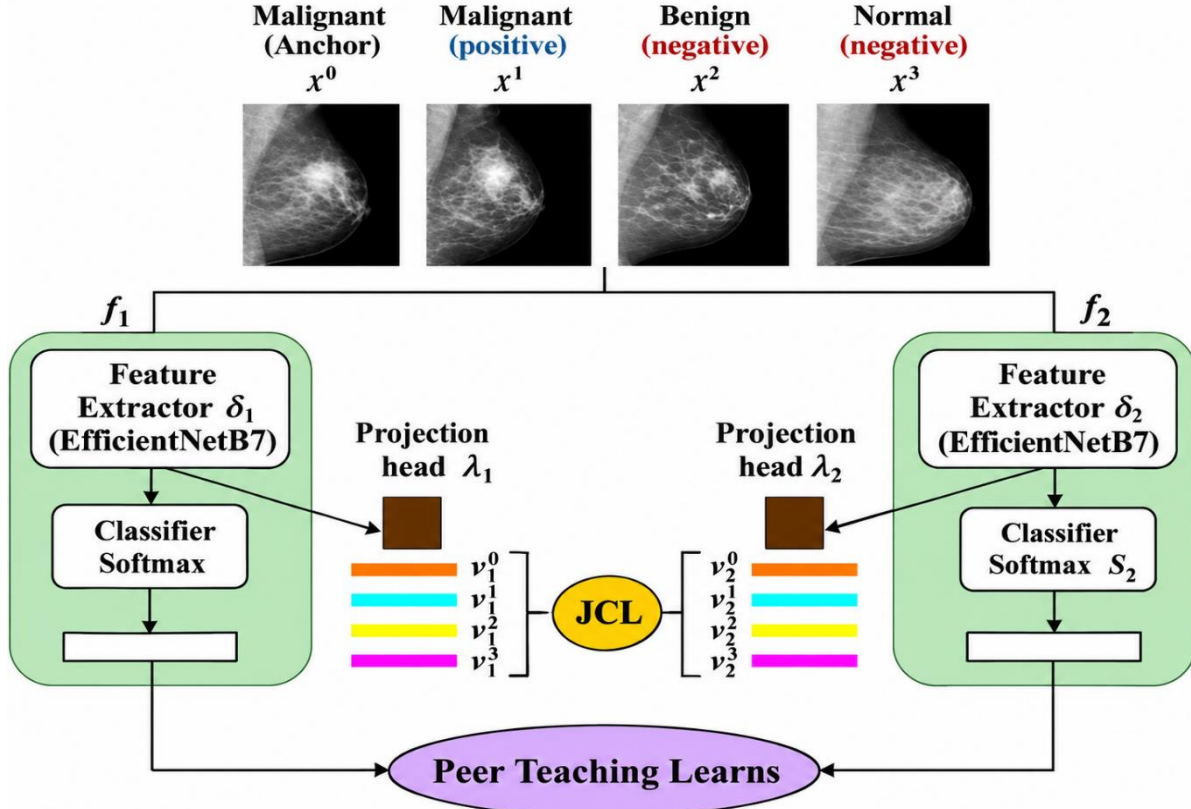


Fig. 2 ENB7-JCL model

3.2. Design of ENB7-JCL

ENB7-JCL is a DL model designed for the prediction of breast cancer, which is the integration of EfficientNet-B7 and JCL. In this model, EfficientNet-B7 is a core network to extract high-level attributes from the input images. Simultaneously, JCL refines the contrastive and classification

learning objectives to produce more discriminative feature representations of relationships amongst samples. By combining these components in this model, the separation between the classes is improved, and the misidentification of similar benign and malignant tissues is decreased. The structure of the ENB7-JCL model is shown in Figure 2.

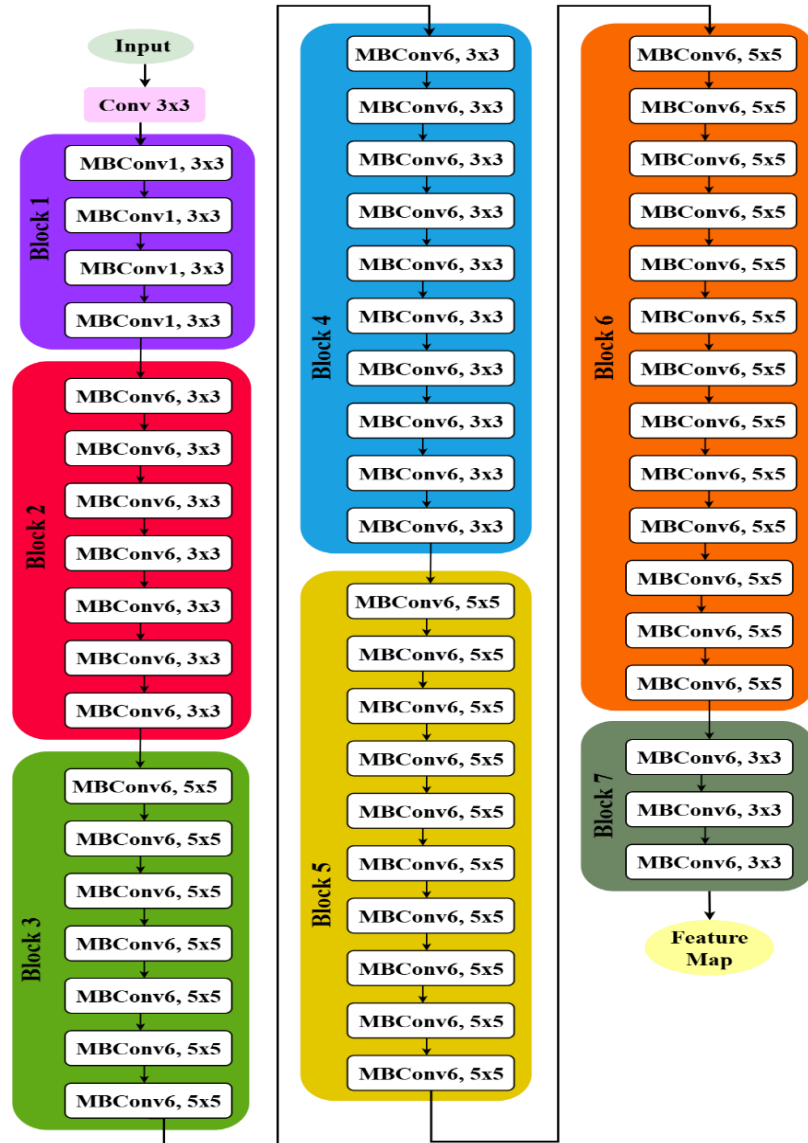


Fig. 3 Structure of the EfficientNet-B7 model

3.2.1 EfficientNet-B7

EfficientNet is one of the CNN architectures used in this work for discriminative feature extraction from high-dimensional images. EfficientNet uses a fusion scaling method to alter the depth, width, and quality of input in a consistent manner, utilizing constant scaling factors. Among the EfficientNet models, EfficientNet-B7 is the most advanced model due to its appropriate representation capacity, achieved through the construction of many layers and more

input resolutions. The convolution layer detects basic attributes of the input image, like outlines and fundamental textures. It was accomplished by utilizing a sequence of Mobile Inverted Bottleneck Convolution (MBConv) blocks, which act as the network's core.

Each MBConv block contains a depth-wise segmented convolution, a Squeeze-and-Excitation (SE) mechanism, and a projection layer. The SE units dynamically change the

distinctive actions of channels, allowing the system to concentrate on channels with a high diagnostic value. To stabilize multi-layer system training, residual (skip) connections are employed between compatible blocks for effective gradient flow. A batch normalization and the Swish activation function from EfficientNet-B7 are applied to improve the model's training with good convergence speed and stability. As the network's depth increases, the hierarchical feature extraction process enables the simulation to obtain fine-grained, intricate features as well as high semantic features simultaneously. Following the last stage of MBConv, global average pooling is utilized to combine spatial data into a compact feature representation.

The model is optimized using stochastic gradient-based backpropagation and a large number of pretrained weights. Figure 3 shows the structure of EfficientNet-B7. EfficientNet-B7 can extract discriminative features from weakly annotated medical images using TL. EfficientNet-B7 has demonstrated great generalization capabilities because of its balanced design and rapid decisions. It is particularly efficient at capturing minor texture variations and fine-grained patterns, which are crucial in accurately diagnosing breast cancer.

JCL is a representation learning methodology that makes maximum utilization of the contrastive and the classification objectives for better feature discrimination. It models similarities and differences connections between samples in an explicit manner, which is conducive to intra-class tightness, and inter-class separation in the learned trait domain. This simultaneous optimization makes it possible to benefit from better and more discriminative feature representations, particularly those representing classes that are visually similar. To get effective JCL uses four complementing loss functions like logit learning loss, KL. Divergence loss, cosine similarity loss and discriminative loss. The combined loss functions promote the model to learn features that ensure reliable classification and maintain proximity among related samples.

Logit Learning Loss: To effectively exploit the interaction among learned representations, JCL incorporates logit learning, which explicitly models embedding interactions across multiple networks to obtain more discriminative feature representations.

This learning strategy is initially formulated for two parallel networks (EfficientNetB7), denoted as f_a and f_b in which $a, b \in \{1, 2, \dots, M\}$ and $a \neq b$ is subsequently extended to a group of M systems $\{f_m\}_{m=1}^M = 1$. During logit learning, the network f_a is fixed while network f_b is enumerated.

For a given anchor embedding v_a^0 extracted from f_a , a positive embedding v_b^1 and a set of negative embeddings $\{v_a^k\}_{k=0}^{K+1}$ and $\{v_b^k\}_{k=0}^{K+1}$ corresponds to the same set of samples

$\{x_a^k\}_{k=0}^{K+1}$, where K indicates the quantity of negative samples. The contrastive probability distribution from f_a to f_b is defined in Equation (1),

$$q_{a \rightarrow b} = \text{softmax} \left(\left[\left(v_a^0 \cdot v_b^1 / \tau \right) \right], \left(v_a^0 \cdot v_b^2 / \tau \right), \dots, \left(v_a^0 \cdot v_b^{K+1} / \tau \right) \right) \quad (1)$$

where τ is the temperature parameter, the logit learning loss is then computed using cross-entropy over the contrastive distribution. $q_{a \rightarrow b}$ as in Equation (2),

$$\mathcal{L}_{logit}^{a \rightarrow b} = -\log q_{a \rightarrow b}^1 = -\log \frac{\exp \left(v_a^0 \cdot v_b^1 / \tau \right)}{\sum_{k=1}^{K+1} \exp \left(v_a^0 \cdot v_b^k / \tau \right)} \quad (2)$$

where, $q_{a \rightarrow b}^k$ is the k^{th} element of $q_{a \rightarrow b}$. Unlike conventional contrastive learning that constructs contrastive distributions within a single embedding space, the proposed logit learning leverages embeddings from different networks. This cross-network interaction explicitly models dependencies among representations, facilitating effective information exchange during training and yielding more discriminative feature embeddings.

For a group of M systems, logit learning is applied to every pair of systems to capture completely interconnected relationships. The general logit learning loss is defined as in Equation (3),

$$\mathcal{L}_{logit}^{1 \sim M} = \sum_{1 \leq a \leq b \leq M} (\mathcal{L}_{logit}^{a \rightarrow b} + \mathcal{L}_{logit}^{b \rightarrow a}) \quad (3)$$

KL Divergence Loss: To further enhance collaborative learning and ensure mutual calibration among networks, a KL divergence-based consistency loss is incorporated. In a given two networks f_a and f_b , an engaging contrast distribution $q_{a \rightarrow b}$ and $q_{b \rightarrow a}$ are obtained through logit learning. Enforcing consistency between these bidirectional distributions encourages aligned and stable representation learning. Accordingly, the KL divergence loss is defined to minimize the discrepancy between $q_{a \rightarrow b}$ and $q_{b \rightarrow a}$. When extended to a group of M systems, the overall KL divergence loss is formulated as in Equation (4),

$$\mathcal{L}_{KL}^{1 \sim M} = \sum_{a=1}^M \sum_{b=1, b \neq a}^M KL(q_{a \rightarrow b} || q_{b \rightarrow a}) \quad (4)$$

In this setting, $q_{a \rightarrow b}$ represents the soft label that is removed from slope back propagation to ensure consistency.

Cosine Similarity Loss: In the proposed framework, cosine similarity is employed to measure the semantic alignment between two correlated representations learned from the same mammography image. Cosine similarity

evaluates the angular distance between feature vectors and is well-suited for comparing normalized embeddings in a shared latent space.

Two representations are considered for contrastive learning. The first representation encodes the interaction-based features determined by concatenating the learned components, as denoted in Equation (5),

$$h_r = [u, v, a] \quad (5)$$

Here, u denotes the global deep features extracted from the EfficientNet-B7 backbone, capturing overall breast tissue structure and contextual information; v represents local or region-level features derived from deeper convolutional layers, focusing on fine-grained textures, masses, and micro-calcifications, and a corresponds to attention-enhanced discriminative features that highlight diagnostically relevant regions while suppressing background noise.

To capture the intricate relationship between these features, the interaction representation is processed through a transformer encoder, resulting in refined representations. H_u, H_v and H_a . The refined features are concatenated to form a final interaction-based embedding as in Equation (6),

$$h_{alu,v} = [H_u, H_v, H_a] \quad (6)$$

The second representation captures complementary semantic information extracted from the same mammography image and is obtained by encoding feature sequences using a transformer-based encoder as in Equation (7),

$$h_{u,v} = [H_{L_1}, \dots, H_{L_s}] \quad (7)$$

Although both $h_{alu,v}$ and $h_{u,v}$ represents the same underlying image that resides in different embedding spaces. To enable cosine similarity computation, projection functions are employed to map these representations into a shared latent space, as in Equation (8) and Equation (9),

$$h_r = \lambda_r(h_{alu,v}) = W_r \sigma(W_t h_{alu,v}) \quad (8)$$

$$h_e = \lambda_e(h_{u,v}) = W_e \sigma(Bert(h_{u,v})) \quad (9)$$

where, λ_r and λ_e denote projection heads, W_r , W_e and W_t are learnable weight matrices and $\sigma(\cdot)$ represents the activation function. After projection, both representations lie in the same latent space with equal dimensionality.

Cosine similarity is then used to construct positive and negative pairs. For a given anchor sample i , the pair $\langle h_{r,i}, h_{e,i} \rangle$ is treated as a positive pair, while $\langle h_{r,i}, h_{e,j} \rangle$, where $j \neq i$ forms a negative pair selected from the same mini-batch.

The cosine similarity-based contrastive loss is defined as in Equation (10),

$$\mathcal{L}_{cos} = -\log \frac{\exp(CS(h_{r,i}, h_{e,i})/\tau)}{\sum_{j=1}^N \exp(CS(h_{r,i}, h_{e,i})/\tau)} \quad (10)$$

where, $CS(\cdot, \cdot)$ indicates the cosine resemblance operation, and N represents mini-batch length. By minimizing this loss, representations corresponding to the same mammography image are pulled closer in the embedding space, while representations from different images are pushed apart, thereby improving feature consistency and discriminative learning.

Discriminative Loss: In the proposed ENB7-JCL framework, discriminative contrastive learning is employed to enhance the separability of mammography feature representations by explicitly distinguishing positive and negative sample pairs. This learning strategy utilizes anchor, positive and negative samples to enforce similarity between representations extracted from the same mammography image while increasing discrimination from representations of different images.

In conventional methods, sample pairs $\{x_i^1, x_i^2\}$ are treated as positives and $\{x_i^1, x_j^2\}$ ($i \neq j$) as negatives. In this approach, the negative set is further expanded by incorporating intra-negative pairs. $\{x_i^1, x_j^{neg}\}$ where j may be equal to i , thereby strengthening the discrimination, a discriminative operation, h_θ , is employed to guarantee that favorable pairings yield high scores, whilst unfavorable pairings produce lower levels. For an anchor representation v_i^1 , the discriminative contrastive loss is defined as in Equation (11),

$$\mathcal{L}_{disc}^{v_i^1} = -\log \frac{h_\theta(\{v_i^1, v_i^2\})}{\sum_{j=1}^{k+1} h_\theta(\{v_i^1, v_j^2\}) + \sum_{j=1}^{k+1} h_\theta(\{v_i^1, v_j^{neg}\})} \quad (11)$$

where k denotes the number of negative samples. The discriminative function $h_\theta(\cdot)$ is implemented using a non-parametric softmax formulation, which is computed as in Equation (12),

$$h_\theta(\{v_i^1, v_i^2\}) = \exp\left(\frac{v_i^1 \cdot v_i^2}{\|v_i^1\| \|v_i^2\| \tau}\right) \quad (12)$$

Equation (12) formulates the first network branch as the anchor. To ensure the symmetric learning and consistent representation alignment across both branches, the same loss is computed by treating the second branch as the anchor.

The final discriminative contrastive loss is obtained as in Equation (13),

$$\mathcal{L}_{disc}^{1 \sim M} = \mathcal{L}_{contrast}^{v^1} + \mathcal{L}_{contrast}^{v^2} \quad (13)$$

Overall Loss function in JCL

To fully exploit collaborative learning, all contrastive loss terms are aggregated to form the overall JCL loss as in Equation (14) across a group of M systems,

$$\mathcal{L}_{1 \sim M}^{MCL} = \alpha \mathcal{L}_{logit}^{1 \sim M} + \beta \mathcal{L}_{KL}^{1 \sim M} + \gamma \mathcal{L}_{CS}^{1 \sim M} + \Omega \mathcal{L}_{disc}^{1 \sim M} \quad (14)$$

In the proposed JCL framework, the weighting coefficients are set to $\alpha = \beta = 0.1$ for the logit learning loss and KL divergence loss, while $\gamma = \Omega = 1$ for the cosine similarity loss and discriminative contrastive loss, respectively, to balance collaborative representation learning effectively.

3.3. LMOA-based Hyper-Parameter Optimization EfficientNetB7

LMOA is applied to EfficientNet-B7 to automatically tune critical hyperparameters, including learning rates, batch size, and dropout values. By guiding the search toward optimal parameter settings, LMOA improves training stability and convergence.

The LMOA is employed in this work to optimize the hyperparameters of EfficientNet-B7, like Dropout, Activation, Momentum, Batch size, Training rate, Optimizer and Epochs. LMO is designed to overcome the limitations of conventional metaheuristic algorithms by enhancing convergence behaviour and preventing premature stagnation in local optima through logarithmic mean-based updates.

Mathematical Formulation of LMOA for Hyperparameter Optimization

Problem Definition

Consider the hyperparameter optimization problem, formulated as in Equation (15),

$$\min_{x \in S} f(x) \text{ or } \max_{x \in S} f(x) \quad (15)$$

where $f(x)$ represents the objective function (e.g., validation loss or classification accuracy of EfficientNet-B7) and $S = [lb, ub]$ specifies the search space defined by its lower bound and upper bound.

Initialization

These considered hyperparameter are defined by

N : Number of candidates hyperparameter solutions (Dropout, Activation, Momentum, Batch size, Training rate, Length, Optimizer, Epochs and Loss Function)

$Iter_{max}$: Maximum iterations

α, β : Control variables for exploitation with exploration

Each candidate solution x_i represents a set of EfficientNetB7 hyperparameters initialize as in Equation (16),

$$x_i = lb + rand(.). (ub - lb), i = 1, 2, \dots, N \quad (16)$$

Fitness Evaluation

Every candidate solution x_i , EfficientNetB7 is trained using the corresponding hyperparameters and the fitness is evaluated as in Equation (17),

$$f_i = f(x_i), \quad i = 1, 2, \dots, N \quad (17)$$

where, f_i reflects the model's performance.

Logarithmic Mean-Based Update Rule

Considering every combination of candidate solutions x_i and x_j , the logarithmic average is computed as in Equation (18) and Equation (19),

$$L(x_i, x_j) = \frac{\ln|x_i| - \ln|x_j|}{x_i - x_j}, x_i \neq x_j \quad (18)$$

If $x_i = x_j$, then

$$L(x_i, x_j) = x_i \quad (19)$$

A new candidate solution is generated using Equation (20),

$$x_{new} = L(x_i, x_j) + \beta \cdot (rand(-0.5)) \cdot (ub - lb) \quad (20)$$

where, $rand$ represents the random number between 0 and 1.

Adaptive Exploration-Exploitation Control

An adaptive control parameter (α_t) is defined as in Equation (21),

$$\alpha_t = \frac{t}{Iter_{max}} \quad (21)$$

The final updated solution is obtained by combining exploration and exploitation as in Equation (22),

$$x_{final} = (1 - \alpha_t) \cdot x_{new} + \alpha_t \cdot x_{best} \quad (22)$$

where the best hyperparameter is defined as in Equation (23),

$$x_{best} = \operatorname{argmin} f(x_i) \quad (23)$$

The best objective value is defined as in Equation (24),

$$f_{best} = f(x_{best}) \quad (24)$$

Boundary Handling

To ensure feasibility, the updated solution is replaced if the new solution improves the objective, as in Equation (25),

$$x_{final} = \min(\max(x_{final}, lb), ub) \quad (25)$$

Selection

The most ineffective approach is replaced if the new solution improves the objective, as in Equation (26),

$$x_{worst} = \operatorname{argmin} f(x_i) \quad (26)$$

Stopping Criterion:

The optimization process terminates when

- $t \geq Iter_{max}$, or
- $|f_{best}(t) - f_{best}(t-1)| < \epsilon$

where ϵ represents the minimal threshold, based on these steps, the LMOA effectively optimizes the hyperparameters of EfficientNet-B7, leading to improved convergence and classification performance. Figure 4 depicts the flow work of LMOA for EfficientNet hyperparameter optimization.

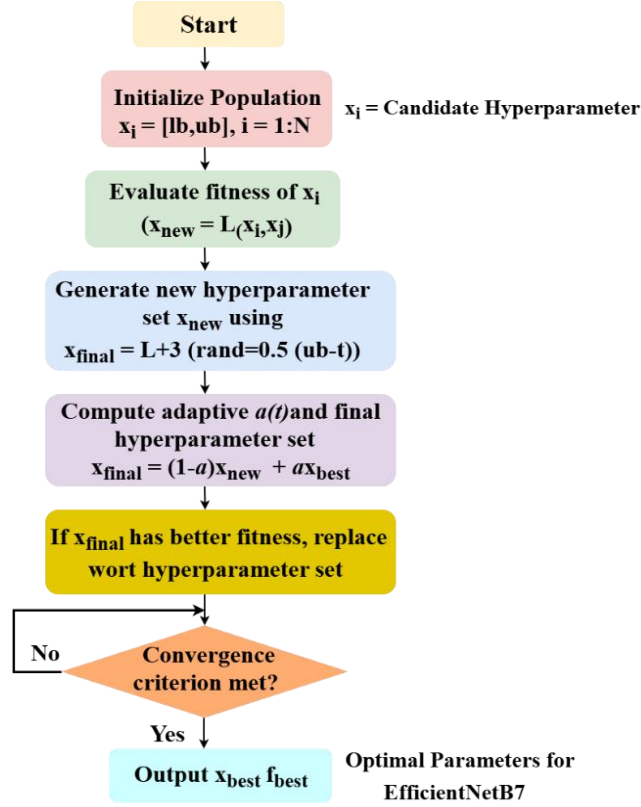


Fig. 4 Flow work of LMOA for EfficientNet hyperparameter optimization

Algorithm 1. LMO-based Hyperparameter Optimization

Input: Objective function $f(x_i)$, Lower and upper bounds (lb, ub), Population size (N), Maximum iterations $Iter_{max}$
Output: Optimal hyperparameter values (x_{best}), Best objective value (f_{best})

1. Start
2. \Initialization
3. Generate N random candidate hyperparameter solutions within $[lb, ub]$ using Equation (16).
4. Evaluate the objective function $f(x_i)$ for each solution using Equation (17).
5. Identify the current best solution. x_{best} using Equation (23).
6. Compute the corresponding best objective value. f_{best} using Equation (24).
7. \Iterative Optimizations
8. For $t = 1$ to $Iter_{max}$
9. For each selected combination of resolutions, (x_i, x_j) :
 10. Calculate logarithmic average $L(x_i, x_j)$ using Equation (18)
 11. If $x_i = x_j$, apply Equation (19)
 12. Generate a new solution (candidate hyperparameters) x_{new} using Equation (20)
 13. Modify the equilibrium between exploration and exploitation, utilising α_t as delineated in Equation (21)
 14. Update the final solution as per Equation (22)
 15. Apply boundary handling to ensure feasibility using Equation (25)
 16. Evaluate the objective function $f(x_{final})$ using Equation (17)
 17. Identify the worst-performing solution using Equation (26) and replace it if x_{final} yields a better objective value.
 18. Update x_{best} and f_{best} If a better solution is found.
 19. \Stopping Criterion
 20. Stop when the highest quantity of iterations $Iter_{max}$ achieved or convergence is satisfied.

21. Return the Best Solution
22. Output optimal hyperparameter set (x_{best}) and Best objective value (f_{best})

3.4. Classification

ENB7-JCL extracts high-level discriminative feature embeddings from mammography images by combining deep feature extraction of EfficientNet-B7 with JCL. These embeddings capture both semantic information and similarity-dissimilarity relationships, making them suitable for accurate breast cancer classification. The feature representation (h) of ENB7-JCL is denoted as in Equation (27),

$$h = \delta(x) \quad (27)$$

where x is the input mammography image and $\delta(\cdot)$ represents the feature extractor. These features are processed via an entirely linked tier to produce a logit vector as in Equation (28),

$$z = Wh + b \quad (28)$$

Here, W and b signify the weight matrix and bias vector, respectively. The softmax function is then applied to determine the class probabilities as in Equation (29),

$$P(y = c|x) = \frac{\exp(z_c)}{\sum_{j=1}^C \exp(z_j)} \quad (29)$$

where C and y represent the total number of classes and true class labels (benign and malignant). z_c denote the logit equivalent to class c . The predicted class label is determined as in Equation (30),

$$\hat{y} = \underset{c}{\operatorname{argmax}} P(y = c|x) \quad (30)$$

$\hat{y}$ is the predicted class label produced by the model.

Algorithm 2. OENB7-JCL for breast Cancer prediction

Input: Collected breast cancer images

Output: Corresponding class labels (Benign / Malignant)

1. Acquire the image data
2. \Augmentation
3. Implement augmentation methods such as rotation, horizontal flipping, random cropping, and normalize pixel intensity values.
4. \ENB7-JCL (EfficientNetB7)
5. Initialize EfficientNet-B7 with pretrained ImageNet weights.
6. Pass augmented images through EfficientNet-B7.
7. Extract deep feature representations from the final convolutional layers.
8. Apply global average pooling to obtain compact feature vectors.

9. \ENB7-JCL (JCL)
10. Generate paired feature representations from the same image.
11. Construct positive pairs (same image) and negative pairs (different images).
12. Compute the following losses.
13. Apply logit learning loss to learn discriminative relationships between samples by modelling inter-sample similarity across network branches.
14. Use KL divergence loss to enforce consistency and alignment between contrastive distributions learned by different network branches.
15. Perform cosine similarity loss to align feature representations of the identical image while distinguishing depictions of other images.
16. Apply discriminative contrastive loss to maximize inter-class separability and minimize intra-class variation between benign and malignant features.
17. Combine all losses to form the overall JCL loss.
18. Update EfficientNet-B7 weights using backpropagation
19. \Hyperparameter Optimization using LMOA
20. Generate a population of candidate hyperparameter sets.
21. Measure model performance for each candidate.
22. Modify hyperparameters using logarithmic mean-based update rules.
23. Determine the best hyperparameter configuration using validation performance.
24. \Classification
25. Feed the optimized feature vectors to a fully connected layer
26. Employ the Softmax algorithm to compute the probability for each class.
27. Assign the class label (Benign / Malignant) with the highest probability

4. Result and Discussion

4.1. Dataset Description

- Mammographic Image Analysis Society (MIAS) Dataset [28]: This MIAS dataset is a popular public mammography dataset for breast cancer research and algorithm benchmarking. It contains 322 digitized mammogram images from 161 patients with one image per breast, acquired from film mammograms, digitizing resolution of approximately 50-micron and in PGM format. Each image is annotated with ground truth information of what the image contains (normal tissue or some abnormality such as mass or calcifications), along with some metadata information such as the location of the lesion, size of the lesion, and benign or malignant nature of the lesion.
- INbreast dataset [29]: It is a database of full-field online mammography extensively built for advanced breast cancer research. It includes 410 DICOM images from 115

cases, such as 90 cases of bilateral with four images each and 25 cases of mastectomy, each with two images each. Acquired from the latest digital mammography systems, the dataset contains several kinds of diseases, such as mass, calcification, imbalance, and structural disorders. Detailed annotations by experts with exact contours of lesions are available in XML files, with high image quality and accurate ground truth. This dataset is one of the most reliable open mammography datasets. Moreover, it can be found at: [29]. With the help of this dataset, an Augmented INbreast Dataset was derived and expanded to a total of 7,632 images. However, it still possessed class bias, which was mitigated in this research through additional augmentation. It is derived from [28].

4.2. Experimental Setup and Performance Metrics

The proposed and existing models are executed on a Windows 11 system using a laptop equipped with an Intel® Core™ i5-4210 CPU @ 3 GHz, 4 GB RAM, and a 1 TB HDD, running 64-bit Windows 10. The software environment was configured with Python 3.10, employing PyTorch 2.2 and TorchVision 0.17 as the primary DL.

The parameters used for experiments are listed in Table 1. The collected dataset was divided into 70% for training and 30% for testing, with the corresponding number of samples in each subset reported in Table 2, Table 3 and Table 4, respectively. The evaluation measures employed for evaluating the performance of the proposed OENB7-JCL model are illustrated below.

Accuracy: It quantifies the entire ratio of accurately categorized events within each category.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (31)$$

In Equation (31), True Positives (TP): Dangerous breast cancer instances accurately classified as cancerous. True Negatives (TN): Cancerous breast tissue samples properly classified as abnormal.

False Positives (FP): Mild instances misclassified as cancerous. False Negatives (FN): Carcinoma instances erroneously categorized as harmless.

Precision: It measures the proportion of accurately identified tumours among the favourably anticipated instances, indicating the validity of carcinoma identification. This may be articulated in Equation (32),

$$Precision = \frac{TP}{TP+FP} \quad (32)$$

Specificity: It evaluates the capacity to accurately recognize harmless instances, hence reducing counterfeit positives. This may be articulated as demonstrated in Equation (33),

$$Specificity = \frac{TN}{TN+FP} \quad (33)$$

F1-Score: The harmonic mean of precision and recall provides a singular metric that equilibrates sensitivity and provides a singular metric that equilibrates sensitivity and specificity. This may be articulated in Equation (34),

$$F1 - score = 2 \times \frac{Precision \times Sensitivity}{Precision + Sensitivity} \quad (34)$$

Table 1. List of optimal hyperparameters of EfficientNetB7

Parameters	Search Space	Optimal Range
Activation	[Tanh, Sigmoid, ReLU]	ReLU
Batch size	[32, 64, 96]	64
Dropout	[0.2, 0.3, 0.4]	0.3
Epochs	[50, 100, 150, 200]	100
Length	[25, 50, 100]	50
Loss Function	[Cross-entropy, Mean Squared Error]	Cross Entropy
Momentum	[0.6, 0.7, 0.8]	0.7
Optimizer	[Adam, NAdam, RMSprop, Stochastic Gradient Descent]	NAdam
Training rate	[0.01, 0.001, 0.0001]	0.001

Table 2. Datasets splitting

Dataset	Training (70%)	Testing (30%)	Total
MIAS dataset	225	97	322
INbreast dataset	287	123	410

Table 3. Data splitting of the MIAS dataset

Classes	Before Augmentation			After Augmentation		
	Training (70%)	Testing (30%)	Total	Training (70%)	Testing (30%)	Total
Normal	145	62	207	161	69	230
Benign	48	21	69	147	63	210
Malignant	38	16	54	140	60	200
Total			330	Total		640

Table 4. Data splitting of the INbreast dataset

Classes	Before Augmentation			After Augmentation		
	Training (70%)	Testing (30%)	Total	Training (70%)	Testing (30%)	Total
Benign	1764	756	2520	3640	1560	5200
Malignant	3578	1534	5112	3710	1590	5300
Total			7632	Total		10500

4.3. Performance Evaluation

The confusion matrix of the test split for the MIAS and INbreast dataset is given in Figure 5(a) and Figure 5(b) accordingly.

4.3.1 Training, Testing Accuracy and Loss

Figure 6(a) illustrates the accuracy of training and testing of the OENB7-JCL framework on the MIAS dataset for 100 epochs. The model has a steady improvement in performance, overall training accuracy improving from around 70% at epoch 10 to around 99% at epoch 100, while the testing accuracy improves from around 55% and stabilizes around 98%, indicating a good level of generalization and little overfitting.

Figure 6(b) plots the training and testing loss values for the MIAS dataset, which shows a steady decrement in the loss values at each epoch. The training loss is reduced from around 0.20 at epoch 10 to 0.02 by epoch 100, while the testing loss is reduced from around 0.15 to 0.03, which denotes stable convergence and minimal overfitting. Figure 7(a) shows the

accuracy trends on the INbreast dataset, and it is observed that training accuracy goes from nearly 65% at epoch 10 to about 99% at epoch 100, and testing accuracy goes from approximately 55% to about 98%, which confirms the robustness of the proposed approach. Figure 7(b) shows the loss trends for the INbreast dataset, showing the reduction of training loss from ~0.22 to 0.02 and also tests loss trends from ~0.17 to 0.03 till the end of the last epoch, which further validates the affectivity and generalization ability of the proposed OENB7-JCL technique. The proposed OENB7-JCL model shows stable development consisting of stable enhancements in acquisition and categorization over the epochs of training. Both the accuracy for training and testing data increase smoothly, and the loss value corresponding to training and testing data decreases steadily, indicating steady convergence and effective optimization of the features. The good agreement between the training and test accuracy and loss curves indicates good generalization ability with little overfitting. These results ensure the robustness of the proposed framework. Overall, OENB7-JCL is effective in breast cancer classification on both datasets.

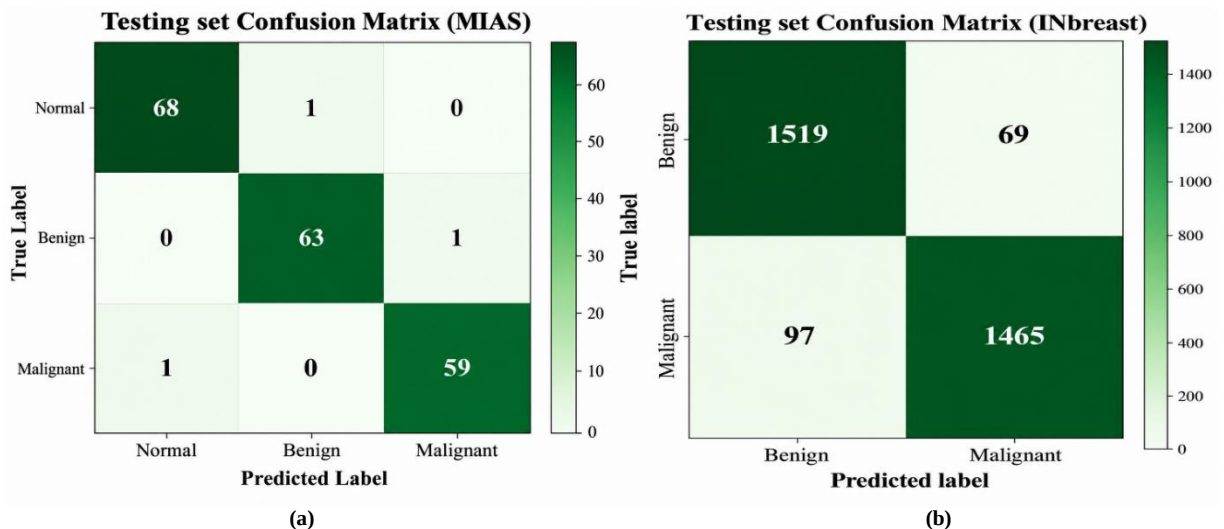


Fig. 5 Confusion matrix for: (a) MIAS dataset, and (b) INbreast dataset.

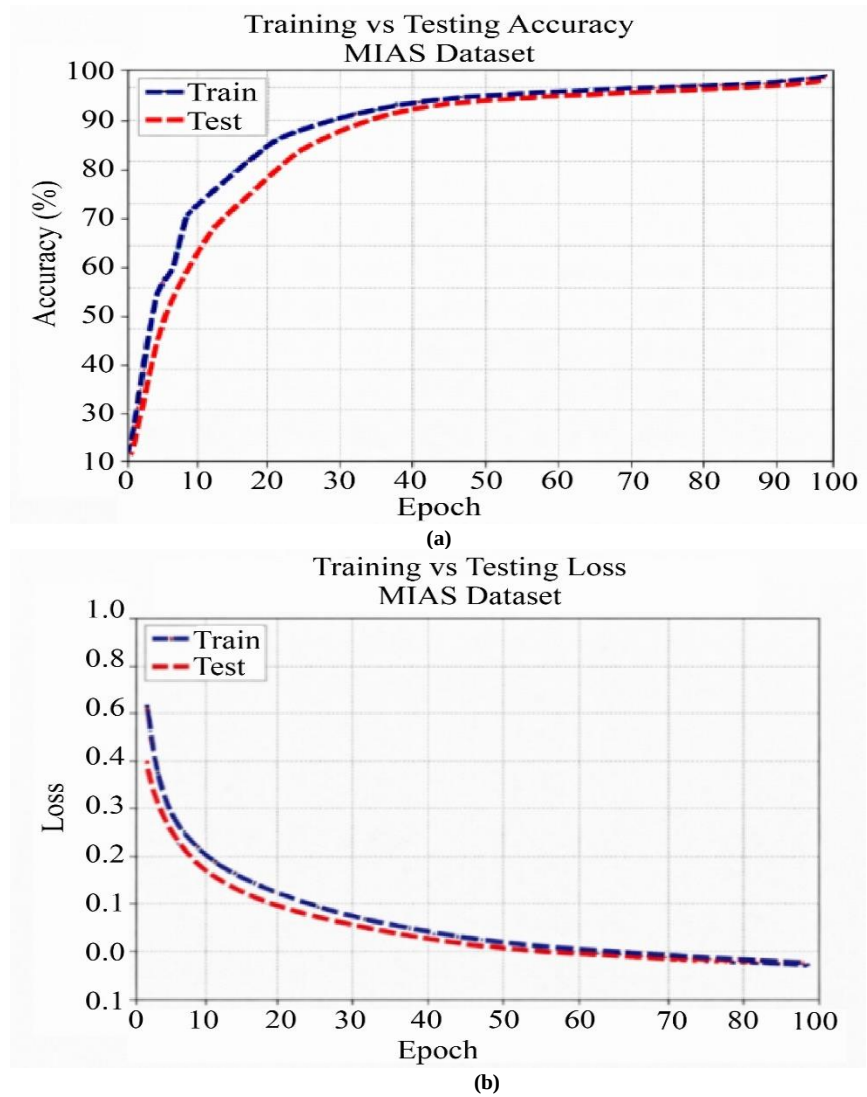
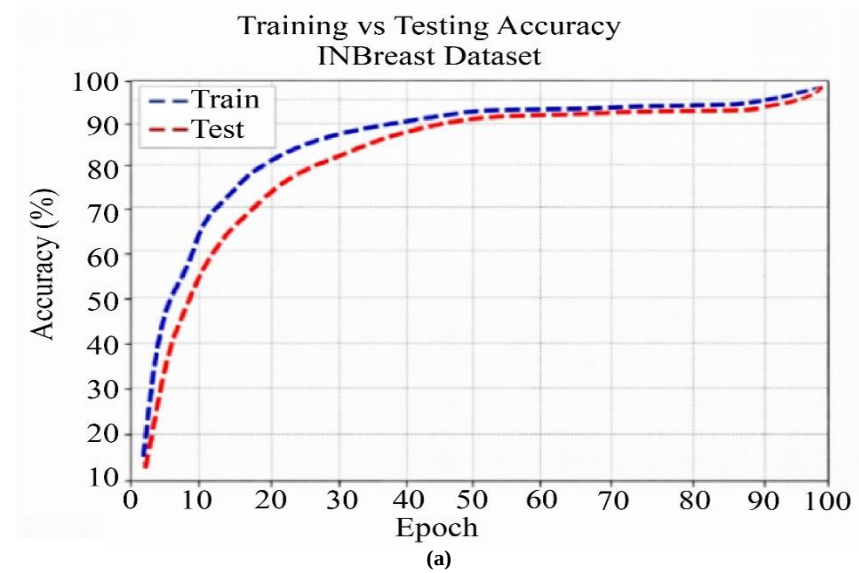


Fig. 6 Proposed model's training and testing: (a) Accuracy, and (b) Loss on MIAS dataset.



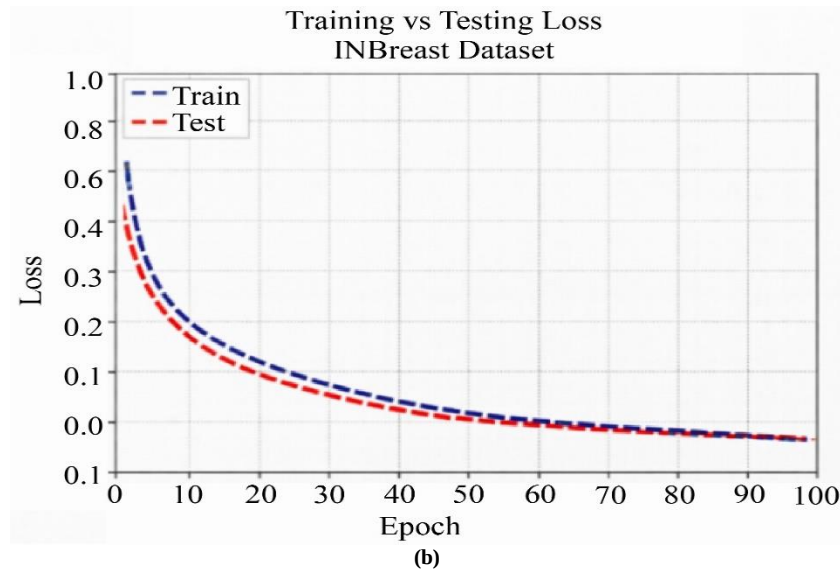


Fig. 7 Proposed model's training and testing: (a) Accuracy, and (b) Loss on INBreast dataset.

4.3.2 Comparison of the Proposed Model with Related Models

This part provides an unbiased evaluation of the proposed OENB7-JCL model against existing approaches, namely TL-ABERO [17], MFAN [19], ECS-A-Net [21] and MAE-AM-XGB [27], employing the datasets and evaluation criteria detailed in the dataset description and Experimental setup and performance metrics.

Figure 8 presents a comparison between the proposed approach and existing methods using the MIAS dataset. The proposed OENB7-JCL model achieves significantly higher performance across all evaluation metrics. In this analysis, the accuracy of OENB7-JCL is 20.16%, 13.78%, 9.11%, and 4.24% higher than TL-ABERO, MFAN, ECS-A-Net and MAE-AM-XGB, respectively.

Similarly, precision improves by 20.61%, 14.32%, 9.47%, and 4.08%; specificity by 20.58%, 14.03%, 9.44% and

4.98%; sensitivity by 20.27%, 13.98%, 9.31%, and 4.12%; and F1-score by 20.46%, 14.17%, 9.38%, and 4.29% compared to the same models on the MIAS dataset.

Figure 9 displays an analysis of proposed and existing approaches using the INbreast dataset. Specifically, its accuracy exceeds that of TL-ABERO, MFAN, ECS-A-Net, and MAE-AM-XGB by 18.89%, 14.81%, 10.06%, and 6.30%, respectively. Similarly, improvements in precision are 18.36%, 14.24%, 9.57%, and 4.76%, while specificity increases by 19.63%, 15.40%, 10.73%, and 6.54%.

Furthermore, OENB7-JCL demonstrates higher sensitivity by 18.14%, 13.85%, 9.29%, and 6.06% and superior F1-scores of 18.77%, 15.58%, 9.92%, and 5.21%, respectively. The outcomes unequivocally demonstrate that the proposed model consistently outperforms existing approaches on the INbreast dataset.

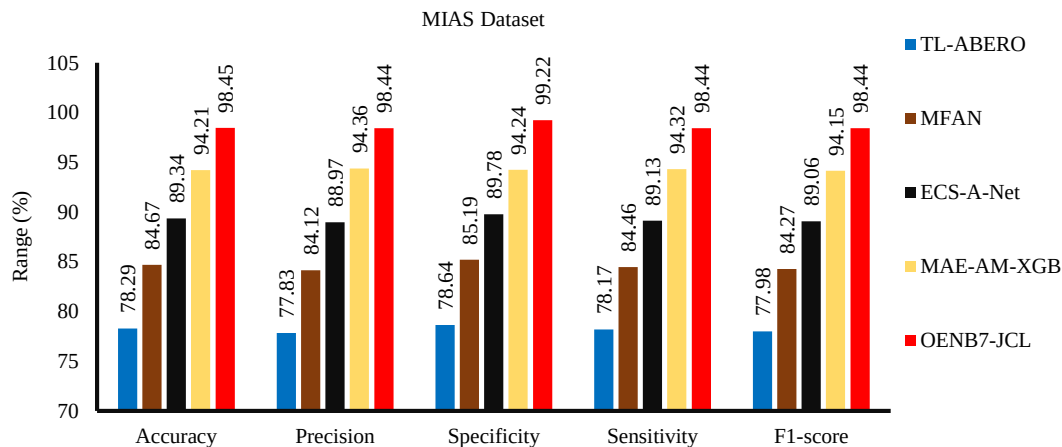


Fig. 8 Performance comparison of the proposed and existing models on the MIAS dataset

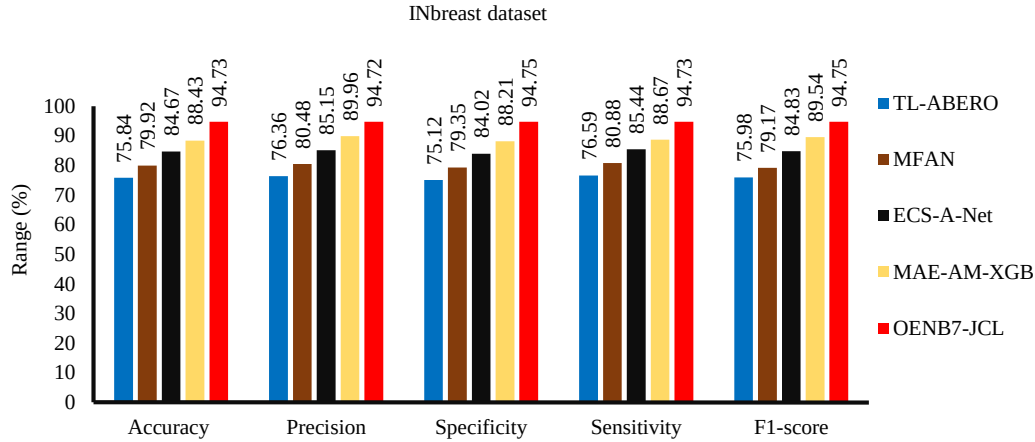


Fig. 9 Performance comparison of the proposed and existing models on the INbreast dataset

4.4. Ablation Study

Table 5 shows the ablation study on MIAS and INbreast datasets for different model variants. This ablation results highlight the contribution of individual optimization and learning components across both MIAS and INbreast datasets. The EfficientNet-B7 performance at the baseline level is moderate, which is greatly improved by LMOA, indicating the importance of hyperparameter optimization.

Logit learning also achieves performance improvements by learning about relationships between samples. The addition of KL loss and cosine similarity helps in better representation consistency and better alignment. Discriminative learning enhances the class separation and effectiveness. The OENB7-JCL model achieves the best accuracy and F1-score, making it useful to incorporate optimization and contrastive learning objectives.

4.5. T-Test Validation

Table 6 provides the combined t-test evaluation for proposed and current approaches as regards accuracy. These results indicate that the proposed OENB7-JCL model achieves statistically significant performance improvements over most existing methods on both MIAS and INbreast datasets. For TL-ABERO, MFAN and ECS-A-Net, the obtained p-values

($p < 0.05$) and positive confidence intervals confirm meaningful accuracy gains. The relatively high t-values reflect great and consistent improvements across folds. In contrast, comparisons with MAE-AM-XGB show marginal significance due to smaller mean differences and limited fold size. Overall, the statistical analysis validates the reliability of the proposed model's performance.

4.6. Discussion

The proposed OENB7-JCL model achieves better performance than existing methods because it combines efficient feature extraction, discriminative representation learning, and automatic hyperparameter optimization in a single framework. EfficientNet-B7 extracts rich features from mammography images, while JCL enhances the distinction between non-cancerous and cancerous tissues by learning similarity and dissimilarity relationships among samples.

In addition, LMOA automatically selects optimal hyperparameters, resulting in stable training and better generalization. Data augmentation further reduces overfitting by increasing the diversity of training samples. As a result, the proposed model achieves higher accuracy and more reliable breast cancer classification than the prevailing advanced methodologies.

Table 5. Ablation study on both datasets for different model variants

Model Variants	Accuracy (%)	F1-Score (%)
MIAS Dataset		
EfficientNet-B7 (Baseline)	82.32	81.89
EfficientNet-B7 + LMOA	85.54	85.12
EfficientNet-B7 + Logit Learning	87.08	86.72
EfficientNet-B7 + KL Loss	89.46	90.03
EfficientNet-B7 + Cosine Similarity	88.97	89.58
EfficientNet-B7 + Discriminative Learning	87.26	88.84
ENB7-JCL	92.14	91.82
OENB7-JCL (Proposed Model)	98.45	98.44

INbreast Dataset		
EfficientNet-B7 (Baseline)	83.32	83.89
EfficientNet-B7 + LMOA	87.07	87.56
EfficientNet-B7 + Logit Learning	88.64	89.02
EfficientNet-B7 + KL Loss	89.12	90.48
EfficientNet-B7 + Cosine Similarity	88.54	88.89
EfficientNet-B7 + Discriminative Learning	90.76	89.08
ENB7-JCL	93.68	93.35
OENB7-JCL (Proposed Model)	94.73	94.73

Table 6. Paired t-test analysis of the accuracy of the proposed vs existing models (four independent runs)

Existing models	t-value	df	Sig. (2-tailed)	Mean Difference (%)	Std. Dev	95% Confidence Interval	
						Lower	Upper
MIAS dataset							
TL-ABERO	5.91	3	0.010	12.53	4.21	10.86	15.73
MFAN	4.88	3	0.016	9.31	3.82	7.23	11.25
ECS-A-Net	3.74	3	0.033	6.39	3.41	4.96	8.19
MAE-AM-XGB	2.96	3	0.059	4.11	2.77	2.29	6.22
INbreast dataset							
TL-ABERO	5.34	3	0.013	11.80	4.06	9.63	15.23
MFAN	4.46	3	0.021	8.94	3.71	6.84	11.01
ECS-A-Net	3.42	3	0.041	6.12	3.32	4.84	7.95
MAE-AM-XGB	2.71	3	0.072	4.38	2.69	2.77	5.37

5. Conclusion

In this paper, OENB7-JCL is proposed for efficient breast cancer prediction. Initially, the collected mammography images are augmented to enhance data diversity and reduce overfitting. Then, ENB7-JCL is developed to learn discriminative visual representations. In ENB7-JCL, EfficientNet B7 is employed for deep feature extraction from mammography images, while JCL jointly optimizes contrastive and classification objectives by modelling similarity and dissimilarity relationships between samples, thereby improving class separability and reducing misclassification of visually similar benign and malignant

tissues. Further, LMOA is used to fine-tune the hyperparameters of EfficientNet B7 for great training works. To conclude, breast cancer classification is carried out by using a Softmax classifier. The results obtained from the research show that the proposed model achieves 94.85% as well as 95.12% accuracies using MIAS and INBreast datasets, respectively, as compared to all the existing models for comparison.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

References

- [1] Adrienne G. Waks, and Eric P. Winer, "Breast Cancer Treatment: A Review," *JAMA*, vol. 321, no. 3, pp. 288-300, 2019. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [2] Joshi Zankhna Rajeshbhai, Mayank Bapna, and Bhagirath Patel, "Breast Cancer and its Treatments: A Comprehensive Review," *Asian Pacific Journal of Health Sciences*, vol. 8, no. 2, pp. 90-94, 2021. [[CrossRef](#)] [[Publisher Link](#)]
- [3] Tracy-Ann Moo et al., "Overview of Breast Cancer Therapy," *PET Clinics*, vol. 13, no. 3, pp. 339-354, 2018. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [4] Ganesh N. Sharma et al., "Various Types and Management of Breast Cancer: An Overview," *Journal of Advanced Pharmaceutical Technology and Research*, vol. 1, no. 2, pp. 109-126, 2010. [[Google Scholar](#)] [[Publisher Link](#)]
- [5] Christoph Engel, and Christine Fischer, "Breast Cancer Risks and Risk Prediction Models," *Breast Care*, vol. 10, no. 1, pp. 7-12, 2015. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [6] Subbhuraam Vinitha Sree et al., "Breast Imaging: A Survey," *World Journal of Clinical Oncology*, vol. 2, no. 4, pp. 171-178, 2011. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [7] William E. Barlow et al., "Prospective Breast Cancer Risk Prediction Model for Women Undergoing Screening Mammography," *Journal of the National Cancer Institute*, vol. 98, no. 17, pp. 1204-1214, 2006. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]

- [8] Lars J. Grimm et al., "Benefits and Risks of Mammography Screening in Women Ages 40 to 49 Years," *Journal of Primary Care and Community Health*, vol. 13, pp. 1-6, 2022. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [9] Jocelyn Lippey et al., "Impact of a Risk based Breast Screening Decision Aid on Understanding, Acceptance and Decision Making," *NPJ Breast Cancer*, vol. 9, no. 1, pp. 1-11, 2023. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [10] Mostafa Alabousi et al., "Performance of Digital Breast Tomosynthesis, Synthetic Mammography, and Digital Mammography in Breast Cancer Screening: A Systematic Review and Meta-Analysis," *Journal of the National Cancer Institute*, vol. 113, no. 6, pp. 680-690, 2021. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [11] Rachel Farber et al., "Impact of Full-Field Digital Mammography Versus Film-Screen Mammography in Population Screening: A Meta-Analysis," *Journal of the National Cancer Institute*, vol. 113, no. 1, pp. 16-26, 2021. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [12] Dooman Arefan et al., "Deep Learning Modeling using Normal Mammograms for Predicting Breast Cancer Risk," *Medical Physics*, vol. 47, no. 1, pp. 110-118, 2020. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [13] Lulu Wang, "Mammography with Deep Learning for Breast Cancer Detection," *Frontiers in Oncology*, vol. 14, pp. 1-16, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [14] Zahra Jafari, and Ebrahim Karami, "Breast Cancer Detection in Mammography Images: A CNN-based Approach with Feature Selection," *Information*, vol. 14, no. 7, pp. 1-14, 2023. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [15] Olasubomi J. Omoleye et al., "External Evaluation of a Mammography-based Deep Learning Model for Predicting Breast Cancer in an Ethnically Diverse Population," *Radiology: Artificial Intelligence*, vol. 5, no. 6, pp. 1-11, 2023. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [16] Salman Zakareya, Habib Izadkhah, and Jaber Karimpour, "A New Deep-Learning-based Model for Breast Cancer Diagnosis from Medical Images," *Diagnostics*, vol. 13, no. 11, pp. 1-23, 2023. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [17] Amel Ali Alhussan et al., "Classification of Breast Cancer using Transfer Learning and Advanced AI-Biruni Earth Radius Optimization," *Biomimetics*, vol. 8, no. 3, pp. 1-24, 2023. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [18] Safak Kayikci, and Taghi M. Khoshgoftaar, "Breast Cancer Prediction using Gated Attentive Multimodal Deep Learning," *Journal of Big Data*, vol. 10, no. 1, pp. 1-11, 2023. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [19] Inzamam Mashood Nasir, Masad A. Alrasheedi, and Nasser Aedh Alreshidi, "MFAN: Multi-Feature Attention Network for Breast Cancer Classification," *Mathematics*, vol. 12, no. 23, pp. 1-15, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [20] Rengarajan Karthiga et al., "A Novel Exploratory Hybrid Deep Neural Network to Predict Breast Cancer for Mammography based on Wavelet Features," *Multimedia Tools and Applications*, vol. 83, no. 24, pp. 65441-65467, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [21] Su Myat Thwin et al., "Attention-based Ensemble Network for Effective Breast Cancer Classification Over Benchmarks," *Technologies*, vol. 12, no. 2, pp. 1-19, 2024. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [22] Santosh Kumari Dhiman, and Rakesh Yadav, "Patch-based Mammogram Cancer Detection using Attention-Driven CNNs," *Advances in Artificial Intelligence and Machine Learning*, vol. 5, no. 4, pp. 4797-4818, 2025. [[CrossRef](#)] [[Google Scholar](#)]
- [23] Dilawar Shah et al., "Dual-View Deep Learning Model for Accurate Breast Cancer Detection in Mammograms," *International Journal of Intelligent Systems*, vol. 2025, no. 1, pp. 1-14, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [24] S. Vijayalakshmi et al., "Innovative Deep Learning Classifiers for Breast Cancer Detection through Hybrid Feature Extraction Techniques," *Scientific Reports*, vol. 15, no. 1, pp. 1-16, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [25] Fajar Walayat et al., "Deep Learning-based Analysis of Mammographic Images for Breast Cancer Detection using Transfer Learning," *Healthcare Technology Letters*, vol. 12, no. 1, pp. 1-20, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [26] Long Jun Yan et al., "A Hybrid Deep Learning Model for Mammographic Breast Cancer Detection: Multi-Autoencoder and Attention Mechanisms," *Journal of Radiation Research and Applied Sciences*, vol. 18, no. 3, pp. 1-17, 2025. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [27] Aniruddha Deka et al., "An Improved Hybrid Feature Selection and Classification Framework for Breast Cancer Detection using Mammography Images," *Ain Shams Engineering Journal*, vol. 17, no. 1, pp. 1-13, 2026. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [28] Kevin Scott Mader, MIAS Mammography Dataset, Kaggle, 2017. [Online]. Available: <https://www.kaggle.com/datasets/kmader/mias-mammography>
- [29] Ramanathan Subramanian, INbreast Dataset, Kaggle, 2020. <https://www.kaggle.com/datasets/ramanathansp20/inbreast-dataset>