



Cross-Shuffled Fusion Based Multi-Modal Hybrid Deep Learning Network for Breast Cancer Detection

J. Nalifabegam^{1*}, C. Ganesh Babu²

¹ Department of ECE, Bannari Amman Institute of Technology, Sathyamangalam, Erode 638401, India

² Department of ECE, PPG Institute of Technology, Coimbatore 641035, India

Corresponding Author Email: deanacademics.it@ppg.edu.in

Copyright: ©2026 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/ts.430427>

ABSTRACT

Received: 13 June 2026

Revised: 10 August 2026

Accepted: 18 August 2026

Available online: 31 August 2026

Keywords:

breast cancer, histopathology, mammogram images, Shuffled-MobileNet, Bidirectional Gated Recurrent Unit, Stacked Autoencoder

Breast cancer (BrC) detection requires accurate analysis of both structural and cellular characteristics of breast tissue. However, conventional single-modality methods are often affected by noise, data imbalance, and limited diagnostic capability. Moreover, many existing approaches focus primarily on binary classification and provide limited differentiation between benign and malignant abnormalities, reducing their clinical applicability. To address these challenges, this study proposes a Cross-Shuffled Fusion based Multi-Modal Hybrid Deep Learning Network (CSFHD-Net) that integrates mammogram and histopathology images for accurate BrC diagnosis. During preprocessing, a 3D Gabor filter is employed to suppress noise, enhance texture information, and preserve critical tumor characteristics. The proposed Shuffled-MobileNet extracts discriminative and computationally efficient features from both imaging modalities. These modality-specific features are subsequently integrated using a Cross-Shuffled Feature Fusion mechanism that promotes inter-modality interactions while preserving complementary information. The fused representation is further processed using Bidirectional Gated Recurrent Unit (Bi-GRU) networks and Stacked Autoencoder (SAE) to refine feature representations and reduce dimensionality. The resulting features are classified into normal and abnormal tissue, followed by benign and malignant sub-classification of abnormal cases. Experimental results demonstrate that CSFHD-Net achieves 99.69% accuracy and 96.20% recall, outperforming DenseNet161, LMHistNet, MPa-DCAE, and a custom Convolutional Neural Network (CNN) by 5.32%, 13.28%, 1.35%, and 6.27%, respectively.

1. INTRODUCTION

Globally, over 2.3 million women will be diagnosed with breast cancer (BrC) by 2020, representing more than a quarter of all cancers among women. In its diagnosis, treatment, and prognosis, BrC is a highly heterogeneous disease [1]. Tumor histology techniques are enabling the detection of prognostic and phenotypic information about tumors. It is difficult to determine histology objectively, as different observers can come to differing conclusions [2]. Patients' quality of life may be negatively affected by inappropriate treatment and complications that arise from conventional pathologic analysis. BrC recurrences are common in the early stages of the disease, which necessitates proactive and appropriate treatments [3]. In spite of this, predicting whether a patient will recur remains one of the biggest challenges patients facing early-stage BrC face. For diagnosis and prognosis of BrC patients, molecular tests are not readily available because of cost and other barriers [4]. A cost-effective method for stratifying patients for surgery may be available through digital pathology analysis of their histopathology images [5].

The heterogeneity of BrC means that there is a need for timely detection of any anomaly in the tissue in order to make correct clinical decisions. Mammography can provide

information about the structural characteristics of breast abnormalities while histopathological analysis provides information regarding cellularity of such abnormalities. However, analyzing these complementary images independently may restrict the potential of an automated system to learn all characteristics of the breast lesion. Integration of data obtained from mammography and histopathology can provide a more comprehensive understanding of the breast abnormalities and can help in enhancing the reliability of computer-aided detection of BrC. Driven by such clinical requirements, this work introduces Cross-Shuffled Fusion based Multi-Modal Hybrid Deep Learning Network (CSFHD-Net), a novel multimodal deep learning approach to fuse complementary information from mammogram and histopathology images.

BrC diagnosis, treatment, and prognosis can be greatly improved by deep learning (DL). Diagnostic accuracy and efficiency are greatly enhanced by the use of DL models through enhancement of image quality and rapid segmentation of lesions [6]. Automated and accurate tumor detection and segmentation can be accomplished using these models utilizing Convolutional Neural Networks (CNNs) [7]. Treatments based on DL technology include staging and subtyping BrC based on image-based techniques [8]. By

identifying different tumor features and classifying them into different cancer subtypes, these models provide a foundation for developing personalized treatment plans [9]. In addition, multi-omics data can also be incorporated into DL models to help doctors choose the most suitable treatment regimen for their patients. DL technology was used to build robust and generalizable prognostic models based on large-scale databases [10].

BrC remains one of the most complex and challenging diseases to diagnose and manage. Although DL has advanced diagnostic, prognostic, and treatment planning strategies, several key limitations persist [11]. Conventional single-modality approaches often suffer from noise, class imbalance, and poor generalizability, reducing their robustness in real-world clinical settings. Moreover, most existing methods are confined to binary classification (normal vs. abnormal) and fail to provide consistent sub-classification into benign and malignant categories, which is crucial for clinical decision-making. Histopathological evaluation, while essential for assessing tumor heterogeneity is subjective and prone to inter-observer variability, whereas molecular tests for precise stratification remain expensive and inaccessible in many healthcare systems. These challenges highlight a cost-effective and reliable diagnostic framework of conventional single-modality approaches. The proposed CSFHD-Net integrate complementary information from mammography and histopathology to enable accurate and clinically meaningful BrC detection and classification. The main contribution of the research are as follows:

(1) An Advanced 3D Gabor filter is employed in the preprocessing phase to simultaneously reduce noise and enhance texture quality ensuring that critical cancer features are preserved for reliable analysis.

(2) A Shuffled with MobileNet is introduced for separate feature extraction from mammography and histopathology images by enabling lightweight yet highly discriminative representation of modality-specific patterns.

(3) A novel Cross-Shuffled Fusion strategy is proposed to effectively integrate multimodal feature sets while preserving inter-modality correlations, reducing redundancy, and generating a unified and robust feature representation.

(4) The fused features are further optimized using a combination of Bidirectional Gated Recurrent Unit (Bi-GRU) and Stacked Autoencoder (SAE), which capture sequential dependencies, and enable accurate classification of normal, benign and malignant cases.

The rest of the work was pre-systematized as follows: Section 2 examines the recent works on BrC classification. Section 3 contains the clear description of the proposed CSFHD-Net for different BrC cases. The outcomes of the proposed CSFHD-Net along with a contrast to the proposed and existing methods are determined in Section 4. The conclusion encloses with Section 5.

2. LITERATURE SURVEY

Over the year, tested various classification algorithms for detecting BrC using pathology images with the objective of improving their accuracy and efficiency. This literature review aims to summarize the existing methodologies and advancements in image-based classification techniques for detecting BrC.

In 2024, an extended deep-learning network incorporating

residual features was proposed for multiclass BrC classification using histopathological images [12]. The framework uses residual features to enhance the representation of discriminative information from the input images. The study demonstrated the effectiveness of the proposed network for binary and multiclass BrC classification.

In 2024, a hybrid CNNs architecture called LMHistNet was introduced for classifying microscopic breast tumor tissue images [13]. The architecture combines asymmetric convolutions with Levenberg–Marquardt optimization and incorporates a convolutional block attention module for adaptive feature refinement. Batch normalization was also employed to improve convergence during training. For multiclass classification into eight subtypes, the model achieved accuracy, precision, recall, and F1-score values of 88%, 89%, 88%, and 88%, respectively.

In 2025, a multi-patch-based DL approach incorporating VGG19 was developed for BrC classification from pathology images [14]. The MPa-DCAE model utilizes the hierarchical feature extraction capability of VGG19 within a deep convolutional autoencoder framework to capture complex patterns in histopathological images. The autoencoder component enables unsupervised feature learning, improving adaptability to variations in image characteristics.

In 2025, a deep CNN-based framework was investigated for early BrC diagnosis using hybrid classifiers [15]. Transfer learning was combined with classifiers such as K-nearest neighbors, decision trees, and support vector machines to improve classification performance. The experimental results demonstrated high classification accuracy, with the SVM-based approach achieving 99.5% accuracy.

In 2025, several DL architectures, including a custom CNN, MobileNet, DenseNet201, and pre-trained VGG16, were evaluated for BrC classification [16]. Particle swarm optimization was employed to optimize the final dense layer, and the models were evaluated using breast histopathology images.

In 2025, CNN- and local binary pattern-based approaches were investigated for the preliminary diagnosis of BrC from histopathology images [17]. A 20-layer CNN based on a proposed star-like LBP structure, referred to as Quick Star-like Local Binary Pattern (QS-LBP), was developed for feature representation. The resulting approach achieved an Area Under the Receiver Operating Characteristic Curve (AUC/ROC) of 97.9%, an F1-score of 92.3%, and an accuracy of 94.58%.

In 2025, ResNet-based deep learning approaches were investigated for the multi-class classification of BrC subtypes using histopathological images [18]. ImageNet-pretrained ResNet-18, ResNet-34, and ResNet-50 architectures were employed with data augmentation to classify eight benign and malignant BrC subtypes using the Break His dataset at different magnification levels. Among the evaluated models, ResNet-50 achieved an accuracy of 92.42% and an AUC/ROC of 99.86%, demonstrating the effectiveness of residual learning for automated histopathological BrC subtype classification.

In 2021, a patch-based DL framework using a Deep Belief Network was developed for BrC classification from histopathological images [19]. The approach performed unsupervised pre-training followed by supervised fine-tuning to automatically extract features from image patches. Logistic regression was subsequently used for classification, and the method achieved an accuracy of 86%.

In 2022, CNN-based approaches were evaluated for classifying mammogram images into benign, malignant, and normal categories [20]. Different architectures, including AlexNet, VGG16, and ResNet50, were investigated using transfer learning and training from scratch on the mini-DDSM dataset. The results showed that VGG16 and ResNet50 achieved accuracies of 65% and 61%, respectively, when pretrained weights were used, while AlexNet achieved an accuracy of 65%.

In 2022, an improved AlexNet-based framework was developed for breast disease classification [21]. The model was initially pretrained using ImageNet and subsequently fine-tuned using an enhanced dataset. An improved cross-entropy loss function was also introduced to reduce overconfident predictions. The proposed approach was evaluated using the BreaKHis, IDC, and UCSB datasets and demonstrated improved performance across different magnification levels.

Recent multimodal research has investigated the fusion of complementary BrC data to overcome the limitations of single-modality approaches. A comprehensive investigation of multimodal DL fusion strategies considered histopathological, mammographic, magnetic resonance imaging, ultrasonography, clinical, and genomic information for BrC classification [22]. Similarly, multimodal fusion with dimensionality adjustment has been investigated to improve feature representation and BrC diagnosis [23].

Overall, the reviewed studies demonstrate the effectiveness of DL for BrC classification; however, most approaches primarily rely on a single imaging modality, limiting their ability to exploit complementary structural and cellular information. Existing multimodal approaches improve information integration, but their effectiveness remains strongly dependent on the adopted fusion strategy. Early fusion combines different modalities at the input or low-level feature stage, which may integrate heterogeneous information before adequate modality-specific representation learning. Late fusion independently processes each modality and combines high-level representations or prediction results, potentially limiting intermediate cross-modal interactions. In contrast, the proposed Cross-Shuffle Feature Fusion Network

(CSFN) integrates mammographic and histopathological features after modality-specific feature extraction. The CSFN mechanism is designed to promote interaction and redistribution of modality-specific feature channels while preserving complementary information before subsequent feature refinement and classification.

3. PROPOSED METHODOLOGY

In this research, a novel CSFHD-Net has been proposed for leverages mammogram and histopathology images for accurate diagnosis. In the pre-processing phase, noise is reduced using a 3D Gabor filter by enhancing texture quality and preserving critical tumor features. For training, modality-specific feature extraction is performed by Shuffled-MobileNet separately for histopathology and mammograms.

These features are integrated using a Cross-Shuffled Fusion mechanism, designed to preserve inter-modality correlations and generate a unified fused vector representation. During testing, images are processed through Bi-GRU networks with SAE, which refine feature representations and reduce dimensionality. These SAE branches are concatenated and classified normal and abnormal tissue, followed by sub-classification of abnormal cases is performed to differentiate benign or malignant. The overall workflow of the proposed CSFHD-Net shown in Figure 1.

3.1 Dataset description

In this research Curated breast imaging subset of digital database for screening mammography (CBIS-DDSM) is used for BrC classification. CBIS-DDSM is better suited for DDSM mammography datasets 33 Once decompressed, images can be viewed on a computer in DICOM format. DICOM data was converted to PNG files and processed according to CBIS DDSM guidelines. This research aimed to develop a classification scheme to separate benign from malignant images.

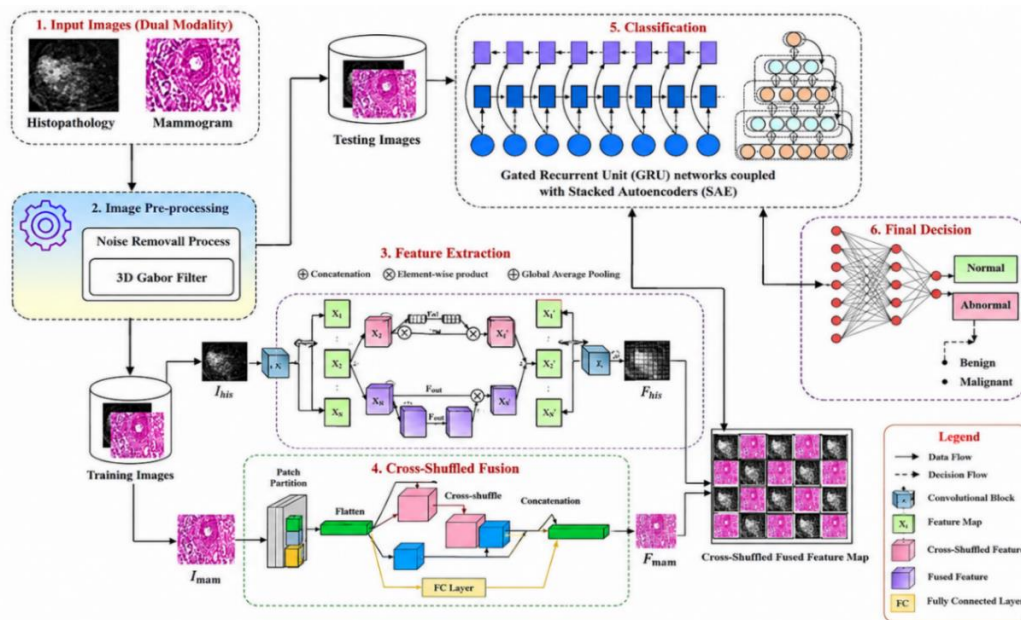


Figure 1. Overall workflow of the proposed Cross-Shuffled fusion based Multi-Modal Hybrid Deep Learning Network (CSFHD-Net)

The BACH dataset contains of microscopy images meticulously categorized by medical experts. The annotation process included two medical experts assessing each image, and any disagreements led to the image being excluded from the dataset. Microscopy images consisted of 400 images in total, evenly distributed between four classes, each containing 100 images.

.tiff images contained specific characteristics: red–green–blue (RGB) color model, 2048×1536 -pixel dimensions, and $0.42 \mu\text{m} \times 0.42 \mu\text{m}$ pixel scale. Furthermore, each image required between 10 and 20 MB of memory. The dataset was labelled image-wise by providing detailed information regarding each cancer type. An image collection of BrC histopathology images was utilized for training and evaluating DL models.

BACH Dataset Class Distribution: The BACH dataset consists of 400 histopathological images equally distributed among four initial diagnostic categories: Normal, Benign, In situ carcinoma, and Invasive carcinoma, with 100 images in each category. In this study, the final classification task was formulated as a binary classification of Benign and Malignant. The Benign category was retained as the benign class, while in situ carcinoma and Invasive carcinoma were grouped into the malignant class. The Normal category was excluded because the objective of the study is to distinguish benign from malignant BrC tissue. Thus, the resulting BACH classification comprises 100 benign and 200 malignant images. No additional class-balancing technique was applied, and the resulting class distribution was retained during the experiments.

Dataset Selection: The Curated Breast Imaging Subset of the Digital Database for Screening Mammography (CBIS-DDSM) and Breast Cancer Histology images (BACH) datasets were chosen to complement each other in terms of mammographic and histopathological image modalities. While CBIS-DDSM offers mammographic images of breast lesions, BACH offers microscopic tissue-level images from histopathological images. The two datasets, thus, offer modality-specific image representations, which can be used for extracting features to be fused using the proposed CSFHD-Net framework. It is important to note that CBIS-DDSM and BACH are independent public datasets and do not offer paired mammographic and histopathological images for the same patients. Thus, the proposed framework performs feature-level fusion of modality-specific features instead of patient-level multimodality fusion. While these datasets offer useful benchmarks for evaluating the effectiveness of the BrC classification algorithms, they might not fully represent the real-world conditions, where there could be differences in patient demographic characteristics, imaging instruments, acquisition procedures, institutional practices, and histopathological preparation methods. Thus, evaluation using larger and multiple institutional datasets is required.

Diversity of the Datasets and Their Drawbacks: Despite the availability and popularity of the CBIS-DDSM and BACH datasets, they have limitations in representing the full real-world clinical diversity of BrC patients. These datasets do not provide sufficiently detailed demographic information to assess their representativeness across different age groups and racial or ethnic populations. In addition, variations in imaging equipment, acquisition protocols, institutional practices, and data-labelling procedures may not be fully represented. CBIS-DDSM consists of mammographic examinations, whereas BACH contains expertly labelled histopathological images,

enabling the investigation of complementary structural and cellular information. Nevertheless, the use of these datasets does not encompass the full variability observed across different hospitals, patient populations, scanners, staining procedures, and image acquisition conditions. Other publicly available datasets, such as BreaKHis for histopathological images and mini-DDSM for mammographic images, can also be considered above in Table 1; however, these datasets may have similar limitations regarding demographic diversity and acquisition variability. Therefore, the high performance achieved by CSFHD-Net should be interpreted within the context of the datasets used in this study. Validation using larger, more diverse, and multi-institutional datasets is necessary to further establish the generalizability, robustness, and clinical applicability of the proposed framework.

Table 1. Description for the Curated Breast Imaging Subset of the Digital Database for Screening Mammography (CBIS-DDSM) dataset

Dataset	Type	Training Image (70%)	Testing Image (30%)	Total Image
CBIS-DDSM	Benign	1890	811	2701
	Malignant	1947	834	2781
Total	-	3837	1645	5482

3.2 Three-dimensional Gabor filter formulation for 2D multimodal images

In the preprocessing stage, a 3D Gabor filter is employed to suppress noise while enhancing texture information and preserving critical tumor features in mammography and histopathology images. While conventional 2D Gabor filters effectively extract spatial frequency features, they process multi-channel information independently, thereby failing to capture inter-channel cross-correlations and multi-scale depth relationships. To overcome this limitation, a 3D Gabor filter is utilized to capture joint spatial, chromatic, and multi-scale textural patterns.

For 2D multimodal medical images, the 3D tensor representation $I(x, y, z)$ is constructed as follows:

Histopathology Images: The 2D RGB biopsy images ($H \times W \times 3$) are mapped such that (x, y) represent spatial coordinates and $z \in \{R, G, B\}$ represents the color channel axis. The 3D Gabor filter performs volumetric spatial-spectral convolution, extracting coherent structural boundaries while preserving colorimetric stain distributions essential for distinguishing hematoxylin-stained nuclei from eosin-stained stroma.

Mammogram Images: The 2D grayscale mammograms ($H \times W$) are transformed into a multi-scale tensor ($H \times W \times D$), where the z -axis ($z \in \{1, 2, \dots, D\}$) corresponds to multi-resolution scale decomposition layers. This enables simultaneous multi-frequency directional texture enhancement and quantum noise suppression.

The mathematical formulation of the three-dimensional Gabor filter $G(x, y, z)$ is computed as follows:

$$G(x, y, z) = \frac{1}{(2\pi)^{3/2} \sigma_x \sigma_y \sigma_z} \exp \left(-\frac{1}{2} \left[\left(\frac{x'}{\sigma_x} \right)^2 + \left(\frac{y'}{\sigma_y} \right)^2 + \left(\frac{z'}{\sigma_z} \right)^2 \right] \right) \exp(j2\pi(f_x x + f_y y + f_z z) + \phi) \quad (1)$$

Here, $(\sigma_x, \sigma_y, \sigma_z)$ denote the Gaussian envelope bandwidths along the spatial and channel dimensions, (f_x, f_y, f_z) represent the spatial-spectral frequencies, and ϕ is the phase offset. The enhanced filtered representation $I_{texture}(x, y, z)$ is obtained by convolving the 3D image tensor with the 3D Gabor filter:

$$I_{texture}(x, y, z) = I(x, y, z) \otimes G(x, y, z) \quad (2)$$

This 3D Gabor formulation ensures that both macro-scale mammographic spicules/micro calcifications and micro-scale cellular textures retain their critical structural and textural characteristics during preprocessing. This directly strengthens the feature quality for subsequent multimodal fusion in CSFHD-Net, thereby improving classification accuracy.

3.3 Pre-processing module Cross-Shuffled Fusion based Multi-Modal Hybrid Deep Learning Network

The proposed CSFHD-Net Employs Shuffled-MobileNet as a lightweight backbone to extract discriminative features separately from mammogram and histopathology images. It ensures efficiency while reducing computational complexity through depth-wise separable convolutions (DSC) and channel shuffle attention. A CSFN module is then used to integrate modality-specific features and enhance inter-modality correlations. It also reduces redundancy and enhances cross-attention. Lastly, the classification network receives the optimized feature representation, enabling high accuracy differentiation between benign and malignant BrC. Architecture of CSFHD-Net shown in Figure 2.

3.3.1 Shuffled-MobileNet for feature extraction

Shuffled-MobileNet extracts features for histopathology and mammograms separately, resulting in feature sets F_{his} and F_{mam} . A pre-trained feature extraction network with numerous characteristics that drastically reduce the model size without sacrificing effectiveness is the MobileNet model. Convolutions that form the foundation of a Mobile-Net are depth-wise separable and point-wise separable. It reduces the computational requirements to separate the spatial and channel dimensions. In order to balance model size and performance, the MobileNet model relies heavily on the usage of width multipliers to regulate the network's width. Using an

additional option, a resolution multiplier regulates the input resolution of MobileNet.

MobileNet was pre-trained as a transfer learning model using an ImageNet and other large-scale datasets. The MobileNet model is a factorized convolution that converts a conventional convolution into a depth-wise, one-dimensional convolution called a pointwise convolution by using DSC. The depth-wise convolution trains each input channel separately using a single filter. Utilize dimension pointwise convolution to merge these distinct filters. A traditional convolution creates a new set of outputs by combining and filtering inputs in a single step. These outputs are then split into two layers by the DSC. This process produces a large reduction in both processing and model size. Two square feature maps are produced by a conventional convolution layer: one for input F with size $DF \times DF \times N$ and one for output G with size $DG \times DG \times M$. The DF and DG stand for the input and output feature map spatial dimensions (width and height). The convolution kernel K of size $Dk \times Dk \times N \times M$ describes the CNN layer parameters, where Dk stands for the kernel's spatial breadth and output channels' height, respectively. The output feature map for standard convolution is calculated using the formula below:

$$G_{k,l,m} = \sum_{i,j,n} K_{i,j,m,n} \cdot F_{k+i-1,i+j-1,n} \quad (3)$$

The input and output channels N and M , the feature map size (D_F), the kernel size (D_k), and the computational cost are all multiplicative. All of these elements and their interconnections are covered by MobileNet models. First, DSC were used to isolate the relationship between kernel size and output channel numbers. Depth-wise and point wise convolutions are the two tiers of DSC. Depth-wise convolutions are used to obtain the filter for each input channel. Pointwise convolution, a straightforward 1×1 convolution, is then used to linearly aggregate the output from the depth wise layer. For every layer, MobileNet uses either the batch norm or the ReLU nonlinearity.

$$DSC_{cost} = D_k \times D_k \times N \times D_F \times D_F + N \times M \times D_F \times D_F \quad (4)$$

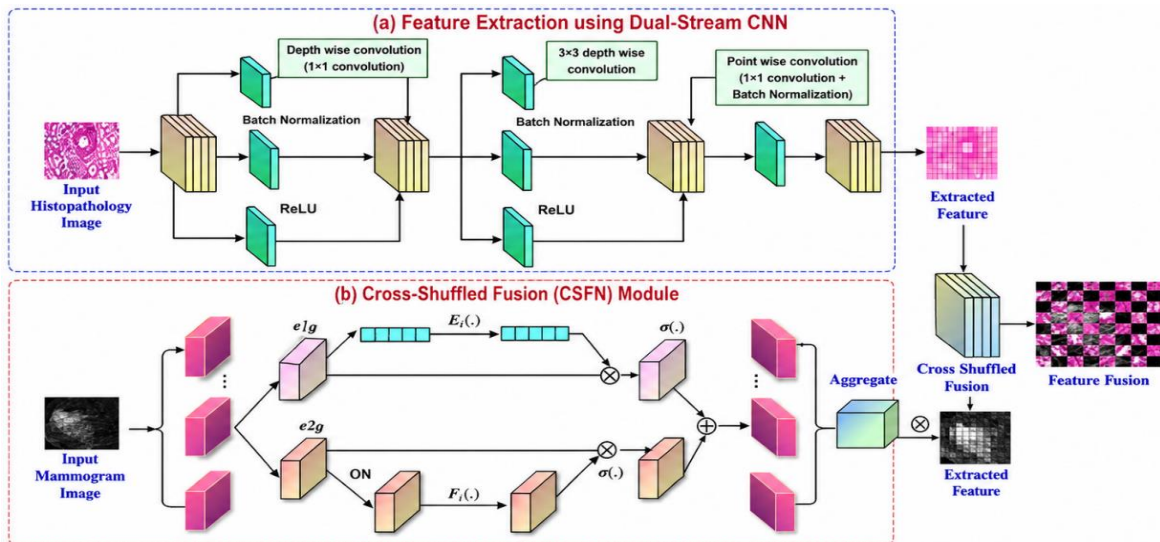


Figure 2. Architecture of Cross-Shuffled Fusion based Multi-Modal Hybrid Deep Learning Network (CSFHD-Net)

Convolution is a two-phase process that involves filtering and combining, which lowers the computing cost. Channel-wise dependencies were fully captured by using the SE block. In terms of balancing speed and accuracy, it will introduce an excessive number of factors, which is detrimental to the creation of a more lightweight attention module. Furthermore, faster 1-D convolutions of size k , ECA are not suitable for producing channel weights since k is typically greater. To improve the global information, channel-wise statistics like $s \in \mathbb{R}^{C/2G \times 1 \times 1}$ was obtained by employing global averaging pooling (GAP), which was calculated by reducing X_{k1} through spatial dimension $H \times W$.

$$s = F_{gp}(X_{k1}) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W (i, j) \quad (5)$$

Additionally, to facilitate assistance for accurate and flexible choosing, a compact feature is developed. This is accomplished by sigmoid activation and a straightforward gating mechanism. The final channel attention output can then be obtained by:

$$X'_{k1} = \sigma(F_c(s)) \cdot X_{k1} = \sigma(W_1 s + b_1) \cdot X_{k1} \quad (6)$$

Efficiently extract lightweight yet highly discriminative features from both mammogram and histopathology images. The proposed feature sets F_{his} and F_{mam} are designed to enhance the accuracy and robustness of BrC classification through a Cross-Shuffled Fusion process.

The CSFHD-Net framework processes mammogram and histopathology images through separate Shuffled-MobileNet branches to extract modality-specific features. The CSFN module integrates these features using cross-attention, feature interaction, channel shuffling, and concatenation for effective multimodal fusion in Figure 3. The fused features are refined using Bi-GRU, compressed through SAE, and classified into Normal/Abnormal and Benign/Malignant categories.

3.3.2 Cross-Shuffle Feature Fusion

The extracted features are integrated using a Cross-Shuffled Fusion Network (CSFN) designed to preserve inter-modality correlations and generate a unified fused vector representation. The CSFN enhances reciprocal learning between histopathology and mammography feature representations by combining cross-attention with channel shuffle operations.

Given a histopathology feature vector F_{his} and a mammography feature vector F_{mam} , their element-wise product is first computed as $mth = f_{his} \otimes f_{mam}$.

$$s_{th} = \text{SeLU}(W_{th}(a_{th} \otimes m_{th}) + b_{th}) \quad (7)$$

The cross-modal correlation score is then obtained through a dense transformation with SeLU activation.

$$S' = \text{softmax}(S) \quad (8)$$

To achieve diverse feature interactions, the histopathology and mammography feature representations are updated using the normalized cross-modal affinity matrix along with channel shuffle options

$$\begin{aligned} F_{his} &*= \text{Shuffle}(F_{his} \cdot S) \\ F_{mam} &*= \text{Shuffle}(F_{mam} \cdot S^T) \end{aligned} \quad (9)$$

Table 2 ablation results demonstrate the effectiveness of the proposed Cross-Shuffled Feature Fusion (CSFN) module. Early fusion achieved an accuracy of 96.84%, while late fusion improved the accuracy to 97.46%. Attention-based fusion further increased the accuracy to 98.31%. In comparison, the proposed CSFN achieved the highest accuracy of 99.69%, together with a specificity of 97.95%, precision of 98.67%, recall of 96.20%, and F1-score of 97.93%. These results indicate that the Cross-Shuffled Fusion strategy effectively preserves complementary information between mammographic and histopathological modalities while reducing redundant representations and improving inter-modality feature interaction. Bi-GRU with SAE for Classification.

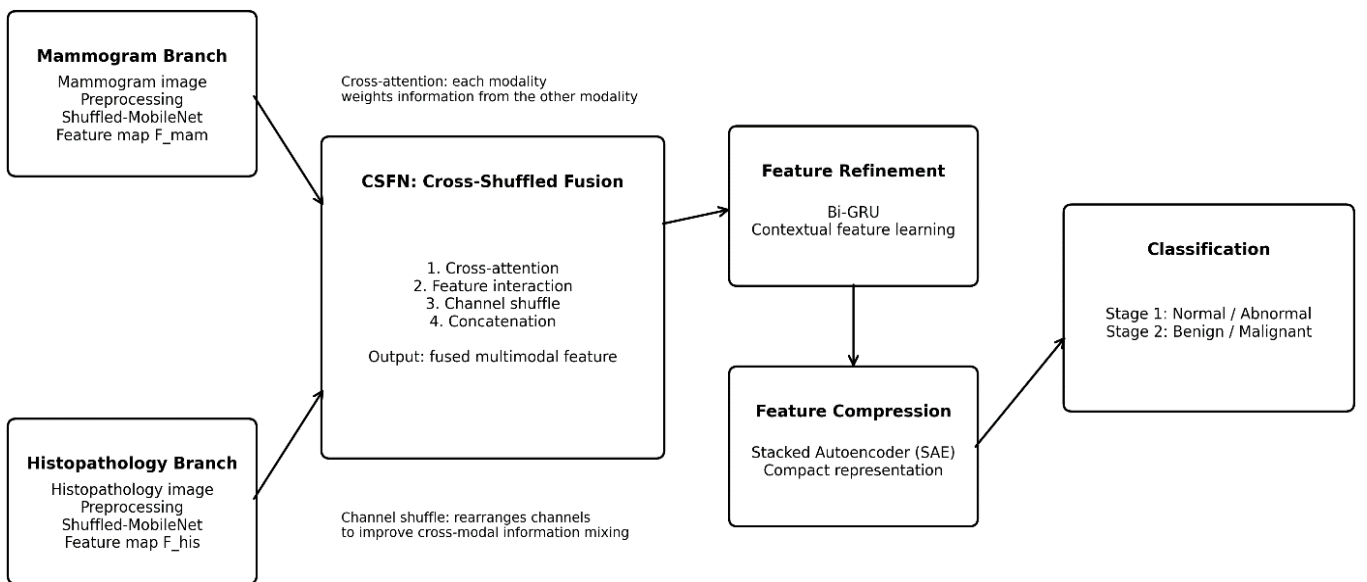


Figure 3. Annotated architecture of the Cross-Shuffled Feature Fusion (CSFN)

Table 2. Ablation study of different fusion strategies in the proposed Cross-Shuffled Fusion based Multi-Modal Hybrid Deep Learning Network (CSFHD-Net)

Fusion Method	Accuracy (%)	Specificity (%)	Precision (%)	Recall (%)	F1-Score (%)
Early Fusion	96.84	94.72	95.31	93.86	94.58
Late Fusion	97.46	95.83	96.12	94.91	95.51
Attention based Fusion	98.31	96.72	97.24	95.68	96.45
Proposed	99.69	97.95	98.67	96.20	97.93

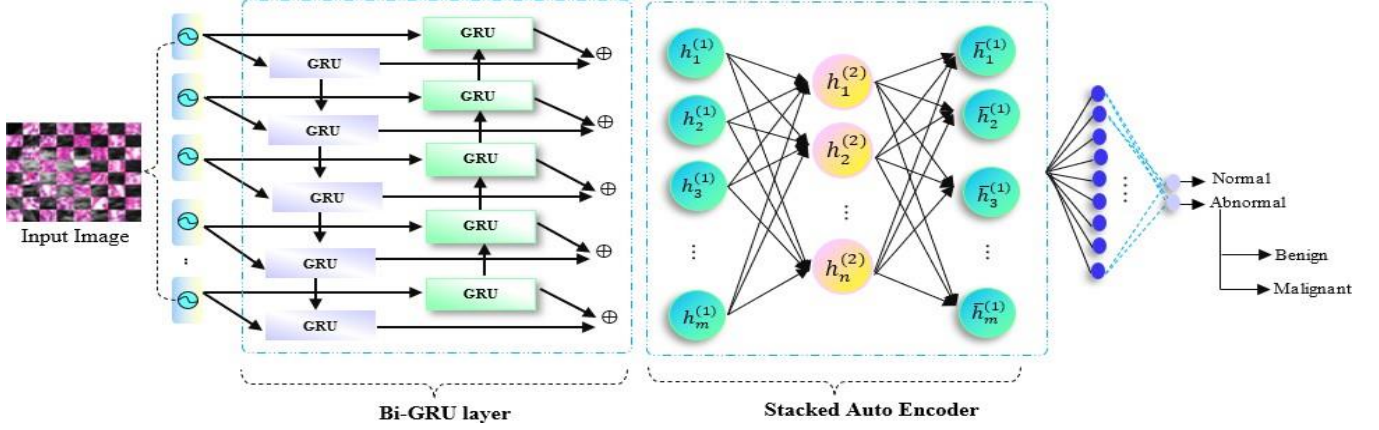


Figure 4. Architecture of Bidirectional Gated Recurrent Unit (Bi-GRU) with Stacked Autoencoder (SAE)

During testing, images are processed through Bi-GRU networks with SAE, which refine feature representations and reduce dimensionality. These SAE branches are concatenated and classified normal and abnormal tissue followed by sub-classification of abnormal cases is performed to differentiate benign or malignant. The Bi-GRU is employed to effectively model sequential dependencies in the extracted multimodal features. Bi-GRU processes information both forward and backward, allowing it to record dependencies from past and future sequences. This bidirectional learning is essential for identifying complex spatial-temporal patterns in histopathological and mammography images.

The number of hidden units in the GRU layers and scaled dot-product attention are two other attention techniques that have been modified in addition to the number of GRU layers. To balance computational cost and recognition accuracy, a single Bi-GRU layer with a scaled dot-product attention mechanism was employed. These decisions on the Bi-GRU and attention components are crucial for the approach to properly identify BrC classification. It highlights significant frames or moments, eliminates superfluous information, and highlights vital aspects of the BrC by delicately assigning differential weights.

Formulas for the backward-pass Bi-GRU:

$$m'_r t = \sigma(W'_m * [h'_{(t-1)}, x'(t)]) \quad (10)$$

$$n'_z t = \sigma(W'_z * [h'_{(t-1)}, x'(t)]) \quad (11)$$

$$\hat{h}'_t = \text{ReLU}(W'_h * [m'_r t * h'_{(t-1)}, x'(t)]) \quad (12)$$

$$h'_t = (1 - n'_z t) * h'_{(t-1)} + n'_z t * \hat{h}'_t \quad (13)$$

Here, $m_r t$ and $m'_r t$ represent the reset gate. $n_z t$ and $n'_z t$ denote the update gate where both update gates. $\hat{h}'_t$ denotes the candidate's hidden state. h_t and h'_t denote hidden state. The model can learn complex behaviors from images that usually contain spatial and temporal information because to this mix

of characteristics. Furthermore, adding a nonlinearity aspect to the model by substituting Tanh activation for ReLU activation makes it easier for deep networks like Bi-GRU to converge. Finally, the models are improved by the attention mechanisms.

3.3.3 Stacked Autoencoder

The Stacked Autoencoder (SAE) is integrated into the framework to provide robust feature compression and hierarchical representation learning. By successively encoding and decoding the feature space, SAE is able to extract increasingly abstract representations of the input data. Classification with this technique is particularly beneficial for intrinsic datasets because it can learn hierarchical representations of data. By stacking layers of auto encoders, the model can capture increasingly features and reduce dimensionality. Additionally, SAE were effectively high-dimensional data and are noise-resistant, which makes them appropriate for use in medical imaging applications. As a result, classification tasks are more generalized and accurate. Figure 4 shows a Bi-GRU with SAE for Classification.

The encoder and decoder layers are proportioned and have equivalent hidden neuron counts. The input images are transformed into intricate coding by the encoding layers, which capture multiple high-level representations of multimodal medical images. The decoding layers then reconstruct output signals that closely resemble the original inputs by utilizing these detailed coding. Experimental feature vectors are encoded and decoded using non-linear functions as follows,

$$h_k = f_k(x) = \phi_{k,e}(W_{k,e}x_{k,i} + \beta_{k,e}) \quad (14)$$

$$\hat{h}_l = g_l(x) = \phi_{l,d}(W_{l,d}x_{l,i} + \beta_{l,d}) \quad (15)$$

where, the functions of encoding and decoding in the k-th hidden layer (HL) (h_k), and l-th HL ($\hat{h}_l$) are indicated by $f_k(x)$ and $g_l(x)$, respectively. Weighted matrices in h_k and $\hat{h}_l$ are indicated by the $W_{k,e}$ and $W_{l,d}$, and modified by tasks

such as a hyperbolical tangent, linear, Maxout unit, sigmoid, step, ramp, and ReLU. In order to achieve an experienced network of SAE using gradient background approaches based on the error results during fine-tuning. The Bi-GRU with SAE modules combine to contextualize multimodal features sequentially and compress them into robust hierarchical embeddings. Combining these two networks provides highly discriminative feature vectors to reliably classify BrC as normal, benign, and malignant.

4. RESULTS AND DISCUSSION

Deployment Feasibility: CSFHD-Net as proposed in the current paper has been tested with the help of MATLAB 2020b software with an Intel Xeon CPU and an NVIDIA GPU as hardware resources. For training purposes, approximately 12 GB of RAM was needed, whereas an NVIDIA RTX 3090 GPU has been used for efficient execution. The proposed architecture was able to achieve a computational time of about 1.5 s, which is less processing time than those of the compared existing methods. This computational efficiency reflects that CSFHD-Net can be used for deployment on workstations

equipped with a GPU for fast multimodal image analysis. Yet, the current research does not have an experimental evaluation of the proposed framework on low-resource or edge devices. As such, deployment on resource-limited hardware can be regarded as a limitation of the current research.

Figure 5 demonstrates an experimental outcome of the proposed CSFHD-Net model. The first column denotes the input mammograms and histopathology images. The second column displays the enhanced versions of these images after applying noise reduction, and contrast adjustment to improve visual quality. The third column (Feature Extraction) highlights the process of extracting important patterns, textures, and structures from the pre-processed images, represented in pixel-grid form. The fourth column (Feature Fusion) shows the integration of extracted features from different modalities (mammogram + histopathology) into a unified representation for better decision-making. The fifth column (Classification 1) provides the first stage of classification, identifying whether the image is Normal or Abnormal. Finally, the sixth column (Classification 2) gives the second stage of classification for abnormal cases, further categorizing them as Benign or Malignant.

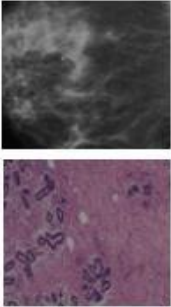
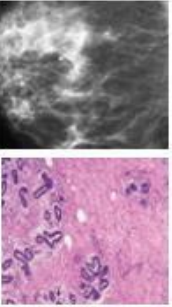
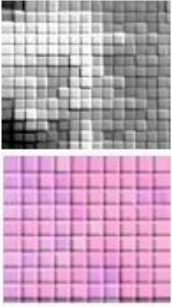
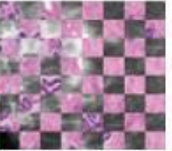
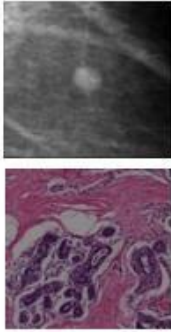
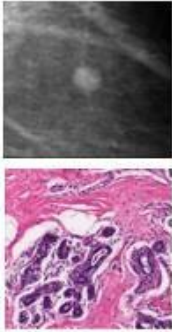
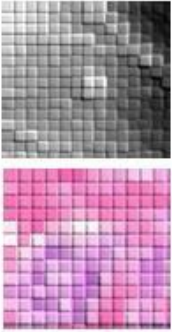
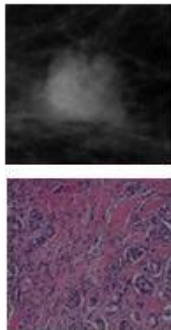
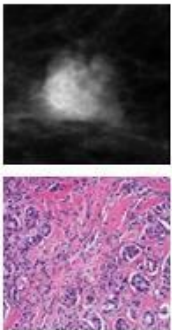
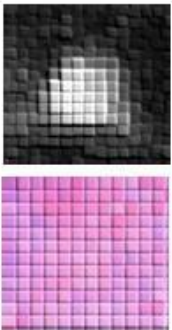
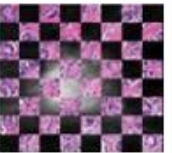
Input Image	Pre-Processing	Feature Extraction	Feature Fusion	Classification 1	Classification 2
				Normal	----
				Abnormal	Benign
				Abnormal	Malignment

Figure 5. Experimental outcome of the proposed Cross-Shuffled Fusion based Multi-Modal Hybrid Deep Learning Network (CSFHD-Net) model

4.1 Deployment challenges and clinical practicality

Although the proposed CSFHD-Net demonstrates promising classification performance, several challenges need to be considered for real-world clinical deployment. Image quality can vary considerably due to differences in acquisition conditions, image resolution, compression, noise levels, scanner manufacturers, imaging protocols, and histopathological staining procedures. Such variations may affect the quality and consistency of extracted features and consequently influence model predictions. In addition, deployment across different clinical environments may involve differences in computational hardware, GPU/CPU availability, memory capacity, and processing time. The proposed framework combines multiple DL components, including Shuffled-MobileNet, CSFN, Bi-GRU, and SAE, which may require adequate computational resources for efficient inference. Therefore, model optimization, hardware-specific benchmarking, and efficient inference strategies would be required before clinical deployment. Furthermore, external validation on images acquired using different scanners, institutions, and acquisition protocols is necessary to assess robustness and generalizability. These considerations highlight the need for prospective multi-institutional evaluation and integration with existing clinical imaging workflows before practical clinical adoption.

4.2 Performance analysis

The efficiency of the proposed CSFHD-Net was evaluated based on the specified measures namely like accuracy (AY), specificity (SY), recall (RL), precision (PN) and F1-score (FS) respectively.

The AY curve in Figure 6 and loss curve in Figure 7 of the proposed CSFHD-Net over a period of 100 epochs to classify BrC cases. A training curve is slightly higher than the testing curve indicates effective learning and a steady improvement in accuracy over time. The proposed CSFHD-Net consistently reduces classification errors by demonstrating its effectiveness. Testing and training curves are nearly aligned that indicates good generalization without overfitting. From this, the proposed CSFHD-Net appears to be well-trained with minimal loss and excellent accuracy.

$$SY = \frac{TN}{TN + FP} \quad (16)$$

$$PN = \frac{TP}{TP + FP} \quad (17)$$

$$RL = \frac{TP}{TP + FN} \quad (18)$$

$$AY = \frac{TP + TN}{TP + FP + FN + TN} \quad (19)$$

$$FS = 2 \left(\frac{P + R}{P + R + 1} \right) \quad (20)$$

where, true positive (TP), true negative (TN), false positive (FP), and false negative (FN) respectively.

Table 3 shows the efficacy analysis of the proposed CSFHD-Net for classifying benign and malignant stages of BrC. The parameters include Accuracy, Specificity, recall, F1-score and Precision for each BrC classes. The benign and

malignant classes attain the AY of 99.63% and 99.75%, with the recall of 96.93% and 95.48%. RL and FS scores consistently ensure low levels of FP and FN. Further, the proposed CSFHD-Net attains high accuracy across all levels with an average accuracy of 99.69%. From this evaluation, the proposed CSFHD-Net performs better in the classification of BrC classes.

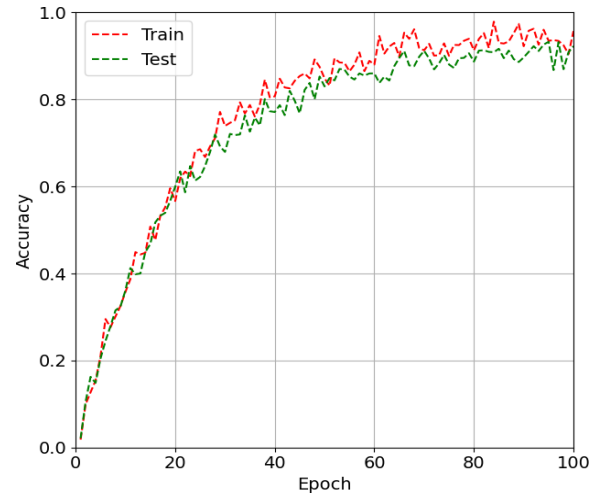


Figure 6. Accuracy curve of the proposed Cross-Shuffled Fusion based Multi-Modal Hybrid Deep Learning Network (CSFHD-Net)

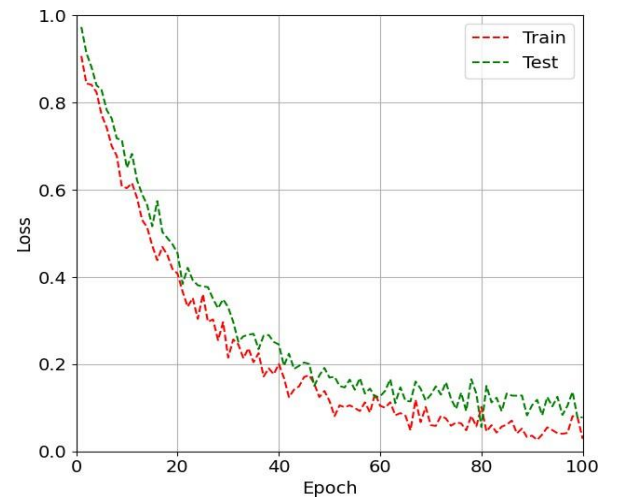


Figure 7. Loss curve of the proposed Cross-Shuffled Fusion based Multi-Modal Hybrid Deep Learning Network (CSFHD-Net)

Table 3. Efficiency analysis of the proposed Cross-Shuffled Fusion based Multi-Modal Hybrid Deep Learning Network (CSFHD-Net)

Classes	AY	SY	PN	RL	FS
Benign	99.63	98.34	99.05	96.93	97.28
Malignant	99.75	97.57	98.29	95.48	98.59
Average	99.69	97.95	98.67	96.20	97.93

4.3 Clinical importance of evaluation metrics

The evaluation metrics offer different perspectives regarding the performance of the proposed CSFHD-Net. The accuracy metric is the ratio of total correctly classified instances. On the other hand, the precision metric shows the

ratio of samples classified as malignant which are actually malignant. Recall or sensitivity is the ratio of the total actual malignant instances correctly identified by the algorithm. It is crucial in BrC screening because a false negative would result in the missed instance and this will defer further diagnosis. Specificity is the ratio of the total benign instances which are accurately classified by the algorithm. It is vital in minimizing false positives and thus reducing unnecessary follow-ups.

In the current experiment, the recall of the malignant-class is 95.48%, and that of the benign-class is 96.93%. Though the difference is quite small, a lower malignant-class recall is important as it would have a higher clinical impact than a false positive. Thus, in addition to accuracy, recall is considered along with precision, specificity, and F1-score.

In order to compare the proposed approach with existing solutions, the recall values of the proposed approach are compared with those of other methods, which are: DenseNet161 [12] (90.37%), LMHistNet [13] (88%), MPa-DCAE [14] (94.85%), and the custom CNN model [16] (89.37%). These serve as a basis for the comparison of the results obtained in this research. However, direct comparison should be done carefully as there might be differences between the studies in terms of dataset used, classes used, preprocessing procedure, training process, etc.

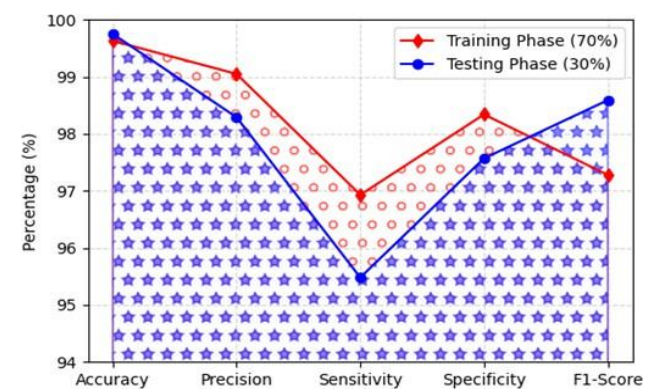


Figure 8. Performance comparison of training and testing phases across evaluation metrics

Figure 8 illustrates the performance of a classification model during the training phase (70% data) and testing phase (30% data). The red line with triangle represents the training phase, while the blue line with circle markers denotes the testing phase. Both phases demonstrate high performance, with values consistently above 95%. Accuracy is nearly perfect, exceeding 99% in both phases. Testing phase precision, specificity, and F1-Score remain strong, although slightly lower compared with training phase, indicating minor generalization gaps. Testing shows a decrease in sensitivity, suggesting the model may overlook some positive cases when applied to unseen data. Results show that the model is robust with excellent generalization ability and minimal overfitting.

4.4 Comparative analysis

In this section, the proposed CSFHD-Net is assessed against existing BrC classification models using a range of performance metrics. The performance of other existing DL models is also assessed and compared with the proposed model using the gathered dataset. This comparative analysis highlights the superior classification accuracy attained by the proposed CSFHD-Net with the gathered dataset.

The comparison was performed with different DL networks based on several parameters as illustrated in Table 4. The Shuffled-MobileNet increases the overall accuracy by 11.55%, 10.17%, 8.11%, 5.01% and 3.56% better than DenseNet, AlexNet, GoogleNet, MobileNet and ShuffleNet respectively. The Shuffled-MobileNet increases the overall RL by 10.77%, 8.73%, 6.55%, 3.61% and 0.28% better than DenseNet, AlexNet, GoogleNet, MobileNet and ShuffleNet respectively. Though, the traditional DL-networks didn't perform well in contrast to the proposed model.

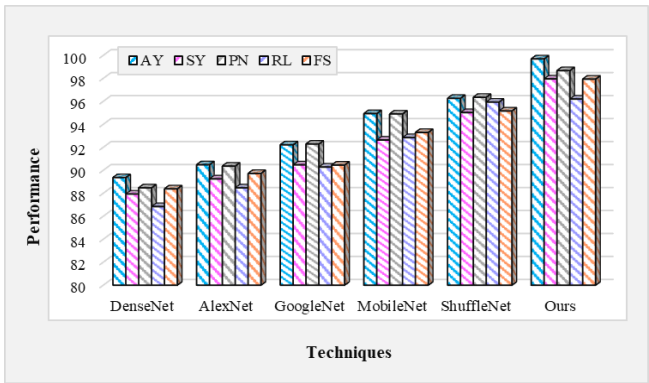


Figure 9. Comparison of feature extraction networks between existing and proposed models

Figure 9 compares the performance of different DL techniques across five evaluation metrics: AY, SY, PN, RL, and FS. Among the evaluated models, the proposed Ours' method achieves the highest overall performance across all metrics. The results demonstrate that the proposed approach provides more consistent and improved classification performance compared with the existing techniques.

Figure 10 illustrates a comparison of feature fusion methods applied to histopathology and mammogram images. The initial column displays the original input images, followed by the ground truth feature fusion. The subsequent columns present the results obtained using various feature fusion methods, including Early Fusion, Late Fusion, and Attention-based Fusion, along with the proposed Cross-Shuffled Fusion approach. The proposed Cross-Shuffled Fusion method produces clearer and more accurate feature fusion results that closely align with the ground truth, thereby demonstrating its effectiveness over existing techniques.

Table 4. Comparison of advanced feature extraction networks

Techniques	Accuracy	Specificity	Precision	Recall	F1-Score
DenseNet	89.36	87.93	88.47	86.84	88.38
AlexNet	90.48	89.25	90.36	88.47	89.72
GoogleNet	92.21	90.47	92.28	90.28	90.45
MobileNet	94.93	92.63	94.90	92.84	93.29
ShuffleNet	96.26	95.02	96.35	95.93	95.16
Shuffled MobileNet	99.69	97.95	98.67	96.20	97.93

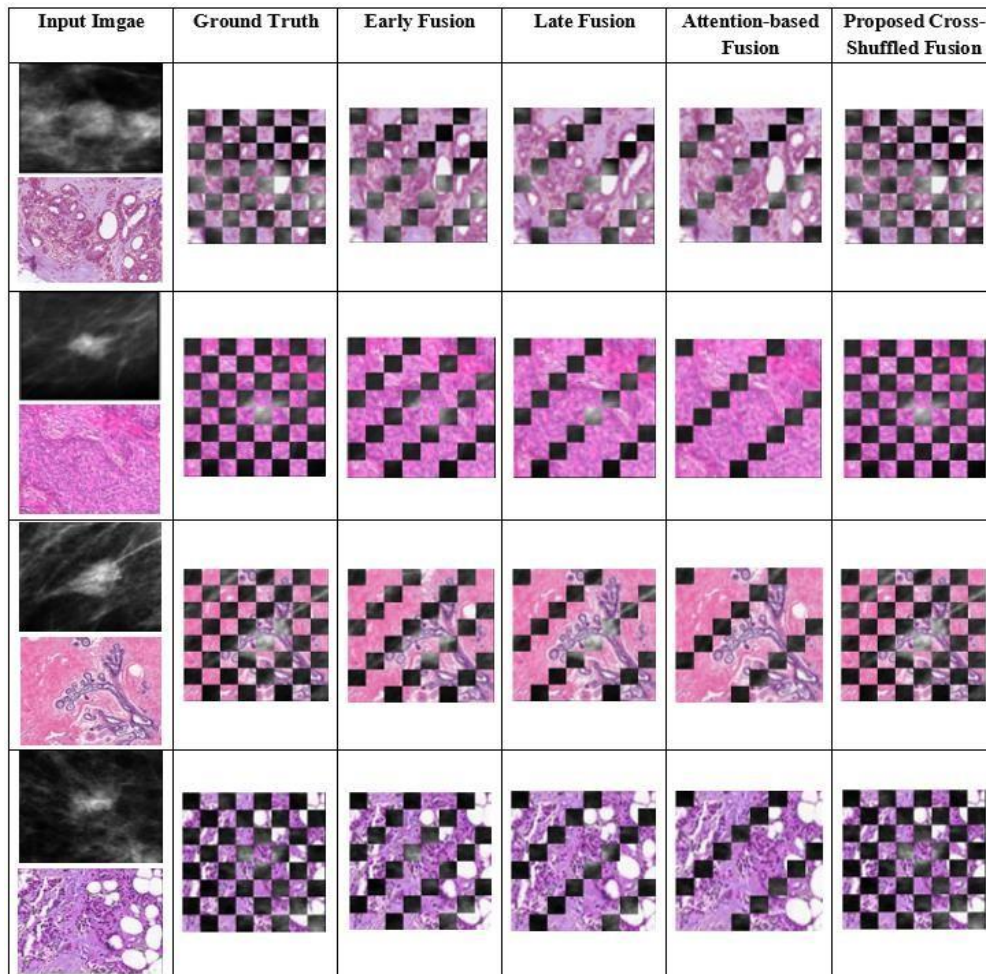


Figure 10. Visual comparison of different technique for feature fusion

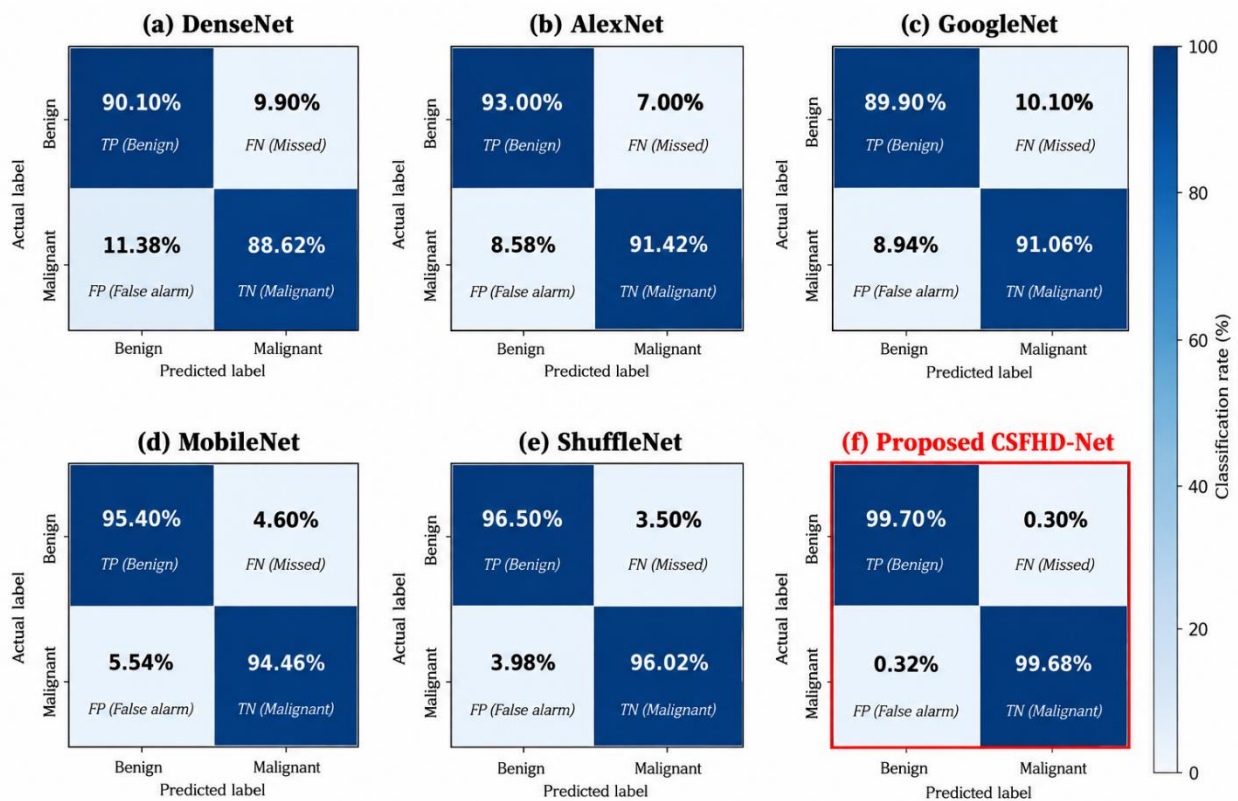


Figure 11. Confusion matrices of deep learning (DL) models for breast cancer (BrC) classification

The confusion matrices for DenseNet, AlexNet, GoogleNet, MobileNet, Shuffle Net, and the CSFHD-Net proposed have been shown in Figure 11 for the classifications of benign and malignant cancers of breast. The diagonal values represent the correct classifications while the off-diagonal values are those that have been misclassified. The numbers present in the figure represent the count of those samples of both the classes. The proposed CSFHD-Net, in comparison to other models, demonstrates lesser number of misclassifications for the two classes.

However, the proposed model demonstrates the best performance with almost perfect classification, reflecting its superior capability in distinguishing malignant tumors from benign ones, which is critical for accurate and early BrC diagnosis.

Figure 12 compares the computational time (in seconds) of different methods with proposed model. Among the existing works, MPa-DCAE [14] required the highest execution time (8.3 s), followed by DenseNet161 [12] (5.2 s), LMHistNet [13] (3.6 s), and the custom CNN [16] (3.1 s). In contrast, the proposed method significantly outperforms the others with the lowest computational time of 1.5 s, highlighting its efficiency in reducing processing overhead. Moreover, minimizing computational time was reduce energy consumption, hardware strain, and latency in clinical decision-making systems by making the proposed method more practical for integration into healthcare systems. Overall, the figure emphasizes that the proposed CSFHD-Net achieves both computational efficiency and practicality compared to existing methods.

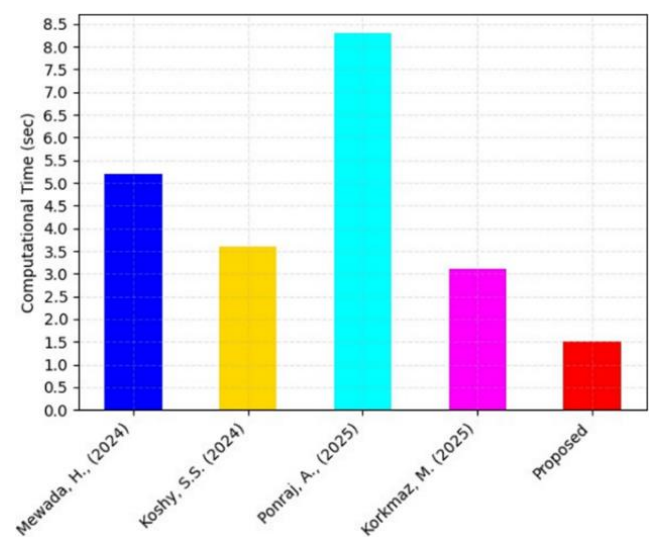


Figure 12. Computational time analysis of Cross-Shuffled Fusion based Multi-Modal Hybrid Deep Learning Network (CSFHD-Net) with existing models [12-14, 16]

Table 5 demonstrates the contrast evaluation of the CSFHD-Net with the existing methods with the gathered dataset. The proposed CSFHD-Net enhances overall accuracy by 5.32%, 13.28%, 1.35% and 6.27% compared to DenseNet161, LMHistNet, MPa-DCAE model, and custom-built CNN model respectively. As shown in Table 5 below, the CSFHD-Net proposed by us shows significantly better results than the chosen existing models. The performance indicators of our model are shown in detail in Section 4.2, whereas Table 5 highlights the performance comparison with existing models. However, these existing networks have not yet achieved the high level of accuracy reached by the proposed CSFHD-Net.

From this evaluation results, the proposed CSFHD-Net was superior with minimal misclassifications for the classification of BrC cases.

Table 5. Efficacy analysis of the proposed Cross-Shuffled Fusion based Multi-Modal Hybrid Deep Learning Network (CSFHD-Net) vs existing methods

Authors	Methods	Accuracy	Recall
Mewada [12]	DenseNet161	94.65%	90.37%
Koshy and Anbarasi [13]	LMHistNet	88%	88%
Ponraj et al. [14]	MPa-DCAE model	98.36%	94.85%
Korkmaz and Kaplan [16]	Custom-built CNN model	93.80%	89.37%
Proposed	CSFHD-Net	99.69%	96.20%

Note: CNN = Convolutional Neural Network.

4.5 Interpretability and clinical trust

Interpretability is an important factor in enhancing clinicians' comprehension and trust in the automatic BrC classification process. In this paper, the proposed CSFHD-Net architecture has used the scaled dot-product attention module in the Bi-GRU layer which assigns different weights to the extracted multimodal features based on their informativeness. Attention weights indicate the significance of each learned feature in the classification process. Nonetheless, the current study has not developed any saliency maps for the clinicians. Generating such attention maps would improve the transparency of the model by indicating the regions of the images or feature patterns that have led to the final prediction of benign or malignant tumors. Accordingly, the next step will be focused on the investigation of attention-based and saliency-based visualization techniques and their evaluation by the clinical experts.

5. CONCLUSIONS

This research proposed a novel CSFHD-Net for early classification BrC stages using mammogram and histopathology images. Shuffled-MobileNet provides a lightweight and efficient method for training by learning discriminative features from both histopathology and mammography images. The Cross-Shuffled Fusion mechanism integrates these features to preserve inter-modality correlations and generate a unified fused vector representation. A GRU combined with SAE refines feature representation and reduces dimensionality during testing. The SAE branches are concatenated and classified as normal or abnormal tissue, followed by sub-classification of abnormal cases for differentiation of benign from malignant. From the experimental analysis, the experimental results provided in Section 4 show that the proposed CSFHD-Net has better classification performance than the current methods, whereas the comparison study proves its effectiveness in multimodal BrC classification. Future work includes extending CSFHD-Net to integrate additional imaging modalities, such as ultrasound and MRI, to enhance early BrC detection. Incorporating explainable AI techniques to provide interpretable and transparent decision-making, improving clinical applicability across diverse patient populations.

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to the Department of Electronics and Communication Engineering, Bannari Amman Institute of Technology, Sathyamangalam, Tamil Nadu, India, for providing the necessary facilities and research environment to carry out this work. The authors also thank the developers of the CBIS-DDSM dataset for making the dataset publicly available, which significantly facilitated this research.

REFERENCES

- [1] Chen, J., Pan, T., Zhu, Z., et al. (2025). A deep learning-based multimodal medical imaging model for breast cancer screening. *Scientific Reports*, 15(1): 14696. <https://doi.org/10.1038/s41598-025-99535-2>
- [2] Brahmareddy, A., Selvan, M.P. (2025). TransBreastNet a CNN transformer hybrid deep learning framework for breast cancer subtype classification and temporal lesion progression analysis. *Scientific Reports*, 15(1): 35106. <https://doi.org/10.1038/s41598-025-19173-6>
- [3] Vijayalakshmi, S., Pandey, B.K., Pandey, D., Lelisho, M.E. (2025). Innovative deep learning classifiers for breast cancer detection through hybrid feature extraction techniques. *Scientific Reports*, 15(1): 22212. <https://doi.org/10.1038/s41598-025-06669-4>
- [4] Abdullah, K.A., Marziali, S., Nanaa, M., Escudero Sánchez, L., Payne, N.R., Gilbert, F.J. (2025). Deep learning-based breast cancer diagnosis in breast MRI: Systematic review and meta-analysis. *European Radiology*, 35(8): 4474-4489. <https://doi.org/10.1007/s00330-025-11406-6>
- [5] Pacal, I., Attallah, O. (2025). InceptionNeXt-Transformer: A novel multi-scale deep feature learning architecture for multimodal breast cancer diagnosis. *Biomedical Signal Processing and Control*, 110: 108116. <https://doi.org/10.1016/j.bspc.2025.108116>
- [6] Oladimeji, O.O., Imran, A.A.Z., Wang, X., Unnikrishnan, S. (2025). Deep learning advances in breast medical imaging with a focus on clinical readiness and radiologists' perspective. *Image and Vision Computing*, 161: 105601. <https://doi.org/10.1016/j.imavis.2025.105601>
- [7] Jiang, B., Bao, L., He, S., Chen, X., Jin, Z., Ye, Y. (2024). Deep learning applications in breast cancer histopathological imaging: Diagnosis, treatment, and prognosis. *Breast Cancer Research*, 26(1): 137. <https://doi.org/10.1186/s13058-024-01895-6>
- [8] Lu, G., Tian, R., Yang, W., et al. (2024). Deep learning radiomics based on multimodal imaging for distinguishing benign and malignant breast tumours. *Frontiers in Medicine*, 11: 1402967. <https://doi.org/10.3389/fmed.2024.1402967>
- [9] Hussain, S., Ali, M., Naseem, U., et al. (2024). Performance evaluation of deep learning and transformer models using multimodal data for breast cancer classification. In *MICCAI Workshop on Cancer Prevention Through Early Detection*, pp. 59-69. https://doi.org/10.1007/978-3-031-73376-5_6
- [10] Li, T., Song, S., Pan, Y., et al. (2025). Deep learning in multi-modal breast cancer data fusion: A literature review. *Quantitative Imaging in Medicine and Surgery*, 15(11): 11578-11610. <https://doi.org/10.21037/qims-2024-2903>
- [11] Adam, R., Dell'Aquila, K., Hodges, L., Maldjian, T., Duong, T.Q. (2023). Deep learning applications to breast cancer detection by magnetic resonance imaging: A literature review. *Breast Cancer Research*, 25(1): 87. <https://doi.org/10.1186/s13058-023-01687-4>
- [12] Mewada, H. (2024). Extended deep-learning network for histopathological image-based multiclass breast cancer classification using residual features. *Symmetry*, 16(5): 507. <https://doi.org/10.3390/sym16050507>
- [13] Koshy, S.S., Anbarasi, L.J. (2024). LMHistNet: Levenberg-marquardt based deep neural network for classification of breast cancer histopathological images. *IEEE Access*, 12: 52051-52066. <https://doi.org/10.1109/ACCESS.2024.3385011>
- [14] Ponraj, A., Nagaraj, P., Balakrishnan, D., et al. (2025). A multi-patch-based deep learning model with VGG19 for breast cancer classifications in the pathology images. *Digital Health*, 11: 1-21. <https://doi.org/10.1177/20552076241313161>
- [15] Pandey, S.K., Rathore, Y.K., Ojha, M.K., Janghel, R.R., Sinha, A., Kumar, A. (2025). BCCHI-HCNN: Breast cancer classification from histopathological images using hybrid deep CNN models. *Journal of Imaging Informatics in Medicine*, 38(3): 1690-1703. <https://doi.org/10.1007/s10278-024-01297-2>
- [16] Korkmaz, M., Kaplan, K. (2025). Effectiveness analysis of deep learning methods for breast cancer diagnosis based on histopathology images. *Applied Sciences*, 15(3): 1005. <https://doi.org/10.3390/app15031005>
- [17] Gül, M. (2025). A novel local binary patterns-based approach and proposed CNN model to diagnose breast cancer by analyzing histopathology images. *IEEE Access*, 13: 39610-39620. <https://doi.org/10.1109/ACCESS.2025.3545052>
- [18] Desai, A., Mahto, R. (2025). Multi-class classification of breast cancer subtypes using ResNet architectures on histopathological images. *Journal of Imaging*, 11(8): 284. <https://doi.org/10.3390/jimaging11080284>
- [19] Hirra, I., Ahmad, M., Hussain, A., et al. (2021). Breast cancer classification from histopathological images using patch-based deep learning modeling. *IEEE Access*, 9: 24273-24287. <https://doi.org/10.1109/ACCESS.2021.3056516>
- [20] Rashmi, R., Prasad, K., Udupa, C.B.K. (2022). Breast histopathological image analysis using image processing techniques for diagnostic purposes: A methodological review. *Journal of Medical Systems*, 46(1): 7. <https://doi.org/10.1007/s10916-021-01786-9>
- [21] Ahmad, J., Akram, S., Jaffar, A., Rashid, M., Bhatti, S.M. (2023). Breast cancer detection using deep learning: An investigation using the DDSM dataset and a customized AlexNet and support vector machine. *IEEE Access*, 11: 108386-108397. <https://doi.org/10.1109/ACCESS.2023.3311892>
- [22] Nakach, F.Z., Idri, A., Goceri, E. (2024). A comprehensive investigation of multimodal deep learning fusion strategies for breast cancer classification. *Artificial Intelligence Review*, 57(12): 327. <https://doi.org/10.1007/s10462-024-10984-z>
- [23] Abdullakutty, F., Akbari, Y., Al-Maadeed, S., Bouridane, A., Talaat, I.M., Hamoudi, R. (2024). Towards improved breast cancer detection via multi-

DDSM	Database for Screening Mammography
BACH	Breast Cancer Histology dataset
RGB	Red–Green–Blue color model
DICOM	Digital Imaging and Communications in Medicine
LBP	Local Binary Pattern
QS-LBP	Quick Star-like Local Binary Pattern
DBN	Deep Belief Network
SVM	Support Vector Machine
KNN	K-Nearest Neighbors
AUC/ROC	Area Under the Receiver Operating Characteristic Curve
HL	Hidden Layer
TP	True Positive count
TN	True Negative count
FP	False Positive count
FN	False Negative count
AY	Accuracy
SY	Specificity
PN	Precision
RL	Recall / Sensitivity
FS	F1-Score

NOMENCLATURE

BrC	Breast Cancer
CSFHD-Net	Cross-Shuffled Fusion based Multi-Modal Hybrid Deep Learning Network
CSFN	Cross-Shuffle Feature Fusion Network
DL	Deep Learning
CNN	Convolutional Neural Network
DSC	Depth-wise Separable Convolution
Bi-GRU	Bidirectional Gated Recurrent Unit
SAE	Stacked Autoencoder
GAP	Global Average Pooling
SE	Squeeze-and-Excitation
ECA	Efficient Channel Attention
ReLU	Rectified Linear Unit
SeLU	Scaled Exponential Linear Unit
CBIS-	Curated Breast Imaging Subset of Digital