Attention Embedded Residual Bottleneck CNN Architecture for Breast Cancer Diagnosis in Ultrasound Images

Mamuna Fatima[1], Muhammad Attique Khan[2,*], Saima Shaheen[1], Seifedine Kadry[3,*], Omar Alqahtani[4], M. Turki-Hadj Alouane[4]

[1] Department of Computer Science, HITEC University, Taxila (Pakistan)
[2] Department of Artificial Intelligence, College of Computer Engineering and Science, Prince Mohammad Bin Fahd University (Saudi Arabia)
[3] Department of Computer Science and Mathematics, Lebanese American University, Beirut, Lebanon; Noroff University College, Kristiansand (Norway)
[4] College of Computer Science, King Khalid University, Abha 62529 (Saudi Arabia)

* Corresponding author: attique.khan@ieee.org (M. Attique Khan), seifedine.kadry@noroff.no (S. Kadry).

Received 19 February 2023 | Accepted 20 February 2025 | Early Access 14 March 2025

Abstract

Breast cancer (BrC) stands as the predominant cancer among women, resulting in a substantial global mortality toll each year. Early detection plays a pivotal role in diminishing mortality rates. Manual diagnosis of BrC is time-intensive, intricate, and prone to errors, emphasizing the necessity for an automated system for timely detection. Various imaging methods have been investigated, underscoring the crucial need for accurate detection to prevent unwarranted treatments and biopsies. Recent years have witnessed substantial exploration and enhancement in the application of DL for efficiently processing medical images. This study aiming to create an effective and resilient DL framework for BrC detection and classification. The steps are contrast enhancement and augmentation, a hybrid CNN network 'BrC-DeepRBNet' is introduced that is built from scratch and incorporates several design elements including residual blocks, bottleneck architecture, and a self-attention mechanism. This framework is employed to construct two networks, one comprising of 107 layers and the other with 149 layers. Moreover, the network capitalizes on the benefits offered by batch normalization (BN) and group normalization (GN), utilizes ReLU and leaky ReLU as activation functions, and integrates Max pooling layer into its architecture in a series of residual-bottleneck blocks. Further, for feature fusion horizontal approach is used and optimization is done using generalized normal distribution optimization (GNDO). The selected features are further classified using neural network classifiers. The introduced framework achieved the highest classification accuracy at 97.05% with publicly available BUS dataset. A detailed ablation study is presented that demonstrates the superior performance of the presented approach, surpassing various pre-trained models (i.e. AlexNet, InceptionV3, ResNet50, and ResNet101) and existing BrC detection and classification techniques.

Keywords

Breast Cancer, Deep Learning, Feature Extraction, Feature Fusion, Feature Optimization.

DOI: 10.9781/ijimai.2025.03.005

I. Introduction

BREAST cancer (BrC) poses a significant risk to the well-being and lives of women. It ranks as one of the most widespread forms of cancer affecting women on a global scale. Statistics presented by the American Cancer Society pertaining to the year 2022 confirmed that 287,850 invasive BrC cases, and 51,400 non-invasive BrC cases were diagnosed among female in US out of which, 43,250 women have passed away. In 2023, 297,790 new cases of BrC will likely be diagnosed among women in the US, resulting in an anticipated 43,170 fatalities. BrC makes up an estimated 19% of all types of cancer diagnoses and accounts for approximately 30% of all reported instances of cancer in the female demographic [1]. However, early detection of this ailment is vital in facilitating the timely administration of suitable treatment [2], [3]. The origin of BrC can be traced back to the irregular and excessive growth of cells within the breast tissue, resulting in the development

Please cite this article as: M. Fatima, M. Attique Khan, S. Shaheen, S. Kadry, O. Alqahtani, M. Turki-Hadj Alouane. Attention Embedded Residual Bottleneck CNN Architecture for Breast Cancer Diagnosis in Ultrasound Images, International Journal of Interactive Multimedia and Artificial Intelligence, vol. 9, no. 5, pp. 108-118, 2025, http://dx.doi.org/10.9781/ijimai.2025.03.005

of diverse forms of lesions. Benign anomalies usually lack harmful characteristics and typically do not necessitate a biopsy. Nevertheless, they undergo monitoring via mammographic examinations due to the potential risk of infiltrating adjacent tissue or exerting harmful pressure on essential structures like blood vessels and nerves. On the contrary, malignant anomalies, which are inherently unstable, are closely associated with BrC. Whenever a suspicious anomaly is detected during a mammogram, its seriousness is invariably confirmed through a biopsy procedure [4].

The most effective method for identifying BrC is through medical imaging analysis [4], [5]. This diagnostic process involves the use of various imaging modalities, including digital mammography [6], ultrasound (US) [7], magnetic resonance imaging (MRI) [8], and infrared thermography [9]. Among these, mammography and ultrasound are recognized as the primary modalities for medical imaging that are utilized for the purpose of identifying BrC in early stage [3]. But mammography effectiveness is reduced when dealing with dense breasts. However, ultrasound remains unaffected by breast density and offers a radiation-free, non-invasive, cost-efficient, and rapid method for BrC detection [10], [11].

The manual analysis and diagnosis of BrC from these modality images demand specialized expertise, prove to be arduous, time-consuming, resource-intensive, and susceptible to errors, in addition to incurring substantial costs [12], [13]. Hence, computer-aided diagnosis systems (CADs) that rely on convolutional neural networks (CNNs) provide significant support to medical practitioners in the identification of anomalies through automated feature extraction and classification techniques [10], [14]. These CNN models exhibit distinctive qualities including automated feature extraction, proficiency in capturing high-level features, and the capacity for transfer learning [15]. Through the utilization of CNN-based CAD systems, accuracy in detection is notably heightened, offering radiologists assistance in the swift assessment of suspicious lesions [16]. Consequently, this diminishes the occurrence of interpretational errors by reducing instances of false positives and false negatives, ultimately contributing to an enhancement in patient survivability rates [17], [5].

However, designing a credible CAD system to classify BrC based on medical modality images is intricate due to the inherent challenges posed by the blurriness and low contrast prevalent in these images [18]. The existence of variations among distinct cases and the limited availability of publicly accessible BrC datasets, coupled with the complication of segregating irrelevant features, contribute to the complexity [3]. Despite the numerous CAD systems that have been formulated by researchers, which involve the integration of diverse medical image modalities and the utilization of various deep learning methodologies for BrC detection, there is an ongoing need for further exploration and experimentation. The areas primed for enhancement encompass refining pre-processing methodologies, advancing techniques for model training and feature extraction, optimizing computational efficiency, and, most significantly, elevating the accuracy of the system.

So, the objective of this study is to present an innovative framework intended for the efficient identification and categorization of BrC. This framework employs ultrasound images as input and delivers a diagnostic result that classifies the observations into benign or malignant categories. The major contributions of our work include:

- To design a hybrid CNN network *'BrC-DeepRBNet'* build from scratch, incorporating diverse architectural components such as residual blocks, bottleneck architecture, and a self-attention mechanism.
- Build two networks based on the framework and trained using contrast-enhanced augmented images.
- Fusion of two feature vectors generated from presented networks using horizontal features fusion approach.
- Feature optimization using GNDO approach.
- A detailed ablation study is performed that demonstrates that our approach outperforms existing methods in terms of improvement in accuracy of 97.05%.

The organization of this paper is as follows: Section II encompasses an exploration of the relevant literature. The detailed explanation of the introduced model is provided within Section III. The outcomes and subsequent discourse are presented in Section IV. Finally, Section V contains the ultimate conclusions drawn from the study.

II. Literature Review

Numerous research studies on BrC diagnosis utilizing different datasets have resulted in the development of various approaches, which are discussed in this section. Sagar et al. [19] introduced an ensemble fuzzy rank model for BrC detection that leverages four base models namely VGG-Net, DenseNet, Xception, and Inception by pre-training the initial layers with ImageNet and fine-tune the last five layers of each model using BU dataset. Each model produces a confidence score which is then combined utilizing fuzzy ensemble technique considering uncertainty as well to get fuzzy-ensemble score on basis of which final prediction is made. The presented ensemble model accuracy is 85.23% which is suboptimal. Gupta et al. [20] presented a new transfer learning based approach that utilizes pre-trained ResNet50 and feature extraction is done by addition of custom layers at the front of the model. The attained accuracy, recall and precision with BU dataset is 97.8, 97.68, and 99.21% respectively. Ahila et al. [21] presented a CAD system that involves image enhancement through sigmoid filter and de-speckling through IDAD. After segmentation is done by using hybrid approach utilizing k-Means with FCM algorithm. In feature selection texture and morphological features are joined into one feature vector and reduced through PCA approach. In this work NN is joined with Haar wavelet theory and GWO algorithm is applied to fine-tune its parameters attaining accuracy of 98%, respectively. Adyasha et al. [22] experimented five hybrid CNNs that are based on combination of various CNN models. The approach takes leverage of each individual model and utilizes probability- weight value and threshold value by empirically attempting various values and choosing the optimal one. The hybrid network ShuffleNet-ResNet performs superior with accuracy of 96.52 for anomaly and 93.18% for malignity detection for BU dataset. Yuhao et al. [23] offered a HoVer-Transformer model which employs the prior understanding of malignant and non-cancerous tumors and connects CNN with the transformer. The model uses patch embedding, vertical strip embedding, and horizontal strip embedding to obtain the within- and between-layer spatial knowledge. There are four stage modules in the model, whereas every stage module contains various HoVer-Trans blocks, 1 Conv block, and one pool layer. The model generates explainable attention maps and attains better performance. Nasim et al. [5] introduced a new DL method in which the Inception-V3 architecture is altered by shifting Inception modules to residual ones, adjusting module count, and modifying hyperparameters. However, the obtained 81% accuracy falls below optimal levels. Lyu et al. [2] introduced a novel system for BrC diagnosis based on a Hierarchical Extreme Learning Machine (H-ELM), comprising two phases. In the initial phase, an ELM sparse auto-encoder is employed for feature extraction, while in the subsequent phase, the extracted features undergo scattering by a randomly generated matrix. The final classification is performed by an original single-layer ELM. The approach is prominent for low resource consumption and a brief 5.31(s) training time but presents drawback with lower accuracy of 86.13% using BUS images.

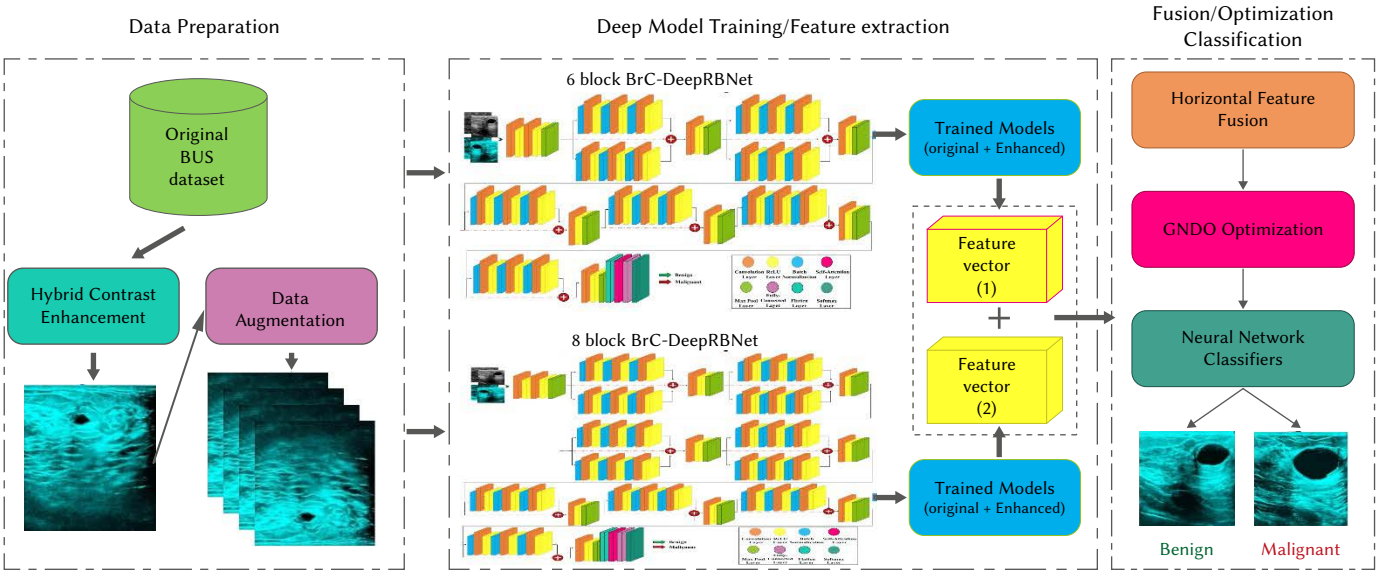


Fig. 1. Presented methodology for BrC classification.

III. Proposed Methodology

The work presented a novel BrC classification approach, illustrated in Fig. 1. At first, a hybrid approach is used to enhance contrast of images on which data augmentation is applied. After that, the presented hybrid CNN architecture *BrC-DeepRBNet* incorporating residual blocks, bottleneck architecture, and a self-attention mechanism is created from scratch and trained with the original augmented images in addition to contrast-enhanced augmented ones. The horizontal concatenation approach is used to fuse the extracted features that were obtained from two deep networks. The fused vector that is obtained is subjected to selection and optimization using the Generalized Normal Distribution approach. For the final classification, neural network classifiers are fed with these optimized features.

A. Data Preparation

This study utilizes the BUS dataset for validation, comprising 780 images with an average size of 500 × 500 pixels. The dataset encompasses three categories: normal-133 images, malignant-210 images-, and benign-487 images [11]. The BUS database images have weak and low contrast, which could lead to misclassification concerns [24]. To overcome this, we used a novel method for contrast enhancement based on segmented histogram equalization (HE) and concatenation of HSV and RGB channels to improve image quality as described in Algorithm 1. Recognizing the inadequacy of the dataset for training a deep learning model, data augmentation is applied which helps minimizing overfitting problem, elevate dataset diversity, and boost the model robustness [3], [25]. The contrast-enhanced BUS images are subjected to numerical computations such as vertical flip, horizontal flip, and 90-degree rotation. Fig. 2 depicts some sample images of the process. These calculations are iterated until each class reaches 8000 images, resulting in a post-augmentation dataset of 24,000 images.

B. Presented BrC-DeepRBNet Architecture

The presented hybrid CNN network *'BrC-DeepRBNet'* is presented in this section. It is widely acknowledged that deeper DL models enhance network's ability to classify by capturing intricate and vital deep features [10], [26]. However, deeper networks pose computational complexities and can suffer from the vanishing gradient issue [3]. Nevertheless, the elements comprising *BrC-DeepRBNet* presented in this study are deliberately selected to enhance the accurate and efficient detection of BrC from ultrasound images, despite the network's depth.

Algorithm 1. Hybrid Contrast Enhancement approach

I (a, b) ← Input image

I_{enh} (a, b) ← Output image

H ← histogram of I1

$\boldsymbol{\mu H}$ ← **mean intensity value**

N ← number of histogram bins

$\boldsymbol{\chi}$ ← transformed image

$\boldsymbol{T}$ ← lookup table transformation

I = 1 to N // for all images

Step 1: I1 = histeq (I) //apply HE

Step 2: $\boldsymbol{\mu_H} = \frac{1}{N}\sum_{k=1}^{N} \boldsymbol{H(k)}$ //Histogram Mean calculation

Step 3: $\boldsymbol{H_{low}} \leq \boldsymbol{\mu_H}$ // Histogram into subsets based on Mean

$\boldsymbol{H_{high}} > \boldsymbol{\mu_H}$

Step 4: $\boldsymbol{\iota_{low}} = \mathbf{histeq}(\boldsymbol{H_{low}})$ //Apply HE to each subset

$\boldsymbol{\iota_{high}} = \mathbf{histeq}(\boldsymbol{H_{high}})$

Step 5: $\boldsymbol{\iota_p} = \boldsymbol{\iota_{low}} + \boldsymbol{\iota_{high}}$ //Merge Sub-Images

Step 6: $\boldsymbol{\chi} = \boldsymbol{T}(\mathbf{I1}, \boldsymbol{\iota_p})$ //Apply Intensity Transformation on I1 based on $\boldsymbol{\iota_p}$

Step 7: $(\boldsymbol{\hbar, s, v}) = \boldsymbol{rgb2hsv}(\boldsymbol{\chi})$ // RGB to HSV conversion

Step 8: $\boldsymbol{\chi_1, \chi_2, \chi_3}$ // Get RGB Channels of $\boldsymbol{\chi}$

Step 9: $\mathbf{I}_{\boldsymbol{enh}} = \boldsymbol{cat}(\boldsymbol{v}, \boldsymbol{\chi_1}, \boldsymbol{\chi_2})$ //Concatenate Channels to get $\mathbf{I}_{\boldsymbol{enh}}$

The proposed network is adept at processing images with a resolution of 227 by 227 pixels. It is meticulously crafted from the scratch, incorporating various architectural components, including residual blocks [27], bottleneck architecture [28], and a self-attention mechanism [29]. Furthermore, the network leverages the advantages of batch normalization (BN) and group normalization (GN), employs ReLU and leaky ReLU as activation functions, and incorporates Max pooling layer. This thoughtful combination of architectural components and normalization techniques contributes to the network's robustness and efficacy in detecting BrC in ultrasound images.

The details of main architecture components are as under:

1. Residual Blocks

The integration of residual blocks within CNNs aids in effective training of deep networks by addressing issues related to vanishing

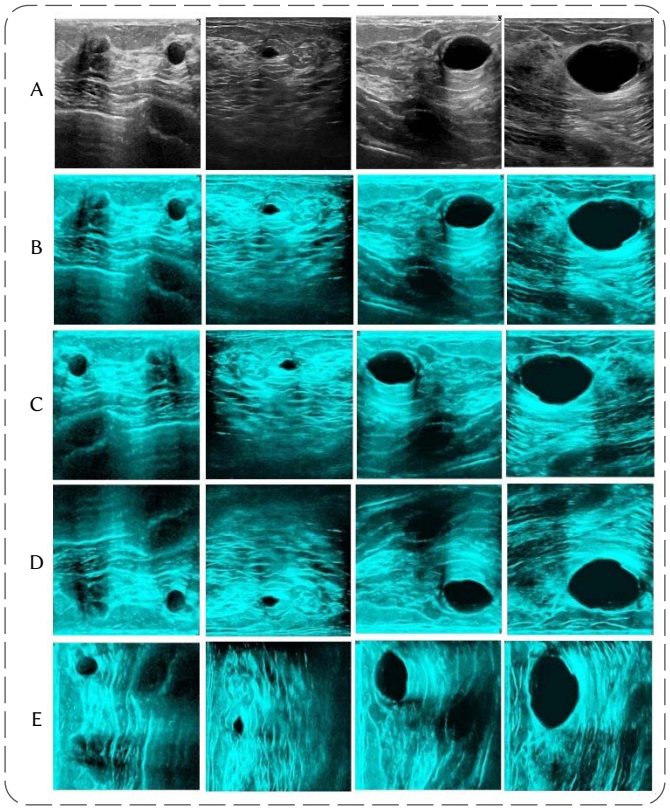


Fig. 2. Sample images (a) Original images, (b) Contrast enhanced images, (c) Horizontal flip image, (d) Vertical flip image, (e) 90-degree rotation.

gradients. This network component adds a shortcut connection, which bypass all layers within the main path by directly channeling the input to the block's output represented mathematically in Equation (1).

$$RB_{op} = RB_{ip} + fun(RB_{ip}) \tag{1}$$

Where RB_{ip} and RB_{op} denote the initial input and resultant output. Equation (2) denotes the residual function $fun(RB_{ip})$ encapsulating the applied transformation within the block. Mathematically, it is defined by Equation (2).

$$fun(RB_{ip}) = \omega_2 \sigma(\omega_1 RB_{ip}) \tag{2}$$

ω_1 and ω_2 represent trainable weights, while σ denotes the activation function.

Thus, this component promotes feature reuse from preceding layers, augments information flow, accelerates training convergence, and elevates overall network performance.

2. Bottleneck Architecture

The bottleneck architecture commences with an initial convolution, referred to as the "bottleneck" layer, wherein the input tensor undergoes a ***1x1 convolution*** operation characterized by a reduced number of filters compared to the previous convolutional layers. After the bottleneck layer, an ***intermediate 3x3 convolutional*** layer with an increased number of filters is applied to capture intricate data patterns. The ***final 1x1 convolution*** layer, featuring a larger number of filters in contrast to the bottleneck layer. This final convolutional operation effectively expands the channel dimensions while transforming the feature representations.

Thus, this component is employed to streamline computational complexity, enhancing the training dynamics of deep networks.

3. Self-Attention Mechanism

Self-attention, a distinctive type of attention mechanism, applied to each channel of feature map separately to create Query (Q), Key (K), and Value (V) vectors as defined in Equation (3).:

$$\begin{aligned} \mathrm{Q}_t &= \mathbb{W}_{Q,t} \times \mathbb{X}_t \\ \mathrm{K}_t &= \mathbb{W}_{k,t} \times \mathbb{X}_t \\ \mathrm{V}_t &= \mathbb{W}_{v,t} \times \mathbb{X}_t \end{aligned} \tag{3}$$

In each channel t, learnable weight matrices $\mathbb{W}_{Q}$, $\mathbb{W}_{k}$, and $\mathbb{W}_{v}$ are used to compute Q, K, and V, where $\mathbb{X}_t$ is the input feature map.

Equation (4) refers that these vectors are pivotal in computing attention scores ($\mathcal{A}s_k$) using Q-K dot product scaled by the square root of the dimensionality of the Key vectors d_k, which guides the model's emphasis on particular elements within the same sequence. Mathematically, it is defined in Equation (4).

$$\mathcal{A}s_t = softmax\left(\frac{\mathrm{Q}_t.K_t^T}{\sqrt{d_k}}\right) \tag{4}$$

Equation (5) represent the output for each channel ($\mathcal{O}_k$) which is obtained by weighted summation of Value vectors using attention scores. Equation (6) depicts the further processing of $\mathcal{O}_k$ by a linear layer with trainable weights ($\mathbb{W}_{o,t}$) to get final output of each channel $\mathcal{O}\mathfrak{p}_k$.

$$\mathcal{O}_k = \mathcal{A}s_k \,.\, \mathrm{V}_t \tag{5}$$

$$\mathcal{O}\mathfrak{p}_k = \mathbb{W}_{o,t} \,.\, \mathcal{O}_k \tag{6}$$

The incorporation of self-attention mechanisms in CNNs showcases adaptability and efficacy across diverse applications, rendering it a valuable inclusion to model architectures [30].

This presented '*BrC-DeepRBNet*' takes leverage of each architectural component and is utilized to create two networks.

a) Architecture of Presented 6 Block BrC-DeepRBNet

The first network configuration comprises of 10.9M total learnable and 107 layers consisting of six residual-bottleneck blocks (6block-RB), encompassing two parallel residual-bottleneck (PRB) blocks and four single residual-bottleneck (SRB) blocks.

As illustrated in Fig. 3, the initial layer of the network presented functions as the input layer, enabling the processing of images sized at 227 × 227×3 as input. Following the input layer, there are Convolution layers with a filter size of 3×3 and a stride of 2, and ReLU layers preceded by series of residual-bottleneck (RB) blocks. Within each RB block, the feature map inherited from the preceding layer is processed through the main bottleneck architecture path as described in section 3.2.2, where 3 convolution layers along with batch normalization and ReLU activation functions are applied. The network incorporates a shortcut connection, which enables the merging of feature representations learned within the main path with the original input, achieved through an element-wise addition operation. The outcome of the addition operation yields a feature map, which is then subjected to processing through a Convolution layer with a 3×3 filter size and a stride of 2, pursued by a ReLU layer and a maxpool layer. This resulting feature map serves as the output of the RB block and is subsequently transmitted to the next layer in the network. Finally, at the culmination of the sequence of residual-bottleneck blocks, a flatten layer, a single self-attention layer, two fully connected layers and a softmax activation function layer is placed. The feature vector is extracted from fully connected layer having dimension of N×2048.

Table I offers additional specifics regarding the showcased network 1.

b) Architecture of Presented 8 Block BrC-DeepRBNet

The second network is deeper and more complex as compared to 6 block BrC-DeepRBNet. It comprises of 149 layers with 12.6M total learnable and has eight residual-bottleneck blocks (8block-RB),

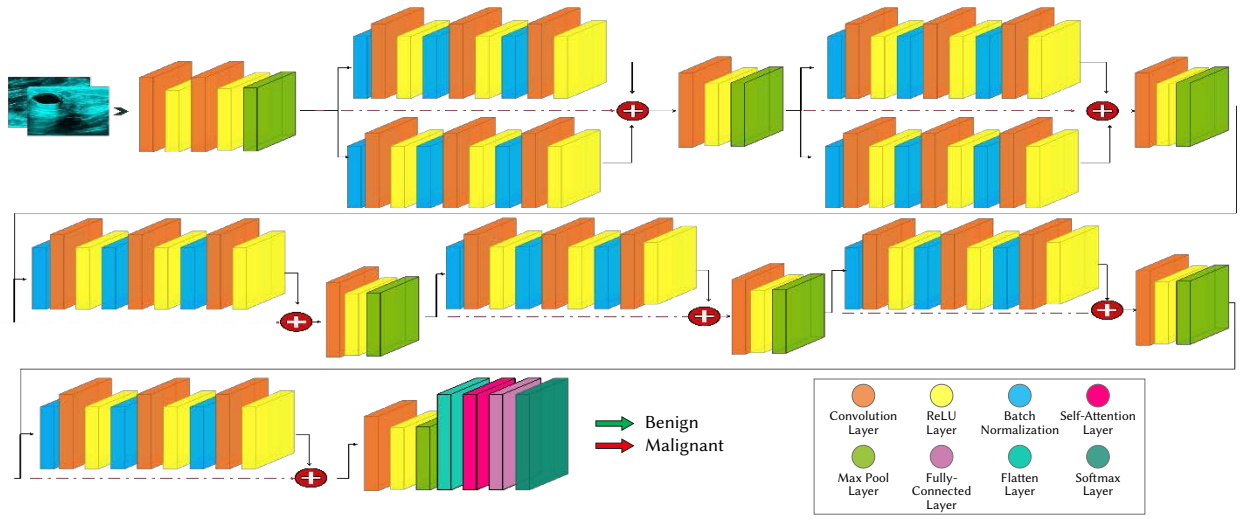


Fig. 3. The detailed structure of presented 6 block BrC-DeepRBNet.

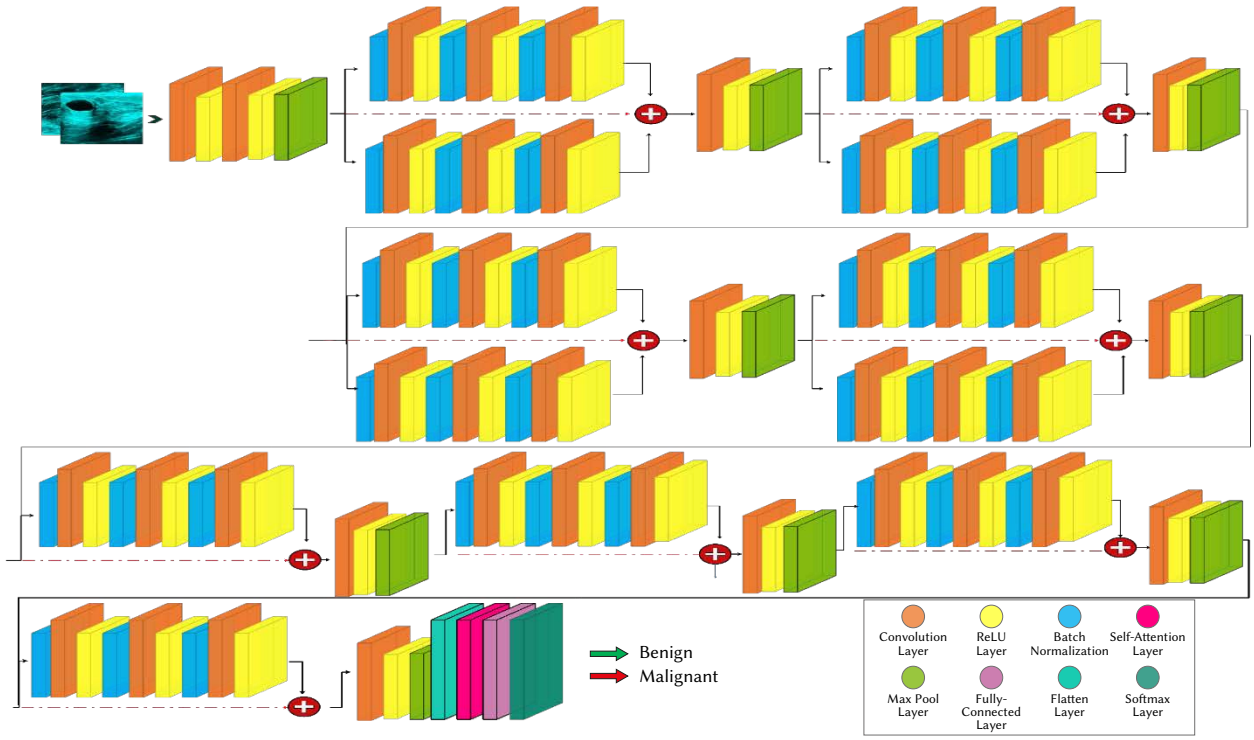


Fig. 4. The detailed structure of presented 8 block BrC-DeepRBNet.

encompassing four (PRB) blocks and four (SRB) blocks.

As illustrated in Fig. 4, the subsequent setup of network2 closely resembles that of network1, yet several modifications have been implemented. Specifically, the convolution layer now incorporates a 2×2 filter size. Normalization is achieved through the utilization of a GN layer instead of BN, and the activation function is replaced with leaky ReLU as opposed to ReLU. Additionally, the feature vector is extracted from the self-attention layer, possessing a dimensionality of N×2048.

For further details on the specifics of network2, Table II provides comprehensive information.

C. Feature Fusion

Feature fusion enables the amalgamation of varied feature vectors, leading to a fused feature set characterized by more comprehensive, discriminative, and robust feature representations [3]. Consider two vectors U and V, each characterized by the dimension $N \times 2048$ where N denotes the number of rows. These dimensions are derived by employing two deep residual-bottleneck networks. The fusion of these vectors is accomplished using the horizontal concatenation approach, resulting in a new vector τ with dimensions $N \times (KU + KV)$, where KU and KV represent the number of columns in vectors U and V, respectively. Equation (7) represents this process in mathematical terms.

$$\tau = \begin{bmatrix} U_{11} & U_{12} & \cdots & U_{1kU} & V_{11} & V_{12} & \cdots & V_{1kV} \\ U_{21} & U_{22} & \cdots & U_{2kU} & V_{21} & V_{22} & \cdots & V_{2kV} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ U_{N1} & U_{N2} & \cdots & V_{NkU} & V_{N1} & V_{N2} & \cdots & V_{NkV} \end{bmatrix} \tag{7}$$

This approach entails the combining of features from distinct feature vectors through their concatenation along the horizontal dimension (columns), contingent upon the condition that the number of rows in both feature vectors are same. The resulting fused vector has dimension N×4096.

TABLE I. Characteristics of the Presented 6 Block BrC-DeepRBNet

S.No	Layer type	No of filter	Filter size
1	Image input		
2	Conv2D+ReLU	32	3×3
3	Conv2D+ReLU	64	3×3
4	2DMaxPool		3×3
	--First PRB block--		
5	BN+Conv2D+ReLU	32	1×1
6	BN+Conv2D+ReLU	64	3×3
7	BN+Conv2D+ReLU	64	1×1
8	BN+Conv2D+ReLU	32	1×1
9	BN+Conv2D+ReLU	64	3×3
10	BN+Conv2D+ReLU	64	1×1
11	Addition(S.No 4, S.No 7, S.No 10)		
12	Conv2D+ReLU	128	3×3
13	2DMaxPool		3×3
	--Second PRB block--		
14	BN+Conv2D+ReLU	256	1×1
15	BN+Conv2D+ReLU	128	3×3
16	BN+Conv2D+ReLU	128	1×1
17	BN+Conv2D+ReLU	256	1×1
18	BN+Conv2D+ReLU	128	3×3
19	BN+Conv2D+ReLU	128	1×1
20	Addition(SNo 13, S.No 16, S.No 19)		
21	Conv2D+ReLU	256	3×3
22	2DMaxPool		3×3
	--1st SRB block--		
23	BN+Conv2D+ReLU	512	1×1
24	BN+Conv2D+ReLU	256	3×3
25	BN+Conv2D+ReLU	256	1×1
26	Addition(S.No 22, S.No 25)		
27	Conv2D+ReLU	512	3×3
28	2DMaxPool		3×3
	--2nd SRB block--		
29	BN+Conv2D+ReLU	512	1×1
30	BN+Conv2D+ReLU	256	3×3
31	BN+Conv2D+ReLU	256	1×1
32	Addition(S.No 28, S.No 31)		
33	Conv2D+ReLU	1024	3×3
	--3rd SRB block--		
34	BN+Conv2D+ReLU	1024	1×1
35	BN+Conv2D+ReLU	512	3×3
36	BN+Conv2D+ReLU	1024	1×1
37	Addition(S.No 33, S.No 36)		
38	Conv2D+ReLU	2048	3×3
	--4th SRB block--		
39	BN+Conv2D+ReLU	1024	1×1
40	BN+Conv2D+ReLU	512	3×3
41	BN+Conv2D+ReLU	2048	1×1
42	Addition(S.No 38, S.No 41)		
43	Conv2D+ReLU	2048	3×3
44	2DMaxPool		3×3
45	Flatten Layer		
46	Self-attention layer		
47	FC Layer		
48	FC Layer		
49	Softmax		

TABLE II. Characteristics of the Presented 8 Block BrC-DeepRBNet

S.No	Layer type	No of filter	Filter size
1	Image input		
2	Conv2D+leakyReLU	32	2×2
3	Conv2D+leakyReLU	64	2×2
4	2DMaxPool		3×3
	--First PRB block--		
5	GN+Conv2D+ leakyReLU	32	1×1
6	GN+Conv2D+ leakyReLU	64	3×3
7	GN+Conv2D+ leakyReLU	64	1×1
8	GN+Conv2D+ leakyReLU	32	1×1
9	GN+Conv2D+ leakyReLU	64	3×3
10	GN+Conv2D+ leakyReLU	64	1×1
11	Addition(S.No 4, S.No 7, S.No 10)		
12	Conv2D+ leakyReLU	64	2×2
13	2DMaxPool		3×3
	--2nd PRB block--		
14	GN+Conv2D+ leakyReLU	32	1×1
15	GN+Conv2D+ leakyReLU	64	3×3
16	GN+Conv2D+ leakyReLU	64	1×1
17	GN+Conv2D+ leakyReLU	32	1×1
18	GN+Conv2D+ leakyReLU	64	3×3
19	GN+Conv2D+ leakyReLU	64	1×1
20	Addition(S.No 13,S.No 16, S.No 19)		
21	Conv2D+ leakyReLU	64	2×2
22	2DMaxPool		3×3
	--3rd PRB block--		
23	GN+Conv2D+ leakyReLU	32	1×1
24	GN+Conv2D+ leakyReLU	64	3×3
25	GN+Conv2D+ leakyReLU	64	1×1
26	GN+Conv2D+ leakyReLU	32	1×1
27	GN+Conv2D+ leakyReLU	64	3×3
28	GN+Conv2D+ leakyReLU	64	1×1
29	Addition(S.No 22, S.No 25, SNo 28)		
30	Conv2D+ leakyReLU	128	2×2
31	2DMaxPool		3×3
	--4th PRB block--		
32	GN+Conv2D+ leakyReLU	256	1×1
33	GN+Conv2D+ leakyReLU	128	3×3
34	GN+Conv2D+ leakyReLU	128	1×1
35	GN+Conv2D+ leakyReLU	256	1×1
36	GN+Conv2D+ leakyReLU	128	3×3
37	GN+Conv2D+ leakyReLU	128	1×1
38	Addition(SNo 31, S.No 34, S.No 37)		
39	Conv2D+ leakyReLU	256	2×2
40	2DMaxPool		3×3
	--1st SRB block--		
41	GN+Conv2D+ leakyReLU	512	1×1
42	GN+Conv2D+ leakyReLU	256	3×3
43	GN+Conv2D+ leakyReLU	256	1×1
44	Addition(S.No 40, S.No 43)		
45	Conv2D+ leakyReLU	512	2×2
46	2DMaxPool		3×3
	--2nd SRB block--		
47	GN+Conv2D+ leakyReLU	1024	1×1
48	GN+Conv2D+ leakyReLU	512	3×3
49	GN+Conv2D+ leakyReLU	512	1×1
50	Addition(S.No 46, S.No 49)		
51	Conv2D+ leakyReLU	1024	2×2
	--3rd SRB block--		
52	GN+Conv2D+ leakyReLU	1024	1×1
53	GN+Conv2D+ leakyReLU	512	3×3
54	GN+Conv2D+ leakyReLU	1024	1×1
55	Addition(S.No 51, S.No 54)		
56	Conv2D+ leakyReLU	2048	2×2
	--4th SRB block--		
57	GN+Conv2D+ leakyReLU	1024	1×1
58	GN+Conv2D+ leakyReLU	512	3×3
59	GN+Conv2D+ leakyReLU	2048	1×1
60	Addition(S.No 56, S.No 59)		
61	Conv2D+ leakyReLU	2048	2×2
62	2DMaxPool		3×3
63	Flatten Layer		
64	Self-attention layer		
65	Flatten layer		
66	FC Layer		
67	FC Layer		
68	Softmax		

D. Generalized Normal Distribution Optimization

GNDO [31], [32], inspired by normal distribution theory, is characterized by a random variable, denoted as $\mathcal{R}$, conforms to a probability distribution governed by two critical variables: location parameter α signifying the mean value, and scale parameter β denoting the standard variance of random variables. Equation (8) expresses the associated probability density function of $\mathcal{R}$.

$$f(\mathcal{R}) = \frac{1}{\sqrt{2\pi}\,\beta}\exp\left(-\frac{(\mathcal{R}-\alpha)^2}{2\beta^2}\right) \tag{8}$$

The structure of GNDO is straightforward, incorporating information sharing strategies that encompass local exploitation and global exploration. Local exploitation is grounded in the established generalized normal distribution model, directed by the current mean position and optimal position. On the other hand, global exploration entails interactions with three randomly selected individuals. A comprehensive elucidation of the two learning strategies is provided below:

1. Local Exploitation

Local exploitation encompasses the quest for improved solutions within the present positions of individuals across the search space. Equation (9) depicts the creation of a generalized normal distribution model for optimization which is facilitated by the correlation between the distribution of individuals in the population and a normal distribution.

$$v_k^t = \alpha_k + \beta_k + \gamma, k = 1,2 \ldots T \tag{9}$$

v_k^t represents the trail vector of the individual indexed at k at time t. The details of three parameters $(\alpha_k, \beta_k, \gamma)$ are as under:

2. Generalized Mean Position

Equation (10) represents computation of the generalized mean position of the kth individual (α_k).

$$\alpha_k = \frac{1}{3}(\mathcal{R}_k^t + \mathcal{R}_{best}^t + \mu) \tag{10}$$

where $\mathcal{R}_{best}^t$ denotes the current optimal position, while μ is indicative of the mean position in the present population. Equation (11) shows its computation.

$$\mu = \frac{\sum_{k=1}^{T}\mathcal{R}_k^t}{T} \tag{11}$$

In this context, $\mathcal{R}_{best}^t$, encapsulates crucial insights pertaining to the global optimal solution. Each individual $\mathcal{R}_k^t$ gravitates towards $\mathcal{R}_{best}^t$ with the aim of augmenting the probability of discovering superior solutions. If $\mathcal{R}_{best}^t$, becomes ensnared in local optima, all individuals persist in converging towards it, resulting in premature convergence. To overcome this, we introduce the mean position μ of the current population. The evolving nature of μ across generations significantly enhances solution refinement. Its incorporation into the local exploitation strategy mitigates the risk of falling into local optima, thereby increasing the overall potential for discovering better solutions.

3. Generalized Standard Variance

Equation (12) refers β_k which functions as a random sequence to enhance local search capabilities by facilitating localized exploration around α_k.

$$\beta_k = \sqrt{\frac{1}{3}[(\mathcal{R}_k^t - \alpha)^2 + (\mathcal{R}_{best}^t - \alpha)^2 + (\mu - \alpha)^2]} \tag{12}$$

where, a greater distance between the individual $\mathcal{R}_k^t$, and both the mean position μ and the best individual $\mathcal{R}_{best}^t$ results in a random sequence with more pronounced fluctuations. This indicates that individuals with lower fitness values have an increased probability of discovering improved solutions when exposed to a highly fluctuating random sequence. Conversely, individuals with higher fitness values benefit from random sequences with weaker fluctuations, thereby augmenting their chances of attaining superior solutions.

4. Penalty Factor

Equation (13) refers computation of the penalty factor γ which enhances the random nature of the β_k. The resulting variances are consistently positive, broadening its overall search capability.

$$\gamma = \begin{cases} \sqrt{-\log(\Re_1)} \times \cos(2\pi\Re_2), & if \;\; p \le q \\ \sqrt{-\log(\Re_1)} \times \cos(2\pi\Re_2 + \pi), & else \end{cases} \tag{13}$$

where p, q, $\mathcal{R}_1$, and $\mathcal{R}_2$ are random values within the range of 0 to 1.

5. Global Exploration

Equation (14) refers the global exploration aspect within GNDO that involves the selection of three individuals at random:

$$v_k^t = \mathcal{R}_k^t + \underbrace{\epsilon \times (|\Re_3| \times v_1)}_{\text{Local information sharing}} + \underbrace{(1-\epsilon) \times (|\Re_4| \times v_2)}_{\text{Global information sharing}} \tag{14}$$

Equation (15) and Equation (16) states the computation of trail vectors v_1 and v_2:

$$v_1 = \begin{cases} \mathcal{R}_k^t - \mathcal{R}_{\mathbb{R}1}^t, & if \;\; f(\mathcal{R}_k^t) < f(\mathcal{R}_{\mathbb{R}1}^t) \\ \mathcal{R}_{\mathbb{R}1}^t - \mathcal{R}_k^t, & else \end{cases} \tag{15}$$

$$v_2 = \begin{cases} \mathcal{R}_{\mathbb{R}2}^t - \mathcal{R}_{\mathbb{R}3}^t, & if \;\; f(\mathcal{R}_{\mathbb{R}2}^t) < f(\mathcal{R}_{\mathbb{R}3}^t) \\ \mathcal{R}_{\mathbb{R}3}^t - \mathcal{R}_{\mathbb{R}2}^t, & else \end{cases} \tag{16}$$

Where, three distinct random integers $\mathbb{R}1$, $\mathbb{R}2$, $\mathbb{R}3$, chosen from the range of 1 to T, satisfy the condition $\mathbb{R}1 \neq \mathbb{R}2 \neq \mathbb{R}3 \neq k$. In Equation (14) the local learning term represents solution $\mathbb{R}1$ sharing information with k while the global information sharing, individual k receives information from $\mathbb{R}2$ and $\mathbb{R}3$. $\mathcal{R}_3$ and $\mathcal{R}_4$ are numbers at random from a standard normal distribution that expand the search space in GNDO, and ϵ, an adjust parameter, is a randomly generated value between 0-1 that balances these local and global sharing strategies.

6. GNDO for Optimization

In GNDO, the population initializes using Equation (17):

$$\mathcal{R}_{kl}^t = \underline{\mathrm{I}}_l + \left(\overline{\mathrm{I}}_l - \underline{\mathrm{I}}_l\right) \times \Re_5, \;\; k = 1,2 \ldots T, l = 1,2 \ldots Q \tag{17}$$

Where Q are design variables. $\mathcal{R}_5$ is random number in range 0-1. The lth design variable Eis bounded by $\underline{\mathrm{I}}_l$ and $\overline{\mathrm{I}}_l$. If the kth individual doesn't achieve a better solution through local or global strategies, a screening mechanism is implemented to introduce improved solutions into the next generation's population, expressed by Equation (18):

$$\mathcal{R}_k^{t+1} = \begin{cases} v_k^t, & if \;\; f(v_k^t) < f(\mathcal{R}_k^t) \\ \mathcal{R}_k^t, & else \end{cases} \tag{18}$$

IV. Experimental Results and Analysis

A. Dataset and Experimental Measures

The deep learning framework introduced in this study is subjected to evaluation and testing using an augmented BUS images dataset, as outlined in Section 3. The dataset undergoes a 50:50 split into training and testing sets. For critical medical datasets, selecting a larger test set allows for a more thorough evaluation of the model's ability to generalize to unseen data, offering more reliable insights into its performance. In the k-fold cross-validation procedure, the value of k is set to 10. The hyperparameters for the training process are defined as follows: epochs = 100, learning rate = 0.0001, mini-batch size = 32, momentum = 0.6, and the chosen loss function is cross-entropy.

To obtain classification results, various neural network classifiers are employed, such as the wide neural network (W-NN), medium neural network (M-NN), bilayered neural network (B-NN), trilayered neural network (T-NN), and narrow neural network (N-NN). The performance metrics considered include Sensitivity, Precision, F1-Score, Matthews's correlation coefficient (MCC) [33], accuracy, and computation time.

B. Experimental Setup

The simulations are executed using MATLABR2023a on a system with a configuration comprising 16GB RAM, 512 SSD, 8GB graphics card, and an Intel Core i7 10th Gen CPU. This thorough evaluation underscores the robustness and reliability of the proposed deep learning framework.

C. Numerical Results

The following experiments have been conducted to validate the methodology presented in this exposition:

1. Classification employing deep features extracted from six block deep residual-bottleneck network applied to contrast-enhanced images.
2. Classification employing deep features extracted from eight block deep residual-bottleneck network applied to contrast-enhanced images.
3. Classification outcomes following the application of feature fusion through a horizontal concatenation approach.
4. Classification outcomes after the implementation of optimization using Generalized Normal Distribution on the fused vector.

1. Experiment 1

This experiment displays classification outcomes by utilizing six block deep residual-bottleneck network on contrast-enhanced augmented BU dataset as indicated in Table III that result in 2048 deep features extracted from FC layer. These features are then inputted into different neural network-based classifiers to assess classification accuracy. It is evident that M-NN classifier excels with an accuracy of 98.20%, recall (98.02%), precision (98.34%), F1-Score (98.18%), and MCC of 96.38, supported by the accompanying CM as depicted in Fig. 5 (a). Furthermore, B-NN classifier requires 558.67 (s) for computation, while the N-NN classifier completes training in 362.89 (s), showing faster processing.

TABLE III. Classification on Contrast-Enhanced Images Utilizing Six Block Deep Residual-Bottleneck Network

Classifiers	Sen (%)	Pre (%)	F1-S (%)	MCC	Acc (%)	Time(s)
W-NN	97.98	98.17	98.07	96.15	98.15	1338.2
M-NN	98.02	98.34	98.18	96.38	**98.20**	529.12
B-NN	97.72	97.90	97.81	95.63	97.83	658.67
T-NN	97.88	97.66	97.77	95.53	97.83	727.51
N-NN	97.88	97.83	97.85	95.70	97.95	**462.89**

2. Experiment 2

The experiment demonstrates classification outcomes employing eight-block deep residual-bottleneck network on contrast-enhanced augmented BU dataset as listed in Table IV, extracting 2048 deep features from the self-attention layer. These features are subsequently used to assess classification accuracy with various classifiers. The W-NN demonstrated superior performance, attaining 95.50% accuracy along with other metrics, including precision (96.67%), recall (94.33%), F1-score (95.48%), and MCC (91.10) supported by the accompanying CM as depicted in Fig. 5 (b). These metrics are also computed for other classifiers. Computation times are documented for all classifiers, revealing that the M-NN exhibited a shorter duration of 91.51 (s), whereas T-NN incurred the lengthiest time of 611.81 (s).

TABLE IV. Classification on Contrast-Enhanced Images Utilizing Eight Block Deep Residual-Bottleneck Network

Classifiers	Sen (%)	Pre (%)	F1-S (%)	MCC	Acc (%)	Time(s)
W-NN	94.33	96.67	95.48	91.10	**95.50**	129.6
M-NN	94.05	96.04	95.04	90.19	95.10	**91.51**
B-NN	93.77	95.25	94.51	89.11	94.55	590.5
T-NN	93.05	94.64	93.84	87.79	93.90	611.81
N-NN	93.05	95.71	94.36	88.91	94.45	321.06

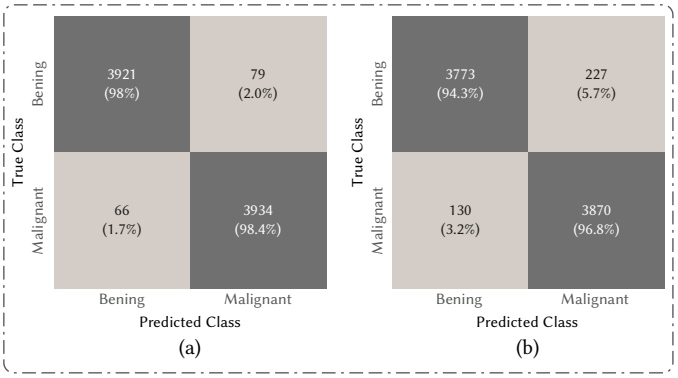


Fig. 5. Confusion matrix using contrast-enhanced images of presented (a) 6 block BrC-DeepRBNet, (b) 8 block BrC-DeepRBNet.

3. Experiment 3

Table V displays the results following feature fusion by applying horizontal concatenation approach on both original augmented and enhanced augmented images using both presented deep networks. The feature vector as result of feature fusion has dimension N×8192 which are then fed to classifiers. Across all classifiers, B-NN attains the highest accuracy at 97.15% but its training time is 334.1(s). W-NN and M-NN secure the second-highest accuracy of 97.05% with training time of 289.08 (s) and 135.07 (s), while T-NN achieves the third highest accuracy at about 96.8%. The values of various metrics of B-NN can be validated using confusion matrix as shown in Fig. 6 (a).

TABLE V. Classification Outcomes of Feature Fusion

Classifiers	Sen (%)	Pre (%)	F1-S (%)	MCC	Acc (%)	Time(s)
W-NN	96.63	97.40	97.01	94.05	97.05	289.08
M-NN	96.95	97.05	97.0	94.0	97.05	**135.07**
B-NN	96.85	97.39	97.12	94.25	**97.15**	334.1
T-NN	96.47	97.13	96.80	93.63	96.80	492.39
N-NN	96.43	96.47	96.45	92.90	96.50	211.28

4. Experiment 4

In this experimental setup, feature optimization is conducted on fused features, resulting in feature vector of dimension N×3929. The obtained accuracies with various classifiers align with those achieved during the feature fusion phase as depicted in Table VI. W-NN achieves the highest accuracy 0f 97.05%, with a prediction time of 221.6 (s). The values of various metrics of W-NN can be validated using confusion matrix as shown in Fig. 6 (b). M-NN and N-NN are the second classifiers with the highest accuracy of 96.95% along with prediction time of 73.06 (s) and 75.88 (s). Thus, when evaluating the combined performance in terms of accuracy and prediction time, M-NN stands out as the optimal choice among the considered approaches.

TABLE VI. Classification Outcomes of Optimization Using GNDO on the Fused Vector

Classifiers	Sen (%)	Pre (%)	F1-S (%)	MCC	Acc (%)	Time(s)
W-NN	96.88	97.19	97.03	94.08	**97.05**	221.6
M-NN	96.92	96.95	96.94	93.88	96.95	**73.06**
B-NN	96.57	96.70	96.64	93.28	96.60	173.11
T-NN	96.17	96.73	96.45	92.93	96.50	192.04
N-NN	96.92	96.92	96.92	93.85	96.95	75.88

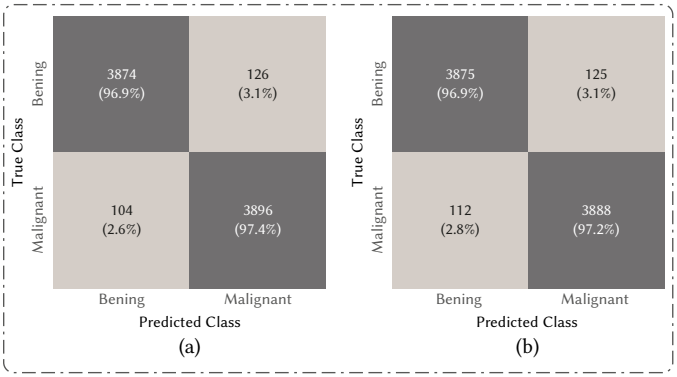


Fig. 6. Confusion matrix of presented (a) feature fusion approach, (b) GNDO based optimization.

D. Ablation Studies

Ablation studies are performed in which proposed networks are compared with pre-trained models on basis of performance, followed by time comparison and comparison with SoA. The selected models include AlexNet, InceptionV3, ResNet50, and ResNet101. Fig. 7 illustrates an accuracy comparison, demonstrating that the proposed 6-block BrC-DeepRBNet and 8-block BrC-DeepRBNet architectures achieved superior performance with accuracies of 98.2 and 95.5% using contrast-enhanced images. The accuracy of remaining pre-trained networks was 93.2, 92.3, 94.9, and 95.3%, respectively.

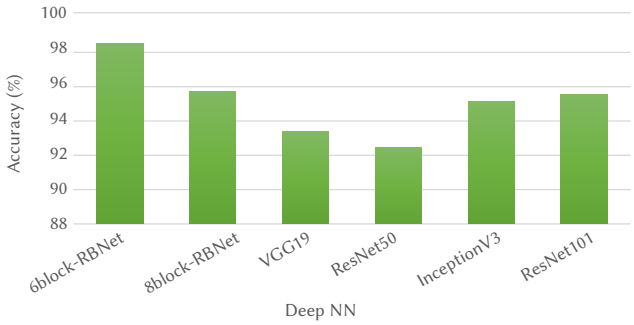


Fig. 7. Analysis of presented networks with pre-trained models.

Furthermore, several optimization algorithms such as PSO, Jaya Optimization, Whale Optimization, and Ant Lion Optimization were replaced in Fig. 1 with GNDO optimization to assess their impact on accuracy. The results depicted in Fig. 8 indicate that GNDO optimization achieves higher accuracy compared to the other optimization approaches.

The time-based comparison in Fig. 9 and Tables V, VI demonstrates how feature selection using GNDO maintains almost consistent accuracy with feature fusion step while cutting down on computation time drastically.

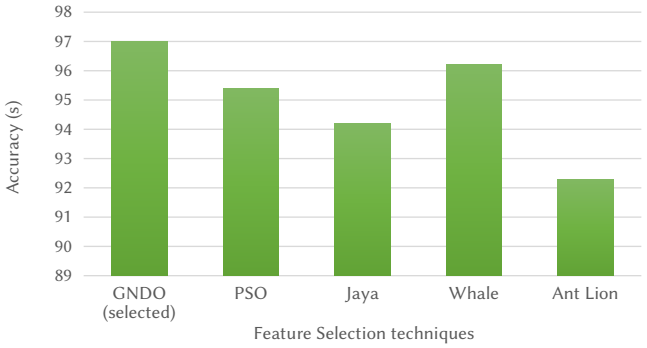


Fig. 8. Analysis of GNDO with other optimization techniques.

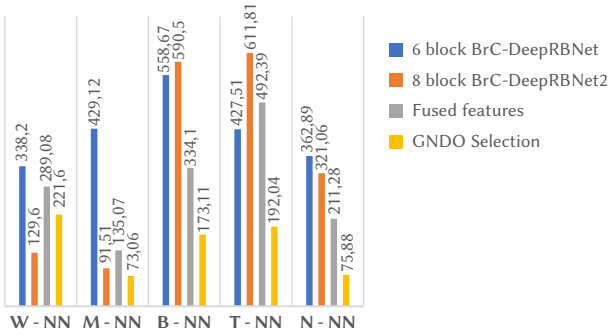


Fig. 9. Comparison of intermediate steps over time BUS dataset.

As demonstrated in Table VII, the same BUS dataset has been studied in recent methods that combined a variety of DL techniques with distinct pre- and post-steps to elevate classification accuracy. In this regard, Sahu et al. [14] achieved maximum accuracy of 94.62%. On the same dataset, our presented approach outperforms these previously presented methods with an accuracy of 97.05%.

TABLE VII: Comparison With SoA

Reference	Year	Methods	Acc (%)
Lyu et al. [2]	2023	Hierarchical extreme learning machine (H-ELM)	86.13
Meng et al. [34]	2023	RMTL-Net ResNet-101 along with regional attention (RA) block.	91.02
Mo et al. [23]	2023	HoVer-Transformer model	85.50
Sagar et al. [19]	2023	Fuzzy ensemble based model	85.23
Islam et al. [35]	2024	EDCNN: integrates MobileNet and Xception	87.82
Sahu et al. [14]	2024	AlexNet, ResNet, and MobileNetV2	94.62
Proposed Methodology			**97.05**

V. Conclusion

This paper presents a novel framework for BrC detection and classification. The framework involves enhancing original images, augmenting both original and enhanced images, and introducing the *'BrC-DeepRBNet'* hybrid CNN model for efficient feature extraction, fusion, and GNDO-based feature selection. These selected features are classified using neural network classifiers. In experiments on the BUS dataset, an enhanced accuracy of 97.0% was achieved.

The noteworthy advancements of this research can be outlined as follows:

- A rapid and adaptive DL based feature extraction model is presented that incorporates residual structures, bottleneck architecture, and attention mechanisms to effectively capture intricate statistical patterns in digital images.
- The utilization of the GNDO approach in feature selection ensures the preservation of fusion accuracy while simultaneously minimizing computational costs.

Furthermore, we performed detailed comparison, and results show that the presented approach outperforms conventional and existing DL techniques in terms of precision, recall, F1-score, and accuracy.

In the future, we aim to apply Generative Adversarial Networks (GANs) for data augmentation. Further, we aim to employ cross-dataset validation to assess the generalization capability of our framework, also more complex datasets will be used for the experimental process.

Acknowledgment

The authors extend their appreciation to the Deanship of Research and Graduate Studies at King Khalid University for funding this work through small group research under grant number RGP1/71/45.

References

[1] A. A. C. Society. (2023, 25 Aug, 2023). *Cancer Facts & Figures 2023.* Available: https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/2023-cancer-facts-figures.html

[2] S. Lyu and R. C. Cheung, "Efficient and Automatic Breast Cancer Early Diagnosis System Based on the Hierarchical Extreme Learning Machine," *Sensors,* vol. 23, p. 7772, 2023.

[3] M. Fatima, M. A. Khan, S. Shaheen, N. A. Almujally, and S. H. Wang, "B2C3NetF2: Breast cancer classification using an end-to-end deep learning feature fusion and satin bowerbird optimization controlled Newton Raphson feature selection," *CAAI Transactions on Intelligence Technology,* vol. 8, pp. 1374-1390, 2023.

[4] M. Nasser and U. K. Yusof, "Deep Learning Based Methods for Breast Cancer Diagnosis: A Systematic Review and Future Direction," *Diagnostics,* vol. 13, p. 161, 2023.

[5] N. Sirjani, M. G. Oghli, M. K. Tarzamni, M. Gity, A. Shabanzadeh, P. Ghaderi, *et al.*, "A novel deep learning model for breast lesion classification using ultrasound Images: A multicenter data evaluation," *Physica Medica,* vol. 107, p. 102560, 2023.

[6] J. M. Lewin, "Digital mammography," *Cancer imaging,* pp. 455-458, 2008.

[7] Y. Kojima and H. Tsunoda, "Mammography and ultrasound features of triple-negative breast cancer," *Breast Cancer,* vol. 18, pp. 146-151, 2011.

[8] M. Kriege, C. T. Brekelmans, C. Boetes, P. E. Besnard, H. M. Zonderland, I. M. Obdeijn, *et al.*, "Efficacy of MRI and mammography for breast-cancer screening in women with a familial or genetic predisposition," *New England Journal of Medicine,* vol. 351, pp. 427-437, 2004.

[9] X. Yao, W. Wei, J. Li, L. Wang, Z. Xu, Y. Wan, *et al.*, "A comparison of mammography, ultrasonography, and far-infrared thermography with pathological results in screening and early diagnosis of breast cancer," *Asian Biomedicine,* vol. 8, pp. 11-19, 2014.

[10] M. Fatima, M. Attique Khan, S. Shaheen, H. M. Albarakati, S. Wang, S. F. Jilani, *et al.*, "Breast Lesion Segmentation and Classification using U-Net Saliency Estimation and Explainable Residual Convolutional Neural Network," *Fractals,* 2024.

[11] X. Yu, Q. Zhou, S. Wang, and Y. D. Zhang, "A systematic survey of deep learning in breast cancer," *International Journal of Intelligent Systems,* vol. 37, pp. 152-216, 2022.

[12] E. M. El Houby and N. I. Yassin, "Malignant and nonmalignant classification of breast lesions in mammograms using convolutional neural networks," *Biomedical Signal Processing and Control,* vol. 70, p. 102954, 2021.

[13] D. kaba Gurmessa and W. Jimma, "Explainable machine learning for breast cancer diagnosis from mammography and ultrasound images: a systematic review," *BMJ Health & Care Informatics,* vol. 31, 2024.

[14] A. Sahu, P. K. Das, and S. Meher, "An efficient deep learning scheme to detect breast cancer using mammogram and ultrasound breast images," *Biomedical Signal Processing and Control,* vol. 87, p. 105377, 2024.

[15] K. Atrey, B. K. Singh, and N. K. Bodhey, "Multimodal classification of breast cancer using feature level fusion of mammogram and ultrasound images in machine learning paradigm," *Multimedia Tools and Applications,* vol. 83, pp. 21347-21368, 2024.

[16] M. Chegini and A. Mahlooji Far, "Uncertainty-aware deep learning-based CAD system for breast cancer classification using ultrasound and mammography images," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization,* vol. 12, p. 2297983, 2024.

[17] D. Muduli, R. Dash, and B. Majhi, "Automated breast cancer detection in digital mammograms: A moth flame optimization based ELM approach," *Biomedical Signal Processing and Control,* vol. 59, p. 101912, 2020.

[18] K. Atrey, B. K. Singh, and N. K. Bodhey, "Integration of ultrasound and mammogram for multimodal classification of breast cancer using hybrid residual neural network and machine learning," *Image and Vision Computing,* vol. 145, p. 104987, 2024.

[19] S. D. Deb and R. K. Jha, "Breast UltraSound Image classification using fuzzy-rank-based ensemble network," *Biomedical Signal Processing and Control,* vol. 85, p. 104871, 2023.

[20] S. Gupta, S. Agrawal, S. K. Singh, and S. Kumar, "A Novel Transfer Learning-Based Model for Ultrasound Breast Cancer Image Classification," in *Computational Vision and Bio-Inspired Computing: Proceedings of ICCVBIC 2022*, ed: Springer, 2023, pp. 511-523.

[21] S. Bourouis, S. S. Band, A. Mosavi, S. Agrawal, and M. Hamdi, "Meta-heuristic algorithm-tuned neural network for breast cancer diagnosis using ultrasound images," *Frontiers in Oncology,* vol. 12, p. 834028, 2022.

[22] A. Sahu, P. K. Das, and S. Meher, "High accuracy hybrid CNN classifiers for breast cancer detection using mammogram and ultrasound datasets," *Biomedical Signal Processing and Control,* vol. 80, p. 104292, 2023.

[23] Y. Mo, C. Han, Y. Liu, M. Liu, Z. Shi, J. Lin, *et al.*, "Hover-trans: Anatomy-aware hover-transformer for roi-free breast cancer diagnosis in ultrasound images," *IEEE Transactions on Medical Imaging,* 2023.

[24] H. Zou, X. Gong, J. Luo, and T. Li, "A robust breast ultrasound segmentation method under noisy annotations," *Computer Methods and Programs in Biomedicine,* vol. 209, p. 106327, 2021.

[25] K. Jabeen, M. A. Khan, M. Alhaisoni, U. Tariq, Y.-D. Zhang, A. Hamza, *et al.*, "Breast cancer classification from ultrasound images using probability-based optimal deep learning feature fusion," *Sensors,* vol. 22, p. 807, 2022.

[26] A. Raza, N. Ullah, J. A. Khan, M. Assam, A. Guzzo, and H. Aljuaid, "DeepBreastCancerNet: A Novel Deep Learning Model for Breast Cancer Detection Using Ultrasound Images," *Applied Sciences,* vol. 13, p. 2082, 2023.

[27] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770-778.

[28] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV 14*, 2016, pp. 630-645.

[29] Z. Niu, G. Zhong, and H. Yu, "A review on the attention mechanism of deep learning," *Neurocomputing,* vol. 452, pp. 48-62, 2021.

[30] A. Rao, J. Park, S. Woo, J.-Y. Lee, and O. Aalami, "Studying the effects of self-attention for medical image analysis," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 3416-3425.

[31] Y. Zhang, Z. Jin, and S. Mirjalili, "Generalized normal distribution optimization and its applications in parameter extraction of photovoltaic models," *Energy Conversion and Management,* vol. 224, p. 113301, 2020.

[32] A. Chempak Kumar and D. Muhammad Noorul Mubarak, "Classification of Esophageal Cancer Using Ensembled CNN with Generalized Normal Distribution Optimization Model and Support Vector Machine Classifier," in *Congress on Smart Computing Technologies*, 2022, pp. 83-111.

[33] D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," *BMC genomics,* vol. 21, pp. 1-13, 2020.

[34] M. Xu, K. Huang, and X. Qi, "A Regional-Attentive Multi-Task Learning Framework for Breast Ultrasound Image Segmentation and Classification," *IEEE Access,* vol. 11, pp. 5377-5392, 2023.

[35] M. R. Islam, M. M. Rahman, M. S. Ali, A. A. N. Nafi, M. S. Alam, T. K.

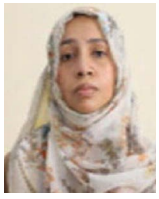
Godder, *et al.*, "Enhancing breast cancer segmentation and classification: An Ensemble Deep Convolutional Neural Network and U-net approach on ultrasound images," *Machine Learning with Applications,* vol. 16, p. 100555, 2024.

Mamuna Fatima

Mamuna Fatima received a Master's degree in Software Engineering from National University of Sciences and Technology, Islamabad, Pakistan in 2014 and a Ph.D degree in Computer Science from HITEC University, Taxila, Pakistan in 2024. She serves as a Lecturer in the Department of Computer Science at COMSATS University Islamabad, Pakistan. Her primary research focuses on medical image processing, computer vision, medical image analysis, and the application of deep learning techniques.

Muhammad Attique Khan

Muhammad Attique Khan (Member IEEE) earned his Master's and Ph.D. degrees in Human Activity Recognition for Application of Video Surveillance and Skin Lesion Classification using Deep Learning from COMSATS University Islamabad, Pakistan in 2018 and 2022. He is currently an Assistant Professor of the Artificial Intelligence Department, Prince Mohammad Bin Fahd University, Saudi Arabia. Previously, he was affiliated with HITEC University Taxila, Pakistan. His primary research focus in recent years is medical imaging, COVID-19, MRI analysis, Video Surveillance, Human Gait Recognition, Remote Sensing, and Agriculture Plants using Deep Learning. He has above 320 publications that have more than 14,500+ citations and an impact factor of 920+ with h-index 72 and i-Index 210. He is also lead and guest editor of several special issues in the well reputed journals such as Measurement, IET CAAI Transaction, CMC, CSSE, Sustainable Energies, and a few more. Moreover, he worked as an Adjunct Professor at Lebanese American University from September 2023 to July 2024.



Saima Shaheen

Saima Shaheen completed her PhD from NUST in the area of software engineering. She is currently working at HITEC University, Taxila Pakistan. Her research interest include medical imaging, software development, and video surveillance. She is actively working in the area of deep learning research and reviewers in several reputed journals including CAAI Transactions and Mobile Networks.

Seifedine Kadry

Seifedine Kadry received the bachelor's degree in applied mathematics from Lebanese University, in 1999, the M.S. degree in computation from Reims University, France, and EPFL, Lausanne, in 2002, the Ph.D. degree in applied statistics from Blaise Pascal University, France, in 2007, and the HDR degree from the University of Rouen, in 2017. His current research interests include education using technology, system prognostics, stochastic systems, and probability and reliability analysis. He is an ABET Program Evaluator.

Omar Alqahtani

Omar Alqahtani received the master's degree in computer science from the University of Denver, Denver, CO, USA, and the Ph.D. degree in computer science from the University of Colorado Denver, Denver, in 2021.,He is an Assistant Professor with the Department of Computer Science, King Khalid University, Abha, Saudi Arabia, where he also serves as the Program Chair for the Computer Science Program. He has authored and coauthored a couple of papers in SCI high-impact factor journals and conferences as well. His current research projects deal with medical event sequences, Crime Forecasting, and Image processing with Deep Learning techniques. His research interests include data science related to Big Data, spatiotemporal processing, time-series data, event sequence data, and database spatial operators.

M. Turki-Hadj Alouane

M. Turki-Hadj Alouane (Member, IEEE) received the Senior Electrical Engineering Diploma degree from the National Engineering School of Tunis (ENIT), Tunis, Tunisia, in 1989, the M.Sc. degree in systems analysis and signal processing, in 1991, and the Ph.D. Diploma degree in electrical engineering and the National Tenure Diploma degree in telecommunications from ENIT, in 1997 and 2007, respectively.,She is currently a Professor with the College of Computer Science, King Khalid University, Abha, Saudi Arabia. In 1997, she was recruited as an Assistant Professor of electrical engineering with ENIT. In 2007, she was promoted to an Associate Professor of telecommunications at ENIT. From 2010 to 2012, she was a Visiting Associate Professor at the Electricity Department, Polytechnic School of Tunisia (EPT). Since 2012, she has been a Full Professor of telecommunications with the Information and Communication Technologies Department, ENIT. She has coordinated internationally sponsored research projects. Since 1997, she has led more than 20 research master theses and 8 Ph.D. theses. She has authored and coauthored more than 70 papers in impacted journals and conferences. Her research interests include signal processing (speech, image, and video), machine learning, deep learning, and evolutionary algorithms.