



An intelligent healthcare framework for breast cancer diagnosis based on the information fusion of novel deep learning architectures and improved optimization algorithm

Kiran Jabeen^a, Muhammad Attique Khan^{g,*}, Robertas Damaševičius^{b,**}, Shrooq Alsenan^c, Jamel Baili^d, Yu-Dong Zhang^e, Amit Verma^f

^a Department of Computer Science, HITEC University, Taxila, Pakistan

^b Faculty of Applied Mathematics, Silesian University of Technology, 44-100, Gliwice, Poland

^c Information Systems Department, College of Computer and Information Sciences, Princess Nourah Bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia

^d Department of Computer Engineering, College of Computer Science, King Khalid University, Abha 61413, Saudi Arabia

^e Department of Informatics, University of Leicester, Leicester, UK

^f University Centre for Research and Development, Department of Computer Science, Chandigarh University, Gharuan, Mohali, Punjab, India

^g Department of Artificial Intelligence, College of Computer Engineering and Science, Prince Mohammad Bin Fahd University, Al Khobar, Saudi Arabia

ARTICLE INFO

Keywords:

Healthcare
Breast cancer
Mammography
Residual blocks
Deep learning
Features fusion
Feature selection
Classification

ABSTRACT

Breast cancer is diagnosed using mammography imaging. Mammography is an effective screening tool for diagnosing and managing breast cancer. This task is highly time-consuming due to the high similarity between benign and malignant cells. However, for medical intervention, it is important to diagnose breast cancer at an early stage. Recently, deep learning has shown remarkable success in the area of medical imaging for the diagnosis of several cancer types. The DL-based computerized techniques assist in detecting and classifying breast cancer correctly. This article proposes a new computerized architecture based on two novel CNN architectures with Bayesian Optimization and feature selection techniques. Initially, two convolutional neural network (CNN) architectures were designed, named 2-Residual Blocks CNN and 3-Residual Blocks CNN. Both designed architectures have been trained using Mammography images, where hyperparameters have been initialized using Bayesian Optimization (BO). After the training from scratch, deep features are extracted from the average pool layer. Extracted deep features are optimized using an improved optimization algorithm named Simulated Annealing controlled Position Shuffling (SAcPS). The fitness of each iteration is computed using an extreme learning machine (ELM) classifier instead of a fine k-nearest neighbor. The selected features of both CNN architectures are finally fused using a novel serial-controlled Reyni Entropy technique. The fused feature vector is passed to neural networks for final classification. The experimental process was conducted on two publicly available datasets and obtained improved accuracy of 97.7% and 97.3%, respectively. In addition, a detailed comparison is conducted with several recent techniques and shows improved performance.

1. Introduction

According to studies, breast cancer is one of the most common cancers in women worldwide. Breast cancer seriously threatens women's health and way of life (Ragab et al., 2021). It is a disorder where the breast's cells proliferate erratically. Women are far more likely than men

to develop BC, although both genders are susceptible (Elkorany and Elsharkawy, 2023). It has a 90% chance of being treated, the highest chance of any cancer type. Because cancer does not cause immediate discomfort, it is not discovered until major health complications appear (Jabeen et al., 2023). Recent years have seen an increase in incidence rates of 0.5% annually; nevertheless, there has been a consistent decline

* Corresponding author.

** Corresponding author.

E-mail addresses: attique.khan@ieee.org (M.A. Khan), robertas.damasevicius@polsl.pl (R. Damaševičius), shaalsenan@pnu.edu.sa (S. Alsenan), Jabaili@kku.edu.sa (J. Baili).

in breast cancer fatalities, with a 43% overall decline from 1989 to 2020 (Ayana et al., 2023). In the United States, there will be over 640,000 cancer deaths and 2,370,000 new instances of breast cancer in 2022 (Xia et al., 2022). In 2023 and 2024, almost 43,170 women lost their lives to breast cancer. The greatest incidence of breast cancer is shown in non-Hispanic White women, but the highest death rate is seen in non-Hispanic Black women (Nicholson et al., 2024). The initially noticed site of the malignant development is the duct or lobule, which frequently shows no symptoms and has a low chance of spreading (Elkorany and Elsharkawy, 2023). Early detection of a malignant breast tumor and prompt treatment could increase the patient's chance of surviving for five years (Iqbal and Sharif, 2023). The causes for the drop in the death rate are thought to be better treatment choices, earlier discovery through screening, and awareness efforts (Ayana et al., 2023). Consequently, it is essential to implement breast cancer prevention techniques (Lou et al., 2022). Surgery (Bleicher et al., 2016), radiation (Balaji et al., 2016), chemotherapy and hormone (Hassan et al., 2010), treatment are typically used to treat breast cancer. Different imaging methods, such as Mammography (Couto et al., 2024), magnetic resonance imaging (MRI) (Eisen et al., 2024), ultrasound (Mo et al., 2023), or histology, can be used to diagnose breast cancer (Karthiga et al., 2022). The American Cancer Society claims that a 30% reduction in mammography sensitivity increases the chance of breast cancer (Loizidou et al., 2023).

Mammography is the primary screening method for breast cancer, using low-energy X-rays to find any breast abnormalities (Loizidou et al., 2023). Digital mammogram analysis is a trustworthy early detection method (Elkorany and Elsharkawy, 2023); therefore, Mammography is used to identify specific lesions, such as masses, calcifications, architectural abnormalities, or localized asymmetries, to discover breast cancer early (Cantone et al., 2023). High-quality images from Mammography can be used to see the internal anatomy of the breast (Ragab et al., 2021). Mammogram images provide a wealth of information about the breast, including its density, size, shape, and suspected abnormalities such as masses and calcifications (Xie et al., 2020). Well-trained clinicians can use the facts mentioned above to make a preliminary diagnosis. Typically, some hospitals lack qualified doctors despite having Mammography scanning technology and operating clinicians (Xie et al., 2020). The ease of use, low cost, and much-improved imaging findings of calcified breast cancer, which facilitates diagnosis, are just a few benefits of mammogram screening (Xie et al., 2020). These factors have made Mammography screening the most widely used method for diagnosing breast disorders in women over 40 (Xie et al., 2020). The biggest drawback of mammography imaging is how challenging it is to distinguish between thick tissue and malignant tissue. Breast cancer analysis is time-consuming, dependent on the pathologist's experience, and might result in errors owing to variables including weariness and drowsiness (Aslan, 2023).

The computer-aided diagnostic (CAD) approach uses image processing to analyze digital mammograms and is very well-liked and effective (Beura et al., 2015). A crucial component of the CAD systems used to diagnose mammography images is mammogram categorization (Garg and Singh, 2022). Earlier CAD systems mostly used conventional machine-learning techniques for mammography classification to provide higher accuracy and early detection (Xia et al., 2023). The deep learning models show the significant performance in medical imaging tasks such as segmentation and classification. Several pre-trained models employed in the recent years for segmentation and classification medical diseases. The most useful deep learning models are AlexNet, ResNet, MobileNet, and Efficientnet. A few other models are also popular for supervised deep learning when the number of images are huge like ConvNets (Lopes and Valiati, 2017; Spolaor et al., 2024). The use of deep learning techniques for Mammography has now opened up a wide range of new study areas for developers and academics to investigate. Convolution Neural Networks (CNNs), in particular, have been utilized for mammograms' lesion localization, detection, and

classification tasks. It supports radiologists in making precise diagnoses and analyses of concerning lesions (Mohapatra et al., 2022). Several deep learning (DL) techniques have recently been employed for better learning and feature extraction from complex data. These models have been applied in several medical specialties, such as breast, skin, colorectal, and lung cancer (Houssein et al., 2022).

1.1. Challenges and major contributions

Although Deep Neural Networks (DNN) designs have produced good results, training is frequently time-consuming and necessitates sophisticated and expensive hardware platforms. These designs also need sufficient training data to function well (Nicholson et al., 2024). Exploring the numerous potential architectures can take time, and finding the best deep-learning model for a specific dataset might be difficult (Iqbal and Sharif, 2023). We also need to investigate a sizable set of hyper-parameters to optimize the DL model. Finally, because DNNs are viewed as a "black box", it can be challenging to evaluate classification findings. The major contributions of this work are defined as follows:

- Designed a novel 2-Residual Block Convolutional Neural Network (CNN) that includes two residual blocks and several layers for classification. These blocks reduced the parameters and consume less computational time.
- We designed a novel lightweight 3-Residual Block CNN that includes several layers under the three residual blocks.
- Instead of manual initialization of hyperparameters of CNN models, we utilized the Bayesian Optimization algorithm. Through BO, hyperparameters are initialized for the learning of proposed models.
- Proposed an improved feature selection algorithm for the best feature selection.
- We have proposed a novel fusion method named serially controlled reyni entropy.

The rest of the manuscript is organized in the following order. Literature review of this manuscript is presented in Section 2. Proposed method including models training, models testing, and classification are presented in Section 3. Following that, the result section is presented in Section 4 that follows the conclusion of this work in Section 5.

2. Literature review

Mass classification in Mammography is still a big problem even though it is crucial for helping radiologists make correct diagnoses. Many researchers contributed to classifying mass images of breast cancer as benign and malignant (Saini and Susan, 2022). Hariraj et al. (2018) suggested that the FMSVM version showed promising results and may pave the way for more complex statistical feature-based cancer prediction models. The suggested method was assessed on a set of images taken from Mini MIAS databases. The analysis of the images reveals that the suggested method was effective and efficient for identifying malignant, benign, and normal tumors, with an accuracy rate of nearly 98%. Chen et al. (2019) developed a fine-tuned residual network that shortened the training times and extracted the most prominent features. After that, a data augmentation process was conducted to increase the training data. A publicly available dataset, CBIS-DDSM, was used for the experimental process and obtained improved accuracy and sensitivity rates of 93.15 and 93.83%, respectively. Setiawan et al. (2015) introduced a Law's Texture Energy Measure (LAWS) technique for extracting texture features from mammogram images. Artificial neural network (ANN) was employed as image classifiers for distinguishing between benign and malignant conditions. The MIAS database was used to retrieve training information for the mammography classification model and achieved 93.90% accuracy for normal-abnormal and 83.30% for the classification of benign tumors.

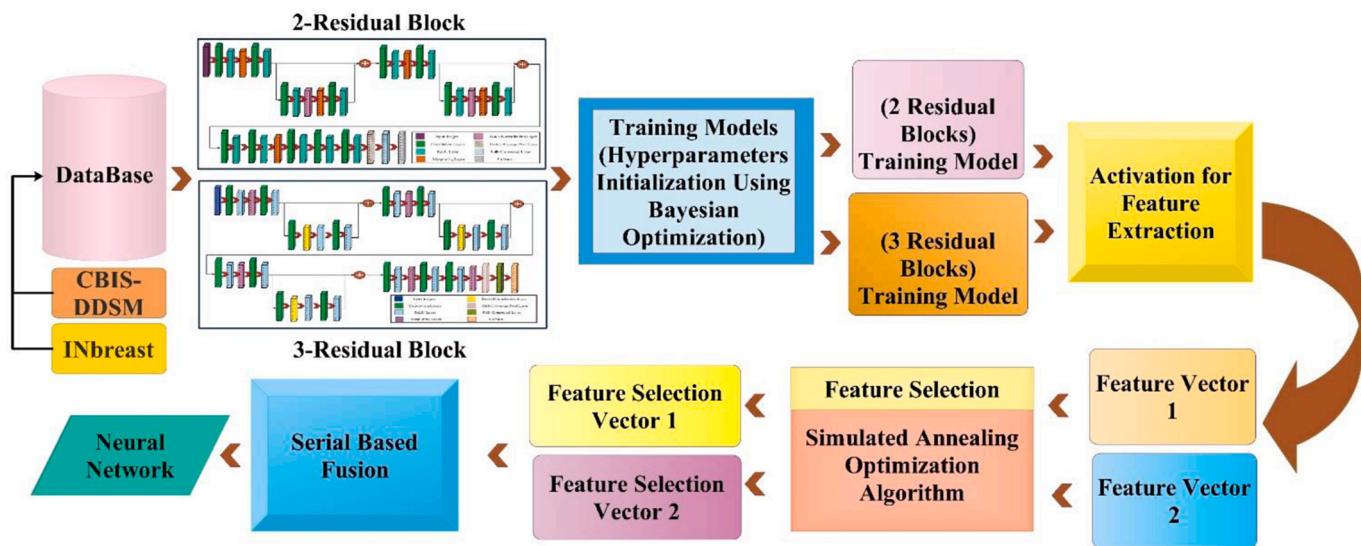


Fig. 1. Proposed methodology of breast cancer classification and diagnosis.

Lou et al. (2022) presented a breast database preprocess (BDP) method for important feature extraction of mammograms. The presented method also addressed the issue of unbalanced categories that were later helpful in the accurate classification. On the presented method, the obtained accuracy, recall, and AUC values on INbreast were 0.960, 0.929, and 0.928, respectively. Heenaye-Mamode Khan et al. (2021) presented a Deep Convolutional Neural Network (CNN) to classify and segment the various types of breast abnormalities. First, transfer learning using the pre-trained model ResNet50 was done on the selected dataset. Then, a new model is designed to control the learning rate adaptively based on changes in error curves throughout the learning process. The performance of the suggested deep learning model in classifying these four forms of breast cancer abnormalities was 88%. Pramanik et al. (2023) presented an approach with two primary stages for classifying breast mass using mammograms. They first incorporated an attention mechanism into the well-known VGG16 model to extract deep features from the input mammograms. To get an ideal features subset, employed a meta-heuristic named the Social Ski-Driver (SSD) method embedded with Adaptive Beta Hill Climbing-based local search. The K-Nearest Neighbours (KNN) classifier performed the classification using the best features subset and obtained the best accuracy of 96.07% on the DDSM database.

Sahu et al. (2024) presented a deep learning-based framework combining the pre-trained models. The authors preprocessed the data by employing the Gaussian-based modified Laplacian high-boosting filter algorithm after that, they utilized mobilenetv2, AlexNet, and ReNet to train the preprocessed images. Through this experiment, the authors achieved the highest accuracy of 97.50%. The limitation of this study was the trained models give a high false prediction. Raiaan et al. (2024) presented a lightweight framework for the enhancement and noise removing mammography breast cancer images. The authors implemented the augmentation process to increase the number of samples and they trained the lightweight CNN. From this experimental process, the authors achieved 98.42% accuracy on the MIAS dataset, the limitation of this work was the extensive use of the augmentation process.

Jayandhi et al. (2022) used mammography image samples to train a DCNN with TL to detect breast cancer. They used the CLAHE contrast enhancement approach and extracted Region of Interest (ROI) for infected region detection. The VGG-16 network model was employed with reduced convolution and max-pooling layers, producing more features for successful mammography categorization into benign or malignant cancer for early diagnosis and proper therapy. The obtained accuracy of the designed model was 82.5%, which was improved than

the recent techniques. Chakravarthy et al. (2023) suggested a robust framework that integrates the ideas of deep transfer learning and Extreme Learning Machines (ELM). The performance of ELM was enhanced by utilizing an iterative flight-length-based crowd-search algorithm (iFLCSA). The parameters of the optimized ELM were discovered using the iFLCSA to improve performance, which was strongly dependent on the parameters of the ELM. For evaluation, digital & full-field mammograms from the INbreast dataset and MIAS were used and showed an accuracy of 98.292% and 98.171%, respectively. Vijayan and Lavanya (2023) suggested a CAD system for classifying benign and malignant tumors. The system used an integrated method based on the fusion of global descriptors. The DDSM dataset was used to validate the suggested method, producing classification accuracy for mass characterization of up to 94.6%. Samee et al. (2022) presented a CLAHE technique for pixel-wise intensity modification and a CNN model for breast cancer classification. This technique's main objective was reducing redundant information, which was done through Logistic Regression (LR) and Principle Components Analysis (PCA). Thwin et al. (2024) utilized the Ensemble-based Channel and Spatial Attention Network (ECS-A-Net), a unique architecture, to automatically identify afflicted areas in BC images. Extensive comparisons performed between multiple competitive state-of-the-art techniques over two benchmarks, DDSM and MIAS, in order to further validate the ECS-A-Net. The suggested model achieved 96.50% accuracy for the DDSM dataset and 95.33% accuracy for the MIAS dataset.

2.1. Summary

In summary, the above-discussed techniques focus on deep models' learning process using augmentation and contrast enhancement. In addition, they implemented a few methods for feature reduction, such as PCA and evolutionary techniques. In addition, they did not consider the concept of information fusion, which can help improve accuracy. Also, they did not focus on the automated selection of automated hyperparameters selection. To address the problem of pre-trained models hyperparameters selection, concept of information fusion, and optimal information selection, we proposed a new fully automated deep learning fusion framework for breast cancer classification and diagnosis.

3. Proposed methodology

This section introduces a proposed methodology for classifying mammograms into benign and malignant. Fig. 1 shows the proposed

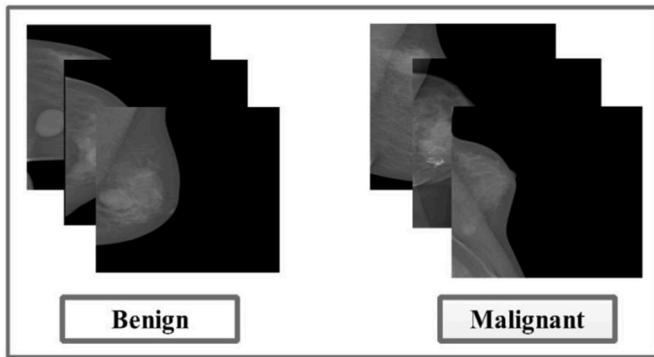


Fig. 2. Sample images of the INbreast dataset.

classification architecture of benign and malignant mammograms. Two datasets, CBIS-DDSM and INbreast, are utilized to evaluate the proposed architecture. In the proposed work, two deep learning models have been designed with residual blocks named the 2-Residual Block Network and the 3-Residual Block Network. Bayesian optimization is utilized to initialize the hyperparameters during training instead of using a manual concept. After that, features are extracted and optimized using an improved version of the simulated annealing optimization algorithm. The selected features are finally fused using a novel parallel convolutional approach that is later classified using neural networks. The description of each step is given in the below subsections.

3.1. Datasets

Digital Database for Screening Mammography (DDSM), which has been updated and standardized, as well as the Curated Breast Imaging Subset of DDSM (CBIS-DDSM) (Ayana et al., 2023), is utilized in this work for the experimental process. A trained mammographer chose and curated a portion of the DDSM data for the dataset.

3.1.1. INbreast dataset

The INbreast dataset consists of Full-Field Digital Mammogram (FFDM) images, including screening, diagnostic, and follow-up cases. The dataset comprises a total of 115 cases. Of these cases, 90 have both cranio-caudal (CC) and medio-lateral oblique (MLO) views for both breasts. The remaining 25 cases have both views for only one breast. Therefore, the dataset contains 410 mammogram images in total (Sathyam et al., 2020). Fig. 2 shows the sample images of the INbreast dataset. The description of the number of images in this dataset has been shown in Fig. 3.

3.1.2. CBIS-DDSM dataset

The Digital Database for Screening Mammography (DDSM) comprises 2620 scanned film mammography studies. A subset of DDSM known as the Curated Breast Imaging Subset of DDSM (CBIS-DDSM) is an updated and standardized version of the dataset. CBIS-DDSM is a comprehensive collection of digitized mammograms that includes

annotations for both mass and calcification (Sathyam et al., 2020). Fig. 4 shows the images of the CBIS-DDSM dataset. Fig. 5 displays the number of images of the CBIS-DDSM dataset.

3.2. Proposed CNN models

3.2.1. Proposed 2-Residual Block Network

A shortcut connection, also known as a skip connection that avoids one or more convolutional layers, makes up a residual block. The gradient can pass directly through the shortcut link, making learning identity mappings simpler for the network. This simplifies the training of deeper networks and lessens the vanishing gradient issue. In this work, a 2-Residual Block CNN architecture is designed that consists of a total of 39 layers and 16 layer deep. The 2-Residual Block network has 37.2 M parameters. Two convolutional layers are applied to the input. The original input is subsequently added element-by-element to the output of the second convolutional layer. After receiving the result, the next layer of the network proceeds to the subsequent stage. A convolutional operation is often followed by a non-linear activation function, such as ReLU (Rectified Linear Unit), in each convolutional layer. The activation function's main objective is to make the network less linear. One essential aspect of the residual block is the element-wise addition operation. It enables the model to pick up residual connections, representing the variation between the convolutional layers' input and output. The network can now concentrate on learning the residual rather than entirely relearning the input representation by adding this difference, or residual, back to the input. Deep learning models can learn more complicated and hierarchical representations by combining numerous residual blocks, resulting in increased performance.

Fig. 6 shows a 2-Residual Block CNN architecture for the classification. In this network, the input size is $227 \times 227 \times 3$ followed by the first convolutional layer of depth 32, filter size of 3×3 , and stride 2. After that, a maxpool layer is added of filter size 3×3 and stride is 1. A second convolutional layer of the same depth followed the first residual block. A ReLu layer has been added as an activation layer for each convolutional layer. In this block, a total of 6 layers have been added, including two

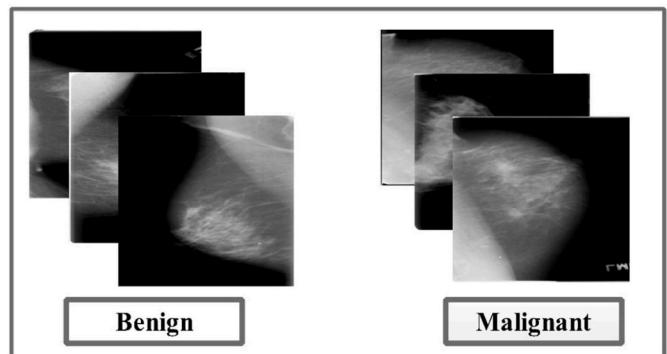


Fig. 4. Sample images of the CBIS-DDSM dataset.

INbreast			
Original Dataset		Augmented Dataset	
Class	#Images	Class	#Images
Benign	76	Benign	1216
Malignant	70	Malignant	1120

Fig. 3. Number of images of the INbreast dataset.

CBIS-DDSM			
Original Dataset		Augmented Dataset	
Class	#Images	Class	#Images
Benign	557	Benign	4939
Malignant	637	Malignant	5096

Fig. 5. Number of images of the CBIS-DDSM dataset.

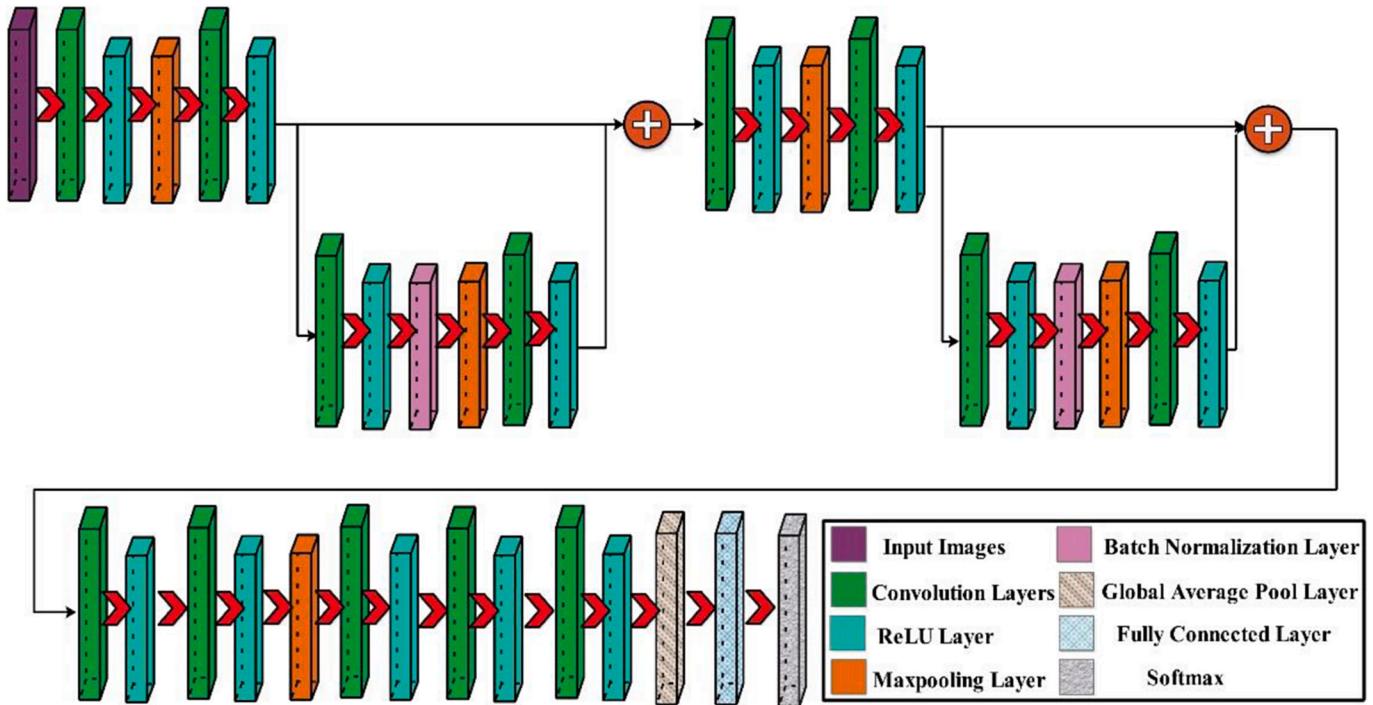


Fig. 6. Proposed 2-Residual Block Architecture for classification of breast cancer.

convolutional, 2 ReLu, one pooling, and one batch normalization. The depth size of each convolutional layer of this block is 64 and 32.

After this block, a convolution layer is added of depth 128, filter size of 3×3 , and stride 2. After that, a max pool layer of filter size 3×3 and stride size two was added. Then, the next convolutional layer is added of depth 256, filter size 3×3 , and stride 2. A new convolutional layer of the same depth that followed the second residual block. In the second block, a total of 6 layers have been added, including two convolutional, 2 ReLu, one pooling, and 1 batch normalization. The depth size of each convolutional layer of this block is 512 and 256.

Following the completion of this block, an additional set of five convolutional layers, which incorporate ReLU activation and max pooling, have been included. The first convolutional layer has a depth size of 1024, while the remaining layers have a depth size 2048. Subsequently, a global average pooling layer, fully connected layer, and softmax layer are appended to the network architecture. Fig. 7 displays the detailed architecture of 2-Residual Blocks. This figure clearly demonstrates the activations, filter sizes, and learnables for each layer.

3.2.2. Proposed 3-Residual Blocks Network

A residual block is a crucial part of deep neural networks, specifically in models such as Residual Networks. The data can explore one or more levels within the block by utilizing a shortcut connection, which carries

through the various layers of the structure. Residual blocks possess the capacity to develop more conceptual and hierarchical features as the network's depth increases (Fang et al., 2021). In our work, we designed 2-residual and 3-Residual Block based CNN because residual blocks can serve as a means to achieve a trade-off between the complexity of a model and its performance. Increasing the depth of a network by adding more residual blocks has the potential to catch more intricate information, but it also carries the danger of overfitting. The second proposed CNN model is 3-Residual Blocks CNN that consists of total 45 layers and 19 layers deep. The 3-Residual Network has 38.6 M parameters. The architecture of this network is shown in Fig. 8. This network accepts the input size of $227 \times 227 \times 3$, that follows the first convolutional layer of depth 32, filter size of 3×3 , and stride 2. After that, a maxpool layer is added of filter size 3×3 and stride is 1. A second convolutional layer of the same depth followed the first residual block. A ReLu layer has been added as an activation layer for each convolutional layer. In this block, a total 5 layers have been added, including two convolutional, 2 ReLu, and one batch normalization. The depth size of each convolutional layer of this block is 64 and 32.

After this block, a convolution layer of depth size 64, filter size of 3×3 , and stride 2 is added. A max pool layer is added after this layer of filter size 3×3 , and the stride is 1. Then, the next convolutional layer is added of depth 64, filter size 3×3 and stride 2. After that, the next

Sr.	Name	Type	Activations	Learnables	Sr.	Name	Type	Activations	Learnables
1	Imageinput 227x227x3 images	Image Input	227x227x3	-	20	relu_7 ReLU	ReLU	15x15x512	-
2	conv_1 32 3x3x3 convolutions with stride[2 2] and padding 'same'	Convolution	114x114x32	Weights 3x3x3x32 Bias 1x1x32	21	batchnorm_2 Batch normalization with 512 channels	Batch Normalization	15x15x512	offset 1x1x512 Scale 1x1x512
3	relu_1 ReLU	ReLU	114x114x32	-	22	maxpool_4 3x3 max pooling with stride [1 1] and padding 'same'	Max Pooling	15x15x512	-
4	maxpool_1 3x3 max pooling with stride [1 1] and padding 'same'	Max Pooling	114x114x32	-	23	conv_8 256 3x3x512 convolutions with stride[1 1] and padding 'same'	Convolution	15x15x256	Weights 3x3x512x256 Bias 1x1x256
5	conv_2 32 3x3x32 convolutions with stride[1 1] and padding 'same'	Convolution	114x114x32	Weights 3x3x32x32 Bias 1x1x32	24	relu_8 ReLU	ReLU	15x15x256	-
6	relu_2 ReLU	ReLU	114x114x32	-	25	addition_2 Element-wise addition of 2 inputs	Addition	15x15x256	-
7	conv_3 64 3x3x32 convolutions with stride[1 1] and padding 'same'	Convolution	114x114x64	Weights 3x3x32x64 Bias 1x1x32	26	conv_9 1024 3x3x256 convolutions with stride[2 2] and padding 'same'	Convolution	8x8x1024	Weights 3x3x256x1024 Bias 1x1x2048
8	relu_3 ReLU	ReLU	114x114x64	-	27	relu_9 ReLU	ReLU	8x8x1024	-
9	batchnorm_1 Batch normalization with 64 channels	Batch Normalization	114x114x64	offset 1x1x64 Scale 1x1x64	28	conv_10 2048 3x3x1024 convolutions with stride[2 2] and padding 'same'	Convolution	4x4x2048	Weights 3x3x2048x2048 Bias 1x1x2048
10	maxpool_2 3x3 max pooling with stride [1 1] and padding 'same'	Max Pooling	114x114x64	-	29	relu_10 ReLU	ReLU	4x4x2048	-
11	conv_4 32 3x3x64 convolutions with stride[1 1] and padding 'same'	Convolution	114x114x32	Weights 3x3x64x32 Bias 1x1x32	30	maxpool_5 3x3 max pooling with stride [2 2] and padding 'same'	Max Pooling	2x2x2048	-
12	relu_4 ReLU	ReLU	114x114x32	-	31	conv_11 2048 3x3x2048 convolutions with stride[2 2] and padding 'same'	Convolution	1x1x2048	Weights 3x3x2048x2048 Bias 1x1x2048
13	addition_1 Element-wise addition of 2 inputs	Addition	114x114x32	-	32	relu_11 ReLU	ReLU	1x1x2048	-
14	conv_5 128 3x3x32 convolutions with stride[2 2] and padding 'same'	Convolution	57x57x128	Weights 3x3x32x128 Bias 1x1x128	33	conv_12 2048 3x3x2048 convolutions with stride[2 2] and padding 'same'	Convolution	1x1x2048	Weights 3x3x2048x2048 Bias 1x1x2048
15	relu_5 ReLU	ReLU	57x57x128	-	34	relu_12 ReLU	ReLU	1x1x2048	-
16	maxpool_3 3x3 max pooling with stride [2 2] and padding 'same'	Max Pooling	29x29x128	-	35	conv_13 2048 3x3x2048 convolutions with stride[1 1] and padding 'same'	Convolution	1x1x2048	Weights 3x3x2048x2048 Bias 1x1x2048
17	conv_6 256 3x3x128 convolutions with stride[2 2] and padding 'same'	Convolution	15x15x256	Weights 3x3x128x256 Bias 1x1x256	36	relu_13 ReLU	ReLU	1x1x2048	-
18	relu_6 ReLU	ReLU	15x15x256	-	37	avgpool2d 3x3 average pooling with stride [1 1] and padding 'same'	Average Pooling	1x1x2048	-
19	conv_7 512 3x3x256 convolutions with stride[2 2] and padding 'same'	Convolution	15x15x512	Weights 3x3x256x512 Bias 1x1x512	38	fc 10 fully connected layer	Fully Connected	10	Weights 10x2048 Bias 10x1
					39	softmax softmax	Softmax	10	-

Fig. 7. Detailed Architecture of proposed 2-Residual Blocks CNN model.

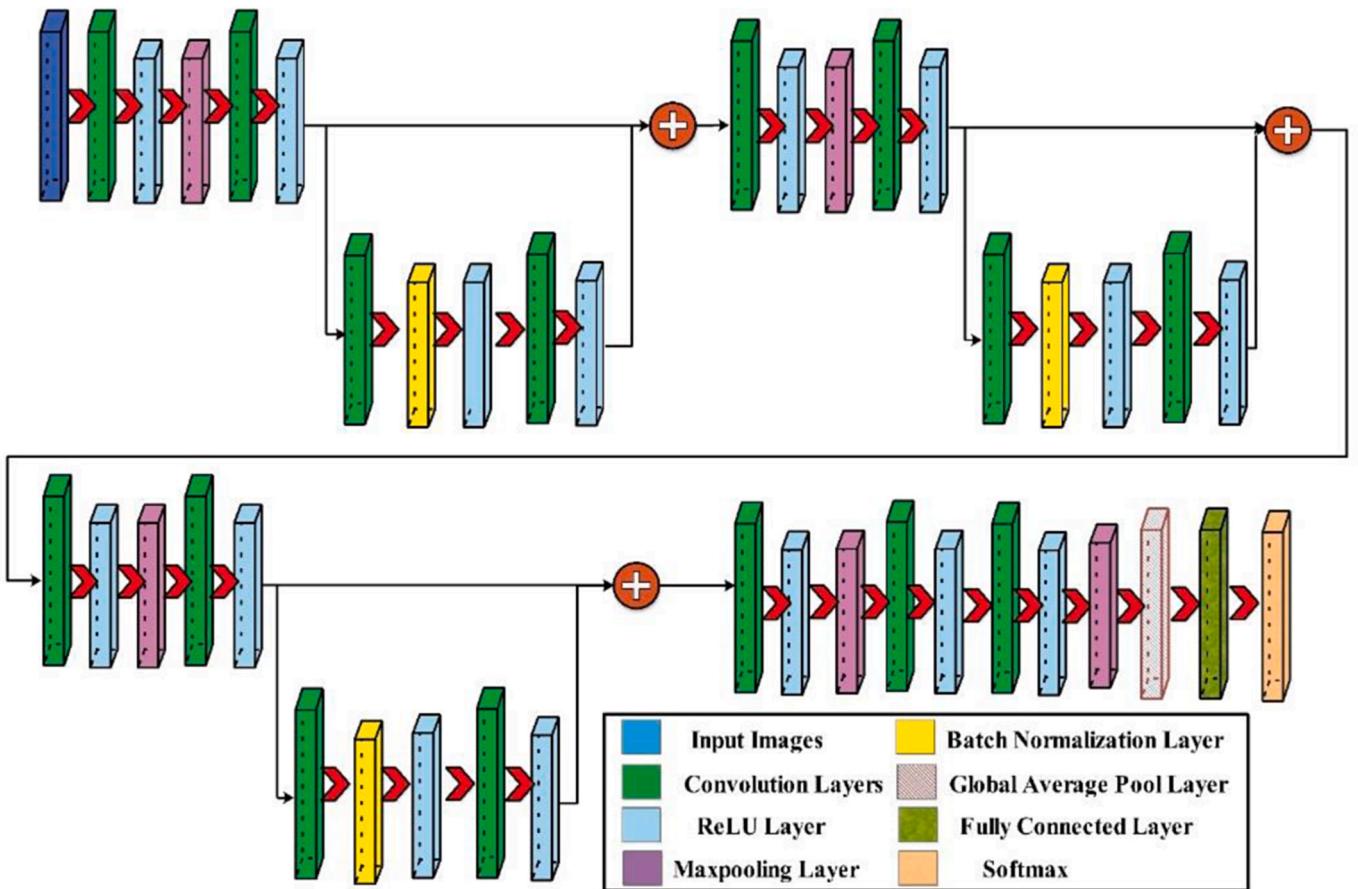


Fig. 8. Proposed 3-residual block architecture.

convolutional layer of the same depth that followed the second residual block is added. In the second residual block, a total of 5 layers have been added, including two convolutional, 2 ReLu, and 1 batch normalization. The depth size of each convolutional layer of this block is 256 and 64. A convolution layer is added outside the second block of depth size 512, filter s of 3×3 , and stride 2. After that, a maxpool layer is added of filter size 3×3 and stride is 1. The convolutional layer is added after depth 1024, filter size 3×3 , and stride 1. Another convolutional layer of the same properties follows the second residual block. The new block contains two convolutional, 2 ReLu, and 1 batch normalization. The depth size of each convolutional layer of this block is 1024.

Following the third residual block, three additional convolutional layers have been incorporated, which include ReLU activation and max pooling layers. The depth size of each convolutional layer is set to 2048. Subsequently, a global average pooling layer, a fully connected layer, and a softmax layer are appended to the network architecture. Fig. 9 displays the detailed architecture of 3-Residual Blocks Network.

3.2.3. Hyperparameters selection

The prediction performance of a model depends on the hyperparameters selection. The proposed networks hyperparameters are learning rate, momentum, L2 Regularization Factor, mini-batch size, and epochs. These parameters controls the network complexity; therefore, it is important to selects the automated values of these hyperparameters. The wrongly selection of these hyperparameters goes for the overfitting of a network. In order to inform the choice of the subsequent

parameter, the Bayesian optimization approach continuously updates the probability distribution indicated by the hyperparameters that have been specified in the objective function. In this paper, we employed the Bayesian Optimization with Gaussian process for the selection of hyperparameters. Mathematically, the BO is defined as follows:

$$\phi(f|Q_{1:t}) = \frac{\phi(Q_{1:t}|f)\phi(f)}{\phi(Q_{1:t})} \quad (1)$$

$$Q_{1:t} = \{(u_1, v_1), (u_2, v_2), (u_3, v_3), \dots (u_t, v_t)\} \quad (2)$$

$$v_t = f(u_t) + e_t \quad (3)$$

Where, f denotes the objective function; v_t denotes the observed value for the t-th step; u_t denotes the hyperparameters of the t-th step; e_t denotes the observation error; $Q_{1:t}$ denotes the aggregation of observation for the previous step; $\phi(Q_{1:t}|f)\phi(f)$ represents the likelihood distribution of the v_t ; $\phi(f)$ denotes the prior distribution of f . The Gaussian process is employed to substitute model, which is mainly determined by the mean and covariance function:

$$f(u) \sim GP(m(u), s(u, u')) \quad (4)$$

$$m(u) = E[f(u)] \quad (5)$$

$$s(u, u') = E[f(u) - m(u)(f(u') - m(u'))] \quad (6)$$

The BO returned the best values after 100 iterations such as a

Sr.	Name	Type	Activations	Learnables	Sr.	Name	Type	Activations	Learnables
1	Imageinput 227x227x3 images	Image Input	227x227x3	-	23	addition_2 Element-wise addition of 2 inputs	Addition	57x57x64	-
2	conv_1 32 3x3x3 convolutions with stride[2 2] and padding'same'	Convolution	114x114x32	Weights 3x3x3x32 Bias 1x1x32	24	conv_9 512 3x3x64 convolutions with stride[2 2] and padding'same'	Convolution	29x29x512	Weights 3x3x64x512 Bias 1x1x2048
3	relu_1 ReLU	ReLU	114x114x32	-	25	relu_9 ReLU	ReLU	29x29x512	-
4	maxpool_1 3x3 max pooling with stride [1 1] and padding'same'	Max Pooling	114x114x32	-	26	maxpool_3 3x3 max pooling with stride [1 1] and padding'same'	Max Pooling	29x29x512	-
5	conv_2 32 3x3x32 convolutions with stride[1 1] and padding'same'	Convolution	114x114x32	Weights 3x3x32x32 Bias 1x1x32	27	conv_10 1024 3x3x512 convolutions with stride[2 2] and padding'same'	Convolution	15x15x1024	Weights 3x3x512x1024 Bias 1x1x1024
6	relu_2 ReLU	ReLU	114x114x32	-	28	relu_10 ReLU	ReLU	15x15x1024	-
7	conv_3 64 3x3x32 convolutions with stride[1 1] and padding'same'	Convolution	114x114x64	Weights 3x3x32x64 Bias 1x1x64	29	conv_11 1024 3x3x1024 convolutions with stride [1 1] and padding'same'	Convolution	15x15x1024	Weights 3x3x1024x1024 Bias 1x1x1024
8	batchnorm_1 Batch normalization with 64 channels	Batch Normalization	114x114x64	offset 1x1x64 Scale 1x1x64	30	batchnorm_3 Batch normalization with 1024 channels	Batch Normalization	15x15x1024	offset 1x1x1024 Scale 1x1x1024
9	relu_3 ReLU	ReLU	114x114x64	-	31	relu_11 ReLU	ReLU	15x15x1024	-
10	conv_4 32 3x3x64 convolutions with stride[1 1] and padding'same'	Convolution	114x114x32	Weights 3x3x64x32 Bias 1x1x32	32	conv_12 1024 3x3x1024 convolutions with stride [1 1] and padding'same'	Convolution	15x15x1024	Weights 3x3x1024x1024 Bias 1x1x1024
11	relu_4 ReLU	ReLU	114x114x32	-	33	relu_12 ReLU	ReLU	15x15x1024	-
12	addition_1 Element-wise addition of 2 inputs	Addition	114x114x32	-	34	addition_3 Element-wise addition of 2 inputs	Addition	15x15x1024	-
13	conv_5 64 3x3x32 convolutions with stride[2 2] and padding'same'	Convolution	57x57x64	Weights 3x3x32x64 Bias 1x1x64	35	conv_13 2048 3x3x1024 convolutions with stride [2 2] and padding'same'	Convolution	8x8x2048	Weights 3x3x1024x2048 Bias 1x1x2048
14	relu_5 ReLU	ReLU	57x57x64	-	36	relu_13 ReLU	ReLU	8x8x2048	-
15	maxpool_2 3x3 max pooling with stride [1 1] and padding'same'	Max Pooling	57x57x64	-	37	maxpool_4 3x3 max pooling with stride [2 2] and padding'same'	Max Pooling	4x4x2048	-
16	conv_6 64 3x3x64 convolutions with stride[1 1] and padding'same'	Convolution	57x57x64	Weights 3x3x64x64 Bias 1x1x64	38	conv_14 2048 3x3x2048 convolutions with stride [2 2] and padding'same'	Convolution	2x2x2048	Weights 3x3x2048x2048 Bias 1x1x2048
17	relu_6 ReLU	ReLU	57x57x64	-	39	relu_14 ReLU	ReLU	2x2x2048	-
18	conv_7 256 3x3x64 convolutions with stride[1 1] and padding'same'	Convolution	57x57x256	Weights 3x3x64x256 Bias 1x1x256	40	conv_15 2048 3x3x2048 convolutions with stride [2 2] and padding'same'	Convolution	1x1x2048	Weights 3x3x2048x2048 Bias 1x1x2048
19	batchnorm_2 Batch normalization with 256 channels	Batch Normalization	57x57x256	offset 1x1x256 Scale 1x1x256	41	relu_15 ReLU	ReLU	1x1x2048	-
20	relu_7 ReLU	ReLU	57x57x256	-	42	maxpool_5 3x3 max pooling with stride [2 2] and padding'same'	Max Pooling	1x1x2048	-
21	conv_8 64 3x3x256 convolutions with stride[1 1] and padding'same'	Convolution	57x57x64	Weights 3x3x256x64 Bias 1x1x256	43	avgpool2d 3x3 average pooling with Stride [1 1] and padding'same'	Average Pooling	1x1x2048	-
22	relu_8 ReLU	ReLU	57x57x64	-	44	fc 10 fully connected layer	Fully Connected	10	Weights 10x2048 Bias 10x1
					45	softmax softmax	Softmax	10	-

Fig. 9. Detailed architecture of 3-Residual Blocks for breast cancer classification.

learning rate of 0.0026, mini-batch size of 128, L2-Regularization Factor of 0.057, and momentum value is 0.628. Based on these hyperparameters, both models are trained and further utilized for feature extraction. In this BO, the Gaussian process is used to guide the search space for the optimal hyperparameters and the stopping condition of BO was the number of iterations. In our case, The BO algorithm was stopped when it completed its 200 iterations. The average pooling layer is selected for the deeper feature extraction of both models that returned a feature vector of dimension $N \times 2048$ and $N \times 2048$, respectively.

3.2.4. Models training and features extraction

Both proposed deep learning models have been trained on the selected dataset from scratch. In the training process, 50:50 approach has been opted and selected the hyperparameters using BO (see section 3.2.3). After the training process, deep features are extracted from the trained proposed deep learning models using activation on average pooling layer. The entire process is mathematically defined as follows:

Given, two newly designed CNN models: $\Delta_1 \in 2\text{-Residual Blocks CNN}$ and $\Delta_2 \in 3\text{-Residual Blocks CNN}$. Both models consists of many layers as shown in Figs. 6 and 7. The training of both models is performed from the scratch and initialized the hyperparameters using BO (section 3.2.3). After the training process, average pool layer is selected from both models and extracted deep features.

$$\xi_{fv1} = ACT(\Delta_1, AP) \quad (7)$$

$$\xi_{fv2} = ACT(\Delta_2, AP) \quad (8)$$

where, ξ_{fv1} and ξ_{fv2} denotes the extracted feature vectors of Δ_1 and Δ_2 , AP denotes the average pool layer, and ACT denotes the activation function. The dimensions of resultant feature vectors of ξ_{fv1} and ξ_{fv2} are $N \times 2048$ and $N \times 2048$, respectively. The resultant vectors are optimized in the later step using an improved version of the Simulated Annealing Optimization algorithm. For CBIS-DDSM, the size of feature vector is 5017×2048 for both models. For INbreast, the size of feature vector is 1168×2048 for both models. Here 2048 are number of extracted features, while 5017 and 1168 are number of images of CBIS-DDSM and INbreast datasets respectively.

3.3. Proposed improved version of SAO algorithm

To model the annealing of metals, Kirkpatrick et al. (1983) created the single-solution method known as SA. Metals are physically hardened by annealing, which begins at a high temperature and gradually cools down. The SA's parameters are first initialized, including the cooling rate (T), final temperature (τ_{final}), and initial starting temperature (τ_0). The highest temperature is the beginning temperature, which decreases over time as a result of the cooling rate to become the final temperature. The algorithm starts with a randomly generated solution. It depends on the current solution being gradually improved. During iterations, a new neighboring solution to the existing solution is chosen. If the new neighboring solution is superior, the current solution is modified. Additionally, the best solution is updated if a neighboring solution is superior. When the target temperature is reached, the algorithm terminates (Abdel-Basset et al., 2021). Each iteration updates the current temperature τ using:

$$\tau = \tau * T, 0 < T < 1 \quad (9)$$

A worse adjacent solution can be accepted by the probabilistic SA algorithm in place of the existing solution to overcome the local optima. According to the definition below, the likelihood of accepting a worse

option depends on how much worse it is and how high the current temperature is.

$$\exp\left(\frac{-\Delta}{\tau}\right), \leq rand \quad (10)$$

where the fitness difference between the present fitness and the new fitness of the surrounding solution is determined by. τ is the current temperature. The exponent to raise e to is $\left(\frac{-\Delta}{\tau}\right)$, where \exp is the exponential function.

The SA enhances the effectiveness further and avoids entering local optima. Because the SA algorithm always favors better solutions, it might also accept less desirable ones depending on the likelihood that they are worse and the current temperature value. In SA, a mutation operation is used to improve the current/present solution throughout iterations incrementally. Based on the current answer, it will produce a new one. In the current solution, the modification operation stores the indices of the chosen characteristics. The mutation then tries to eliminate any unnecessary or irrelevant features from the chosen traits to increase accuracy. The mutation is carried out based on probability to avoid becoming time-consuming, especially for the huge data dimensions. The mutation process improves the SA's effectiveness while also assisting in escaping local optima.

3.3.1. Improved version

In the improved version, the fitness is computed using the ELM (Chorowski et al., 2014) classifier instead of KNN. We used a single layered feed-forward neural network ELM classifier as a fitness function. The fitness value is computed as:

$$fns = \frac{1}{1 - accuracy} \quad (11)$$

Based on the fitness value, the selected features are shuffled again using the following equation:

$$h_k^i = g_{best}^i + S_k^i(g_{p1}^i - g_{p2}^i) \quad (12)$$

where, h_k^i denotes the mutation array, $i = 1, 2, 3, \dots, N$, the variable $p1$ and $p2$ denotes the random feature values, and $p1 \neq p2$. The notation S_k^i denotes the scale parameter and formulated as follows:

$$S_k^i = S_{ini} + (S_{final} - S_{ini}) \times \frac{S(g_k^i) - S(g_{best}^i)}{S(g_{worst}^i) - S(g_{best}^i)} \quad (13)$$

Where, g_{worst}^i and g_{best}^i denotes the worst and best output during each iteration. The features that have greater accuracy they are selected as the best features. The crossover process is carried out to generate a feature matrix. $Q_{kl}^i = (Q_{k1}^i, Q_{k2}^i, \dots, Q_{kd}^i)$:

$$Q_{kl}^i = \begin{cases} h_{kl}^i & \text{if } i = j, \text{ and } rand(0, 1) \leq CR \\ g_{kl}^i & \text{else} \end{cases} \quad (14)$$

$$g_k^{i+1} = \begin{cases} Q_k^i & \text{if } S(Q_k^i) < S(g_k^i) \\ g_k^i & \text{else} \end{cases} \quad (15)$$

The stopping condition of the improved SOA algorithm is several iterations. The algorithm was stopped when it approached 200 iterations. The returned features are $N \times 1,229$. The selected features of both deep models are further fused using a novel fusion approach. The entire selection process is given under Algorithm 1.

Algorithm 1. A proposed improved version of SOA algorithm for feature selection

3.4. Feature fusion and classification

Feature fusion boosts the discriminative power of the features, improving the accuracy of a classifier. In this work, we proposed a feature fusion technique, which improves accuracy and reduces computational time. The proposed fusion technique is based on a two-

Input: \mathbf{So}_c in SA

$$F = \text{fitness}(\mathbf{So}_c)$$

To record the indices of the selected attributes in \mathbf{So}_c , define the vector "*sel_attri*."

$$\mathbf{S}_{MU} = \mathbf{So}_c$$

for j=1 to length of *sel_attri* // for every chosen attribute in \mathbf{So}_c

 Generate a random number using *rand_num*

if (*rand_num* < *mp*)

$\mathbf{S}_{mu}[\text{one_posi}[j]] = 0$ while keeping the other attributes same

$$F_{mut} = \text{fitness}(\mathbf{S}_{mu})$$

if ($F_{mut} < F$)

$$F = F_{mut}$$

$$\mathbf{So}_c = \mathbf{S}_{mu}$$

end if

end if

end for

return \mathbf{So}_c

for i=1 to N

 Calculate the fitness using Eq. 11

 Rotate the positions using Eq. 12-13

 Compute best matrix using Eq. 14-15

End for

 Return \mathbf{g}_k^{i+1} // Best features

Table 1

Proposed classification results of 2-Residual Blocks CNN on CBIS-DDSM dataset.

Classifiers	Accuracy (%)	SR (%)	PR (%)	F1S (%)	FNR (%)	AUC	Time Complexity (s)
WNN	95.9	95.95	95.95	95.95	4.05	0.99	130.27
TNN	95.7	95.75	95.7	95.72	4.25	0.98	471.93
NNN	95.6	95.65	95.6	95.62	4.35	0.98	177.97
B NN	95.6	95.6	95.6	95.6	4.4	0.98	290.56
M NN	95.5	95.45	95.45	95.45	4.55	0.98	164.83

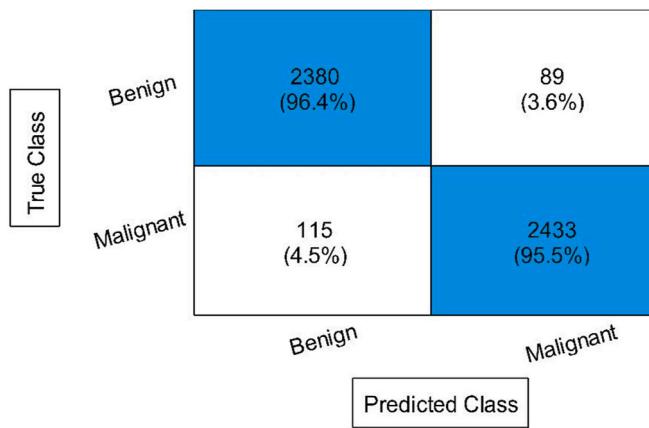


Fig. 10. The confusion matrix and observations for WNN using CBIS-DDSM dataset.

step process. In the first step, all features are aggregated using a serial-based nature, and then, an entropy operation is performed that selects the maximum value features. Serial-based fusion can be defined as follows:

Given two feature vectors of both CNN models defined by φ_1 and φ_2 of dimension $N \times L1$ and $N \times L2$. The length of selected features for both datasets has been different in this work. Serial fusion is done using the following equation:

$$F_k^i = \begin{pmatrix} \varphi_1 \\ \varphi_2 \end{pmatrix}_{N \times L1 + N \times L2} \quad (16)$$

After that, the fused feature vector is employed to reduce some irrelevant information. Mathematically, this process is defined as follows:

$$E_a(F) = \frac{1}{1 - \alpha} \log \left(\sum_{i=1}^n p_i^\alpha \right) \quad (17)$$

$$E_a(F) = \lim_{x \rightarrow \alpha} E_x(F) \quad (18)$$

The features that fall under the range of E_a , will be fused, and the rest of them will be removed. The dimension of resultant feature vector is $N \times 2,458$. The final fused feature vector is finally classified using neural network-based classifiers such as wide and narrow neural networks and named a few more.

4. Experimental results and discussion

For the classification of Breast cancer, the CBIS-DDSM and INbreast datasets have been employed to conduct experiments. Both datasets have been split into 50:50 ratios, with 50% of the images from each class used for training and the remaining 50% used for testing. The 10-fold cross-validation is utilized during both the training and testing processes. In 10-fold cross-validation, every data point in tenfold cross-validation appears once in the test set and nine times in the training set. With respect to different data subsets, this provides a trustworthy evaluation of the model's performance. Additionally, various

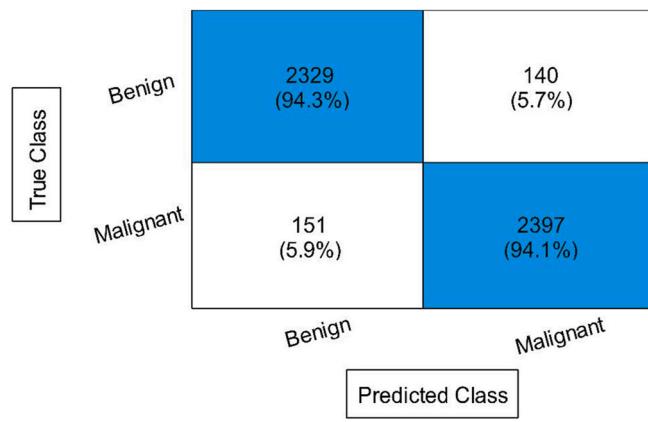


Fig. 11. The confusion matrix and observations for WNN using CBIS-DDSM dataset on proposed 3-Residual CNN.

hyperparameters are applied while training the Residual block network, including learning rate, epochs, momentum, mini-batch size, and optimizer. The values of these hyperparameters have been discussed in the proposed section. The dimension of the feature is large which is $N * 2048$. As a result, we choose neural network classifiers for the classification process because of their capacity to efficiently handle large datasets with millions of data points and high-dimensional feature spaces. They can efficiently train on large datasets using parallel processing and distributed computing infrastructures, which may be impossible for some traditional classifiers such as KNN or SVM. The entire experimental process is carried out on MATLAB2022a using a personal computer with a Core i7 processor having 256 GB of RAM, 12 GB 3060 GPU, and 512 SSD.

The performance measures demonstrated by the classifiers included an accuracy, a sensitivity rate, a precision rate, F1 score, FNR and Time complexity. Several neural network classifiers have been selected, such as Wide neural network (WNN), Medium neural network (MNN), Narrow neural network (NNN), Bi-layered neural network (BNN), and Tri-layered neural network (TNN), for the classification results. The main purpose to utilize these classifiers is their efficiency and better performance compared to other classifiers. WNN, MNN and NNN each have one hidden layer, while Bi-layered Neural Network has two hidden layer and the Tri-layered Neural Network has three hidden layers. All the results are computed using the following steps:

1. Classification using 2-Residual Block network features
2. Classification using 3-Residual Block network features
3. Classification using improved SAO feature selection on 2-Residual Block network features
4. Classification using improved SAO feature selection on 3-Residual Block network features
5. Proposed features fusion

4.1. CBIS-DDSM results

[Table 1](#) demonstrates the classification results of the CBIS-DDSM

Table 2
Proposed classification results of 3-Residual Blocks CNN on CBIS-DDSM dataset.

Classifiers	Accuracy (%)	SR (%)	PR (%)	F1S (%)	FNR (%)	AUC	Time Complexity (s)
WNN	94.2	94.2	94.2	94.2	5.8	0.97	192.6
TNN	93.4	93.35	93.35	93.35	6.65	0.96	196.57
MNN	93.3	93.35	93.35	93.35	6.65	0.97	46.51
NNN	93.1	93.15	93.1	93.12	6.85	0.96	87.7
BNN	93	93	93	93.00	7.00	0.96	133.05

Table 3

Proposed classification results after employing improved feature selection algorithm on 2-Residual Blocks CNN features using CBIS-DDSM dataset.

Classifiers	Accuracy (%)	SR (%)	PR (%)	F1S (%)	FNR (%)	AUC	Time Complexity (s)
WNN	95.2	95.2	95.2	95.2	4.8	0.98	65.161
TNN	95.5	95.45	95.45	95.45	4.55	0.97	219.66
MNN	95.3	95.3	95.3	95.3	4.7	0.98	48.165
NNN	95.3	95.3	95.3	95.3	4.7	0.98	78.718
BNN	95.3	95.25	95.25	95.25	4.75	0.98	155.76

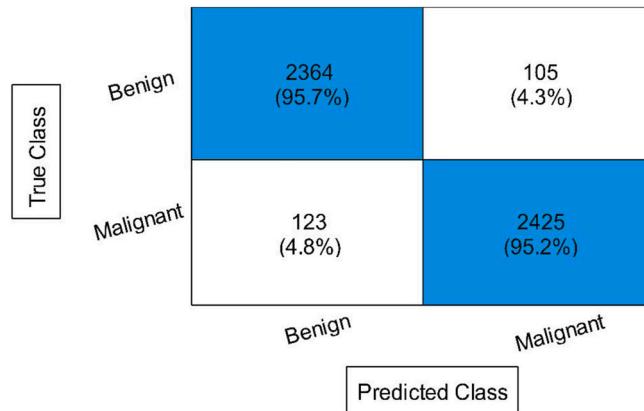


Fig. 12. The confusion matrix and observations for TNN using the CBIS-DDSM dataset on proposed feature selection algorithm for 3-Residual CNN features.

dataset using deep features of 2-Residual Block CNN. The Wide NN is shown as a top-performing classification, achieving an accuracy of 95.9%. This classifier also demonstrates a sensitivity rate (SR) of 95.95%, a precision rate (PR) of 95.95%, and an F1 score (F1S) of 95.95%, along with an FNR of 4.05%. Tri-layered NN obtains the second-best accuracy of 95.7%. The computational time for each classifier is also noted and included in this table. Based on the time, the Wide NN execution time is comparatively less than the other classifiers. Notably, Tri-layered NN records the longest execution time of 471.93 s. **Fig. 10** displays the confusion matrix alongside the recorded observations.

Table 2 shows the classification outcomes of the CBIS-DDSM dataset using deep features from the original dataset. The top-performing classifier is WNN, achieving an accuracy of 94.2%. This classifier also demonstrates a sensitivity rate of 94.2%, a precision rate of 94.2%, an F1 score of 94.2%, and an FNR of 5.8%, respectively. The second-best accuracy of 93.4% is obtained by the TNN classifier of 93.4%. The computational time for each classifier is also noted, and shown that the Medium NN's execution time is comparatively shorter than the other listed classifiers. Notably, TNN records the longest execution time of 196.57 s. **Fig. 11** displays the confusion matrix alongside the recorded observations.

Table 3 displays the classification outcomes of the CBIS-DDSM dataset using the proposed feature selection algorithm on 2-residual CNN features. The top-performing classifier is TNN, achieving an accuracy of 95.5%. This classifier also demonstrates a sensitivity rate of 95.45%, precision rate of 95.45%, and F1 score of 95.45%, along with an FNR of 4.55%. The second-best accuracy of 95.3% is obtained by MNN,

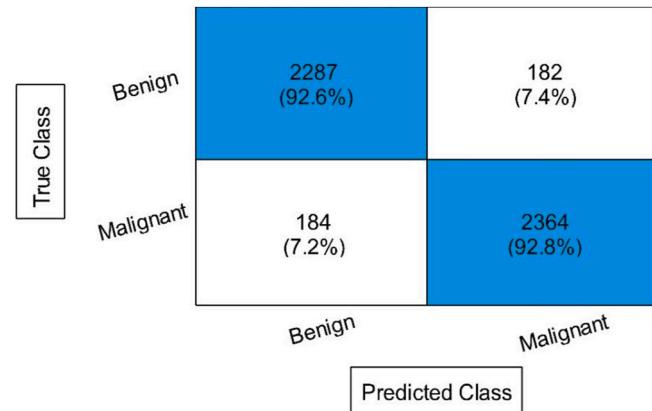


Fig. 13. The confusion matrix and observations for NNN using CBIS-DDSM dataset on proposed feature selection algorithm for 3-Residual CNN features.

NNN, and BNN. The table includes the computational time for each classifier, where NNN execution time is comparatively shorter than the other classifiers. Notably, TNN records the longest execution time of 219.66 s. **Fig. 12** displays the confusion matrix alongside the recorded observations. Compared to the accuracy of **Table 1**, this experiment's performance is improved, showing the proposed optimization algorithm's importance.

Table 4 shows the results of the proposed feature selection on 3-Residual Blocks CNN features. The top-performing classifier is NNN, achieving an accuracy of 92.7%. This classifier also demonstrates a sensitivity rate of 92.7%, a precision rate of 92.75%, an F1 score of 92.72%, and an FNR of 7.3%. The computational time of each classifier is also noted, and it is observed that the MNN execution time is comparatively shorter than the other classifiers. Notably, Tri-layered NN records the longest execution time of 200.27 s. **Fig. 13** displays the confusion matrix alongside the recorded observations.

Table 5 presents the classification results of the proposed feature fusion technique on the CBIS-DDSM dataset. The MNN classifier obtained the highest classification accuracy of 97.3%. This classifier also demonstrates a sensitivity rate of 97.3%, a precision rate of 97.35%, an F1 score of 97.32, and an FNR of 2.7%. The second-best accuracies of 97.3% and 97.1 are obtained by Wide NN and Tri-layered NN, respectively. The table includes the computational time for each classifier, where Medium NN's execution time is comparatively shorter than the other classifiers. Notably, Tri-layered NN records the longest execution time of 35.021 s. **Fig. 14** displays the confusion matrix alongside the recorded observations.

Table 4

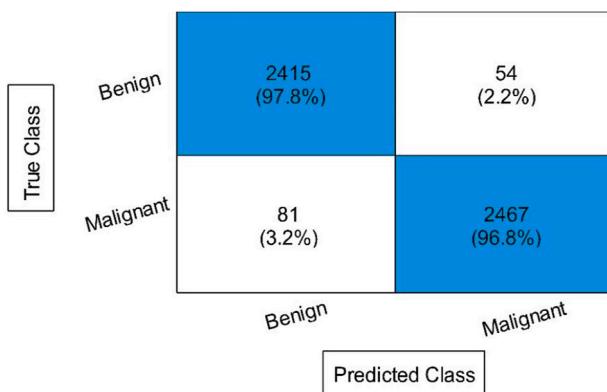
Proposed classification results after employing improved feature selection algorithm on 3-Residual Blocks CNN features using CBIS-DDSM dataset.

Classifiers	Accuracy (%)	SR (%)	PR (%)	F1S (%)	FNR (%)	AUC	Time Complexity (s)
WNN	92.6	92.65	92.6	92.62	7.35	0.96	81.113
TNN	92.6	92.6	92.6	92.6	7.4	0.95	200.27
MNN	92.3	92.35	92.35	92.35	7.65	0.96	80.873
NNN	92.7	92.7	92.75	92.72	7.3	0.96	113.52
BNN	92.4	92.35	92.35	92.35	7.65	0.96	179.61

Table 5

Classification results of proposed feature fusion technique using CBIS-DDSM dataset.

Classifiers	Accuracy (%)	SR (%)	PR (%)	F1S (%)	FNR (%)	AUC	Time Complexity (s)
WNN	97.3	97.30	97.30	97.30	2.70	1.0	30.205
TNN	97.1	97.15	97.1	97.12	2.85	0.99	35.021
MNN	97.3	97.35	97.35	97.32	2.7	0.99	21.636
NNN	97.0	97	96.95	96.97	3.0	0.99	32.585
BNN	97.0	96.95	96.95	96.95	3.05	0.99	32.352

**Fig. 14.** The confusion matrix of MNN after employing the proposed fusion technique.

4.2. INbreast results

The results of INbreast dataset are presented in this section. [Table 6](#) describes the results of first experiment that is deep features extraction from the original dataset. The WNN classifier obtained the highest accuracy of 96.7%. Moreover, the SR of 96.7%, PR of 96.85%, and F1 score of 96.77, respectively. For each classifier, time is also noted and it is observed from this table that the MNN execution time is comparatively shorter than the other classifiers.

[Table 7](#) presents the classification results of the proposed 3-Residual Blocks CNN features. The MNN obtained a maximum accuracy of 97.8%. The obtained accuracy is improved than the performance attained by 2-Residual Blocks CNN features (see [Table 6](#)). Due to adding a few more layers, the performance in terms of computational time is increased for this model. However, the strength of this network is increased by the precision rate (97.8) and F1-Score (97.8). The second highest attained accuracy of 97.8% by BNN is also improved than the performance noted in [Table 6](#). To further optimize the performance of each model feature,

an improved optimization method is proposed and applied to both CNN model's features.

The classification results of the proposed feature selection technique for 2-Residual Blocks CNN are presented in [Table 8](#). In this table, the MNN achieved the highest accuracy of 96.4%, whereas the sensitivity rate of 96.35%, precision rate of 96.5%, and F1 score of 96.42%, respectively. The computational time of each classifier is noted, and it is observed that the time is reduced after employing the proposed technique. In addition, a minor change occurred in the accuracy, which is also a strength of this technique. [Fig. 15](#) illustrated a confusion matrix of this experiment.

The same technique is employed on the features of 3-Residual Blocks CNN and obtained the maximum accuracy of 97.6%. This classifier also demonstrates a sensitivity rate of 97.6%, precision rate of 97.6%, and F1 score of 97.6, which is improved than the performance of [Table 7](#). Time is reduced for all classifiers, which is a strength of this technique. [Fig. 16](#) shows the TNN confusion matrix that can be utilized to confirm true positive rates.

Finally, the classification results of the proposed fusion approach are presented in [Table 10](#). The MNN classifier obtained the highest accuracy of 97.7%, which is improved than the previous experiments. The sensitivity rate of 97.65%, precision rate of 97.7%, and F1 score of 97.67% are also computed and show improvement. The rest of the classifiers also show improved accuracy, and there is another strength: lessens the computational time. Overall, the proposed framework shows improved accuracy after the optimization and fusion. [Fig. 17](#) shows the confusion matrix of the MNN classifier that includes the true positive rates and positive predictive observations. Using these values, the sensitivity rate and precision rate values of the MNN classifier can be confirmed.

4.3. Discussion

A brief discussion of the proposed framework has been presented in this section. [Fig. 1](#) shows the proposed framework that consists of important steps including custom CNN models, models optimization,

Table 6

Proposed classification results of newly designed 2-Residual Blocks CNN on INbreast dataset.

Classifiers	Accuracy (%)	SR (%)	PR (%)	F1S (%)	FNR (%)	AUC	Time Complexity (s)
WNN	96.7	96.70	96.85	96.77	3.3	0.99	65.179
TNN	96.0	95.90	96.05	95.97	4.1	0.98	99.493
MNN	96.3	96.25	96.40	96.32	3.75	0.98	32.389
NNN	96.5	96.45	96.55	96.49	3.55	0.98	63.403
BNN	96.2	96.15	96.3	96.22	3.85	0.98	71.145

Table 7

Proposed classification results of newly designed 3-Residual Blocks CNN on INbreast dataset.

Classifiers	Accuracy (%)	SR (%)	PR (%)	F1S (%)	FNR (%)	AUC	Time Complexity (s)
WNN	97.7	97.7	97.7	97.7	2.3	1	88.871
TNN	97.7	97.7	97.7	97.7	2.3	0.99	94.451
MNN	97.8	97.8	97.8	97.8	2.2	0.99	61.571
NNN	97.2	97.15	97.2	97.17	2.85	0.99	79.485
BNN	97.8	97.75	97.75	97.75	2.25	1	92.917

Table 8

Proposed classification results of improved feature selection technique for 2-Residual Blocks CNN on INbreast dataset.

Classifiers	Accuracy (%)	SR (%)	PR (%)	F1S (%)	FNR (%)	AUC	Time Complexity (s)
WNN	96.1	96.05	96.1	96.07	3.95	0.98	57.953
TNN	95.0	94.95	95.2	95.07	5.05	0.97	55.819
MNN	96.4	96.35	96.5	96.42	3.65	0.98	38.203
NNN	96.3	96.3	96.35	96.32	3.7	0.98	50.245
BNN	95.9	95.8	95.95	95.87	4.2	0.98	51.463

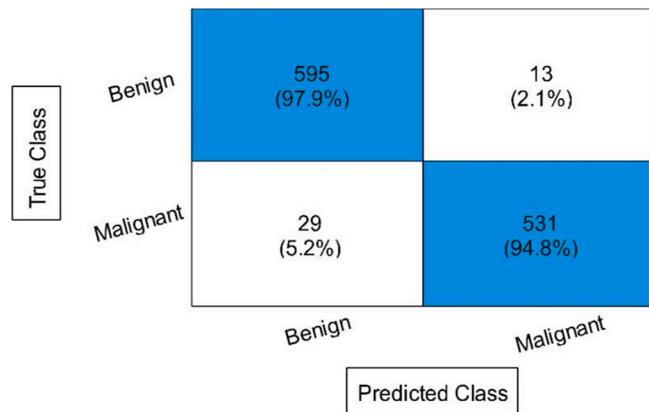


Fig. 15. The confusion matrix and observations of MNN for proposed feature selection on 2-Residual Blocks CNN features.

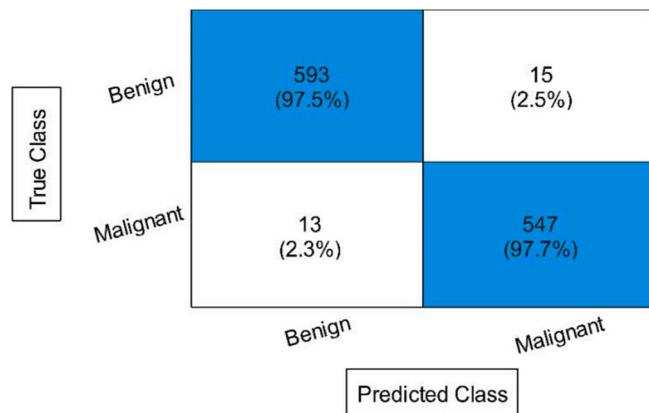


Fig. 16. The confusion matrix and observations of MNN for proposed feature selection on 3-Residual Blocks CNN features.

and fusion. The proposed custom CNN models such as 2-RBCNN and 3-RBCNN results on both datasets are presented in [Tables 1, 2, 6 and 7](#). The performance of these models further improved by the optimization technique and results are given in [Tables 3, 4, 8 and 9](#). Finally, these models are fused and accuracy is improved as shown in [Tables 5 and 10](#). Confusion matrices are given in [Figs. 10–17](#) that can be utilized to confirm the computed measures of the best classifier in each table.

A comparison among proposed custom models and pre-trained models is also conducted, as illustrated in [Fig. 18](#). In this figure, several pre-trained networks are employed for the comparison on both datasets and shows the proposed models better performance. The SqueezeNet model obtained the better precision rate of 93.75 and 95.6% that is best after the proposed methods. These visual and numerical results shows the strength of proposed models. Also, the optimization and fusion process shows the improved accuracy that is strength of this work.

4.3.1. Student's t-test analysis

The student's t-test based analysis is conducted to determine if a noteworthy difference exists between the means of two groups (classifiers). After completing the experiments, the outcomes were evaluated using the student's t-test. Initially, we selected two classifiers from the dataset, each with accuracies aligned across all experiments. This selection allowed us to carry out our experiment using a high-accuracy classifier and a low-accuracy classifier.

The defined hypothesis is h_0 = the accuracy of the chosen classifiers varies significantly in a meaningful way and h_1 = the accuracy of the chosen classifiers has no significant difference. A t-test for the CBIS-DDSM dataset has been conducted comparing two classifiers such as WNN and BNN. The corresponding accuracies of these classifiers are presented in [Tables 11 and 12](#) for CBIS-DDSM and INbreast datasets.

The difference is computed at the initial stage using the below equation. The mean accuracy is computed using Eq. (20) based on the difference value.

$$Diff = |Acc(C1) - Acc(C2)| \quad (19)$$

$$Mean = \mu = \frac{1}{N} \sum_{i=1}^N (Diff_i) \quad (20)$$

where $Acc(C1)$ stands for the uppermost accuracy, and $Acc(C2)$ denotes the lowermost accuracy. μ is the mean values which is calculated from the sum of differences $Diff$ among the classifiers $Acc(C1)$ and $Acc(C2)$. The obtained mean values are 0.42 for the CBIS-DDSM dataset and 0.52 for the INbreast dataset. The symbol N denotes total number of methods. After that, the standard deviation is computed based on Eq. (21). The resultant values are 0.443 for the CBIS-DDSM dataset and 0.524 for the INbreast dataset. Based on the standard deviation value, a t-selection formula is applied that is defined in Eq. (22). The resultant values are 2.114 and 2.212 for CBIS-DDSM and INbreast. These values are checked against the t table based on the degree of freedom (df). The df value is four and the confidence level is 0.05. On this interval, the range is $(-2.776, +2.776)$.

$$Standard Deviation = \sigma = \sqrt{\frac{\sum_{i=1}^N (Diff_i - \mu)^2}{N - 1}} \quad (21)$$

$$t - selection = t = \frac{\sqrt{N} \times \mu}{\sigma} \quad (22)$$

So, according to this confidence interval, our t-selection value falls in this interval, which shows no significant difference among the accuracies of the two selected classifiers. Hence, our h_0 hypothesis is rejected.

4.3.2. Comparison with SOTA

A comparison is conducted at the end with several state-of-the-art techniques (SOTA), as presented in [Table 13](#). This table shows that [Lou et al. \(2022\)](#) presented a method for breast cancer classification and obtained an accuracy of 92.9% using the INbreast dataset. [Melekodappattu et al. \(2022\)](#) introduced a computerized technique and obtained a classification accuracy of 97.5% on the CBIS-DDSM dataset. [Kumari and Jagadesh \(2022\)](#) introduced an ensemble classifier-based approach and obtained an accuracy of 94.3% using the CBIS-DDSM

Table 9

Proposed classification results of improved feature selection technique for 3-Residual Blocks CNN on INbreast dataset.

Classifiers	Accuracy (%)	Sensitivity Rate (%)	Precision Rate (%)	F1-Score	FNR (%)	AUC	Time Complexity (s)
WNN	97.3	97.35	97.35	97.35	2.65	1	41.554
TNN	97.6	97.6	97.6	97.6	2.4	0.99	47.642
MNN	97.3	97.35	97.35	97.35	2.65	0.99	27.077
NNN	97.1	97.05	97.1	97.07	2.95	0.99	24.144
BNN	97.2	97.2	97.15	97.17	2.8	1	45.828

Table 10

Classification results of proposed feature fusion on the INbreast dataset.

Classifiers	Accuracy (%)	SR (%)	PR (%)	F1S (%)	FNR (%)	AUC	Time Complexity (s)
WNN	97.2	97.15	97.25	97.19	2.85	1.00	17.392
TNN	97.2	97.15	97.20	97.17	2.85	0.99	11.924
MNN	97.7	97.65	97.70	97.67	2.35	1.00	12.893
NNN	96.6	96.55	96.55	96.55	3.45	0.99	13.576
BNN	96.6	96.55	96.55	96.55	3.45	0.99	12.209

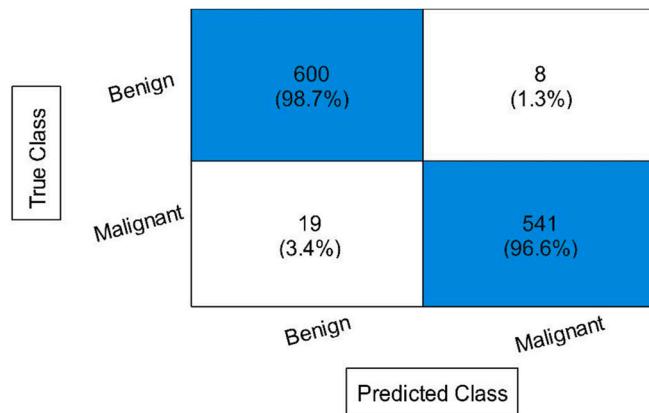


Fig. 17. The confusion matrix of the MNN classifier using proposed features fusion.

dataset. The proposed method obtained an accuracy of 97.7% for the INbreast dataset and 97.3% for the CBIS-DDSM dataset. In 2024, the (Jayandhi et al., 2022) and (Chakravarthy et al., 2023) articles implement the deep learning frameworks on the Mias and INbreast dataset. They achieved 94.87% and 95.86% accuracy. In 2023, the (Vijayan and Lavanya, 2023) authors implemented the automatic deep learning based methodology for the classification of breast cancer by employing the INbreast dataset and gained 91.34% accuracy.

5. Conclusion

In this work, two new CNN architectures have been designed and optimized using an improved optimization algorithm, SAcPS, for accurate breast cancer classification (benign and malignant). The proposed 3-Residual Blocks CNN obtained better accuracy at the initial stage and was further improved by employing the SAcPC feature selection algorithm. The individual performance was insufficient; therefore, a fusion technique (features level) was proposed and obtained enhanced accuracy. The obtained accuracy on the INbreast and CBIS-DDSM datasets using the proposed architecture was 97.7% accuracy, the parameters were sensitivity, precision, f1-score, FNR, and AUC having values was 97.65%, 97.70%, 97.67%, 2.35, and 1.00 and 97.3% accuracy, 97.35% sensitivity, 97.35% precision, 97.32% f1-score, 2.7 FNR, and 0.99 AUC respectively. Choosing the best features improved the accuracy and reduced the computational time. The proposed feature fusion technique improved the accuracy and significantly reduced the computational

time, which is a strength of this step. The limitations of this work are as follows:

- The proposed custom CNN models includes higher parameters that consumes more computational cost and resources during the training process.
- The extracted features from the global average pool layer not fully contains the information of an image. The features extraction from the middle layers can be useful for the better accuracy and precision rates.
- In the optimization process, some important features are also lost due to the fitness function that in returned degrade the accuracy and increase the loss.

Based on these points, the future work is also defined as follows:

- Few advanced transfer learning schemes will be employed to use the pre-trained deep networks to classify benign and malignant breast cancer.
- Network level based fusion architecture will be designed to improve classification performance.
- A hybrid feature selection technique will be proposed to improve the selection capability of essential features.
- The proposed architecture will also be tested on ultrasound images for detection and classification.
- To improve accuracy of the classification model, Generative AI will be used such as Auto-encoders, GAN, and Self-Attention models.

CRediT authorship contribution statement

Kiran Jabeen: Writing – original draft, Software, Methodology, Data curation, Conceptualization. **Muhammad Attique Khan:** Writing – original draft, Software, Methodology, Supervision, Investigation, Formal analysis, Data curation, Conceptualization. **Robertas Damasevicius:** Writing – review & editing, Writing – original draft, Supervision, Software, Conceptualization. **Yu-Dong Zhang:** Writing – review & editing, Supervision, Software, Project administration, Methodology, Formal analysis. **Amit Verma:** Writing – review & editing, Validation, Software, Formal analysis, **Shrooq Alsenan:** Writing – review & editing, Supervision, Software, Project administration, Funding acquisition, Formal analysis, Methodology, Conceptualization, **Jamel Baili:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Conceptualization.

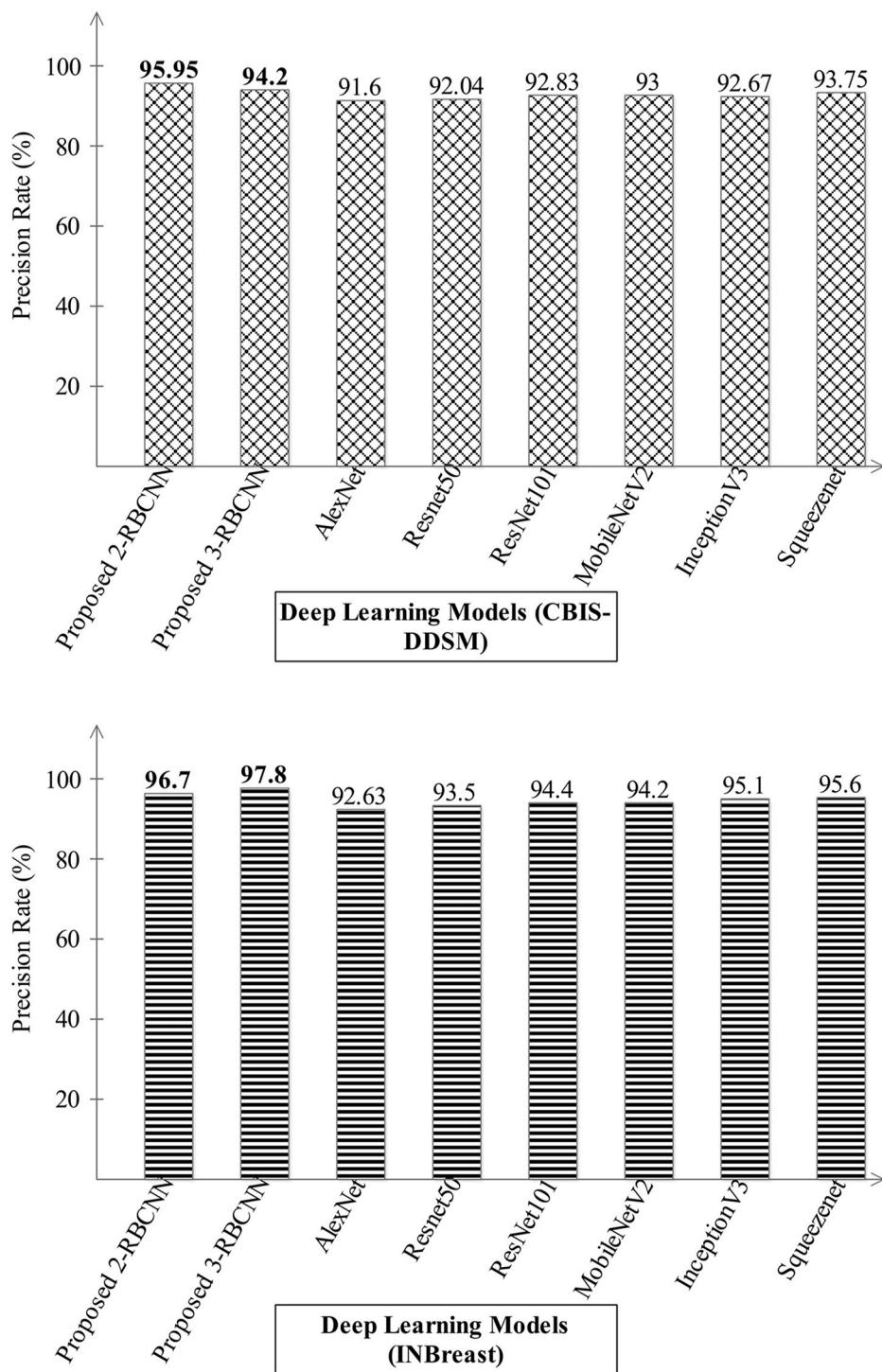


Fig. 18. Comparison among proposed designed deep models vs pre-trained deep networks.

Table 11
Accuracy values of selected classifiers for *t*-test using CBIS-DDSM dataset.

Classifiers	2-Residual Block	3-Residual Block	Proposed feature selection for 2-Residual Block	Proposed feature selection for 3-Residual Block	Proposed Fusion
WNN	95.9	94.2	95.2	92.6	97.3
BNN	95.6	93	95.3	92.4	97
Difference	0.3	1.2	0.1	0.2	0.3

Table 12
Accuracy values of selected classifiers for *t*-test using INbreast dataset.

Classifiers	2-Residual Block	3-Residual Block	Simulated Annealing 2-Residual Block	Simulated Annealing 3-Residual Block	Fusion
Medium NN	96.3	97.8	96.4	97.3	97.7
Tri-layered NN	96	97.7	95	97.6	97.2
Difference	0.3	0.1	1.4	0.3	0.5

Table 13

Comparison of the proposed method accuracy with SOTA techniques.

Reference	Year	Dataset	Accuracy (%)
Lou et al. (2022)	2022	INbreast	92.9
Melekooodappattu et al. (2022)	2022	MIAS, CBIS-DDSM	98, 97.5
Kumari and Jagadish (2022)	2022	CBIS-DDSM	94.3
Sannasi Chakravarthy et al. (2022)	2022	INbreast	96.64
Shukla and Behera (2024)	2024	MIAS dataset	94.87%
Boudouh and Bouakkaz (2024)	2024	INbreast	95.86%
Bouzar-Benlabiod et al. (2023)	2023	INbreast	91.34%
Proposed		INbreast	97.7
Proposed		CBIS-DDSM	97.3

Declaration of competing interest

All authors declared no conflict of interest in this work. All authors read and agreed to be submitting work in this well reputed journal.

Data availability

Data is added in the manuscript

Acknowledgments

This work is funded by Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2024R506), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. The authors also extend their appreciation to the Deanship of Research and Graduate Studies at King Khalid University for funding this work through Large Research Project under grant number RGP.2/379/45.

References

- Abdel-Basset, M., Ding, W., El-Shahat, D., 2021. A hybrid Harris Hawks optimization algorithm with simulated annealing for feature selection. *Artif. Intell. Rev.* 54, 593–637.
- Aslan, M.F., 2023. A hybrid end-to-end learning approach for breast cancer diagnosis: convolutional recurrent network. *Comput. Electr. Eng.* 105, 108562.
- Ayana, G., et al., 2023. Vision-Transformer-based transfer learning for mammogram classification. *Diagnostics* 13 (2), 178.
- Balaji, K., et al., 2016. Radiation therapy for breast cancer: literature review. *Med. Dosim.* 41 (3), 253–257.
- Beura, S., Majhi, B., Dash, R., 2015. Mammogram classification using two dimensional discrete wavelet transform and gray-level co-occurrence matrix for detection of breast cancer. *Neurocomputing* 154, 1–14.
- Bleicher, R.J., et al., 2016. Time to surgery and breast cancer survival in the United States. *JAMA Oncol.* 2 (3), 330–339.
- Boudouh, S.S., Bouakkaz, M., 2024. Breast cancer: new mammography dual-view classification approach based on pre-processing and transfer learning techniques. *Multimed. Tool. Appl.* 83 (8), 24315–24337.
- Bouzar-Benlabiod, L., et al., 2023. A novel breast cancer detection architecture based on a CNN-CBR system for mammogram classification. *Comput. Biol. Med.* 163, 107133.
- Cantone, M., et al., 2023. Convolutional networks and transformers for mammography classification: an experimental study. *Sensors* 23 (3), 1229.
- Chakravarthy, S.S., Bharanidharan, N., Rajaguru, H., 2023. Processing of digital mammogram images using optimized ELM with deep transfer learning for breast cancer diagnosis. *Multimed. Tool. Appl.* 1–25.
- Chen, Y., et al., 2019. Fine-tuning ResNet for breast cancer classification from mammography. In: Proceedings of the 2nd International Conference on Healthcare Science and Engineering 2nd. Springer.
- Chorowski, J., Wang, J., Zurada, J.M., 2014. Review and performance comparison of SVM-and ELM-based classifiers. *Neurocomputing* 128, 507–516.
- Couto, H.L., et al., 2024. Cost-effectiveness analysis of digital breast tomosynthesis added to synthetic mammography in breast cancer screening in Brazil. *PharmacoEconomics-Open* 1–14.
- Eisen, A., et al., 2024. Breast magnetic resonance imaging for preoperative evaluation of breast cancer: a systematic review and meta-analysis. *Can. Assoc. Radiol. J.* 75 (1), 118–135.
- Elkorany, A.S., Elsharkawy, Z.F., 2023. Efficient breast cancer mammograms diagnosis using three deep neural networks and term variance. *Sci. Rep.* 13 (1), 2663.
- Fang, W., et al., 2021. Deep residual learning in spiking neural networks. *Adv. Neural Inf. Process. Syst.* 34, 21056–21069.
- Garg, S., Singh, P., 2022. Transfer learning based lightweight ensemble model for imbalanced breast cancer classification. *IEEE ACM Trans. Comput. Biol. Bioinf* 20 (2), 1529–1539.
- Hariraj, V., et al., 2018. Fuzzy multi-layer SVM classification of breast cancer mammogram images. *Int. J. Mech. Eng. Technol.* 9 (8), 1281–1299.
- Hassan, M., et al., 2010. Chemotherapy for breast cancer. *Oncol. Rep.* 24 (5), 1121–1131.
- Heenaye-Mamode Khan, M., et al., 2021. Multi-class classification of breast cancer abnormalities using Deep Convolutional Neural Network (CNN). *PLoS One* 16 (8), e0256500.
- Houssein, E.H., Emam, M.M., Ali, A.A., 2022. An optimized deep learning architecture for breast cancer diagnosis based on improved marine predators algorithm. *Neural Comput. Appl.* 34 (20), 18015–18033.
- Iqbal, A., Sharif, M., 2023. BTS-ST: swin transformer network for segmentation and classification of multimodality breast cancer images. *Knowl. Base Syst.* 267, 110393.
- Jabeen, K., et al., 2023. BC2NetRF: breast cancer classification from mammogram images using enhanced deep learning features and equilibrium-Jaya controlled regula Falsi-based features selection. *Diagnostics* 13 (7), 1238.
- Jayandhi, G., Jasmine, J.L., Joans, S.M., 2022. Mammogram image classification system using deep learning for breast cancer diagnosis. In: AIP Conference Proceedings. AIP Publishing LLC.
- Karthiga, R., Narasimhan, K., Amirtharajan, R., 2022. Diagnosis of breast cancer for modern mammography using artificial intelligence. *Math. Comput. Simulat.* 202, 316–330.
- Kirkpatrick, S., Gelatt Jr., C.D., Vecchi, M.P., 1983. Optimization by simulated annealing. *Science* 220 (4598), 671–680.
- Kumari, L.K., Jagadish, B., 2022. Classification of mammograms using adaptive binary TLBO with ensemble classifier for early detection of breast cancer. *Int. J. Inf. Technol.* 1–12.
- Loizidou, K., Elia, R., Pitris, C., 2023. Computer-aided breast cancer detection and classification in mammography: a comprehensive review. *Comput. Biol. Med.*, 106554.
- Lopes, U., Valiati, J.F., 2017. Pre-trained convolutional neural networks as feature extractors for tuberculosis detection. *Comput. Biol. Med.* 89, 135–143.
- Lou, Q., et al., 2022. Mammogram classification based on a novel convolutional neural network with efficient channel attention. *Comput. Biol. Med.* 150, 106082.
- Melekooodappattu, J.G., et al., 2022. Breast cancer detection in mammogram: combining modified CNN and texture feature based approach. *J. Ambient Intell. Hum. Comput.* 1–10.
- Mo, Y., et al., 2023. Hover-trans: anatomy-aware hover-transformer for roi-free breast cancer diagnosis in ultrasound images. *IEEE Trans. Med. Imag.*
- Mohapatra, S., et al., 2022. Evaluation of deep learning models for detecting breast cancer using histopathological mammograms Images. *Sustain. Oper. Comput.* 3, 296–302.
- Nicholson, W.K., et al., 2024. Screening for breast cancer: US preventive services task force recommendation statement. *JAMA*.
- Pramanik, P., et al., 2023. Deep feature selection using local search embedded social skidriver optimization algorithm for breast cancer detection in mammograms. *Neural Comput. Appl.* 35 (7), 5479–5499.
- Ragab, D.A., et al., 2021. A framework for breast cancer classification using multi-DCCNNs. *Comput. Biol. Med.* 131, 104245.
- Raiyan, M.A.K., et al., 2024. Mammo-Light: a lightweight convolutional neural network for diagnosing breast cancer from mammography images. *Biomed. Signal Process Control* 94, 106279.
- Sahu, A., Das, P.K., Meher, S., 2024. An efficient deep learning scheme to detect breast cancer using mammogram and ultrasound breast images. *Biomed. Signal Process Control* 87, 105377.
- Saini, M., Susan, S., 2022. Vggin-net: deep transfer network for imbalanced breast cancer dataset. *IEEE ACM Trans. Comput. Biol. Bioinf* 20 (1), 752–762.
- Samee, N.A., et al., 2022. A hybrid deep transfer learning of CNN-based LR-PCA for breast lesion diagnosis via medical breast mammograms. *Sensors* 22 (13), 4938.
- Sannasi Chakravarthy, S., Bharanidharan, N., Rajaguru, H., 2022. Multi-deep CNN based experiments for early diagnosis of breast cancer. *IETE J. Res.* 1–16.
- Sathyam, A., Martis, D., Cohen, K., 2020. Mass and calcification detection from digital mammograms using unets. In: 2020 7th International Conference on Soft Computing & Machine Intelligence (ISCM). IEEE.
- Setiawan, A.S., Wesley, J., Purnama, Y., 2015. Mammogram classification using law's texture energy measure and neural networks. *Procedia Comput. Sci.* 59, 92–97.
- Shukla, P.K., Behera, A.R., 2024. A framework for breast cancer prediction and classification using deep learning. *Int. J. Comput. Vis. Robot* 14 (2), 154–169.
- Spolaor, N., et al., 2024. Fine-tuning pre-trained neural networks for medical image classification in small clinical datasets. *Multimed. Tool. Appl.* 83 (9), 27305–27329.
- Thwin, S.M., et al., 2024. Attention-based ensemble network for effective breast cancer classification over benchmarks. *Technologies* 12 (2), 16.
- Vijayan, D., Lavanya, R., 2023. Integration of global and local descriptors for mass characterization in mammograms. *Procedia Comput. Sci.* 218, 393–405.
- Xia, C., et al., 2022. Cancer statistics in China and United States, 2022: profiles, trends, and determinants. *Chin. Med. J.* 135 (5), 584–590.
- Xia, L., et al., 2023. Neural network model based on global and local features for multi-view mammogram classification. *Neurocomputing* 536, 21–29.
- Xie, L., et al., 2020. Neural networks model based on an automated multi-scale method for mammogram classification. *Knowl. Base Syst.* 208, 106465.