International Journal of Electrical and Computer Engineering (IJECE)
Vol. 15, No. 3, June 2025, pp. 2820~2833
ISSN: 2088-8708, DOI: 10.11591/ijece.v15i3.pp2820-2833

Genetic algorithm-adapted activation function optimization of deep learning framework for breast mass cancer classification in mammogram images

Noor Fadzilah Razali[1], Iza Sazanita Isa[1], Siti Noraini Sulaiman[1,2], Muhammad Khusairi Osman[1], Noor Khairiah A. Karim[3], Dayang Suhaida Awang Damit[1]
[1]Electrical Engineering Studies, Universiti Teknologi MARA, Cawangan Pulau Pinang, Kampus Permatang Pauh, Pulau Pinang, Malaysia
[2]Integrative Pharmacogenomics Institute (iPROMISE), UiTM Puncak Alam Campus, Bandar Puncak Alam, Selangor, Malaysia
[3]Department of Biomedical Imaging, Advanced Medical and Dental Institute, Universiti Sains Malaysia Bertam, Pulau Pinang, Malaysia

Article Info

Article history:

Received Aug 9, 2024
Revised Feb 16, 2025
Accepted Mar 4, 2025

Keywords:

Activation function
Convolutional neural network
Deep learning
Mammogram
Optimization

ABSTRACT

The convolutional neural network (CNN) has been explored for mammogram cancer classification to aid radiologists. CNNs require multiple convolution and non-linearity repetitions to learn data sparsity, but deeper networks often face the vanishing gradient effect, which hinders effective learning. The rectified linear unit (ReLU) activation function activates neurons only when the output exceeds zero, limiting activation and potentially lowering performance. This study proposes an adaptive ReLU based on a genetic algorithm (GA) to determine the optimal threshold for neuron activation, thus improving the restrictive nature of the original ReLU. We compared performances on the INbreast and IPPT-mammo mammogram datasets using ReLU and leakyReLU activation functions. Results show accuracy improvements from 95.0% to 97.01% for INbreast and 84.9% to 87.4% for IPPT-mammo with ReLU and from 93.03% to 99.0% for INbreast and 84.03% to 91.06% for IPPT-mammo with leakyReLU. Significant accuracy improvements were observed for breast cancer classification in mammograms, demonstrating its potential to aid radiologists with more robust and reliable diagnostic tools.

Corresponding Author:

Noor Fadzilah Razali
Electrical Engineering Studies, College of Engineering, Universiti Teknologi MARA Cawangan Pulau Pinang, Kampus Permatang Pauh
13500 Permatang Pauh, Pulau Pinang, Malaysia
Email: fadzilah708@uitm.edu.my

1. INTRODUCTION

Breast cancer is the most common cancer among women worldwide, with incidence rates increasing as women age [1]. Breast cancer cases have risen by 31%, making it the second leading cause of cancer-related deaths after lung cancer [1]. Early detection is possible through mammogram screenings, but evaluating mammogram images typically involves manual tasks such as adjusting contrast, detecting masses, and making initial diagnoses. Automating these processes with computer-aided detection (CAD) systems can improve diagnosis accuracy and patient care.

The advancement of deep learning through convolutional neural network (CNN) has been used to develop the mammogram CAD system. Its architecture involves convolution, non-linearity, and pooling layers to improve the overall CAD performance [2], [3]. However, CNNs can be affected by the vanishing gradient

problem [4], where weights stop updating, failing to capture complex features. Multiple activation functions in CNN architecture obtain the sparsity of the learned weights on each layer to learn a complex feature representation.

Figure 1 shows multiple activation functions available to be used within the CNN architecture. It introduced non-linearity and helped mitigate vanishing gradient problem by deactivating neurons in the negative region, resulting in sparsity, allowing complex feature learning, making them crucial in CNNs [5], [6]. Rectified linear units (ReLU) as depicted in Figure 1(a) are the most commonly used activation functions. ReLU is favored for its simplicity, low computational cost, and efficiency. Figure 1(b) shows variants of the ReLU that focus on allowing these outliers to be activated on the negative regions, such as the leakyReLU, parametric ReLU (PreLU), Gaussian linear units (GELU), and exponential linear units (ELU).

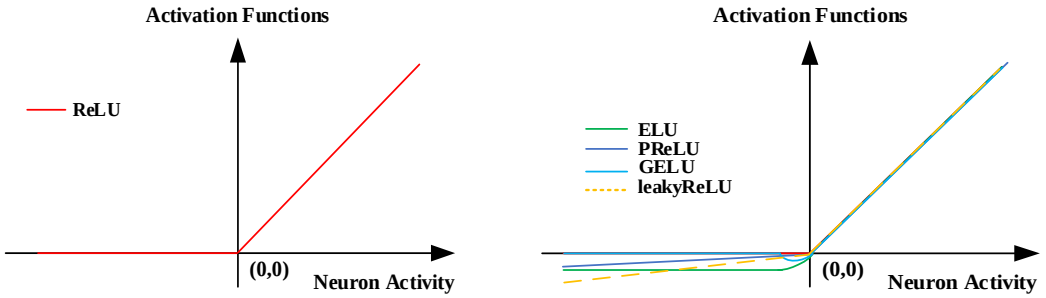


Figure 1. Graphs of (a) ReLU and (b) its derivatives' activation function variants used in CNN

However, due to this slight allowance in the negative region compared to the original ReLU, they are computationally more expensive and require additional memory to store the trained parameters. Choosing the right non-linearity parameter coefficient in ELU and leakyReLU requires prior knowledge about the data and careful tuning to select the correct value. Meanwhile, the newer GELU requires the user to determine whether to apply the *Tanh* approximation method for the underlying error function, which also requires prior knowledge of the dataset behavior. Therefore, continuous research on activation functions remained active, while the more established and the ReLU architecture continued to be used in standard CNN architecture.

Genetic algorithm (GA) [7] is a versatile pioneering metaheuristic optimization method, allowing practitioners to tailor it to specific tasks. Its applications involve image processing modules and computer vision through natural selection and evolution through the survival of the fittest concept [8], [9]. GA is mostly applied for feature selection [9], [10] and hyperparameter optimization [8], [11] in breast-related studies and in biomedical applications, including brain MRI [12] and CNN training hyperparameter selection [9]. It explores the solution space, favors better solutions, and the mutations contribute modest random changes to maintain genetic diversity, while crossover combines genetic information from two parents to produce optimal offspring through successive generations marked by improvements in solution quality [7], [13]. GA's versatility and ability to identify nearly optimal solutions make it a popular optimization technique.

CNNs have significantly advanced CAD systems, but they still face challenges, particularly the vanishing gradient problem and the "dying ReLU" issue. In "dying ReLU," some outlier neurons become inactive which limits the model's ability to learn complex features [14]. Existing solutions, like variants of ReLU, address some of these issues but come with drawbacks, including higher computational costs and the need for extensive parameter tuning. These challenges reduce their practicality in clinical applications.

This study offers a solution by introducing a novel adaptation of the ReLU activation function in pre-trained CNNs, optimized using GA. The adaptive function adjusts dynamically to allow limited negative or positive outputs based on an automatically determined threshold. This approach preserves the benefits of non-linearity while addressing vanishing gradients. Unlike traditional methods that require manual tuning, GA automatically fine-tunes activation parameters, making the system more efficient and better suited for handling complex data. Building on prior research that applied GA to tasks like feature selection and hyperparameter optimization in mammogram analysis, this study takes a pioneering step by using GA to optimize the non-linearity layer itself. This approach directly tackles the "dying ReLU" problem and improves the model's robustness. This study contributes to the following: i) Introduction of metaheuristic GA optimization within established CNN pre-trained layers; ii) Method of adapted activation function optimization on mammogram datasets; and iii) Adapting different activation functions using the proposed method for improved performance.

2. RELATED WORKS

In recent years, CNN-based studies on mammogram-related images have been performed via computer vision in various stages, such as in image pre-processing [15], cancer detection and segmentation [16], [17], and cancer classification [18]–[20]. Since the relevant features were automatically extracted, the convoluted combination of features extracted from the training images makes it difficult to interpret what happened during the decision-making process. This relates to inducing the non-linear functions within the architecture in the form of repeated layers with convolution, which dissociates the original straightforward convolution and filtering process in a CNN.

In the past, researchers have proposed several innovative approaches related to activation functions to enhance the performance of CNNs in network analysis. An adaptive activation function, namely layer-wise and neuron-wise, was introduced with optimization to accelerate model convergence as the effect of slope recovery during training [21]. This author also proposed a Rowdy activation function to be used with multiple types of trainable parameters achieved by injecting sinusoidal fluctuations [22] to lessen the saturation region. Similarly, another study by [23] proposed a modified ReLU-based activation function by introducing a periodic sinusoidal function on the positive region for activating the neuron on a voice order classification, suggesting improved accuracy in a convolutional neural network-long short-term memory (CNN-LSTM) model. In a study for the detection of distributed denial of service attacks on recurrent neural networks, modification of the hyperbolic tangent activation function known as Tanh2 was introduced. The function has a range of values between 0 and 1 in the bell-shaped curve function, which removed negative values that caused the vanishing gradient problem in deep learning [24]. Meanwhile, an adaptive network enhancement using a two-layer ReLU was introduced for the best initialization of its non-linear optimization process on the best least-square to the best approximation of a target function [6]. The method determined the number of neurons for the activation to be added until the best approximation leads to the best model accuracy. Additionally, an adaptive ReLU method was proposed by [25] by optimizing the position of when the neuron in the negative region will be activated by retaining the shape of the ReLU on the early layers of CNN, improving the accuracy of three different image databases. Furthermore, a modified ReLU was introduced by [26], introducing a coefficient to an exponential value on the negative-region slope, creating a small-scale negative slope similar to the leakyReLU with the exception of the negative slope.

Although limited, some studies have focused on tailoring activation functions specifically for mammogram datasets. In a study using a cogent activation function based on the modified Tanh function, introduced to be applied to mammogram images, with the best accuracy at 99% [27]. Next, replacements of activation functions in CNN models, such as placing the leaky ReLU [28], and PReLU [29] were done to deliver promising performance in breast cancer classification compared to their base networks. These functions mitigated the vanishing gradient problem caused by dying ReLU often observed in ReLU-based networks, improving classification accuracy through various methods. These studies also revealed that non-linearity in a network plays a vital role in defining features representing each class and improving the model's generalization through meaningful data sparsity. However, besides requiring in-depth knowledge of developing the mathematical model of the activation functions, these methods are applied to a customized network model to the specific tasks and are limited particularly due to the fixed functions. This makes it more complicated to adapt and generalize to other datasets and tasks, affecting their generalizability. Table 1 summarizes past studies that explored the modification of activation functions within neural networks on various applications.

Table 1. Past related studies on modification or introduction of novel activation functions in neural networks

Study	Activation function novelty	Application/datasets
[6]	Two-layer ReLU architecture	Self-adjoint second-order elliptic partial differential equations
[21]	Layer-wise and neuron-wise activation function	Semeion, CIFAR-10, CIFAR-100 and SVHN
[22]	Rowdy activation function, sinusoidal fluctuations	Semeion, CIFAR-10, CIFAR-100 and SVHN
[23]	Sinosoidal injections, sinusoidal fluctuations	Voice order classification
[24]	Modified *Tanh* activation function	Distributed denial of service attack detection
[25]	Adaptive ReLU on [-1 0] and [0 -1] regions	Cifar10, Cifar100, Tiny ImageNet and Text databases
[26]	Modified ReLU, negative slope on the negative region	Multispectral image for labelled EuroSAT satellite images
[27]	Modified *Tanh* function for cogent activation function	MIAS dataset, mammogram classification

Meanwhile, bio-inspired metaheuristic optimization methods have been applied to mammogram-based studies, demonstrating their ability to enhance the overall performance of systems when combined with CNN architectures. These algorithms offer distinct strengths and have been integrated into various stages of mammogram analysis pipelines to address specific challenges, improve system performance, and contribute to more accurate and reliable breast cancer diagnosis. For instance, in a study focused on classifying breast

abnormalities, a GA-CNN model achieved an accuracy of 98.5% by optimally extracting features from segmented regions [30]. Another feature selection optimization method was applied using multi-objective improved ant colony optimization (ACO) based on correlation coefficient, which showed improved classification rates when applied to machine learning classifiers [10]. Another feature selection optimization utilizing the chaotic-crow-search optimization algorithm was also employed to optimize across multiple mammogram databases, resulting in enhanced overall accuracy performance [31]. Weighted average gravitational search algorithms were used as feature selection to determine the best textural mammogram feature producing good results [19]. In another study, various metaheuristic optimization methods, including GA, particle swarm optimization (PSO) and artificial bee colony (ABC) algorithms, were compared against a novel Ebola optimization search algorithm (EOSA) for CNN hyperparameter optimization [11], which demonstrated improved performance by implementing the disease's susceptible-infection-recovery model. Optimization of the mammogram pre-processing segmentation utilizing the adaptive Kernel-Based Fuzzy Cuckoo Search Optimization Clustering is also introduced, with fuzzy-c-means clustering problems, improving the Dice similarity index on several mammogram image tests [32]. Another study by [33] optimized their classifier by using firefly binary grey optimization and moth flame lion optimization in an ensemble model to classify the INbreast dataset. In summary, the optimization processes in these studies were applied after the extracted features, and they did not directly impact the training path of the network, which led to the tedious process of determining which types of optimization algorithms suited the given task. Moreover, most of the optimization algorithms are newer and less established, requiring extensive knowledge of the method to be applied to other tasks as well. However, this shows that combining optimization techniques and CNN architectures has demonstrated significant potential in improving the effectiveness of mammogram-based CAD for diagnosing breast cancer. Substantial enhancements in accuracy and reliability are achieved by utilizing the advantages of optimizing different phases of the architecture. Continued progress in this field, particularly when applied within the CNN architecture, has the potential to revolutionize the identification and treatment of breast cancer, resulting in improved patient outcomes.

In conclusion, developments of CNN-based analysis of mammogram-related images have highlighted the importance of using different non-linear activation functions in improving the network's capacity to extract complex information. However, novel non-linear models such as those demonstrated by [21], [22] are implemented theoretically and have not been conducted on critical real-world datasets such as medical images. Otherwise, it documented the need and its importance well and provided base knowledge to the mathematical development to improve the network's performance.

3. METHOD

This study proposed a method to improve the final feature extraction path by introducing a GA-based optimization on the last ReLU layer in the CNN feature extraction pathway. The model improvement is demonstrated by comparing the model's performance before and after optimization and validated on two mammogram datasets to prove the hypothesis of this study. This study uses mammogram images as input to an optimized CNN pathway for best feature extraction based on metrics performance. First, the images were pre-processed to obtain the mass location, and the customized optimized CNN using GA was applied during feature extraction. Finally, the mass was classified as benign or malignant. All model development is done in a workstation equipped with an Intel® Core™ i7-12700 2.1 GHz, GPU graphic card of 12GB and RAM 32GB, through MATLAB, Natick, Massachusetts: The Mathworks Inc. platform. The following subsections discuss the details of the proposed method.

3.1. Dataset preparation

Two mammogram datasets are employed in this study to validate the proposed method. First, an established, publicly available digitized mammogram of INbreast widely used in multiple studies [18], [31], [33]–[36] is obtained from its source [37]. The images were collected from Universidade de Porto, Portugal, from Centro Hospitalar de S. João Breast Center, containing 115 cases of breast lesions. Only the mass cases are selected in this study, resulting in 112 images with both benign and malignant masses. The INbreast dataset is comprised of 26.8% denser breasts [37]. Additionally, this study's second mammogram dataset, termed the IPPT-mammo, was retrospectively collected at the Institut Pergigian dan Perubatan Termaju (IPPT), Universiti Sains Malaysia Bertam, Malaysia, according to relevant laws through the approval of the institution's human ethics research committee for this research study (File: JEPeM/21090624 year: 2021). It contained 200 breast mass lesion cases with annotated abnormalities by expert radiologists and comprised 60% dense breast density. Both datasets include ground-truth information such as the masses' Range of Interest (ROI), breast density level, and types of abnormality.

The images contained within both datasets were applied with normalization and other morphological processing techniques. This is done to extract only the breast region to eliminate unnecessary artefacts [20].

An image enhancement process based on histogram equalization from our previous study [38] is applied, and the process is depicted in Figure 2.

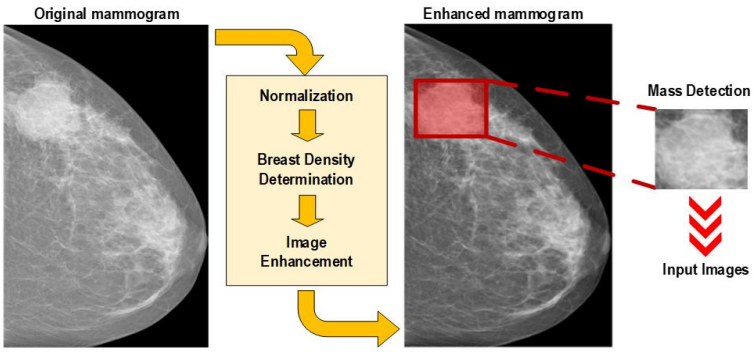


Figure 2. Pre-processing of the input images

The method is performed by adjusting the contrast based on the breast density level that is subjected to the object detection method. To lessen the effect of overfitting due to low trainable images by using this dataset for the deep learning approach, geometrical augmentation, such as rotation ranges of -20 to 350 degrees, flipping, and random scaling of 1.0–1.3 magnitude, are applied to the images to increase the training number and improve model convergence. The dataset is divided into training, validation, and testing datasets at a ratio of 65:15:20 to test the proposed method.

3.2. Optimized CNN feature extraction

Next, the mass input images are fed into the CNN feature extraction network that has been modified with the optimization process proposed in this study, as depicted in Figure 3. Figure 3(a) illustrates the proposed overall design for the CNN training pathway for feature extraction and classification. Meanwhile, Figure 3(b) shows the modification made on the non-linear activation layer for the optimization.

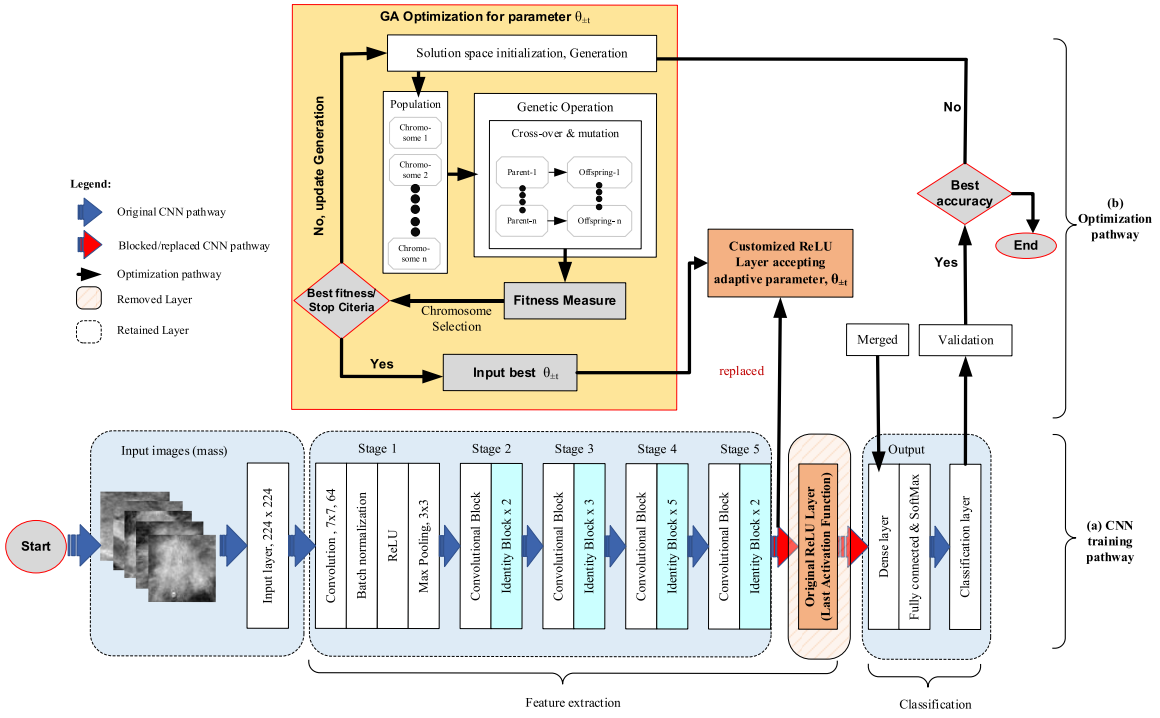


Figure 3. The architecture of the proposed methodology. (a) CNN training and (b) GA optimization

3.2.1. CNN training pathway

In the feature extraction phase, pooling layers and activation functions are utilized to implement spatial down sampling and introduce non-linearity to the model. The estimation of the optimal decimal probability that a given class belonged to a member of a class occurs on the SoftMax function layer during the classification phase. To test the proposed method, this study employed two established pre-trained CNN networks of ResNet50 [4] and ShuffleNet [39]. Both networks were trained on a massive ImageNet database to extract and learn higher and semantic features, including patterns in images where the learned weights are reused as part of the transfer learning process. The two CNNs selected in this study represent lightweight CNN for ShuffleNet and heavyweight CNN for ResNet50 with 1.2 million and 25.5 million training parameters, respectively. Both architectures also differ in terms of specific requirements of specific tasks, such as computational resources and model size constraints and uniquely represent different types of pre-trained CNN architectures established in the neural network field. To utilize the pre-trained networks with the proposed method, both CNN models' fully connected layers are remodelled to output two classes, benign and malignant, and the training parameters are set the same as listed in Table 2. Note that the second IPPT-mammo dataset required a higher maximum epoch and mini-batch size number for training as it contained larger image samples than the INbreast dataset.

Table 2. Hyperparameter settings for CNN training for INbreast and IPPT-mammo datasets

Hyperparameter setting	Set value/dataset	
	INbreast	IPPT-mammo
Maximum epoch	50	80
Mini-batch size	80	128
Learning rate	0.001	
Solver	Adaptive moment estimation (Adam)	
L2 Regularization	0.0001	

3.2.2. Genetic algorithm optimization pathway

In this study, the proposed optimization pathway, as in Figure 3, is conducted on the last activation function during the feature extraction stage. Using the last activation function to be modified allows the default extraction process on the upper level of the network to shape the final decision boundary output for the specific task. The last feature extraction stage's ReLU activation function, layers 173 and 168 for ResNet50 and ShuffleNet, was replaced with customized ReLU and leakyReLU activation functions.

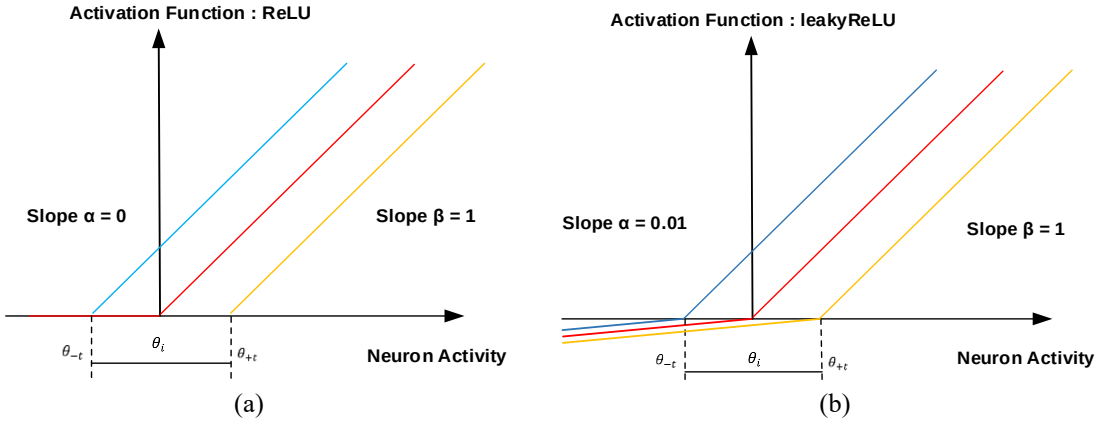
The customized activation functions are built to receive an adaptable parameter value set as the threshold, $\theta_{\pm t}$, that controls the position of the activation graph to be activated by shifting the graph from its default position of (0,0) to the optimized ($\theta_{\pm t}$,0) on the *x*-axis. To optimize the threshold, GA, a metaheuristic optimization approach based on natural genetic selection based on the survival of the fittest concept to find the best solution to a problem, was employed [13]. The best $\theta_{\pm t}$ is determined through genetic operators of crossover and mutation based on the best combination of the initialized parent chromosome as parameters selected at the start of the first population training. The search space is set for the threshold, $\theta_{\pm t}$ with upper and lower bound set to [-1 1]. The GA is evaluated with the validation dataset during the CNN training as the optimal solution of the fitness function. Given the probability, *P*, of drawing individual, *i*, in the initial population where it is defined based on the fitness function *F(x,y)* where *x* and *y* are parameters defining the *i* characteristics or genotype when GA is looking for optimal value in a search space within the total number, N, of *i* in a population, defined as in (1):

$$P_i = \left| \frac{F(x,y)_i}{\sum_i^N F(x,y)_i} \right| \quad (1)$$

The CNN training undergoes forward and backward propagation, as the best $\theta_{\pm t}$ is updated through the customized activation function layer on each mini-batch training. In this study, the setting for the GA is conditioned to allow elite chromosome, mutation, and crossover intermediate in the selection process to allow a diverse population and provide a balance between exploration and exploitation to find the optimal solution. The number of individuals, *i*, in the population, were elite and preserved in each generation is set to the top 2 individuals with the best fitness score.

The proposed modification is made on two established activation functions of ReLU and leakyReLU. The probable threshold shifts in the negative or positive neuron activity range in the *x*-axis are expected to help activate the dead neuron that is possibly dead in the default settings before the proposed optimization. This is especially true at the end of the extraction stage to alleviate the dead neuron problems caused by the upper-level ReLU, where a direct evaluation can be made through the fitness function for optimized performance.

Figure 4 shows the possible location of $\theta_{\pm t}$ on *x*-axis and its subsequent linear equation graph for the proposed method on the original red graphs for Figure 4(a) ReLU and Figure 4(b) leakyReLU, shifted to blue and yellow graphs respectively, according to the optimized $\theta_{\pm t}$ value, with positive and negative slopes α and β retained at their default values.



(a) (b)

Figure 4. Activation functions graphs of. (a) ReLU and (b) leakyReLU. The original ReLU and leakyReLU (red), possible GAa-ReLU and GAa-LeakyReLU on the negative threshold (blue), and positive threshold (orange), where θ_i is the original threshold position

a. Genetic algorithm-adapted rectifier linear unit (GAa-ReLU)

Figure 4(a) shows the ReLU graphs and modifications possible due to the proposed method. Note that the original threshold was fixed at $\theta_i = 0$ and the slope β =1 values for neuron activation $\geq \theta_i$ (positive region) retained for all graphs. The original ReLU activation function $f(x)_{ReLU}$ is as in (2).

$$f(x)_{ReLU} = max(0\ ,\beta x);\ = \begin{cases} 0, if\ x < \theta_i \\ \beta x, if\ x \geq \theta_i \end{cases},at\ \theta_i = (0,0), \beta = 1 \quad (2)$$

In the case of the original ReLU (red) of Figure 4(a), the threshold of θ_i is always = (0,0) position, in which the neuron will be activated when the input $x \geq \theta_i$. The simplicity of ReLU relies on its linear equation, which permanently deactivates the neuron when the input is less than 0. This improves the computation ability to train multiple layers of CNN by introducing sparsity within the weight update [4]. This study proposed the GAa-ReLU activation function, replacing the final ReLU layer in the original CNN architecture, to allow an unrestrictive, albeit controlled threshold for the ReLU that enables weight updates from the neuron activation on the input *x*. The GAa-ReLU function $g(x)_{GAa-ReLU}$ is as in (3), with its possible graphs in Figure 4(b), where the intersection *C* is obtained automatically by retaining the positive region slope β as = 1 and negative region slope α = 0, while the graph is shifted by $\theta_{\pm t}$ value on the *x*-axis optimized through the GA search during the training loop.

$$g(x)_{GAa-ReLU} = max(0\ ,\beta x + \theta_{\pm t});\ = \begin{cases} \theta_{\pm t},\ if\ x < \theta_{\pm t} \\ \beta x + C,\ if\ x \geq \theta_{\pm t} \end{cases},\text{at } \theta_i = (\theta_{\pm t}, 0), \beta = 1 \quad (3)$$

b. Genetic algorithm-adapted leaky rectifier linear unit (GAa-leakyReLU)

Additionally, the proposed method is tested on another established activation function of leakyReLU as depicted in Figure 4(b), where the original leakyReLU is developed to improve the performance of the ReLU [40]. The leakyReLU allows a non-zero gradient to activate specific neurons and improve the dead-neuron problem in the ReLU by introducing a fixed low-value coefficient, α, as the negative slope. The α allows a small skew slope of the negative region, thus 'leaked' activated neuron preventing dead neuron problem in ReLU. The availability of these activated neurons allows the preservation of important feature learning and improves performance. In this work, we take further advantage of leakyReLU and implement our proposed optimization method, GAa-leakyReLU. In this study, the positive region slope, α, is set to default 0.01, and (4) indicates the original leakyReLU equation $f(x)_{leakyReLU}$, whereas (5) indicates the GAa-leakyReLU function $g(x)_{leakyGAa-ReLU}$ when applied to the proposed method.

$$f(x)_{leakyReLU} = max(\alpha x, \beta x); = \begin{cases} \alpha x, if\ x < \theta_i \\ \beta x + C, if\ x \geq \theta_i \end{cases}, at\ \theta_i = (0,0), \alpha = 0.01, \beta = 1 \quad (4)$$

$$g(x)_{leakyGAa-ReLU} = max(\alpha x, \beta x); = \begin{cases} \alpha x, if\ x < \theta_{\pm t} \\ \beta x \pm C, if\ x \geq \theta_{\pm t} \end{cases}, \text{at } \theta_i = (\theta_{\pm t}, 0), \alpha = 0.01, \beta = 1 \quad (5)$$

The CNN-extracted features are then classified to get the best performance scores on the testing dataset. In this study, the customized activation function layer is only applied to the last layer of feature extraction. This step is done to allow strict feature extraction and sparsity on the upper layer during semantic feature extraction using the original ReLU and only allow adaptiveness on the final extraction stage.

3.3. Performance metrics

The evaluation of the proposed model is based on the confusion matrix (CM), which includes four components: True positive (TP), false positive (FP), true negative (TN), and false negative (FN). Each component captures specific aspects of the model's performance: TP identifies correctly predicted positive malignant cases, FP counts incorrectly predicted positive malignant as benign cases, TN represents correctly predicted negative cases, and FN tallies missed positive cases. Using these foundational components, the evaluation proceeds with the calculation of metrics through (6) to (9), enabling a detailed assessment of the model's strengths and weaknesses.

$$\text{Accuracy, Acc} = \frac{\text{TP+TN}}{\text{TP+FN+TN+FP}}. \quad (6)$$

$$\text{Sensitivity / Recall} = \frac{\text{TP}}{\text{TP+FN}}. \quad (7)$$

$$\text{Specificity} = \frac{\text{TN}}{\text{(TP+FN)}}. \quad (8)$$

$$\text{F1-score} = 2\text{x}\frac{\text{Precision X Recall}}{\text{Precision+Recall}}. \quad (9)$$

4. RESULT AND DISCUSSION

4.1. Different pre-trained models

This section presents the results of the proposed method evaluated on the INbreast dataset. Two established pre-trained networks, ResNet50 and ShuffleNet, were used for the evaluation. The comparison highlights the performance of models trained under two conditions: the default ReLU activation of the pre-trained networks and the modified versions trained with GAa-ReLU. The results are summarized in Table 3.

Based on the results in Table 3, ShuffleNet outperformed ResNet50 across all tested metrics for INbreast cancer classification. When trained using the proposed GAa-ReLU activation function, both models showed notable improvements over their default ReLU-based counterparts. ShuffleNet achieved the best performance with an accuracy of 99.00%, an F1-score of 0.9851, sensitivity of 1.000, and specificity of 0.9852. ResNet50 also showed improvement, achieving 97.01% accuracy, an F1-score of 0.9559, and specificity of 0.9559, while maintaining a sensitivity of 0.9848. These results demonstrate that the proposed GAa-ReLU method enhances performance effectively across different pre-trained networks, achieving results on par with or better than default configurations, particularly for ShuffleNet.

Table 3. Classification of breast mass for INbreast dataset using the original ReLU vs. proposed GAa-ReLU on different pre-trained networks

Metric	ShuffleNet		ResNet50	
	ReLU (Default)	GAa-ReLU	ReLU (Default)	GAa-ReLU
Accuracy (%)	97.51	99.00	95.00	97.01
Sensitivity/Recall (TPR)	0.9687	1.0000	0.9848	0.9848
Specificity (TFR)	0.9778	0.9852	0.9333	0.9559
F1-score	0.9624	0.9851	0.9286	0.9559

4.2. Different activation functions

In this experiment, testing and comparison to evaluate the proposed method on another established activation function, leakyReLU. The experiment is tested based on metrics of accuracy, sensitivity, specificity, and F1-scores. The ResNet50 result from Table 3 is then tabulated again in Table 4, where its performance is compared to the proposed method applied to the second activation function of leakyReLU.

Table 4. Classification of breast mass for INbreast dataset using the original ReLU and leakyReLU vs. the proposed GAa-ReLU and GAa-leakyReLU

Metric/Dataset	INbreast			
	ReLU (Default)	*GAa-ReLU*	*leakyReLU*	*GAa-leakyReLU*
Accuracy (%)	95.00	97.01	93.03	99.00
Sensitivity / Recall (TPR)	0.9848	0.9848	0.9697	0.9697
Specificity (TNR)	0.9333	0.9559	0.9111	1.0000
F1-score	0.9286	0.9559	0.9014	0.9846

From Table 4, the substitution of ReLU to leakyReLU for the INbreast dataset decreased its overall performance. Other metrics, such as accuracy, dropped to 93.03%, and F1-score, sensitivity, and specificity values dropped as well. This suggests that simply replacing the ReLU with leakyReLU does not necessarily improve its dying ReLU problem without proper knowledge of the value of positive slope or coefficient α. However, when the proposed method of GAa-leakyReLu is implemented, it can enhance the overall performance for all metrics, surpassing the performance of the original ReLU and GAa-ReLU on a fixed coefficient α. A 1.0000 performance is demonstrated on the specificity value. Other metrics of the accuracy of 99.00%, and an F1-score of 0.9846 is also demonstrated. From this result, it can be concluded that even though when the ReLU is initially substituted with leakyReLU, causing the model to perform lower even with the introduction of α, it is proven that is not the case for this dataset. Simply introducing the coefficient α is unable to allow the outliers within the tested dataset to cope well and cause overcompensation for not being able to be grouped perfectly with the fixed value of α during activation of the testing features on the trained model. However, when optimized using the GAa-leakyReLU, the original leakyReLU, having a network landscape of a typically non-convex, with its own optimization method used, such as Adam, may get stuck in determining its local minima. GA, as employed in this study, with their ability to explore diverse solutions, helps escape this local minimum and find global or near-global optima, as demonstrated in both adapted activation function results.

4.3. Second mammogram dataset

In this experiment, testing and comparison are conducted using the second dataset, IPPT-mammo to further evaluate the proposed method's effectiveness. This experiment also examines the method's performance when applied to another established activation function, leakyReLU. Table 5 shows the result of classification using the IPPT-mammo dataset using the original ReLU vs. proposed GAa-ReLU and leakyReLU vs GAa-leakyReLU.

Table 5. Classification of benign and malignant breast mass for IPPT-mammo dataset using the original ReLU vs. proposed GAa-ReLU and leakyReLU vs GAa-leakyReLU

Metric/Dataset	IPPT-mammo			
	ReLU (Default)	*GAa-ReLU*	*leakyReLU*	*GAa-leakyReLU*
Accuracy (%)	84.90	87.40	84.03	91.60
Sensitivity / Recall (TPR)	0.8986	0.8116	0.9275	0.9130
Specificity (TNR)	0.7800	0.9600	0.7200	0.9200
F1-score	0.8732	0.8819	0.8707	0.9265

The metrics suggest overall lower performance with the established INbreast dataset due to a higher number of denser breasts in the IPPT-mammo dataset. Generally, when applied with the proposed method of GAa-ReLU, the performance pattern exhibits improvements and consistently outperforms the ReLU (default) activation on all evaluation metrics, aside from the sensitivity. The improvement in accuracy is observed at 84.90% to 87.40%. Significant improvement is observed, such as the F1-score relates to the precision and sensitivity of a model. It indicates a better representation of a system with an unbalanced dataset, with a slight increase to 0.8819 from 0.8732 in the IPPT-mammo dataset. Comparing the result to Table 4, a balance of sensitivity and specificity for the proposed model demonstrates the ability of the model to correctly classify the false positives and false negatives of the testing data on the testing INbreast dataset. The specificity was significantly improved with the GAa-ReLU activation, reaching 0.9600 compared to its base version of 0.7800. However, this also simultaneously affected the TPR, as it slightly decreased to 0.8116 for GAa-ReLU and a similar trend is seen on the test using GAa-leakyReLU. Overall, the improvements in performance metrics on the IPPT-mammo dataset for both base activation functions show the proposed method can be implemented on different datasets without parameter changes, suggesting the generalizability of the proposed method. To evaluate the proposed method with several state-of-the-art models from previous studies, Table 6 provides

the performance comparison between the studies using the same public dataset from INbreast and the proposed method of the CNN model using the proposed activation function optimization method.

Table 6. Comparison of evaluation performance for classification studies using the INbreast dataset

Studies	Method	Accuracy	Specificity	Sensitivity	F1-score
[18]	CNN, wavelet scattering	93.50	N/A	N/A	0.9337
[31]	CNN, chaotic map FS optimization	98.46	0.9868	0.9821	0.9839
[34]	CNN, transferable texture	96.82	0.9768	0.9599	N/A
[35]	CNN	93.00	0.9386	0.9383	0.9303
[36]	CNN, bidirectional long short-term memory neural network (BiLSTM)	92.26	0.9295	0.8621	0.8853
This Study	CNN, GA optimized ReLU (ResNet50)	97.01	0.9559	0.9848	0.9559
	CNN, GA optimized ReLU (ShuffleNet)	**99.00**	0.9852	**1.0000**	0.9851
	CNN, GA optimized leakyReLU (ResNet50)	**99.00**	**1.0000**	0.9697	**0.9846**

The overall performance revealed that the proposed method of optimizing the activation function performed best on all the evaluation metrics across three models experimented in this study. Maqsood *et al.* [34] also exhibits good accuracy (96.82%) when using a transferable texture CNN and good specificity (0.9768). However, they did not provide other evaluations to make a sound comparison. The study by [18] has an accuracy of 93.50% and a lower F1-score, possibly due to the non-augmented images used. In addition to the proposed method, Chakravarthy *et al.* [31] performed comparatively well, with 98.46% accuracy and a balanced overall specificity and sensitivity of 0.9868 and 0.9821, respectively. This is due to optimization implementation during its feature selection process, demonstrating the importance of introducing a significant method to improve the final classification performance as proposed in this study. However, the best-performed model is attained by the proposed method on GAa-leakyReLU for ResNet50 and GAa-ReLU for ShuffleNet, with both achieving the best accuracy at 99.00%. As demonstrated from the result, modifying the threshold of the activation functions, such as ReLU and leakyReLU, which originally starts at (0,0), to a value corresponding to the degree of adaptability of the nonlinear function in the final feature extraction phase has substantial significance. The flexibility of the activation function involving negative neuron activation has numerous significant benefits in modelling intricate data and optimizing neural networks. Hence, incorporating an adjustable threshold with the help of an optimization algorithm, as demonstrated in this study, allows flexibility that enables neurons to sustain partial activation even when presented with inputs that fall below the threshold. This approach indirectly addresses the issue of the dying ReLU problem, which results in a higher level of neuron involvement in the learning process and consequently enhances the model's capacity for expression.

5. CONCLUSION

This study introduces a method for optimising classification during feature extraction, focusing on nonlinear activation at the final stage. The GA optimization is used to adjust a CNN's threshold of ReLU and leakyReLU. The results show the best model performed using the optimized leakyReLU with 99% accuracy, having the best specificity at 1.0000 for the INbreast dataset, and 91.60% accuracy for the newly collected IPPT-mammo dataset. This adjustment improves model performance, by shifting the neuron activation region affecting the non-linearity. Tests on public and private mammogram datasets with different CNN models using the modified ReLU and leakyReLU indicate that the method is generalizable across various architectures and datasets. In summary, modifying activation function thresholds helps tailor neural networks to specific tasks, enhancing performance and generalization, addressing issues like the dying ReLU, refining nonlinearity, and increasing robustness against noisy data. A noted limitation is the longer GA training time required for larger populations, though final training is less computationally intensive. Future work includes applying this method to other ReLU-based functions and exploring new metaheuristic optimization approaches to reduce computational load.

ACKNOWLEDGEMENTS

The authors would like to express their gratitude to members of the advanced control system and computing research group (ACSCRG), advanced rehabilitation engineering in diagnostic and monitoring research group (AREDiM), integrative pharmacogenomics institute (iPROMISE), and Centre for Electrical Engineering Studies, Universiti Teknologi MARA, Cawangan Pulau Pinang for their assistance and guidance during the fieldwork. The authors are grateful to Universiti Teknologi MARA, Cawangan Pulau Pinang for their immense administrative support. Special thanks to the Imaging Department, Advanced Medical and

Dental Institute, Universiti Sains Malaysia, Kepala Batas, Pulau Pinang for the professional consultation and expert guidance given.

FUNDING INFORMATION

This research work is financially supported by the Research Management Centre (RMC) through Pembiayaan Yuran Penerbitan Artikel (PYPA) of Universiti Teknologi MARA.

AUTHOR CONTRIBUTIONS STATEMENT

The contributions of each author to this study are detailed below, in accordance with the Contributor Roles Taxonomy (CRediT). The following table describes each author's specific role in the research and writing process.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Noor Fadzilah Razali	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Iza Sazanita Isa	✓	✓	✓	✓		✓		✓	✓	✓	✓	✓		✓
Siti Noraini Sulaiman	✓					✓		✓			✓	✓	✓	
Noor Khairiah A. Karim		✓	✓	✓	✓		✓			✓	✓	✓		
Muhammad Khusairi Osman	✓	✓			✓		✓	✓		✓		✓		
Dayang Suhaida Awang Damit		✓		✓		✓	✓			✓				

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

ETHICAL APPROVAL

The dataset used in this study consists of mammogram images identified by the ethical approval code from Jawatankuasa Etika Penyelidikan Manusia (JEPeM) of USM/JEPeM/21090624, which were retrieved from the Institut Pergigian dan Perubatan Termaju (IPPT) repository at the University Sains Malaysia (USM). All data were obtained with the explicit permission and approval of the relevant ethical review board at USM, ensuring adherence to local ethical guidelines and regulations governing the use of medical imaging data. Furthermore, all patient data used in this study were anonymized to protect patient privacy, in compliance with relevant data protection laws and ethical standards for research involving human subjects.

DATA AVAILABILITY

The INbreast public dataset that support the findings of this study are openly available at *https://www.kaggle.com/datasets/ramanathansp20/inbreast-dataset*, reference number [37]. The IPPT-mammo dataset is available on request from the corresponding author, NFR. The data, which contain information that could compromise the privacy of research participants, is not publicly available dues to ethical document restrictions.

REFERENCES

[1] R. L. Siegel, K. D. Miller, H. E. Fuchs, and A. Jemal, "Cancer statistics, 2022," *CA: A Cancer Journal for Clinicians*, vol. 72, no. 1, pp. 7–33, Jan. 2022, doi: 10.3322/caac.21708.
[2] F. Yan, H. Huang, W. Pedrycz, and K. Hirota, "Automated breast cancer detection in mammography using ensemble classifier and feature weighting algorithms," *Expert Systems with Applications*, vol. 227, Oct. 2023, doi: 10.1016/j.eswa.2023.120282.
[3] E. Justaniah, G. Aldabbagh, A. Alhothali, and N. Abourokbah, "Classifying breast density from mammogram with pretrained CNNs and weighted average ensembles," *Applied Sciences*, vol. 12, no. 11, May 2022, doi: 10.3390/app12115599.

[4] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90.
[5] I. Vallés-Pérez, E. Soria-Olivas, M. Martínez-Sober, A. J. Serrano-López, J. Vila-Francés, and J. Gómez-Sanchís, "Empirical study of the modulus as activation function in computer vision applications," *Engineering Applications of Artificial Intelligence*, vol. 120, Apr. 2023, doi: 10.1016/j.engappai.2023.105863.
[6] M. Liu, Z. Cai, and J. Chen, "Adaptive two-layer ReLU neural network: I. Best least-squares approximation," *Computers & Mathematics with Applications*, vol. 113, pp. 34–44, May 2022, doi: 10.1016/j.camwa.2022.03.005.
[7] D. E. Goldberg, *Genetic algorithms in search, optimization and machine learning*. 1st ed. USA: Addison-Wesley Longman Publishing Co., Inc., 1989.
[8] S. Alsubai, A. Alqahtani, and M. Sha, "Genetic hyperparameter optimization with modified scalable-neighbourhood component analysis for breast cancer prognostication," *Neural Networks*, vol. 162, pp. 240–257, May 2023, doi: 10.1016/j.neunet.2023.02.035.
[9] D. A. Zebari *et al.*, "Breast cancer detection using mammogram images with improved multi-fractal dimension approach and feature fusion," *Applied Sciences*, vol. 11, no. 24, Dec. 2021, doi: 10.3390/app112412122.
[10] N. Sreevidya, P. Siva, G. Sumalatha, S. Pannala, and S. Nelli, "Effective feature selection using multi-objective improved ant colony optimization for breast cancer classification," *International Journal of Intelligent Engineering and Systems*, vol. 16, no. 3, pp. 604–613, Jun. 2023, doi: 10.22266/ijies2023.0630.48.
[11] O. N. Oyelade, A. E.-S. Ezugwu, T. I. A. Mohamed, and L. Abualigah, "Ebola optimization search algorithm: A new nature-inspired metaheuristic optimization algorithm," *IEEE Access*, vol. 10, pp. 16150–16177, 2022, doi: 10.1109/ACCESS.2022.3147821.
[12] S. A. Ali Shah Tirmzi, A. I. Umar, S. H. Shirazi, M. A. H. Khokhar, and I. Younes, "Modified genetic algorithm for optimal classification of abnormal MRI tissues using hybrid model with discriminative learning approach," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, vol. 10, no. 1, pp. 14–21, Jan. 2022, doi: 10.1080/21681163.2021.1956371.
[13] J. H. Holland, "Genetic algorithms," *Scientific American*, vol. 267, no. 1, pp. 66–73, 1992.
[14] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *2015 IEEE International Conference on Computer Vision (ICCV)*, Dec. 2015, pp. 1026–1034. doi: 10.1109/ICCV.2015.123.
[15] A. Hashmi *et al.*, "Contrast Enhancement in Mammograms Using Convolution Neural Networks for Edge Computing Systems," *Scientific Programming*, vol. 2022, pp. 1–9, Apr. 2022, doi: 10.1155/2022/1882464.
[16] A. Baccouche, B. Garcia-Zapirain, C. Castillo Olea, and A. S. Elmaghraby, "Connected-UNets: a deep learning architecture for breast mass segmentation," *npj Breast Cancer*, vol. 7, no. 1, Dec. 2021, doi: 10.1038/s41523-021-00358-x.
[17] K. B. Soulami, N. Kaabouch, M. N. Saidi, and A. Tamtaoui, "Breast cancer: One-stage automated detection, segmentation, and classification of digital mammograms using UNet model based-semantic segmentation," *Biomedical Signal Processing and Control*, vol. 66, Apr. 2021, doi: 10.1016/j.bspc.2021.102481.
[18] N. F. Razali, I. S. Isa, S. N. Sulaiman, N. K. A. Karim, and M. K. Osman, "CNN-Wavelet scattering textural feature fusion for classifying breast tissue in mammograms," *Biomedical Signal Processing and Control*, vol. 83, May 2023, doi: 10.1016/j.bspc.2023.104683.
[19] O. Yinka Ogundepo, I. O. Avazi Omeiza, and J. P. Oguntoye, "Optimized textural features for mass classification in digital mammography using a weighted average gravitational search algorithm," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 12, no. 5, pp. 5001–5013, Oct. 2022, doi: 10.11591/ijece.v12i5.pp5001-5013.
[20] E. Sadeghi Pour, M. Esmaeili, and M. Romoozi, "Breast cancer diagnosis: a survey of pre-processing, segmentation, feature extraction and classification," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 12, no. 6, pp. 6397–6409, Dec. 2022, doi: 10.11591/ijece.v12i6.pp6397-6409.
[21] A. D. Jagtap, K. Kawaguchi, and G. Em Karniadakis, "Locally adaptive activation functions with slope recovery for deep and physics-informed neural networks," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 476, no. 2239, Jul. 2020, doi: 10.1098/rspa.2020.0334.
[22] A. D. Jagtap, Y. Shin, K. Kawaguchi, and G. E. Karniadakis, "Deep Kronecker neural networks: A general framework for neural networks with adaptive activation functions," *Neurocomputing*, vol. 468, pp. 165–180, Jan. 2022, doi: 10.1016/j.neucom.2021.10.036.
[23] M. Chaiani, S. A. Selouani, M. Boudraa, and M. S. Yakoub, "Voice disorder classification using speech enhancement and deep learning models," *Biocybernetics and Biomedical Engineering*, vol. 42, no. 2, pp. 463–480, Apr. 2022, doi: 10.1016/j.bbe.2022.03.002.
[24] O. Yousuf and R. N. Mir, "DDoS attack detection in internet of things using recurrent neural network," *Computers and Electrical Engineering*, vol. 101, Jul. 2022, doi: 10.1016/j.compeleceng.2022.108034.
[25] H. Gao, L. Cai, and S. Ji, "Adaptive convolutional ReLUs," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 04, pp. 3914–3921, Apr. 2020, doi: 10.1609/aaai.v34i04.5805.
[26] R. K. Vasanthakumari, R. V. Nair, and V. G. Krishnappa, "Improved learning by using a modified activation function of a convolutional neural network in multi-spectral image classification," *Machine Learning with Applications*, vol. 14, Dec. 2023, doi: 10.1016/j.mlwa.2023.100502.
[27] G. Rajput, S. Agrawal, K. Biyani, and S. K. Vishvakarma, "Early breast cancer diagnosis using cogent activation function-based deep learning implementation on screened mammograms," *International Journal of Imaging Systems and Technology*, vol. 32, no. 4, pp. 1101–1118, Jul. 2022, doi: 10.1002/ima.22701.
[28] S. G., V. K., and G. R., "A deep convolutional neural network architecture for breast mass classification using mammogram images," *Journal of Autonomous Intelligence*, vol. 7, no. 3, Dec. 2023, doi: 10.32629/jai.v7i3.1288.
[29] T. Wei *et al.*, "Beyond fine-tuning: Classifying high resolution mammograms using function-preserving transformations," *Medical Image Analysis*, vol. 82, Nov. 2022, doi: 10.1016/j.media.2022.102618.
[30] G. Meenalochini and S. Ramkumar, "A deep learning based breast cancer classification system using mammograms," *Journal of Electrical Engineering & Technology*, vol. 19, no. 4, pp. 2637–2650, May 2024, doi: 10.1007/s42835-023-01747-x.
[31] S. Chakravarthy, B. Nagarajan, V. V. Kumar, T. R. Mahesh, R. Sivakami, and J. R. Annand, "Breast tumor classification with enhanced transfer learning features and selection using chaotic map-based optimization," *International Journal of Computational Intelligence Systems*, vol. 17, no. 1, Feb. 2024, doi: 10.1007/s44196-024-00409-8.
[32] A. Srinivas, V. V. K. D. V. Prasad, and B. L. Kumari, "Computer-aided diagnosis for early cancer detection using adaptive kernel based fuzzy cuckoo search optimization clustering from mammogram images," *Computers and Electrical Engineering*, vol. 103, Oct. 2022, doi: 10.1016/j.compeleceng.2022.108343.
[33] P. Kalpana and P. T. Selvy, "A novel machine learning model for breast cancer detection using mammogram images," *Medical & Biological Engineering & Computing*, vol. 62, no. 7, pp. 2247–2264, Jul. 2024, doi: 10.1007/s11517-024-03057-4.

[34] S. Maqsood, R. Damaševičius, and R. Maskeliūnas, "TTCNN: A breast cancer detection and classification towards computer-aided diagnosis using digital mammography in early stages," *Applied Sciences*, vol. 12, no. 7, Mar. 2022, doi: 10.3390/app12073273.
[35] H. Rahman, T. F. N. Bukht, R. Ahmad, A. Almadhor, and A. R. Javed, "Efficient breast cancer diagnosis from complex mammographic images using deep convolutional neural network," *Computational Intelligence and Neuroscience*, vol. 2023, no. 1, Jan. 2023, doi: 10.1155/2023/7717712.
[36] M. F. Aslan, "A hybrid end-to-end learning approach for breast cancer diagnosis: convolutional recurrent network," *Computers and Electrical Engineering*, vol. 105, Jan. 2023, doi: 10.1016/j.compeleceng.2022.108562.
[37] I. C. Moreira, I. Amaral, I. Domingues, A. Cardoso, M. J. Cardoso, and J. S. Cardoso, "INbreast: Toward a full-field digital mammographic database," *Academic Radiology*, vol. 19, no. 2, pp. 236–248, Feb. 2012, doi: 10.1016/j.acra.2011.09.014.
[38] N. F. Razali, I. S. Isa, S. N. Sulaiman, N. K. A. Karim, M. K. Osman, and Z. H. Che Soh, "Enhancement technique based on the breast density level for mammogram for computer-aided diagnosis," *Bioengineering*, vol. 10, no. 2, Jan. 2023, doi: 10.3390/bioengineering10020153.
[39] X. Zhang, X. Zhou, M. Lin, and J. Sun, "ShuffleNet: An extremely efficient convolutional neural network for mobile devices," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Jun. 2018, pp. 6848–6856. doi: 10.1109/CVPR.2018.00716.
[40] B. Xu, N. Wang, T. Chen, and M. Li, "Empirical evaluation of rectified activations in convolutional network," *Prepr. arXiv.1505.00853*.

BIOGRAPHIES OF AUTHORS

Noor Fadzilah Razali received her masters's degree in Biomedical Engineering in 2014 and a bachelor's degree in biomedical engineering in 2010 from the University of Malaya (UM), Malaysia. She is currently a lecturer at Electrical Engineering Studies, Universiti Teknologi MARA (UiTM), Malaysia. Her main research interests include biomedical engineering, signal and image processing, internet of things (IoT) and machine and deep learning applications. She can be contacted at email: fadzilah708@uitm.edu.my.

Iza Sazanita Isa received her bachelor's in electrical engineering from the Universiti Teknologi MARA, Malaysia, in 2004 and her M.Sc. degree from the Universiti Sains Malaysia, Malaysia, in 2008. In 2009, she joined Universiti Teknologi MARA, Penang Campus, Malaysia, as a young lecturer and promoted to senior lecturer with the Electrical Engineering Studies, Universiti Teknologi MARA, in 2013. She pursues her PhD in Electrical Engineering under the SLAB/SLAI scholarship and graduated in 2018. Currently, she is attached to the department of Control System Engineering at the faculty. She is a member of the AREDiM research group and the head of the research group RIDyLT and is actively involved in teaching and learning research. She can be contacted at email: izasazanita@uitm.edu.my.

Siti Noraini Sulaiman obtained her Ph.D. degree in (imaging) from the School of Electrical and Electronics Engineering, Universiti Sains Malaysia in 2012. Currently, she is an associate professor at the Electrical Engineering Studies, Universiti Teknologi MARA, Penang Campus, Malaysia. She is also now attached as the chair of Advanced Rehabilitation Engineering in Diagnostic and Monitoring Research Group (AREDiM) and the Deputy Chair of Advanced Control System and Computing Research Group (ACSCRG), Universiti Teknologi MARA, Penang Campus in Permatang Pauh, Penang, Malaysia. Her main research interest is biomedical engineering focusing on image filtering, image clustering and algorithms for image processing. She can be contacted at email: sitinoraini@uitm.edu.my.

Muhammad Khusairi Osman obtained his B.Eng degree in electrical and electronic engineering in 2000 and MSc in Electrical and Electronic Engineering in 2004 from Universiti Sains Malaysia. In 2014, he obtained his Ph.D. in medical electronic engineering from Universiti Malaysia Perlis (UniMAP), Malaysia. He is currently a senior lecturer at Faculty of Electrical Engineering, Universiti Teknologi MARA (UiTM), Malaysia. His research interest is in image processing, pattern recognition and artificial intelligence. He can be contacted at email: khusairi@uitm.edu.my.

Noor Khairiah A. Karim received the bachelor's degree in medicine, bachelor's degree in surgery (MBBS) and Master of Radiology (MRad) from the University of Malaya, Malaysia. She then obtained her Fellowship in Cardiac Imaging from the National Heart Center Singapore. She is currently a Senior Medical Lecturer of the Regenerative Medicine Cluster, and a Consultant Radiologist at the Advanced Medical and Dental Institute, Universiti Sains Malaysia. Her current research areas include medical image processing and analysis with special interest in cardiac, breast and brain imaging. She can be contacted at email: drkhairiah@usm.my.

Dayang Suhaida Awang Damit received her MSc in electronic design from the School of Electrical and Electronics Engineering, Universiti Sains Malaysia in 2010. Currently, she is doing her Ph.D. in biomedical engineering at Universiti Teknologi MARA, Malaysia. Her research interests include image processing, artificial intelligence and wireless communication. She can be contacted at email: dayang671@uitm.edu.my.