Original Article

An Optimized Deep Learning Techniques for Analysing Mammograms

Satish Babu Bandaru[1], Natarajasivan Deivarajan[2], Rama Mohan Babu Gatram[3]

[1] *Department of Computer Science and Engineering, Annamalai University, Annamalainagar, Tamil Nadu, 608002.*
[2] *Department of Computer Science and Engineering, Faculty of Computer Science and Engineering, Annamalai University, Annamalai Nagar, Tamil Nadu, 608002.*
[3] *Department of Computer Science and Engineering (AI & ML), RVR & JC College of Engineering, Guntur, A.P. 522019*

[1]researchbsbabu@gmail.com

Abstract - *Breast cancer screening makes extensive utilization of mammography. Even so, there has been a lot of debate regarding this application's starting age and screening interval. The deep learning technique of transfer learning is employed for transferring the knowledge learned from the source tasks to the target tasks. For the resolution of real-world problems, deep neural networks have demonstrated superior performance in comparison with the standard machine learning algorithms. The architecture of the deep neural networks has to be defined by considering the problem domain knowledge. Normally, this technique will consume a lot of time and computational resources. This work evaluated the efficacy of the deep learning neural network like Visual Geometry Group Network (VGGNet), Residual Network (ResNet), and the inception network for classifying mammograms. This work proposed optimization of ResNet with Teaching Learning Based Optimization (TLBO) algorithm's to predict breast cancers utilizing mammogram images. The proposed TLBO-ResNet is an optimized ResNet with faster convergence than other evolutionary mammogram classification methods.*

Keywords - *Mammography, Deep Learning, Convolutional Neural Networks (CNNs), Visual Geometry Group Network (VGGNet), Residual Network (ResNet), Teaching Learning Based Optimization Algorithm (TLBO).*

1. Introduction

Amo women, the commonest cancer type is breast cancer. Despite significant therapy advancements, breast cancer remains the chief cause of cancer-related mortalities and globally accounts for around 500,000 deaths per year. To minimize breast cancer-related mortality, the most efficient approaches are the population-based breast cancer screening programmes with mammography. Nevertheless, the existing screening programmes are extremely labor intensive because of the huge number of women screened for per detected cancer, which also costs extra. Furthermore, despite all this, the screening is still unable to detect up to 25% of mammographically visible cancers [1]. Keeping in mind the ever-increasing shortage of radiologists and expert medical personnel, it is essential to have alternate strategies to facilitate the continuation of current screening programmes. Moreover, preventing either overlooking or misinterpreting the visible lesions in Digital Mammography (DM) is critical. Computer-Aided Detection (CAD) automates marking possible suspicious regions in mammograms for radiologists' review. Various reports had emphasized that this, in turn, would boost mammographic sensitivity. Although CAD has been extensively introduced in clinics, radiologists have complained about this technology's high false-positive rate. Moreover, multiple latest investigations have stated the CAD's apparent inability to enhance the mammography's accuracy of diagnosis. However, this was expected to some degree. Most learning algorithms like the CAD [2] are based on pre-defined handcrafted features. Due to this, these algorithms are task-specific and a-priori knowledge-based, which in turn will cause a huge bias towards how human beings consider the performance of a task. Deep learning is a promising image-analysis classifier due to its feature learning and transfer learning ability. Instead of utilizing advanced image processing techniques for extracting handcrafted features, deep learning will simply transfer the feature activation maps for the subsequent convolution layer. These features are supervised by flattening the pixel matrix into the multi-layered fully connected layers in the final stage [3].

The key objective of transfer learning in deep learning is to employ the trained model parameters in new models to reduce the training time. Transfer learning [4] is extensively utilized in training new models with many unrelated instances to the new training model. Migration of parameters to a new model can minimize the training time and achieve a

higher accuracy using limited target data for training the models. Transfer learning will use its benefits to resolve the problems mentioned earlier in deep learning.

Convolutional Neural Networks (CNNs) can learn hierarchical interpretations from images and transmit knowledge implanted in the pre-trained model's weights to the new images. The low-level features like curves and edges are extracted using a convolutional layer, and the later layer's operation can learn more abstract representations for a diverse field of applications. As a result, the lower-level representations are transferred to a new model while only the higher-level representations must be learned. Fine-tuning will refer to updating the higher hidden layers' weights. This procedure's success depends to some degree on the "distance" between the source and the new dataset. The pre-trained weights of deep CNN's [5] training are successfully employed on various distinct tasks. Most existing pre-trained architectures have been trained based on natural image data. As an example, for speech recognition, despite the huge gaps between various languages, certain studies have been able to showcase the effectiveness of employing pre-trained models for speech recognition tasks in diverse languages.

Most research studies have applied deep learning to numerous image types such as time-series, satellite images, radiography images, mammogram images, etc. Deep learning has several new articles on the analysis of mammograms for diagnosing breast cancer, calcification in the breast, and segmentation of cancer lesions through the re-training of well-known pre-trained CNN models such as VGGNet, DenseNet, AlexNet, DenseNet, and GoogleNet on diverse datasets. These models have been able to accomplish high classification performances for breast cancer. Nonetheless, the pre-trained CNN model's most unfavorable attribute is its vast number of classification parameters.

Implementation of transfer learning [6] will involve the following two steps: feature extraction and parameter tuning. The pre-trained model will use the training data to hold the new features from the dataset during the feature extraction stage. Secondly, to optimize a model's performance in the current applied domain, the model architecture must be rebuilt and updated together with the parameter tuning. Metaheuristic methods are used to optimize the parameters. In this work, the TLBO optimized ResNet is proposed to detect cancer in mammography. The paper comprises related works, methodology, experimental results, and conclusion sections.

2. Related Works

Li et al. [7] presented an improved DenseNet-II neural network for mammogram image classification. The proposed model applied image normalization to avoid any interference from light. Secondly, improvement was done on the DenseNet neural network by appending an inception structure in place of the first convolutional layer. In the end, the pre-processed mammogram datasets were offered as inputs to the AlexNet, VGGNet, GoogLeNet, DenseNet network model, and DenseNet-II, neural network model. Then, analysis, as well as comparison of the experimental outcomes, were performed. It was evident from the experimental outcomes that, compared to other network models, the proposed deep learner had better classification performance. The model arrived at 94.55% average accuracy, significantly enhancing the mammogram images' malignant and benign classification accuracy. At the same time, the outcomes were able to confirm the model's robustness as well as good generalization ability.

Agarwal et al. [8] devised a fully automated framework for detecting masses using Faster Region-based CNN (Faster-RCNN). The proposed framework was evaluated using OPTIMAM Mammography Image Database (OMI-DB). For images from the GE scanner, this framework had acquired a TPR of 0.91±0.06 at 1.69 FPI. Furthermore, in comparison with other advanced methods on the INbreast dataset, the proposed framework's superior performance in acquiring higher TPR for benign and malignant masses was able to demonstrate its potentiality in being employed for screening breast cancers.

Chakravarthy & Rajaguru [9] presented a novel customized method that incorporated the deep learning concept with the Extreme Learning Machine (ELM) and then optimized it with Crow-Search Algorithm (CSA). Hence, to boost the technologically advanced operations, a proposal of an enhanced deep feature-based CSA optimized ELM was offered to address the problems in healthcare. At first, the work focused on detecting the input mammograms as either abnormal or normal. Afterward, it was involved with further classification of the abnormal severities into either malignant type or benign type. Eventually, comparison studies were done on this proposed work with other existing SVM, ELM, Particle Swarm Optimization (PSO) optimized ELM, and CSA-ELM. It was found that the proposed method had attained a maximum overall classification accuracy.

Patil & Biradar [10] had optimized region growing segmentation using a hybrid metaheuristic algorithm called the Firefly updated Chicken-based CSO (FC-CSO). It was followed by feature extraction. The two distinct deep learning architectures were referred to as the CNN and the Recurrent Neural Network (RNN). In addition, the tumor segmented binary image was treated as the CNN's input, while both GLCM and GLRM were treated as the RNN's input. It was evident from the proposed study's simulated outcomes that the AND operation of two classifier outputs would yield the overall diagnostic accuracy, surpassing the standard models' performance.

Reenadevi et al. [11] developed an optimized Wrapper-based Chaotic CSA (WCCSA) to improve breast cancer diagnosis. The proposed WCCSA was utilized along with the Probabilistic Neural Network (PNN) for identifying the mammogram images as either normal, benign, or malignant, thus aiding the patients in receiving feasible care and treatment at the earlier stages. A mini-MIAS dataset of 322 images was used to assess the WCCSA with the PNN method's effectivity. Then, its performance was compared with that of other machine learning algorithms. The assessment's outcome showed that the proposed WCCSA with the PNN method would seek an optimal feature subset while retaining its stability and had accomplished 97% accuracy.

Ashok et al. [12] proposed an optimized region growing method, wherein the Grey Wolf Optimization (GWO) was used to accomplish the optimal threshold and seed point selection. The proposed work would consider extracting both local and global features. The duly considered global features were inclusive of shape features. A combination of the local and the global features was fed into the SVM classifier to differentiate the breast mass's nature as either malignant or benign. Through the combination of global texture feature GLCM as well as LBP, the proposed work achieved the highest accuracy of 96%.

The Glowworm Swarm Optimization (GSO) algorithm was best suited for the simultaneous detection of numerous solutions and equivalent or dissimilar objective function values. This GSO feature was feasible for optimization of the feature vector obtained from the multiscale feature extraction technique. However, the unconditioned output matrix of the ELM classifier's hidden stage resulted in the issue of poor generalization performance. The optimization algorithms were employed for this issue's resolution due to their global search abilities. Melekoodappattu et al. [14] suggested the ELM with Fruit fly Optimization Algorithm (ELM-FOA) together with the GSO (that is, the GSO-ELM-FOA) to alter the input weight for the accomplishment of maximal performance at the ELM's hidden node. The proposed GSO-ELM-FOA had 100% testing precision of 100% and 97.91% sensitivity. Moreover, the devised system could identify tumors and calcifications with 99.15% accuracy.

Frazer et al. [15] presented ResNet deep learning-based AI techniques for classifying mammograms. The ResNet extracts the local features from cancer and non-cancer local regions. These are combined with the global features extracted, and the mammogram is classified. The proposed method was evaluated using 930 mammogram images. An AUC of 0.8979 [95% confidence interval (CI) 0.873, 0.923] and ACC of 0.8178 was achieved.

Kanya Kumari et al. [16] used Contrast-Limited Histogram Equalization (CLAHE) for pre-processing and Advanced Gray-Level Co-occurrence Matrix (AGLCM) for extracting features like intensity, texture, and shape. A new feature selection method called Weighted Adaptive Binary Teaching Learning Based Optimization (WA-BTLBO) was proposed to optimize the feature set. Experimental results showed that the XGBoost classifier with the proposed feature selection achieved the maximum accuracy in classifying the mammograms.

Shivhare & Saxena [17] proposed Velocity Updated Lion Algorithm (VU-LA) optimization for both feature selection and classifier. The neural network weights were chosen to maximize the accuracy in classifying benign and malignant breast cancer. The optimal feature selection and the weight optimization of NN are achieved by hybridizing the LA and PSO.

Table 1. Summary of Literature Reviews

| S. No	Authors	Techniques	Merits	Demerits
1	Li et al., [7]	DenseNet-II neural network	Robustness as well as good generalization ability	Any interference from light
2	Chakravarthy & Rajaguru [9]	ELM and CSA	Better accuracy	Healthcare
3	Reenadevi et al., [11]	WCCSA with PNN	Stability and accuracy	Breast cancer and mammogram images
4	Ashok et al., [12]	GWO	Combination of global texture features	Threshold and seed point
5	Melekoodappattu et al., [14]	GSO-ELM-FOA	Better accuracy	ELM-poor generalization
6	Shivhare & Saxena [17]	CLAHE and VU-LA	Detection of breast cancer using mammogram images	To enhance the accuracy of diagnosis

3. Materials and Methods

With the emergence of deep learning or CNN, the image classification task has experienced a lot of enhancements. It is necessary to have a huge training dataset for training the deep neural network model. A deep learning model's performance is extremely dependent on the number of images that are employed for the model's training due to the model's inherent capability to extract features (spatial as well as temporal) with the filter utilization. The transfer learning concept performed deep learning in the domain with small-sized datasets. A CNN model would extract the feature from

the specified data in transfer learning. Later on, this feature would get transmitted for the resolution of associated tasks, including new data (small dataset), in which it is not feasible to construct the CNN from scratch [6]. From amongst the extensively utilized transfer learning methods for the medical domain, training a model with a large-sized dataset, the ImageNet, was a pre-trained model for object identification and classification. For transfer learning, the deep learning model's selection relied on the model's capability to extract the domain's associated features. This section will discuss the VGG network, ResNet, Inception network, and TLBO with ResNet methods. In all these methods, the mammogram is given as input, and the output is the classification of the mammogram as Benign or Malignant.

3.1. VGG-16 Network Learning

Development of the VGG-Net by Simonyan et al. was achieved via the utilization of an extremely small convolution in the network. Despite its simplistic nature, the proposed model's key distinction from the earlier models was its extensive utilization due to the model having a more in-depth structure, along with double or triple convolution layers. The earlier architectures were characterized by the layers of sharing and convolution following one another. Since the VGG is pre-trained using ImageNet dataset with one million images having 1000 distinct classes, it is employed as an effective feature extractor for new images [18].

Figure 1 shows the architecture of the VGG-16.
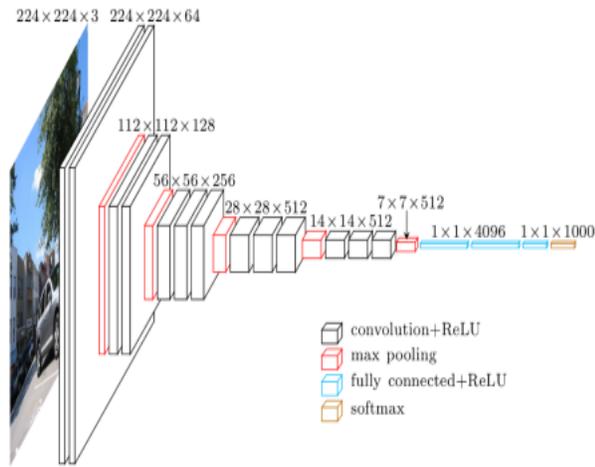
Fig. 1 Architecture of VGG16

For feature extraction, the VGG-16 model will utilize three convolution filters with thirteen convolution layers, a ReLU layer, and maximum pooling layers for sampling. It will have three layers fully connected for the classification, wherein 2 layers will act as hidden layers. In contrast, the last layer will have 1000 units, indicating the ImageNet database's image categories.

This structure can simulate a bigger filter while retaining the merits of smaller filters. The VGGNet has demonstrated better performance with fewer parameters than the older models. Also, rather than just one ReLU layer, it will use 2 ReLU layers for 2 convolution layers. The convolution and partnering layers will result in a decrease in the input volumes' spatial size. Thus, the increased number of filters will lead to an increase in the volumes' depth. It is feasible for both edge detection as well as object classification problems.

The VGG16 network model fine-tunes the model duty. Let it have a dataset with *n* samples *{(a (1), b (1)),…,(a (n), b (n))}* for training. Hence, Eq. (1) will express the network's overall cost function as below:

$$J(W, x) = \left[\frac{1}{n} \sum_{i=1}^{n} \left(\frac{1}{2} \left\| K_{w,x}(y^{(i)} - x^{(i)}) \right\|^2 \right) \right]$$
$$+ \frac{\lambda}{2} \sum_{l=1}^{n_{l-1}} \sum_{i=1}^{s_l} \sum_{j=1}^{s_{l+1}} \left(W_{ji}^{(l)} \right)^2 \qquad (1)$$

For the above equation, $K_{w,b}{}^{(i)}$ will denote the neural network model, $W_{ji}^{(l)}$ will denote the connection weight between the *j*th element of layer *l*+1, and the *i*th element of layer *l*+1, and *x* will denote the bias term of the hidden layer neuron. Eq. (1) is a regulation item on the right-hand side, which can do the following: prevent over-fitting, significantly minimize the weight, and adjust the relative importance of the two terms before and after the cost function, λ.

3.2. VGG-19 Network Learning

The VGG CNN is made up of 6 key structures [19], wherein every structure will primarily have multiple connected convolutional layers as well as full-connected layers. The convolutional kernel's size is set at 3*3 while the input size will be 224*224*3. The VGG-19 model structure is given in Figure 2.
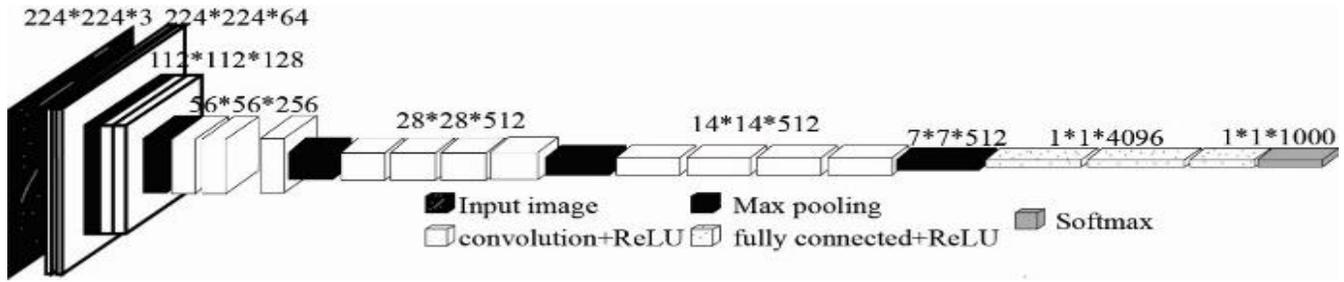
Fig. 2 VGG-19 network model

The VGG-19 CNN has found usage as a pre-processing approach. In contrast to standard CNNs, the VGG-19 CNN has more improvements in the network depth. Rather than a single convolution, this CNN will employ the more improved interchanging structure of multiple convolutional and non-linear activation layers. The layer structure's advantages include better image feature extraction, downsampling through max pooling, and modification of the linear unit (ReLU) as the activation function, in other words, to pick the area of the image's biggest as the area's pooled value. The key purpose of the downsampling layer is to improve the network's anti-distortion capability to the image while preserving the sample's major features and minimizing the number of parameters. This downsampling layer can be expressed as the below-mentioned Eq. (2).

For this equation, $down\left(X_j^{(n-1)}\right)$ will denote the maximum pooling sampling function, τ_j^n will denote the coefficient which corresponds to the nth layer's jth feature map and $f\left(\tau_j^n down\left(X_j^{(n-1)}\right) + b_j^{(n)}\right)$ Will denote the ReLU activation function.

$$X_{p_i}^{(n)} = f\left(\tau_j^n down\left(X_j^{(n-1)}\right) + b_j^{(n)}\right) \qquad (2)$$

3.3. ResNet-18 Network Learning

The original ResNet-18 architecture is depicted in Figure 3. This architecture's network has eighteen layers, i.e., seventeen convolutional layers, one fully-connected layer, and an extra softmax layer to execute the classification. The convolutional layers will employ 3×3 filters. Moreover, it is designed to have a similar number of filters in the layers if the output feature map is of a similar size. If the output feature map is halved, the filters will get doubled in the layers. Convolutional layers with a stride of 2 will carry out the downsampling. An average pooling is included after the convolution layer, followed by a fully connected layer and softmax layer. Residual shortcut connections are introduced between the layers across the network [20]. These connections are of two distinct types. The first connection type, indicated as solid lines, is employed when both the input and the output have similar dimensions. The second connection type, indicated as dotted lines, is employed when there is an increase in the dimensions. Although this connection type will continue to perform identity mapping, it will also perform zero padding for increased dimensions with a stride of 2.
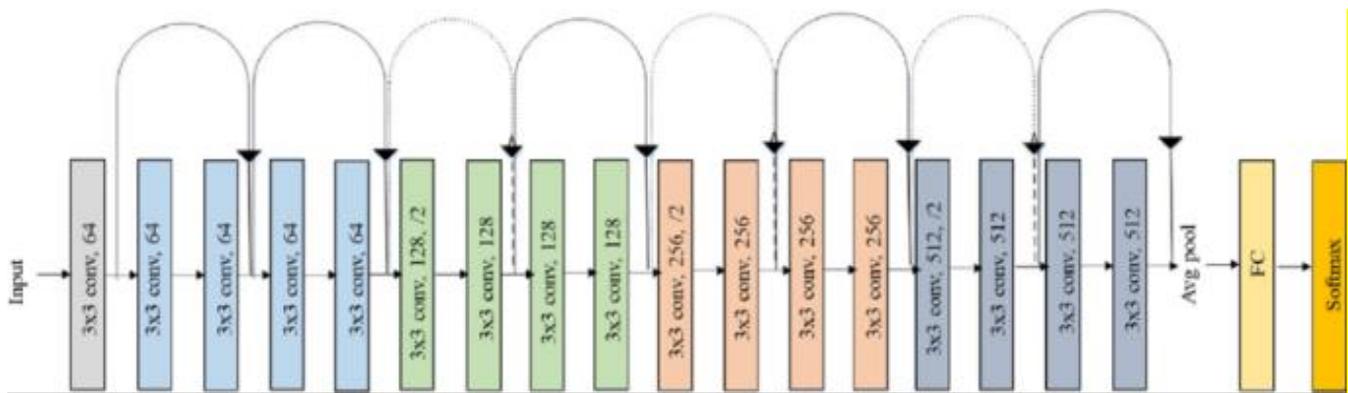

Fig. 3 Architecture of ResNet-18

The following two distinct strategies have been employed in these network training experimentations. At first, there is the usage of a marginally altered version of the ResNet-18 to execute training from scratch through random initialization of the network parameters. Moreover, training with the greyscale images could minimize the number of input channels to merely a single channel.

Secondly, the weight initialization was carried out with a pre-trained network, and then the transfer learning was executed. As the trained model is applied for mammogram classification, it will carry out the task by adapting the network. It will employ two approaches to transfer learning for transmitting knowledge from a pre-trained network. The first approach will perform Off-The-Shelf (OTS) transfer learning by replacing the original network's last dense layer with the new dense layer to match the number of classes for the task. This OTS approach will use all the network layers, with the exclusion of the last layer (classifier), for feature extraction, and the last layer's weights will get adapted to the new task. Fine-Tuning (FT) is the second approach, wherein more than one layer of the network will get re-trained from the new task's samples. This approach will re-train the network's entire convolution layers with the dataset. The starting point of these transfer learning approaches will be the weights of the ResNet-18 network trained on ImageNet.

3.4. ResNet-34 Network Learning

The ResNet-34 infrastructure is the residual building block and the overall network's key constituent. The residual building block will employ a shortcut connection to skip the convolutional layers for effective mitigation of the problem of gradient disappearance or gradient explosion, a result of increasing the depth in neural networks. This, in turn, will aid in constructing more flexible CNN structures and boost the wood knot defects' recognition rate [21].
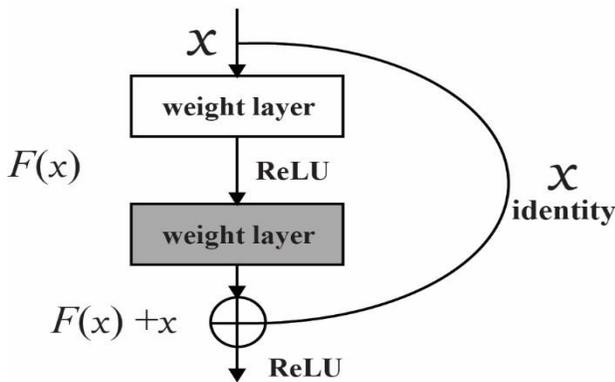
Fig. 4 Building blocks of ResNet-34

Depiction of the basic-block's structure is given in Figure 4. Its utilization is done for the ResNet's 34 layers. The following are the residual building block's constituents: multiple convolutional layers (Conv), Batch Normalizations (BN), a ReLU activation function, and a shortcut. Eq. (3) will formulate the residual building block's output as below:

$$y = F(x) + x \qquad (3)$$

In this equation, F will denote the residual function, x will denote the residual function's input, and y will denote the residual function's output. The overall residual network's constituents will include the first convolutional layer and multiple basic blocks.

The ResNet-34 will constitute 33 convolutional layers, a 3 * 3 size max-pooling layer, an average pool layer, and a fully connected layer. In a standard ResNet-34 model, the back of all the convolution layers in the "Basic Block" block is applied with the ReLU activation and BN, and the final layer is applied with the softmax function. Table 2 will offer an illustration of the ResNet-34's architecture.

Table 2. The structure of ResNet-34

Layer Name	Output Size	34-Layer
Conv1	112*112	7*7, 64, stride 2 3*3 maxpool, stride 2
Conv2_x	56*56	$\begin{bmatrix} 3*3, & 64 \\ 3*3 & 64 \end{bmatrix} * 3$
Conv3_x	28*28	$\begin{bmatrix} 3*3, & 128 \\ 3*3 & 128 \end{bmatrix} * 4$
Conv4_x	14*14	$\begin{bmatrix} 3*3, & 256 \\ 3*3 & 256 \end{bmatrix} * 6$
Conv5_x	7*7 1*1	$\begin{bmatrix} 3*3, & 512 \\ 3*3 & 512 \end{bmatrix} * 3$ Average pool, 1000-d fc, softmax

3.5. Inception/ GoogLeNet Network Learning

GoogLeNet [22] is the first implementation done with the. This module's key concept is based on the authors' findings on using dense components to approximate a sparse local structure. The authors intended to build a multi-layer network by seeking the optimal local structure and then repeating it. The Inception module has four distinct branches, giving each branch the same input (as shown in Figure 5). The first branch will filter the input with a 1 * 1 convolution that serves as a linear transformation on the input channels. Subsequently, 1 * 1 kerneled convolution is carried out in the second and third branches to reduce the dimensionality, followed by convolutional layers with 3*3 and 5*5 sized kernels, respectively. The fourth branch will execute max-pooling followed by convolution with kernels of size 1 * 1. In the end, each branch's outputs will undergo concatenation and will be given as the next block's input.
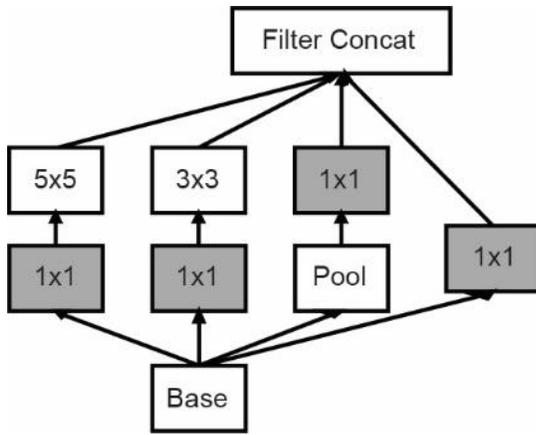
Fig. 5 Inception module of GoogLeNet

GoogLeNet's construction will involve the stacking of nine Inception modules. To minimize the feature maps' dimensionality, a max-pooling layer is included between the inception modules at the chosen locations. The incorporation of auxiliary classifiers is a remarkable feature of GoogLeNet. The authors assumed that a CNN's middle layers must yield discriminative features to append simple classifiers (two fully connected layers and a softmax layer), which will be applied to generated features of an intermediate network point. These classifiers' decisions will evaluate the loss. Late on, this loss will be utilized in the backpropagation to evaluate additional gradients, which will train the respective convolutional layers. The auxiliary classifiers will get discarded at the inference time.

3.6. Proposed Teaching Learning -Based Optimization (TLBO) with ResNet (18 & 34)

In the proposed TLBO Resnet18 and TLBO Resnet34, the architecture of the deep learners is optimized to improve the classification of the mammograms as benign or malignant. The batch size, learning rate, activation, and the number of epochs of Resnet 18 and 34 are optimized. The TLBO algorithm addresses a teacher's impact on the students/learner's accomplishment within a class in terms of ratings or outcomes. The population-based TLBO will simulate the teaching-learning procedure within a classroom. In general, the class teacher is considered a well-educated individual who will share their knowledge with the students. With the aid of a good teacher, the class will often improve the learner's marks or grades.

Moreover, the learner will learn information from other class members to improve their marks. The learners will represent the solutions in this method. The proposed method represents the solutions with batch size, learning rate, activation, and the number of epoch values for Resnet18 and 34. The current population's learners with the best fitness value will get picked as a classroom teachers. The teacher and learner stages are the two distinct stages of the TLBO's

learning procedure. In the teacher stage, all the learners in the class will gain expertise from the teacher. Likewise, during the learner stage, a learner will gain expertise from the class's other students [23].

In the teacher stage, there is the utilization of Eq. (4) as well as Eq. (5) to update the population:

$$Mean\_difference_i = rand_i(Mean_{new} - T_F * Mean_i) \quad (4)$$

$$L_{new,i} = L_{old,i} + Mean\_difference_i \quad (5)$$

In these equations, r and ᵢ will denote a random variable evenly distributed in the [0, 1] range. T_F will denote the teaching factor's value. This value can be either 1 or 2, based on random selection. $L_{new,i}$ will denote i^{th} new population. Eq. (1) and Eq. (2) are used to generate the new solution via a fitness function's evaluation. If the previous solution has lower fitness than the new solution, then this new solution will replace the previous solution.

During the learner stage, the student will acquire expertise from other students through random interactions amongst themselves. The below procedure is used to update the population in the learner phase:

• Randomly pick two distinct students: L_i and L_j.

• If the fitness of f(L_i) is lower than that of f(L_j), then there is the utilization of Eq. (6) to update the population:

$$L_{new,i} = L_{old,i} + rand_i(L_i - L_j) \quad (6)$$

Otherwise, there is the utilization of Eq. (7) to update the population:

$$L_{new,i} = L_{old,i} + rand_i(L_j - L_i) \quad (7)$$

1. The TLBO algorithm's implementation has been depicted in Input parameters :
   Dataset, N: Number of iterations; nPop: population size; nVar: decision variables
2. Define objective(cost) functions: f(x)
3. Create initial population (J = 1,2,3,…………., nPop)
4. Complete the fitness of the initial population using the cost function
5. Select the best solution
6. For I = 1 to N
7. Calculate the mean value of the population
8. Select the teacher
9. For J = 1 to nPop
10. Select the teaching factor (TF)
11. Update the population using Eq. (4) and Eq. (5)
12. Evaluate the population using the cost function

13. Update the best solution
14. End loop J
15. For K = 1 to nPop
16. Randomly select two leaners, L_1 and L_2
17. If $f(L_1) < f(L_2)$
18. Update the population using Eq. (6)
19. Else
20. Update the population using Eq. (7)
21. End if
22. Update the best solution
23. End loop K
24. End loop I
25. Select the optimal architecture

The optimal architecture of TLBO-Resnet18 and 34 to enhance the classification of mammograms is obtained.

The multi-layers-based CNN model suffers from the vanishing gradient, wherein the model's learning in the initial layers will be hindered. This problem will result in a degradation of the model's performance: there will be a decrease in the accuracy and an increase in the probability of over-fitting. However, this problem can be resolved with the residual learning's method of "identity mapping short-connections."

This work has used the TLBO to examine the training's effect on deep residual learning [24]. This cumbersome procedure will involve identifying an accurate model that will offer better results within a reasonable amount of time. It will update the teaching and learning phases of the model and optimization algorithms to demonstrate the incorporation of performance and accuracy. Instead of the classical augmentation method, it will employ the data augmentation technique utilized in technologically advanced residual learning. It can boost the TLBO algorithm's error rate and training time via updating and picking accurate teaching and learning phases.

Initially, the same class's images will get arranged into a list. Later, the order will change such that the least highlighted images are placed at the bottom while the most highlighted features are placed at the top. The similarity of images within the class is evaluated as the degree of the representative images corresponding to its class. The below Eq. (8) will offer the expression of the representation degree of images:

$$degree_{rep}(x_m) = \sum_{l-1}^{n}(l \neq m)similarity(xl, xm) \quad (8)$$

In the second step, the training set will get appended with the top-selected images having more similarities within each class. Also, the BN is utilized after each convolutional layer and before the activation function (ReLU). Through the application of the proposed method, suitable batch size, learning rate, activation, and the number of epochs of the

ResNet network are obtained, it can arrive at the improvement stage in the final training and the validation accuracy. Furthermore, this method can overcome the over-fitting problem and minimize the training's computational complexity.

4. Results and Discussion

This section evaluates the VGG16, VGG19, Resnet18, Resnet34, inception, TLBO-Resnet18, and TLBO-Resnet34 methods for classifying mammograms. The Curated Breast Imaging Subset of Digital Database for Screening Mammography (CBIS-DDSM) is used to evaluate various algorithms. The DDSM is a database of 2,620 scanned film mammography studies. It contains normal, benign, and malignant cases with verified pathology information. This work uses 550 Benign and 625 Malignant mammogram images for evaluation. Python, open CV, tensor flow, and Keras implement the algorithms.

Table 3 shows the summary of the results. The classification accuracy, recall, and precision as shown in figures 6 to 8.
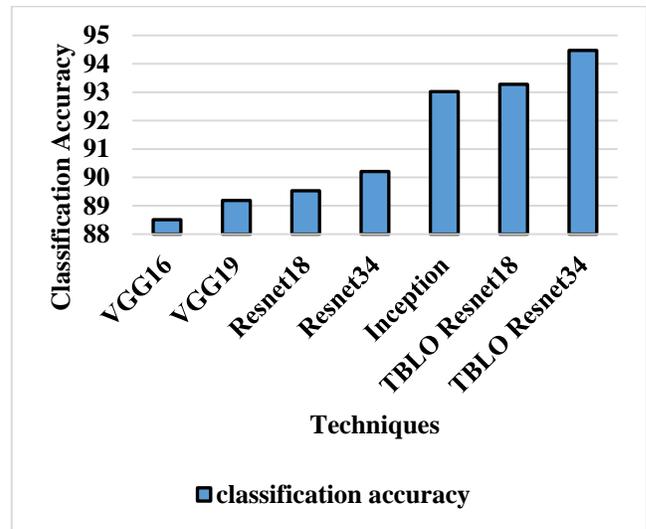

Fig. 6 Classification Accuracy for TLBO Resnet34

Figure 6 shows that the TLBO Resnet34 has higher classification accuracy by 6.06% for VGG16, 5.29% for VGG19, and 4.91% for Resnet18, by 4.16% for Resnet34, by 1.55% for inception, and by 1.27% for TLBO Resnet18 respectively. It is observed that the optimization of the ResNet significantly improves the results. The selection of hyperparameters improves the efficiency of the deep learners.
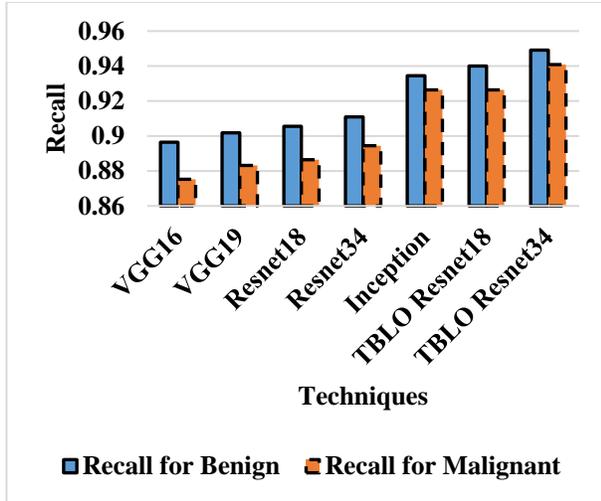
Fig. 7 Recall for TLBO Resnet34

Figure 7 shows that the TLBO Resnet34 has a higher recall by 6.66% for VGG16, 6.06% for VGG19, and 5.65% for Resnet18, by 5.06% for Resnet34, by 1.55% for inception, and by 0.96% for TLBO Resnet18 respectively for benign. The TLBO Resnet34 has a higher recall by 5.51% for VGG16, 4.6% for VGG19, 4.24% for Resnet18, 3.34% for Resnet34, and 1.54% for inception, and 1.54% for TLBO Resnet18 respectively for malignant.

Figure 8 shows that the TLBO Resnet34 has higher precision of 6.14% for VGG16, 5.18% for VGG19, 4.78% for Resnet18, and 3.83% for Resnet34, by 1.72% for inception and 1.67% for TLBO Resnet18 respectively for benign. The TLBO Resnet34 has higher precision of 6.01% for VGG16, 5.42% for VGG19, 5.06% for Resnet18, and 4.49% for Resnet34, by 1.37% for inception and 0.88% for TLBO Resnet18 respectively for malignant.

As observed from the experimental results, the training set with the top-selected images has more similarity within each class, and the selection of appropriate batch size, learning rate, activation, and the number of epochs of the ResNet network obtained through the proposed TLBO ResNet18 and proposed TLBO ResNet34 improves the accuracy.
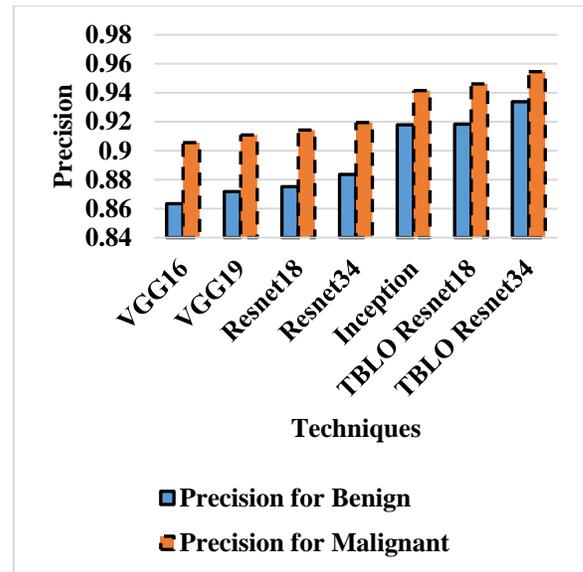
Table 3. Summary of Results

Techniques	VGG16	VGG19	Resnet18	Resnet34
Classification Accuracy	88.51	89.19	89.53	90.21
Recall for Benign	0.8964	0.9018	0.9055	0.9109
Recall for Malignant	0.8752	0.8832	0.8864	0.8944
Precision for Benign	0.8634	0.8717	0.8752	0.8836
Precision for Malignant	0.9056	0.9109	0.9142	0.9194
Techniques	Inception	TLBO Resnet18	TLBO Resnet34	
Classification Accuracy	93.02	93.28	94.47	
Recall for Benign	0.9345	0.94	0.9491	
Recall for Malignant	0.9264	0.9264	0.9408	
Precision for Benign	0.9179	0.9183	0.9338	

Fig. 8 Precision for TLBO Resnet34

5. Conclusion

Transfer learning is vital in applying deep CNNs to medical imaging tasks. This work has presented the implementation of a computer-aided classification approach for mammogram-based breast cancer prediction via an evolutionary system by integrating architectural evolution with deep neural network learning using ResNet (18 & 34) and TLBO. The TLBO is touted as the latest, accurate, reliable, and robust technique for global optimization across continuous spaces. Researchers have also put forward some variations of the TLBO to boost the basic TLBO algorithm's performance. This technique will involve the proposal of diverse variations and their hybridization with the evolutionary algorithm. As a result, this technique will enhance its performance and offer a better prediction of breast cancers (either non-cancerous or cancerous) through the utilization of the TLBO's global search capability and the ResNet's (18 & 34) local search capability. Furthermore, there is the proposal of the latest direct encoding strategy that will partition the ResNet (18 & 34) architecture into two distinct blocks: the first block, which is made up of just the convolutional as well as pooling layers, and the second block which is solely made up the fully connected layers. With this encoding strategy, and almost standard TLBO algorithm can be employed for comparing and combining the variable length ResNet (18 & 34) architectures.

Conflicts of Interest

This section is compulsory. A competing interest exists when professional judgment concerning the validity of research is influenced by a secondary interest, such as financial gain. We require that our authors reveal any possible conflict of interest in their submitted manuscripts. If there is no conflict of interest, the authors should state that "The author(s) declare(s) that there is no conflict of interest regarding the publication of this paper."

Funding Statement

Acknowledgments

References

[1] Rodriguez-Ruiz A, Lång K, Gubern-Merida A, Broeders M, Gennaro G, Clauser P & Sechopoulos I, "*Stand-Alone Artificial Intelligence for Breast Cancer Detection in Mammography: Comparison with 101 Radiologists*," JNCI: Journal of the National Cancer Institute, vol. 111, no. 9, pp. 916-922, 2019.

[2] Kim E. K, Kim H. E, Han K, Kang B. J, Sohn Y. M, Woo O. H & Lee C. W, "*Applying Data-Driven Imaging Biomarker in Mammography for Breast Cancer Screening: Preliminary Study,*" Scientific Reports, vol. 8, no. 1, pp. 1-8, 2018.

[3] Altan G, "Deep Learning-based Mammogram Classification for Breast Cancer," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 8, no. 4, pp. 171-176, 2020.

[4] Chougrad H, Zouaki H & Alheyane O, "Deep Convolutional Neural Networks for Breast Cancer Screening," *Computer methods and programs in biomedicine*, vol. 157, pp. 19-30, 2018.

[5] Zhou J, Yang X, Zhang L, Shao S & Bian G, "Multisignal VGG19 Network with Transposed Convolution for Rotating Machinery Fault Diagnosis Based on Deep Transfer Learning," *Shock and Vibration*, 2020.

[6] Khan I. U & Aslam N, "A Deep-Learning-Based Framework for Automated Diagnosis of COVID-19 Using X-Ray Images," Information, vol. 11, no. 9, pp. 1-13, 2020.

[7] Li H, Zhuang S, Li D. A, Zhao J & Ma Y, "Benign and Malignant Classification of Mammogram Images Based on Deep Learning," *Biomedical Signal Processing and Control*, vol. 51, pp. 347-354.

[8] Agarwal R, Díaz O, Yap M. H, Llado X & Marti R, "Deep Learning for Mass Detection in Full Field Digital Mammograms," *Computers in Biology and Medicine*, vol. 121, pp. 103774, 2020.

[9] Chakravarthy S. S & Rajaguru H, "Automatic Detection and Classification of Mammograms Using Improved Extreme Learning Machine with Deep Learning," *IRBM*, 2021.

[10] Patil R. S & Biradar N, "Automated Mammogram Breast Cancer Detection using the Optimized Combination of Convolutional and Recurrent Neural Network," *Evolutionary Intelligence*, pp. 1-16, 2020.

[11] Kavitha T, Mathai P. P, Karthikeyan C, Ashok M, Kohar R, Avanija J & Neelakandan S, "Deep Learning Based Capsule Neural Network Model for Breast Cancer Diagnosis Using Mammogram Images," *Interdisciplinary Sciences: Computational Life Sciences*, pp. 1-17, 2021.

[12] Reenadevi R, Sathiya T & Sathiyabhama B, "Classification of Digital Mammogram Images using Wrapper based Chaotic Crow Search Optimization Algorithm," *Annals of the Romanian Society for Cell Biology,* pp. 2970-2979, (2021).

[13] Ashok A, Vijayan D & Lavanya R, "Computer Aided Mass Segmentation in Mammogram Images using Grey Wolf Optimized Region Growing Technique," In 2021 5th *International Conference on Trends in Electronics and Informatics (ICOEI), IEEE*, pp. 1082-1087, 2021.

[14] Melekoodappattu J. G, Subbian P. S & Queen M. F, "Detection and Classification of Breast Cancer from Digital Mammograms Using Hybrid Extreme Learning Machine Classifier," *International Journal of Imaging Systems and Technology*, vol. 31, no. 2, pp. 909-920, 2021.

[15] Frazer H. M, Qin A. K, Pan H & Brotchie P, "Evaluation of Deep Learning Based Artificial Intelligence Techniques for Breast Cancer Detection on Mammograms: Results from a Retrospective Study Using a Breastscreen Victoria Dataset," *Journal of Medical Imaging and Radiation Oncology*, vol. 65, no. 5, pp. 529-537, 2021.

[16] Kanya Kumari L & Naga Jagadesh B, "An Adaptive Teaching Learning Based Optimization Technique for Feature Selection to Classify Mammogram Medical Images in Breast Cancer Detection," *International Journal of System Assurance Engineering and Management*, pp. 1-14, 2022.

[17] Shivhare E & Saxena V, "Breast Cancer Diagnosis from Mammographic Images Using Optimized Feature Selection and Neural Network Architecture*," International Journal of Imaging Systems and Technology,* vol. 31, no. 1, pp. 253-269, 2021.

[18] Sahinbas K & Catak F. O, "Transfer Learning-Based Convolutional Neural Network for COVID-19 Detection with X-Ray Images," In Data Science for COVID-19, *Academic Press*, pp. 451-466, 2021.

[19] Xiao J, Wang J, Cao S & Li B, "Application of a Novel and Improved VGG-19 Network in the Detection of Workers Wearing Masks," In Journal of Physics: Conference Series, *IOP Publishing*, vol. 1518, no. 1, pp. 012041, 2020.

[20] Ramzan F, Khan M. U. G, Rehmat A, Iqbal S, Saba T, Rehman A & Mehmood Z, "A Deep Learning Approach for Automated Diagnosis and Multi-Class Classification of Alzheimer's Disease Stages Using Resting-State FMRI and Residual Neural Networks," *Journal of medical systems*, vol. 44, no. 2, pp. 1-16, 2020.

[21] Gao M, Chen J, Mu H & Qi D, "A Transfer Residual Neural Network Based on ResNet-34 for Detection of Wood Knot Defects," *Forests*, vol. 12, no. 2, pp. 212, 2021.

[22] Tsochatzidis L, Costaridou L & Pratikakis I, "Deep Learning for Breast Cancer Diagnosis from Mammograms—A Comparative Study," *Journal of Imaging*, vol. 5, no. 3, pp. 37, 2019.

[23] Thawkar S, "A Hybrid Model Using Teaching–Learning-Based Optimization and Salp Swarm Algorithm for Feature Selection and Classification in Digital Mammography," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1-16, 2021.

[24] Jafar A & Myungho L, "Hyperparameter Optimization for Deep Residual Learning in Image Classification," In 2020 *IEEE International Conference on Autonomic Computing and Self-Organizing Systems Companion (ACSOS-C), IEEE*, pp. 24-29, 2020.