An Optimized Deep Learning Techniques for Analyzing Mammograms

Satish Babu Bandaru[1†], Natarajasivan. D[2], Rama Mohan Babu. G[3]
researchbsbabu@gmail.com, natarajasivan@gmail.com rmbgatram@gmail.com
[1] Research Scholar, Department of Computer Science and Engineering, Annamalai University, Annamalinagar, Tamil Nadu. India
[2]Assistant Professor, Department of Computer Science and Engineering, Faculty of Computer Science and Engineering Annamalai University, Annamalinagar, Tamil Nadu. India
[3]Professor, Department of Computer Science and Engineering (AI & ML), RVR & JC College of Engineering, Guntur, A.P. India

Abstract
Breast cancer screening makes extensive utilization of mammography. Even so, there has been a lot of debate with regards to this application's starting age as well as screening interval. The deep learning technique of transfer learning is employed for transferring the knowledge learnt 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 taking into account the problem domain knowledge. Normally, this technique will consume a lot of time as well as computational resources. This work evaluated the efficacy of the deep learning neural network like Visual Geometry Group Network (VGG Net) Residual Network (Res Net), as well as inception network for classifying the mammograms. This work proposed optimization of ResNet with Teaching Learning Based Optimization (TLBO) algorithm's in order to predict breast cancers by means of mammogram images. The proposed TLBO-ResNet, an optimized ResNet with faster convergence ability when compared with other evolutionary methods for mammogram classification.

Keywords
Mammography, Convolutional Neural Networks (CNNs), Visual Geometry Group Network (VGG Net), Residual Network (Res Net), Teaching Learning Based Optimization Algorithm (TLBO)

1. Introduction

Among women, the commonest cancer type is breast cancer. Despite of significant advancements in therapy, breast cancer continues to be the chief cause of cancer-related mortalities, and globally accounts for around 500,000 deaths per year. For minimization of breast cancer-related mortality, the most efficient approaches are the population-based breast cancer screening programmes with utilization of mammography. Nevertheless, the existing screening programmes are extremely labor intensive because of the huge number of women to be screened for per detected cancer, that in turn also results in extra economical costs. Furthermore, in spite of 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 so as to facilitate the of current screening programmers' continuation. Moreover, the prevention of either overlooking or misinterpreting the visible lesions in Digital Mammography (DM) is critical.

Computer-Aided Detection (CAD) automates the process of marking possible suspicious regions in the mammograms for the review of radiologists. Various reports had emphasized that this in turn would boost the mammographic sensitivity. Although CAD has been extensively introduced in clinics, radiologists have raised complaints of this technology's high false-positive rate. Moreover, multiple latest investigations have stated on the CAD's apparent inability to enhance the mammography's accuracy of diagnosis. However, this was expected to some degree. Majority of the learning algorithms like the CAD [2] are based on pre-defined hand-crafted features. Due to this, these algorithms are task-specific as well as a-priori knowledge based, that in turn will cause a huge bias towards how human beings consider the performance of a task. Deep learning, a promising classifier on image analysis, due to its feature learning as well as transfer learning ability. In lieu of utilizing advanced level image processing techniques for the extraction of handcrafted features, deep learning will simply transfer the feature activation maps for the subsequent convolution layer. These features are learnt in supervised ways 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 to new models to reduce the training time. Transfer learning [4] is extensively utilized in training new models which have a huge number of unrelated instances to the new training model. Migration of

Manuscript received July 5, 2023
Manuscript revised July 20, 2023
https://doi.org/10.22937/IJCSNS.2023.23.7.6

parameters to a new model can minimize the training time, and also achieve a higher accuracy with the usage of limited target data for training the models. Transfer learning will use its benefits for resolution of the aforementioned problems in deep learning.

Convolutional Neural Networks (CNNs) have the ability of learning hierarchical interpretations from images, and also to transmit knowledge which is implanted in the pre-trained model's weights to the new images. The low-level features like curves and edges are extracted using low-level convolutional layer, and the later layer's operation is able to learn more abstract representations for a diverse field of applications. As a result, for a new model, the lower-level representations is transferred while only the higher-level representations have to be learnt. Fine-tuning will refer to the procedure of updating the higher hidden layers' weights This procedure's success is dependent 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, and majority of the existing pre-trained architectures have been trained based on natural image data. As an example, for speech recognition, in spite of the huge gaps between various languages, certain studies have been able to showcase the effectivity of employing pre-trained models to speech recognition tasks in diverse languages.

Nowadays, most of the research studies have applied deep learning on 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 breast as well as 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, for optimization of a model's performance in the current applied domain, the model architecture must be rebuilt as well as updated together with the parameter tuning. Meta heuristic methods are used to optimize the parameters. In this work, the TLBO optimized ResNet is proposed for detection of cancer in mammography. The rest of the paper is organized into sections of related works, methodology, experimental results and conclusion.

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 as well as DenseNet-II neural network model, and then, analysis as well as comparison of the experimental outcomes were performed. It was evident from the experimental outcomes that, in comparison to other network models, the proposed deep learner had better classification performance. The model was able to arrive at 94.55% average accuracy that significantly enhanced the accuracy of the mammogram images' malignant and benign classification. At the same time, the outcomes were able to confirm the model's robustness as well as good generalization ability.

Agarwal et al., [8] had devised a fully automated framework for the detection of 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 the screening of breast cancers.

Chakravarthy & Rajaguru [9] had presented a novel customized method which incorporated the deep learning concept with the Extreme Learning Machine (ELM), and then optimized it with Crow-Search Algorithm (CSA). Hence, in order to boost the technological advanced operations, proposal of an enhanced deep feature-based CSA optimized ELM was offered to address the problems in healthcare. At first, the work was primarily focussed on detection of the input mammograms as either abnormal or normal. Afterwards, it was involved with further classification of the abnormal severities into either malignant type or benign type. Eventually, comparison studies were done of 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 performed optimized region growing segmentation using a hybrid metaheuristic algorithm called the Firefly updated Chicken based CSO (FC-CSO). This 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 as well as 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 output would yield the overall diagnostic accuracy, which also surpassed the standard models' performance.

Reenadevi et al., [11] had developed an optimized Wrapper-based Chaotic CSA (WCCSA) for improving the breast cancer diagnosis. The proposed WCCSA was utilized along with the Probabilistic Neural Network (PNN) for identification of the mammogram images as either normal, benign or malignant, or thus, aid the patients in receiving feasible care as well as treatment at the earlier stages. A mini-MIAS dataset of 322 images was used to assess the WCCSA with PNN method's effectivity. Then, its performance was compared with that of other machine learning algorithms. The assessment's outcome had showed that the proposed WCCSA with PNN method would seek an optimal feature subset whilst retaining its stability, and also had accomplished 97% accuracy. Ashok et al., [12] had 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 give due consideration extracting both local as well as global features. The duly considered global features were inclusive of shape features. A combination of the local as well as the global features were fed into the SVM classifier for differentiation of the breast mass's nature as either malignant or benign. By means of 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 as well as either equivalent or dissimilar objective function values. This GSO feature was feasible for optimization of the feature vector obtained from multiscale feature extraction technique. However, the unconditioned output matrix of the ELM classifier's hidden stage had resulted in the issue of poor generalization performance. For this issue's resolution, the optimization algorithms were employed due to their abilities of global search. Melekoodappattu et al., [14] had 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 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 had the ability to identify tumors as well as calcifications with 99.15% accuracy.

3. Methodology

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 which 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 was used to perform deep learning in domain with the small-sized datasets. In transfer learning, a CNN model would extract the feature from the specified data. Later on, this feature would get transmitted for the resolution of associated tasks, inclusive of 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, that is, the ImageNet, was a pre-trained model for object identification as well as classification. For the purpose of transfer learning, the deep learning model's selection was reliant on the model's capability to extract the domain's associated features. This section will offer discussions on the VGG network, ResNet, Inception network, TLBO with ResNet methods. In all these methods, the mammogram is given as input and the output is 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 utilization of an extremely small convolution in the network. Despite its simplistic nature, the proposed model's most key distinction from the earlier models was its extensive utilization due to the model having more in-depth structure, and along with double or triple convolution layers. The earlier architectures were characterized by the layers of sharing as well as 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 [15]. Figure 1 shows the architecture of the VGG-16

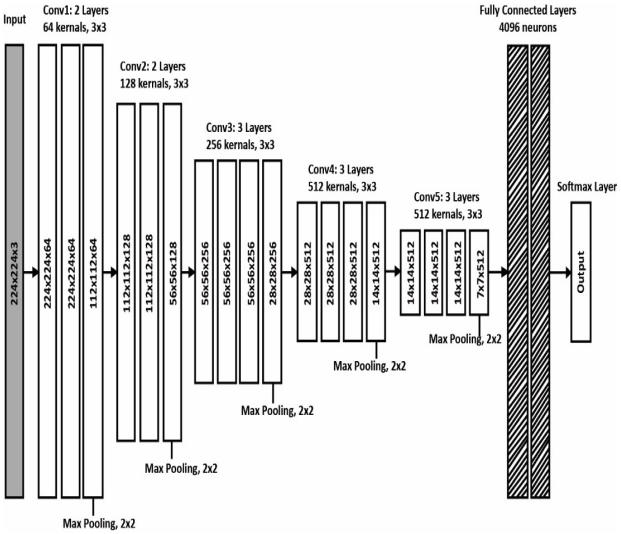

Figure 1 Architecture of the VGG-16

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

This structure is able to simulate a bigger filter whilst retaining the merits of smaller sized filters. Unlike the older models, the VGGNet has been able to demonstrate better performance with fewer parameters. 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 overall cost function as below:

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

For the above equation, Kw, b $((i))$ will denote the neural network model, $W_{ji}^{(l)}$ will denote the connection weight between the jth element of layer l+1, and the ith 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 is able to do the following: to prevent over-fitting, to significantly minimize the weight, and also to adjust the relative importance of the two terms prior to and after the cost function, λ.

3.2 VGG-19 Network Learning

The VGG CNN is made up of 6 key structures [16], 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.

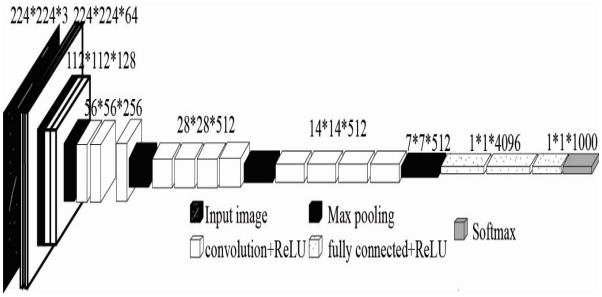

Figure 2 VGG-19 network model

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 are inclusive of better image feature extraction, downsampling by means of max pooling, and also modification of the linear unit (ReLU) as the activation function, in other words, to pick the area of 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 whilst preserving the sample's major features as well as minimizing the number of parameters. This down sampling layer can be expressed as the below-mentioned Eq. (2).

For this equation, $down\left(\chi_j^{(n-1)}\right)$ 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 $f\left(\tau_j^n down\left(\chi_j^{(n-1)}\right)+b_j^{(n)}\right)$ will denote the ReLU activation function.

$$\chi_{p_j}^{(n)}=f\left(\tau_j^n down\left(\chi_j^{(n-1)}\right)+b_j^{(n)}\right) \quad (2)$$

3.3 ResNet-18 Network Learning

The original ResNet-18's architecture is depicted in Figure 3. This architecture's network has eighteen layers in total, i.e., seventeen convolutional layers, one fully-connected layer as well as an extra softmax layer to execute the classification. The convolutional layers will employ 3 × 3 filters. Moreover, it is designed to have similar number of filters in the layers if the output feature map is of similar size. If the output feature map is halved, then the filters will get doubled in the layers. Convolutional layers with a stride of 2 will carry out the down sampling. An average-pooling is included after the convolution layer which is followed by a fully connected layer and softmax layer. Residual shortcut connections are introduced between the layers all across the network [17]. These connections are of two distinct types. The first connection type, indicated as solid lines, are employed when both the input as well as the output have similar dimensions. The second connection type, indicated as dotted lines, are employed used 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.

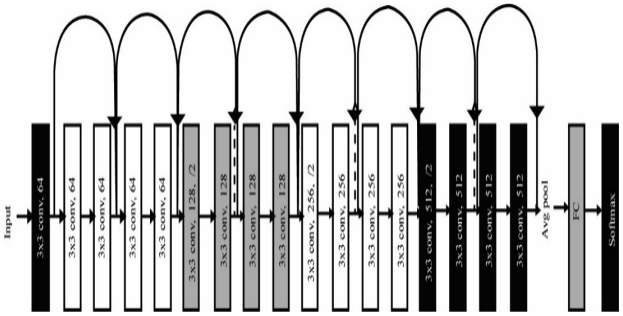

Figure 3 Architecture of ResNet-18

The following two distinct strategies have been employed in these network training experimentations. At first, there is 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 be done by minimizing the number of input channels to merely a single channel.

Secondly, the weight initialization has been carried out with a pre-trained network, and then, there has been execution of the transfer learning. As the trained model is applied for mammogram classification, it will carry out the task by adapting the network. It will employ the following two approaches to conduct the transfer learning for the transmission of knowledge from a pre-trained network. The first approach will perform Off-The-Shelf (OTS) transfer learning through replacement of the original network's last dense layer with the new dense layer in order to match the number of classes for the task. This OTS approach will use all the network layers, with 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 layers 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 given dataset. The starting point of both these transfer learning approaches will be the weights of ResNet-18 network trained on ImageNet.

3.4 ResNet-34 Network Learning

The ResNet-34 network's infrastructure is the residual building block, and also it is 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 the construction of more flexible CNN structures, and also will boost the wood knot defects' recognition rate [18].

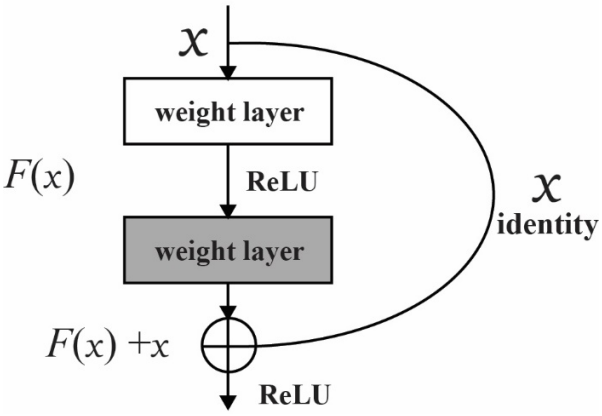

Figure 4 Building blocks of ResNet-34.

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 as well as a shortcut. Eq. (3) will formulate the residual building block's output as below:

$$y = F(x) + x \tag{3}$$

In this equation, F will denote the residual function, x will denote the residual function's input while y will denote the residual function's output. The overall residual network's constituents will be inclusive of the first convolutional layer as well as multiple basic-blocks.

The ResNet-34 will constitute 33 convolutional layers, a 3 * 3 size max-pooling layer, an average pool layer as well as a fully

connected layer. In a standard ResNet-34 model, the back of all the convolution layers in the "Basic Block" block are applied with the ReLU activation as well as BN, and the final layer is applied with the softmax function. Table 1 will offer an illustration of the ResNet-34's architecture.

Table 1 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 [19] is the first implementation done with the. This module's key concept is on the basis of the authors' findings on using dense components to approximate a local sparse structure. It was the intention of the authors to build a multi-layer network by seeking the optimal local structure, and then repeating it. The Inception module has four distinct branches, where each branch is given 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 kernelled convolutions 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.

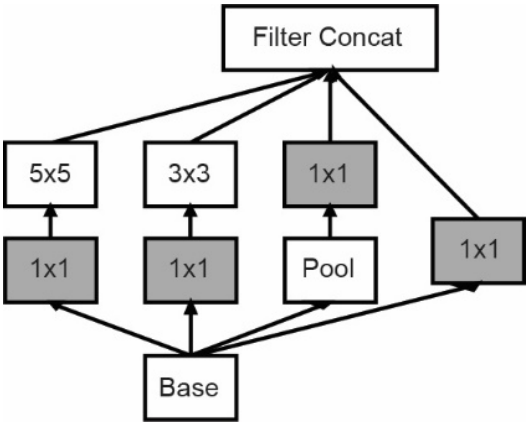

Figure 5 Inception module of GoogLeNet

GoogLeNet's construction will involve the stacking of nine Inception modules. For minimizing 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 the GoogLeNet. The authors assumed that a CNN's middle layers must yield discriminative features so as to append simple classifiers (two fully connected layers as well as a softmax layer) which will be applied on generated features of an intermediate point of the network. These classifiers' decisions will evaluate the loss. Late on, this loss will get utilized in the back propagation to evaluate additional gradients which will provide to training 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, number of epoch 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 to be a well-educated individual who will share their knowledge with the students. With the aid of a good teacher, the class will often have improvements in the learner's marks or grades. Moreover, the learner will learn information from other members of their class in order to improve their marks. The learners will represent the solutions in this method. In the proposed method, the solutions are represented with values of batch size, learning rate, activation and number of epochs for Resnet18 and 34. The current population's learner with the best fitness value will get picked as a classroom's teacher. The teacher stage, and the learner stage 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 [20].
In the teacher stage, there is 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, randi will denote a random variable which has been evenly distributed in the [0, 1] range. TF will denote the teaching factor's value. This value can be either 1 or 2, on the basis of random selection. $L_{new,i}$ will denote ith new population. The Eq. (1) as well as 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: Li and Lj.
• If the fitness of f(Li) is lower than that of f(Lj), then there is 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 utilization of Eq. (7) to update the population:

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

The TLBO algorithm's implementation has been depicted in algorithm 1.

1. 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 pf initial population using 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 L1 and L2
17. If f(L1) < f(L2)
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 is
The multi-layers-based CNN model suffers from the problem of vanishing gradient wherein the model's learning in the initial layers will get hindered. This problem will result in a degradation of the model's performance: there will be a decrease in the accuracy as well as 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 [21]. 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 as well as learning phases of the model and optimization algorithms in order to demonstrate the incorporation of performance and accuracy. Instead of the classical augmentation method, it will employ the data augmentation technique which is utilized on the technologically advanced residual learning. It is able to boost the TLBO algorithm's error rate as well as the training time via updating and picking accurate teaching and learning phases.
Initially, the same class's images will get arranged into a list. Later on, the order will get changed such that the least highlighted images are placed in the bottom while the most highlighted features are placed at the top. There is evaluation of the similarity of images within the class as the degree of the representative images which will correspond to its class. The below Eq. (8) will offer the expression of the representation degree of images:

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

In the second step, the training set will get appended with the top-selected images having more similarity within each class. Also, there is utilization of the BN after each convolutional layer as well as prior to the activation function (ReLU). Through application of the proposed method, suitable of batch size, learning rate, activation and number of epochs of the ResNet network is obtained, it is able to arrive at the

improvement stage in the final training as well as the validation accuracy. Furthermore, this method is able to overcome the over-fitting problem, and also minimize the training's computational complexity.

4. Results and Discussion

In this section, the VGG16, VGG19, Resnet18, Resnet34, inception, TLBO-Resnet18 and TLBO-Resnet34 methods are evaluated for classifying mammograms. The Curated Breast Imaging Subset of Digital Database for Screening Mammography (CBIS-DDSM) is used for evaluating the 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. In this work, 550 Benign and 625 Malignant mammogram images are used for evaluation. Python, open CV, tensor flow, keras are used for the implementation of the algorithms.
Table 2 shows the summary of results. The classification accuracy, recall and precision as shown in figures 6 to 8.

Table 2 Summary of Results

Techniques	VGG16	VGG19	Resnet18	Resnet34	Inception	TLBO Resnet18	TLBO Resnet34
Classification Accuracy	88.51	89.19	89.53	90.21	93.02	93.28	94.47
Recall for Benign	0.8964	0.9018	0.90055	0.91009	0.93445	0.94	0.9491
Recall for Malignant	0.8752	0.8832	0.88064	0.89044	0.92664	0.92664	0.9408
Precision for Benign	0.8634	0.8717	0.8752	0.8836	0.9179	0.9183	0.9338
Precision for Malignant	0.9056	0.9109	0.9142	0.9194	0.9415	0.9461	0.9545

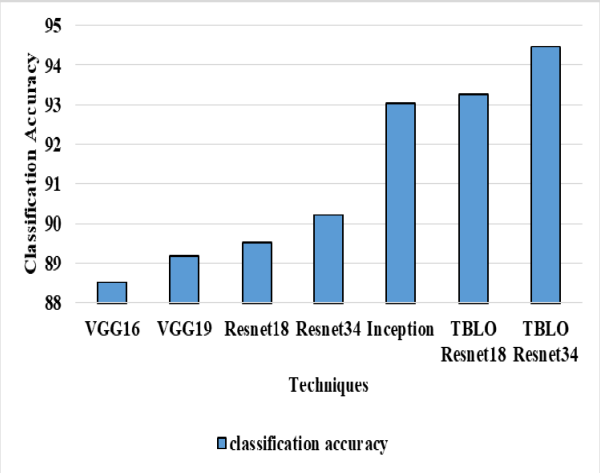

Figure 6 Classification Accuracy for TLBO Resnet34

From the figure 6, it can be observed that the TLBO Resnet34 has higher classification accuracy by 6.06% for VGG16, by 5.29% for VGG19, by 4.91% for Resnet18, by 4.16% for Resnet34, by 1.55% for inception and by 1.27% for TLBO Resnet18 respectively.

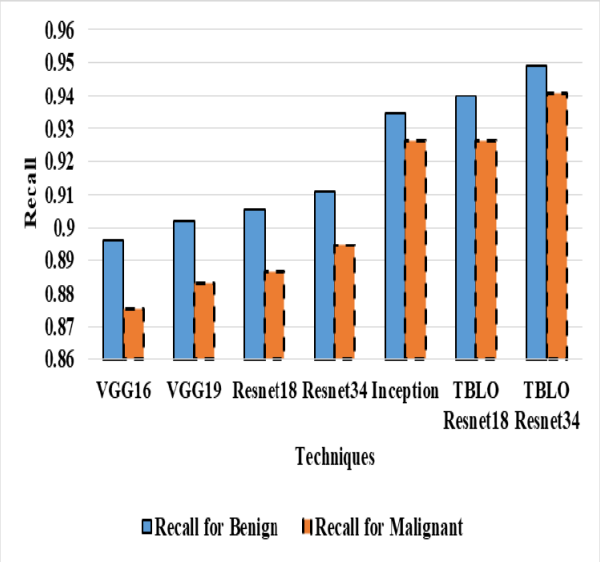

Figure 7 Recall for TLBO Resnet34

From the figure 7, it can be observed that the TLBO Resnet34 has higher recall by 6.66% for VGG16, by 6.06% for VGG19, by 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 higher recall by 5.51% for VGG16, by 4.6% for VGG19, by 4.24% for Resnet18, by 3.34% for Resnet34, by 1.54% for inception and by 1.54% for TLBO Resnet18 respectively for malignant.

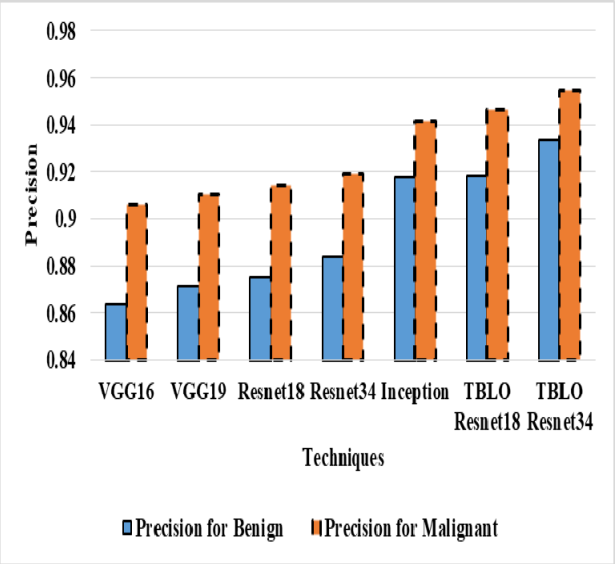

Figure 8 Precision for TLBO Resnet34

From the figure 8, it can be observed that the TLBO Resnet34 has higher precision by 6.14% for VGG16, by 5.18% for VGG19, by 4.78% for Resnet18, by 3.83% for Resnet34, by 1.72% for inception and by 1.67% for TLBO Resnet18 respectively for benign. The TLBO Resnet34 has higher precision by 6.01% for VGG16, by 5.42% for VGG19, by 5.06% for Resnet18, by 4.49% for Resnet34, by 1.37% for inception and by 0.88% for TLBO Resnet18 respectively for malignant.

5. Conclusion

For the application of deep CNNs to medical imaging tasks, transfer learning is a vital step. This work has presented the implementation of a computer aided classification approach for mammogram-based breast cancer prediction via an evolutionary system by means of integrating architectural evolution with deep neural network learning using ResNet (18 & 34) and TLBO is implemented. Of late, the TLBO is being touted as the latest, accurate, reliable as well as robust technique for the global optimization across continuous spaces. Researchers have also put forward some variations of the TLBO so as to boost the basic TLBO algorithm's performance. This technique will involve the proposal of diverse variations as well as its hybridization with the evolutionary algorithm. As a result, this technique will have an enhancement in its performance, and also will offer better prediction of breast cancers (as either non-cancerous or cancerous) through utilization of the TLBO's global search capability as well as the ResNet's (18 & 34) local search capability. Furthermore, there is proposal of the latest direct encoding strategy that will the 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, an almost standard TLBO algorithm can be employed for comparing as well as combining the variable length ResNet (18 & 34) architectures.

References

[1] Rodriguez-Ruiz, A., Lång, K., Gubern-Merida, A., Broeders, M., Gennaro, G., Clauser, P., ... & Sechopoulos, I. (2019). Stand-alone artificial intelligence for breast cancer detection in mammography: comparison with 101 radiologists. JNCI: Journal of the National Cancer Institute, 111(9), 916-922.

[2] Kim, E. K., Kim, H. E., Han, K., Kang, B. J., Sohn, Y. M., Woo, O. H., & Lee, C. W. (2018). Applying data-driven imaging biomarker in mammography for breast cancer screening: preliminary study. Scientific reports, 8(1), 1-8.

[3] Altan, G. (2020). Deep Learning-based Mammogram Classification for Breast Cancer. International Journal of Intelligent Systems and Applications in Engineering, 8(4), 171-176.

[4] Chougrad, H., Zouaki, H., & Alheyane, O. (2018). Deep convolutional neural networks for breast cancer screening. Computer methods and programs in biomedicine, 157, 19-30.

[5] Zhou, J., Yang, X., Zhang, L., Shao, S., & Bian, G. (2020). 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. (2020). A deep-learning-based framework for automated diagnosis of COVID-19 using X-ray images. Information, 11(9), 1-13.

[7] Li, H., Zhuang, S., Li, D. A., Zhao, J., & Ma, Y. (2019). Benign and malignant classification of mammogram images based on deep learning. Biomedical Signal Processing and Control, 51, 347-354.

[8] Agarwal, R., Díaz, O., Yap, M. H., Llado, X., & Marti, R. (2020). Deep learning for mass detection in Full Field Digital Mammograms. Computers in biology and medicine, 121, 103774.

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

[10] Patil, R. S., & Biradar, N. (2020). Automated mammogram breast cancer detection using the optimized combination of convolutional and

recurrent neural network. Evolutionary Intelligence, 1-16.

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

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

[13] Ashok, A., Vijayan, D., & Lavanya, R. (2021, June). 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) (pp. 1082-1087). IEEE.

[14] Melekoodappattu, J. G., Subbian, P. S., & Queen, M. F. (2021). Detection and classification of breast cancer from digital mammograms using hybrid extreme learning machine classifier. International Journal of Imaging Systems and Technology, 31(2), 909-920.

[15] Sahinbas, K., & Catak, F. O. (2021). Transfer learning-based convolutional neural network for COVID-19 detection with X-ray images. In Data Science for COVID-19 (pp. 451-466). Academic Press.

[16] Xiao, J., Wang, J., Cao, S., & Li, B. (2020, April). Application of a novel and improved VGG-19 network in the detection of workers wearing masks. In Journal of Physics: Conference Series (Vol. 1518, No. 1, p. 012041). IOP Publishing.

[17] Ramzan, F., Khan, M. U. G., Rehmat, A., Iqbal, S., Saba, T., Rehman, A., & Mehmood, Z. (2020). 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, 44(2), 1-16.

[18] Gao, M., Chen, J., Mu, H., & Qi, D. (2021). A Transfer Residual Neural Network Based on ResNet-34 for Detection of Wood Knot Defects. Forests, 12(2), 212.

[19] Tsochatzidis, L., Costaridou, L., & Pratikakis, I. (2019). Deep learning for breast cancer diagnosis from mammograms—a comparative study. Journal of Imaging, 5(3), 37.

[20] Thawkar, S. (2021). 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, 1-16.

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