Breast Cancer Detection and Classification using Deeper Convolutional Neural Networks based on Wavelet Packet Decomposition Techniques

Rajakumari R (✉ rajakumarinatarajan@gmail.com)

National Engineering College <https://orcid.org/0000-0002-7810-4737>

L. Kalaivani

National Engineering College

Research Article

Keywords: Deep Learning, Breast cancer, DCNN, GoogleNet, AlexNet, Image Classification

Posted Date: April 13th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-405990/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Breast Cancer Detection and Classification using Deeper Convolutional Neural Networks based on Wavelet Packet Decomposition Techniques

R. Rajakumari^{1*}, L. Kalaivani²

¹Dept. of CSE, National Engineering College, Kovilpatti, Tamilnadu, India

²Dept. of EEE, National Engineering College, Kovilpatti, Tamilnadu, India

{^{1*} rajakumarinatarajan@gmail.com, ² anuprakad@gmail.com,}

ABSTRACT

Breast cancer is a commonly diagnosed disease in women. Early detection, a personalized treatment approach, and better understanding are necessary for cancer patients to survive. In this work, both Deep learning Network and traditional Convolution Network has been employed for Digital Database for Screening Mammography (DDSM) dataset. In this work, breast cancer images are subjected to removal of background images followed by Weiner filtering and Contrast Limited Histogram Equalization (CLAHE) filter for image restoration. Wavelet Packet Decomposition (WPD) using Daubechies wavelet level 3 (db3) is employed to improve the smoothness of the images. In the first part of breast cancer recognition, these preprocessed images are fed to a deep convolution neural network, namely GoogleNet and AlexNetfor ADAM. RMSPROP and SGDM optimizers for different learning rates such as 0.01, 0.001, and 0.0001. As medical image necessitates discriminative features for classification, the pre-trained GoogleNet architectures extract the complicated features from the image and increase the recognition rate. In the latter part of this paper, Particle Swarm Optimization based Multi-Layer Perceptron (PSO-MLP), and Ant Colony Optimization based Multi-Layer Perceptron (ACO-MLP) are employed for breast cancer recognition using statistical features like skewness, kurtosis, variance, entropy, contrast, correlation, energy, homogeneity, mean which are extracted from the preprocessed image. The performance of this GoogleNet has been compared with AlexNet, PSO-MLP, and ACO-MLP in terms of accuracy, loss rate, and runtime and achieves an accuracy of 99%, with less loss rate of 0.1547 and the lowest run time of 4.14 minutes.

Keywords: *Deep Learning, Breast cancer, DCNN, GoogleNet, AlexNet, Image Classification*

1. Introduction

Breast cancer is the commonly occurring cancer, especially in women than other cancer types like lung & bronchus, colon & rectum, and uterine corpus in women [1, 2]. Even with a lot of advancement in diagnosis and treatment methods a large number of annual mortalities occurred [2]. Long-term mortality mainly resulted due to the size of the tumor [3] and the detection of breast cancer less than two cm is essential [4]. Identifying and treating breast cancer at an early stage is needed to reduce the death rate [5]. Mammography, which uses low energy electromagnetic waves, is used to image breast cancer and ranked as the most efficient method for detection at an incipient stage. Hence, an accurate computer-aided classifier is needed to aid medical professionals in the effective recognition of normal, benign, and malignant breast cancer to reduce the wrong verdict due to asthenopia, lethargy, or lack of skillset.

In the past years, a lot of breast cancer recognition was developed, and the performance of the computer aided recognition had been verified for the dataset taken from the UCI machine-learning repository. The recognition rate of breast cancer using the optimized learning vector quantization (optimized-LVQ) method, bigLVQ method, and artificial immune system were 96.7%, 96.8%, and 97.2%, respectively [6]. The classification rate had reached 94.74% accuracy for the decision tree method with C4.5 using a 10-fold cross [7]. In fuzzy clustering method using supervisory control yielded 95.57% and Neuro-Fuzzy based classifier achieved a 95.06% accuracy rate [8]. The breast cancer classification accuracy was compared for Wisconsin Breast Cancer (WBC) datasets for cancer using three classifiers, namely Multi-Layer Perceptron (MLP), Naïve Bayes (NB), and decision tree. The fusion classifier made of MLP and J48 classifier found to be superior [9].

96% recognition rate was achieved for breast cancer classification with RIAC [10], and around 98.53% was achieved for Least Square Support Vector Machines(LSSVM) [11]. The performance of LSSVM was compared in terms of accuracy sensitivity specificity, and the high accuracy rate is due to the robustness nature of LSSVM. The feature reduction-based breast cancer decision system was explored in independent component analysis, and the substantiated single dimension feature was sufficient to increase the radial basis Neural network accuracy up to 90.49 [12]. BPA based rough set method incorporated two phases, with the first phase used to

handle missing values to smoothen the dataset, and the second phase is used for correctly selecting the appropriate attributes from the obscure clinical dataset. Particle Swarm Optimization (PSO) based wavelet network was employed for abnormalities in breast cancer using texture energy measured from mammograms and achieved 93.671%, 92.105%, and 94.167% for accuracy, specificity, and sensitivity, respectively [13]. Detection of breast cancer using deep belief network unsupervised path with Liebenberg Marquardt learning produced an accuracy of 99.68% for WBC Dataset [14].

A knowledge-based system using fuzzy logic was used for breast cancer recognition by employing maximization for clustering the data, and the problem of multicollinearity was avoided using principal component analysis [15]. Extreme Learning Machine classifier classified the benign and malignant breast masses using fused deep features, morphological features, texture features, and density features and achieved an 86.50% recognition rate.

In the first part of the proposed work, Digital Database for Screening Mammography (DDSM) dataset images is used for research, and these images are subjected to removal of noise and image restoration using Weiner filter, CLAHE filter, and Wavelet tree decomposition before it is subjected to feature segmentation. In the first part, the reconstructed images are then fed to Deep Convolutional Neural Network (DCNN) based GoogleNet model, which has an inception approach for feature extraction at the training and testing phase. The performance of GoogleNet has been compared with AlexNet.

In the Second Part, the statistical features of images like skewness, kurtosis, variance, entropy, contrast, correlation, energy, homogeneity, mean are extracted from the preprocessed image and fed as input to PSO-MLP and ACO-MLP for breast cancer recognition.

Section II explains the preprocessing of breast cancer images using the Weiner filter, CLAHE filter, and Wavelet Packet Decomposition. Section III covers the proposed methodology, classifier, and initialization Parameters. Section IV discusses the results and comparison of different classifiers, and Section V concludes the impact of this work.

II. Preprocessing of Breast cancer Images

Digital Database for Screening Mammography (DDSM) dataset images is taken for breast cancer recognition work. This database contains normal, malignant, and benign cancer images with 650 pictures for each category. DDSM images are preprocessed to remove the background, noise, and augment the contrast between the cancer cell and neighborhood areas, which aid in localizing the Region of Interest (ROI). Feature extraction, training, and testing images are fed to DCNN based GoogleNet classifier.

2.1 Preprocessing of DDSM Images

There are so many preprocessing methods are reported in [16-18]. The preprocessing approach adopted in this work is shown in Fig. 1.

Fig. 1 Block diagram of Breast cancer image preprocessing.

The original DDSM images are converted into the binary image shown in Fig. 2 (b), and then to increase the intensity of breast and muscles and suppress the other unwanted information, the binary image is multiplied with the original breast image is shown in Fig. 2 (c). This process is done by increasing and decreasing the luminosity.

Fig. 2 Preprocessing of breast cancer image for breast and muscle identification
(a) Original image (b) binary image (c) identification of breast and muscles

The background is deleted in the image, as shown in Fig. 3 (a), by deleting the zero intensity pixels rows and columns. The gray thresholding based on Ostu's method is applied to the processed image, and the intensity of the image is varied in mid half between minimum and maximum of the original intensity, and resultant muscle removed image is shown in Fig. 3 (b). Then the breast image fed to the Wiener filter to eliminate noise present in the image. Once the spectral power is estimated, a mask is applied to all pixels of the original image with a signal to noise ratio (SNR) of 0.2. By using the mean and variance of the authentic images, new mean, variance, and power are calculated for all pixels of the transformed images.

2.2 Weiner and CLAHE Filtering

Weiner filter has the characteristics of the trade-off between inverse filtering and smoothening of noise, and hence it has the simultaneous capability of removing additive noise and blur inversion. The stochastic nature of the wiener filter estimates the normalized image in a linear manner based on the orthogonality expressed in Equation (1).

$$W(f_1, f_2) = \frac{blur*(f_1, f_2)power_{ori}(f_1, f_2)}{|blur(f_1, f_2)|^2 power_{ori}(f_1, f_2) + power_{noise}(f_1, f_2)} \quad (1)$$

where $blur(f_1, f_2)$ is the blurring filter, $power_{ori}(f_1, f_2)$, and $power_{noise}(f_1, f_2)$ are spectral power of normalized image and noise, respectively. It is evident that the wiener filter performs deconvolution in a high pass filter and compression along with noise removal in a low pass filter. For the purpose of contrast enhancement, CLAHE filtering is employed as it has the

suppleness of choosing a local mapping function for a histogram. CLAHE computes the histogram for the local window for the given pixel and boosts the contrast. Fig. 3 (c) and (d) show the breast cancer images after the Weiner filter and CLAHE filter, respectively.

Fig. 3 Breast cancer image before and after preprocessing
(a) background removed image (b) muscle removed and breast alone present (c) Weiner filter applied image (d) image after applying CLAHE filter

2.3 Wavelet Packet Decomposition

Non-stationary noises and noise in high frequency are eliminated using wavelet packet decomposition. It uses windows of different lengths and efficient in denoising images, preserving the edges and texture of the image. In Two Dimensional (2D) WPD, the breast cancer DDSM image is decomposed into sub-images containing details about edges in horizontal, perpendicular, and transverse orientations. The 2D WPD generates one approximation coefficient and one detail coefficient in horizontal, vertical, and diagonal orientation with a total of four components for each level.

Equation (2) represents the 2D WPD applied to the $I(m, n)$, representing the breast cancer image of size $m \times n$.

$$WPD_n(i, j, k) = \frac{1}{\sqrt{i}} \iint I(m, n) \varphi_m\left(\frac{m-j}{i}\right) \varphi_n\left(\frac{n-k}{i}\right) dm dn \quad (2)$$

where, $\varphi_m\left(\frac{m-j}{i}\right)$, $\varphi_n\left(\frac{n-k}{i}\right)$ represents the wavelet decomposition in m and n direction, respectively, i represent the scaling parameter, and it is always greater than zero, j and k represent translation parameter in x and y direction, respectively. Equations (3), (4), (5), and (6) represent the approximation coefficient, the detailed coefficient in the horizontal orientation, the detailed coefficient in perpendicular orientation, and the detailed coefficient in a transverse orientation.

$$WPD_{approximation}(u, m, n) = \frac{1}{\sqrt{mn}} \sum_{x=0}^{m-1} \sum_{y=0}^{n-1} I(x, y) \varphi_{u,m,n}^{approx}(x, y) \quad (3)$$

$$WPD_{detail\ horizontal}(u, m, n) = \frac{1}{\sqrt{mn}} \sum_{x=0}^{m-1} \sum_{y=0}^{n-1} I(x, y) \varphi_{u,m,n}^{hori}(x, y) \quad (4)$$

$$WPD_{detail\ perpendicular}(u, m, n) = \frac{1}{\sqrt{mn}} \sum_{x=0}^{m-1} \sum_{y=0}^{n-1} I(x, y) \varphi_{u,m,n}^{perp}(x, y) \quad (5)$$

$$WPD_{detail\ transverse}(u, m, n) = \frac{1}{\sqrt{mn}} \sum_{x=0}^{m-1} \sum_{y=0}^{n-1} I(x, y) \varphi_{u,m,n}^{trans}(x, y) \quad (6)$$

Equation 3 to 6 generates the coefficient for the sub-band with one coefficient in low frequency in both horizontal and perpendicular region (L-H, L-H) and three coefficients at high frequency, namely, low frequency in horizontal and high frequency in the perpendicular region (L-H, H-P), high frequency in horizontal and low frequency in the perpendicular region (H-H, L-P) and high frequency in both horizontal and perpendicular region called as transverse orientation (H-H, H-P). Fig. 4 represents the Wavelet packet decomposition of the image and its hierarchy for level 2.

Fig. 4

Wavelet Packet Decomposition of Image for Level 2

In this work, WPD using Daubechies wavelet3 (db3) level 2 is employed, and wavelet tree along with reconstructed output image is shown in Fig. 5. The optimal tree is chosen by finding the cost function at each node using the Shannon entropy function. For each sub-band, if each parent's cost function is greater than the cumulative sum of all children nodes, both parent and child nodes are retained; else, retain parent node only. This process is repeated up to level 2 of decomposition, which results in the quaternary tree. In the quaternary tree, each node represents the image in WPD. Node (0, 0) is the original image and nodes (1,0), (1,1), (1,2), and (1,3) are the images after level 1 decomposition. Likewise, nodes (2,0), (2,1), (2,2), (2,3), (2,4), (2,5), (2,6), (2,7) up to (2,15) and level 2 decomposition images which is shown in Fig. 5 (a) and (b). Fig. 5 (c) and (d) shown the WPD image and wavelet packet coefficients.

(a)

(b)

(c)

(d)

Fig.5 Breast cancer image decomposition using wavelet Packets using db3
(a) Image after applying CLAHE filter (b) Quaternary tree (c) Reconstructed breast cancer image after denoising using WPD (d) Wavelet Packet Coefficients

III. Proposed Methodology

After preprocessing of images, it is fed to the classifier for breast cancer recognition. In the first part of the work, images are resized as 224×224 for GoogleNet and 227×227 for AlexNet architecture. In the second of the work, features such as skewness, kurtosis, variance, entropy, contrast, correlation, energy, homogeneity, and mean are extracted from the preprocessed images and fed as input for PSO-MLP and ACO-MLP for recognition work. The flow of the proposed work for the recognition of breast cancer images is depicted in Fig. 6.

Fig.6. Proposed Methodology for Breast Cancer Recognition

3.1 Pretrained Networks

The pre-trained deep learning Convolution Neural Networks, such as GoogleNet and AlexNet are considered in this work. In both the network, images are resized and fed as input. Training and testing are done for a 70:30 proportion of preprocessed images in both the networks.

3.1.1 GoogleNet

The GoogleNet model proposed in [19] is complex and goes deep than all other CNN architectures. It has a 22-layer DCNN with an inception module, which cascades various dimensions and sizes into a single new filter shown in Fig. 7.

Fig.7. GoogleNet Inception model

The GoogleNet architecture goes deep as it has a 22-layer DCNN with an inception module. GoogleNet has multiple features extractor in a single layer, which aids the network performance by choosing either convolute the input or pool the input directly during training as well as image classification and piling of this inception layer will result in final architecture. Each layer act as a filter and the common features are detected in starting layers and followed by discriminating features identification in final layers. The inception layer covers a large area and retains delicate resolution even for small information in an image. This property is maintained by convolving in a parallel approach with different sizes ranging from 1x1 to 5x5. Inception layers are employed with Gabor filters in series with various sizes having the learning capability. Convolution layer and pooling layers reduce the image features, which are not necessary for training, thus reducing feature dimension and storage space. To avoid overfitting, nine inception modules are used in GoogleNet. The top layer has its own output layer and both joint and parallel training, and this subtlety helps in faster convergence is in most.

In this work, GoogleNet with a total of 144 layers was retrained to classify breast cancer images by inserting four layers to its structure, namely a dropout layer with a 50% probability of dropout, a fully connected layer, a softmax layer, and a classification-output layer with three number of outputs.

3.1.2 AlexNet

AlexNet is also a pre-trained network [20] that has 11×11 , 5×5 , 3×3 , convolution, max pooling, dropout, and fully connected layers with ReLU activation functions after every convolutional and fully connected layer. The dropout layer has a dropout with a 50% probability. The fully connected layers are modified to three as there is three output classes.

3.1.3 Training Optimization

DCNN networks are trained with stochastic gradient descent with momentum (SGDM) optimization algorithm, Root Mean Square Propagation (Rmsprop) optimizer, and Adaptive Momentum (Adam) optimizer along with transfer learning as it does not necessitates too many databases for achieving high accuracy. The learning rate is used to control the weight updation based on the error value. A low learning rate may result in a long time for training may get stuck, whereas a high value may yield a nominal weight value resulting in instability. In training optimization, SGDM, ADAM, and Rmsprop updated their descent algorithm weight based on the loss function value of each iteration

3.1.3.1 Stochastic gradient descent with momentum (SGDM) optimization

SGDM optimizer swing towards optimum through steepest descent path. The magnitude of swing can be limited by momentum [21]. Equation (7) gives the weight updation in SGDM optimization method.

$$V_{j+1} = V_j - \alpha \nabla L(V_j) + \beta(V_j - V_{j-1}) \quad (7)$$

where j is the iteration, α is the learning rate, and it is greater than zero, V is the vector parameter, and $L(V)$ is the loss function. $\nabla L(V)$ is calculated from training data. β indicates the previous gradient step involved in the current iteration step. Table 2 shows the parameter used for SGDM optimizer for breast cancer recognition.

3.1.3.2 Root Mean Square Propagation (Rmsprop) optimization

Rmsprop overcomes the drawback of SGDM optimizer as it used the same learning rate for all parameters. But Rmsprop implements various learning rates for the various parameter.

Weight updation is done using Equation (9), which is normalized by Equation (8) based on loss function optimization [22].

$$X_j = \gamma X_{j-1} + (1 - \gamma)[\nabla L(V_j)]^2 \quad (8)$$

where γ is the decay rate.

$$V_{j+1} = V_j - \frac{\alpha \nabla L(V_j)}{\sqrt{X_j} + \mu} \quad (9)$$

where μ is the minimum value acting as constant to avoid Equation (9) becoming indefinite. This optimizer has the advantage of less learning rate, even for an extensive input database.

3.1.3.3 Adaptive Momentum (Adam) optimization

ADAM also follows like Rmsprop for weight updation using momentum. Equations (10) and (11) are used to find the gradient parameter and the squared gradient parameter.

$$T_j = \sigma_1 T_{j-1} + (1 - \sigma_1) \nabla L(V_j) \quad (10)$$

$$X_j = \sigma_2 X_{j-1} + (1 - \sigma_2) [\nabla L(V_j)]^2 \quad (11)$$

where σ_1 and σ_2 are decay rates from SGDM and Rmsprop optimizer training function. Equation (12) is used for weight updation in Adam optimizer by using average values of SGDM and Rmsprop.

$$V_{j+1} = V_j - \frac{\alpha T_j}{\sqrt{X_j} + \mu} \quad (12)$$

If the gradient remains constant for many iterations, then pick up momentum is used for the gradient's moving average.

3.1.4 Training of GoogleNet and AlexNet

This work is executed on the DDSM breast cancer image dataset with 1950 images with three classes: normal, benign, and malignant, with 650 images for each set. The preprocessed image size is modified to 224*224 and fed to pre-trained GoogleNet, and image resizing is 227*227 for AlexNet.

DDSM images are randomly divided for 70:30 proportion for training and testing. In the training phase, for each iteration minimum batch size used is 20 images. In this work, MiniBatchSize, MaxEpochs, Number of iteration, and InitialLearnRate are assigned as 10, 10, 100, and 0.0001, respectively. A single epoch refers to applying the training algorithm for complete training images. The training is done for three learning rates 0.01, 0.001, and 0.0001, with SGDM, ADAM, and Rmsprop as the optimizer. The simulations are carried out in MATLAB 2020b in an Intel Core Processor i7-7500U CPU at 2.70 GHz, 2904 MHz, two cores, and 64 GB of RAM.

3.2 Multilayer Perceptron

A multilayer perceptron (MLP) is one among the feed-forward artificial neural network (ANN). MLP used nonlinear activation for all nodes except the input node and implemented backpropagation learning during training with the capability to distinguish nonlinearly separable data. The optimization process involved in MLP training is finding the weights to minimize the mean square error. The backpropagation algorithms based on the gradient descent method get locked in local minima as this algorithm efficiency depends on initial weights and bias values, and numerous iterations are required to tune the learning rate. To obtain the optimal weights of MLP networks, evolutionary-based approaches like Particle swarm optimization (PSO) [23] and Ant colony algorithm [24] are implemented as less number of parameters are required to be adjusted for better convergence.

3.2.1 Particle Swarm Optimization based Multi Layer Perceptron

Particle Swarm Optimization (PSO) algorithm is an outcome of the birds' flock behavior in flying pattern proposed by Kennedy and Eberhart in 1995. Each particle of PSO is characterized by position and velocity. In PSO, swarm particle size is taken as 'm' and current instant as 't', then each particle $1 \leq u \leq m$ has the position as $y_u(t) \in \mathbb{R}^n$ with velocity $vel_u(t) \in \mathbb{R}^n$. Both position and velocity control the position and direction of the swarm in PSO. Every particle, as well as the whole swarm, holds the best position in their memory as $ybest_u(t) \in \mathbb{R}^n$ and $\widehat{ybest}(t) \in \mathbb{R}^n$, respectively. Equations (13) and (14) describe the updation of the velocity and new position of each particle.

$$vel_{uv}(t+1) = mw vel_{uv}(t) + g_1 rand_1(ybest_{uv}(t) - y_u(t)) + g_2 rand_2(\widehat{ybest}(t) - y_{uv}(t)) \quad (13)$$

$$y_{uv}(t+1) = y_{uv}(t) + vel_{uv}(t+1) \quad (14)$$

where mw is the inertia momentum is a scalar quantity used to control the degree of exploration in searching. $rand_1$ and $rand_2$ are random variables which is uniformly assigned from $V_{uv1}(0,1)$ and $V_{uv2}(0,1)$, respectively. g_1 and g_2 are fixed global acceleration coefficients whose values in the region $0 < g_1, g_2 \leq 2$ are used to control the distance traveled by a particle in a single iteration. PSO performance can be enhanced by adjusting inertia momentum using Equation (15).

$$mw = mw_{max} - \left(\frac{mw_{max} - mw_{min}}{maxigen} \right) gen \quad (15)$$

where mw_{max} and mw_{min} are maximum and minimum inertia momentum, gen is the current generation, and $maxigen$ is the maximum generation.

3.2.2 Ant Colony Optimization based Multi-Layer Perceptron

ACO algorithm is an outcome of ant foraging behavior of finding the optimum path between food source and destination. ACO can resolve the complex optimization problem when it combines with another method. Based on transition probability and pheromone quantity in that region, the pheromone trajectory is updated. The fitness value is calculated for global ants produced in each iteration. On updating pheromone, if fitness improves, then the local ant is moved to the better region, else the search is directed towards a new direction. ACO continues its local and global search by updating pheromone and evaporating it. Equation (16) express the transition probability in the region r , which is the measure of local ant capability to move in a concealed region.

$$P_r(t) = \frac{tot_r(t)}{\sum_{m=1}^n tot_m(t)} \quad (16)$$

where $tot_r(t)$ is the total pheromone in region r , and n is the number of global ants.

Equation (17) is used to update the pheromone.

$$tot_i(t + 1) = (1 - \gamma)tot_i(t) + \sum_u \Delta tot_i^u \quad (17)$$

where γ is pheromone decay rate, $\sum_u \Delta tot_i^u$ is the pheromone deposited, and it is calculated using Equation (18).

$$\Delta tot_i^u = \begin{cases} \frac{Q}{L_u}, & \text{if ant } u \text{ uses } i^{th} \text{ path} \\ \text{zero}, & \text{otherwise} \end{cases} \quad (18)$$

where L is the cost function for i^{th} path, and Q is a pheromone update constant.

3.2.3 Features Extraction and Training of PSO-MLP and ACO-MLP

The statistical features such as mean, skewness, kurtosis, variance, entropy, contrast, correlation, energy, and homogeneity are extracted from the preprocessed image and fed as input to PSO-MLP ACO-MLP for breast cancer recognition.

Mean indicates the brightness of the image calculated using Equation (19)

$$mean \mu = \frac{1}{m*n} \sum_{i=0}^m \sum_{j=0}^n p(i, j) \quad (19)$$

where $\sum \sum p(i, j)$ is the summation of pixel values, and $(m*n)$ is the image size.

Skewness reflects the disproportionate pixel value distribution used to measure the dark lustrous portion calculated using Equation (20).

$$skewness = -\frac{1}{m*n} \sum_{i=0}^m \sum_{j=0}^n \left[\frac{p(i, j) - \mu}{\sigma} \right]^3 \quad (20)$$

$$standard \ deviation \ \sigma = \sqrt{\frac{1}{m * n} \sum_{i=0}^m \sum_{j=0}^n (p(i, j) - \mu)^2}$$

Kurtosis measures the tendency of the peak with normal distribution, which is calculated using Equation (21).

$$kurtosis = -\frac{1}{m*n} \sum_{i=0}^m \sum_{j=0}^n \left[\frac{p(i,j)-\mu}{\sigma} \right]^4 - 3 \quad (21)$$

Variance: it is a measure of the contrast of the image calculated using Equation (22)

$$Variance = \sigma^2 \quad (22)$$

Entropy: Randomness nature and texture characterization are done using entropy, which is calculated using Equation (23).

$$Entropy = \sum_{i=0}^m \sum_{j=0}^n p(i,j) \log_2(p(i,j)) \quad (23)$$

Contrast calculates the difference between image maximum and minimum pixel intensity values calculated using Equation (24).

$$contrast = \frac{(maximum\ intensity - minimum\ intensity)}{(maximum\ intensity + minimum\ intensity)} \quad (24)$$

Correlation: filter mask is moved over the image, and the sum of products is evaluated in each region used to detect the object in the image irrespective of the position in the image and calculated using Equation (25).

$$Fil(x,y) \circ p(i,j) = \sum_{i=-m}^m \sum_{j=-n}^n Fil(x,y) p(i+x, j+y) \quad (25)$$

where $Fil(x,y)$ is the filter

Energy measures the gray level distribution, and it is based on image normalized histogram is calculated using Equation (26).

$$Energy = \frac{1}{m*n} \sum_{i=0}^m \sum_{j=0}^n (p(i,j))^2 \quad (26)$$

Homogeneity: It describes how values of a pixel vary in the image, and it is calculated using Equation (27).

$$Homogeneity = \sum_{i=0}^m \sum_{j=0}^n \frac{1}{1+(i-j)^2} p(i,j) \quad (27)$$

The structure used for PSO-MLP and ACO-MLP for breast cancer images is shown in Fig. 8, and the parameters used for PSO-MLP and ACO-MLP are listed in Table 1.

Fig.8. PSO/ACO-MLP structure

Table 1. Parameter Setting for PSO-MLP and ACO-MLP

S. No	PSO-MLP		ACO-MLP	
	Parameters	Values	Variable	Values
1	Training	Levenberg Marquardt	Training	Levenberg Marquardt
2	Learning	Gradient Descent	Learning	Gradient Descent
3	Input neurons	8	Input neurons	8
4	Hidden Neurons	19	Hidden Neurons	19
5	Output neurons		Output neurons	
6	Performance function	Mean Square Error	Performance function	Mean Square Error
7	Iteration	100	Iteration	100
8	Activation Functions	Tan sig	Activation Functions	Tan sig
9	Population size	50	Population size	50
10	maximum generation, <i>maxigen</i>	249	maximum generation, <i>maxigen</i>	249
11	number of genes in each population	209	number of genes in each population	209
12	inertia momentum, mw	0.3	initial pheromone value	1e-6

13	elitism parameter	2	elitism parameter	2
14	Global acceleration constant, g_1, g_2	1	pheromone update constant, Q	20
15	Cognitive constant	1	exploration constant	1
16	mw_{max} and mw_{min}	0.8, 0.2	visibility and pheromone sensitivity	5 and 1
17	-	-	The global and local pheromone decay rate, γ	0.9 and 0.5

IV. Experimental Results and Discussion

During the training process, GoogleNet, AlexNet, PSO-MLP, and ACO-MLP classifiers are trained to categorize the DDSM breast cancer images into three categories, namely normal, benign and malignant classifiers, using their extracted features and corresponding labels. The test images, which are underdiagnosis breast cancer images, are classified using this trained network.

4.1 Recognition using GoogleNet and AlexNet

Fig.9 and 10 show the performance such accuracy and loss function value with learning rate as 0.001 of GoogleNet and AlexNet, respectively. The graphs are shown for a single epoch only for better visualization. The performance of both GoogleNet and AlexNet classifier varies with learning rate and highest accuracy achieved for low learning rate. It is evident from Fig. 8 that GoogleNet achieves 98.23 % accuracy in the first epoch for ADAM optimizer and becomes oscillatory. It becomes stable till the maximum iteration for a low learning rate of 0.001, and it is clear the loss function is greatly reduced and remains constant after the 50th iteration. From Fig. 96.19% accuracy is achieved by AlexNet only for ADAM optimizer. Both GoogleNet and AlexNet have low loss function valuation around the 50th iteration and remain stable for ADAM optimizer. Both GoogleNet and AlexNet is not showing a stable learning process for a high learning rate.

(a)

(b)

Fig.9. Performance Curves of GoogleNet during Training with Learning Rate 0.001

(a)accuracy(b) loss function

(a)

(b)

Fig.10. Performance Curves of AlexNet during Training with Learning Rate 0.001

(a)accuracy (b) loss function

Tables 2 and 3 listed the performance of GoogleNet and AlexNet for the classification of DDSM images for breast cancer, respectively. It is clear that network performance increases with a decrease in the learning rate in both networks.

Table 2 GoogleNet Performances for Breast Cancer Image for Training and Testing

Parameters	Training Methods	Learning Rate					
		0.01		0.001		0.0001	
		Training	Testing	Training	Testing	Training	Testing
Accuracy	ADAM	0.8939	0.9009	0.9283	0.9296	0.9823	0.9900
Loss rate		0.3689	0.2195	0.2097	0.1972	0.1830	0.1547
Runtime (minutes)		4.92		4.42		4.14	
Accuracy	RMSPROP	0.7772	0.7541	0.8421	0.8191	0.8391	0.8479
Loss rate		0.9340	0.8687	0.8530	0.9106	0.7943	0.7482
Runtime (minutes)		5.12		4.95		4.79	
Accuracy	SGDM	0.7491	0.7219	0.7814	0.7258	0.7905	0.7651
Loss rate		1.7696	1.7342	1.6085	1.6941	1.4933	1.3665
Runtime (minutes)		5.56		4.96		4.80	

Table 3 AlexNet Performances for Breast Cancer Image for Training and Testing

Parameters	Training Methods	Learning Rate					
		0.01		0.001		0.001	
		Training	Testing	Training	Testing	Training	Testing
Accuracy	ADAM	0.9001	0.9106	0.9561	0.9448	0.9797	0.9880
Loss rate		0.9012	0.9236	0.8912	0.8443	0.8234	0.8176
Runtime (minutes)		4.98		4.72		4.71	
Accuracy	RMSPROP	0.9289	0.9294	0.9390	0.9631	0.9619	0.9891
Loss rate		1.3049	1.3740	1.1483	1.1771	1.2629	1.0427

Runtime (minutes)		5.41		4.86		4.85	
Accuracy	SGDM	0.7040	0.7549	0.8611	0.8219	0.9098	0.8878
Loss rate		1.8211	1.7325	1.8074	1.6628	1.6363	1.5924
Runtime (minutes)		5.91		5.30		4.82	

During training, the maximum accuracy is recorded at 98.23% for GoogleNet. In contrast, it is 97.97% for testing data as the recognition accuracy for GoogleNet and AlexNet for training with ADAM as optimizer and learning rate as 0.001. During testing, GoogleNet yields an accuracy of 99% for learning rate 0.001 and ADAM optimizer. Still, AlexNet, the accuracy achieved is 98.91% for learning 0.001 with optimizer as RMSPROP, which is slightly higher than 98.80% for ADAM with the same learning rate. The performance is also measured in terms of runtime during process and measures the lowest run time of 4.14 minutes for 0.001 learning rate for ADAM and the highest run time of 5.56 recorded SGDM with 0.01 learning rate in GoogleNet whereas for AlexNet it records the lowest time as 4.71 minutes by ADAM optimizer for 0.001 learning and 5.91 minutes as highest run time by SGDM for 0.01 learning rate. Both the network shows superior performance for low learning rate of 0.001 and ADAM optimizer. In comparison, GoogleNet outperforms AlexNet network in terms of accuracy and runtime and show exemplary performance in image processing to detect breast cancer.

4.2 Comparison of Breast Cancer Recognition of GoogleNet with PSO/ACO-MLP

PSO-MLP and ACO-MLP classifiers are trained and tested in the proportion of 70:30 proportion of preprocessed images. The recognition rate of PSO-MLP and ACO_MLP are listed in Table 4.

Table 4 Comparison of GoogleNet Performances over PSO-MLP and ACO-MLP

Parameters	GoogleNet	PSO-MLP	ACO-MLP
Accuracy	0.9900	0.9021	0.8614
Loss rate	0.1547	0.1972	0.2234
Runtime (minutes)	4.14	6.89	8.54

The performance of these networks has been compared with the highest recognition rate of pre-trained network produced by AlexNet with ADAM optimizer for learning rate of 0.001. Fig. 11 shows the recognition rate of PSO-MLP and ACO-MLP.

Fig.11. The recognition rate of PSO-MLP and ACO-MLP

It is found that GoogleNet shows superior performance over PSO-MLP and ACO-MLP with fast convergence due to its robust nature and parallel training in the layer.

4.3 Receiver Operating Characteristic curves of GoogleNet and AlexNet

The Receiver Operating Characteristic (ROC) curves estimate the performance of GoogleNet and AlexNet in breast cancer recognition by establishing a relationship between true positive rate and false positive rate. Figure 12 shows the ROC curves of GoogleNet and AlexNet for 0.0001 learning for ADAM, RMSPROP, and SGDM. Both GoogleNet and AlexNet show a stable learning process of 0.0001 learning rate along with slow convergence. Evaluation metrics, including the sensitivity for both DCNNs, reached around 98% to 100%. There is a complete trade-off between the sensitivity and the specificity for both Networks. In GoogleNet and AlexNet act as a perfect classifier with ADAM optimizer with the true positive rate as one and false positive rate as zero.

Fig.12. ROC Curves for Test data for 0.0001 learning rate
(a) GoogleNet (b) AlexNet

V. Conclusion

In this paper, GoogleNet, a deep neural convolution network, is proposed for breast cancer recognition. The recognition is carried out on DDSM benchmark datasets containing normal, benign, and malignant cancer images. The performance of GoogleNet is assessed in terms of accuracy, loss rate, and runtime. The performance parameters are evaluated for different learning rates 0.01, 0.001, and 0.0001 and also for different optimizers such as ADAM, RMSPROP, and SGDM. DDSM images are preprocessed like identification of breast and muscles, background removal, identification of breast, noise and blur removal using Weiner filter, contrast enhancement by CLAHE filter, and non-stationary noise removal using WPD by db3 wavelet. The proposed methods produce a recognition accuracy of around 99% with ADAM optimizer for learning rate 0.0001. The average recognition rate is high for low learning rate. It also has less loss rate and a runtime of 4.14 minutes. GoogleNet based breast cancer recognition produces significantly high recognition compared to PSO-MLP and ACO-MLP classifier. These classifiers use statistical features like skewness, kurtosis, variance, entropy, contrast, correlation, energy, homogeneity, mean extracted from the preprocessed image as input, and an accuracy of 90.21% and 86.14% for PSO-MLP and ACO-MLP classifiers. Besides, DCNN based GoogleNet

produces a lower runtime of 4.14 minutes, which is significantly lower than PSO-MLP and ACO-MLP classifier run rates of 6.89 minutes and 8.54 minutes, respectively. The performance of the GoogleNet based breast cancer recognition produces a less loss rate of 0.1547 against the 0.1972 and 0.2234 produced by PSO-MLP and ACO-MLP classifiers. Thus, the experimental results show that GoogleNet outperforms AlexNet, PSO-MLP, and ACO-MLP in state-of-the-art testing, with high accuracy, low loss rate, and runtime for breast cancer recognition compared against existing methods for the DDSM dataset.

Funding details

This work is not funded by any agency.

Contributions

The authors contributed to each part of this paper equally. The authors read and approved the final manuscript.

Corresponding author

Correspondence to R. Rajakumari.

Ethics declarations

Conflict of interest

The authors declare that they have no conflict of interest.

Human and animal rights

This article does not contain any studies with animals performed by any of the authors.

Informed consent

Informed consent was obtained from all individual participants included in the study.

References

- [1] S. Tanini, A. D. Fisher, I. Meattini, S. Bianchi, J. Ristori, M. Maggi, and G. L. Russo, “Testosterone and breast cancer in transmen: Case reports, review of the literature, and clinical observation,” *Clin. Breast Cancer*, vol. 19, no. 2, pp. e271–e275, Dec. 2018.
- [2] R. L. Siegel, K. D. Miller, and A. Jemal, “Cancer statistics, 2018,” *CA, Cancer J. Clin.*, vol. 68, no. 1, pp. 7–30, Jan. 2018.
- [3] S. A. Narod, “Tumour size predicts long-term survival among women with lymph node-positive breast cancer,” *Current Oncol.*, vol. 19, no. 5, pp. 249–253, Oct. 2012.
- [4] A. B. Miller, C. Wall, C. J. Baines, P. Sun, T. To, and S. A. Narod, “Twenty-five year follow-up for breast cancer incidence and mortality of the Canadian National Breast Screening Study: Randomised screening trial,” *BMJ*, vol. 348, p. 366, Feb. 2014.
- [5] R. M. Rangayyan, F. J. Ayres, and J. E. L. Desautels, "A review of computer-aided diagnosis of breast cancer: Toward the detection of early signs," *J. Franklin Inst.*, vol. 344, no. 3/4, pp. 312-348, 2007.
- [6] Goodman, D. E., Boggess, L. C., & Watkins, a. B. (2002). Artificial immune system classification of multiple-class problems. In *Proceedings of the intelligent engineering systems* (pp. 179–184). ASME.
- [7] Quinlan, J. R. (1996). Improved use of continuous attributes in C4.5. *Journal of Artificial Intelligence Research*, 4, 77–90.
- [8] Nauck, D., & Kruse, R. (1999). Obtaining interpretable fuzzy classification rules from medical data. *Artificial Intelligence in Medicine*, 16, 149–169.
- [9] Salama, G. I., Abdelhalim, M. B., & Zeid, M. A. (2012). Breast Cancer diagnosis on three different datasets using multi-classifiers. *International Journal of Computer and Information Technology*, 1, 36–43.

- [10] Hamilton, H.J., Shan, N., & Cercone, N. (1996). RIAC: A Rule Induction Algorithm Based on Approximate Classification
- [11] Polat, K., & Günes, S. (2007). Breast cancer diagnosis using least square support vector machine. *Digital Signal Processing*, 17, 694–701.
- [12] Mert, A., Kılıç, N. Z., Bilgili, E., & Akan, A. (2015). Breast cancer detection with a reduced feature set. *Computational and Mathematical Methods in Medicine*, 1–11.
- [13] Dheeba, J., Singh, N. A., & Selvi, S. T. (2014). Computer-aided detection of breast cancer on mammograms: A swarm intelligence optimized wavelet neural network approach. *Journal of Biomedical Informatics*, 49, 45–52.
- [14] Ahmed M. Abdel-Zaher, Ayman M. Eldeib (2016) Breast cancer classification using deep belief networks, *Expert Systems with Applications* 46, 139–144.
- [15] Nilashi, M., Ibrahim, O., Ahmadi, H., & Shahmoradi, L. (2017). A knowledge-based system for breast cancer classification using fuzzy logic method. *Telematics and Informatics*, 34(4), 133–144.
- [16] S. Tiwari, “A variational framework for low-dose sinogram restoration,” *Int. J. Biomed. Eng. Technol.*, vol. 24, no. 4, pp. 356–367, 2017.
- [17] Z. Gao et al., “Motion tracking of the carotid artery wall from ultrasound image sequences: A nonlinear state-space approach,” *IEEE Trans. Med. Imag.*, vol. 37, no. 1, pp. 273–283, Jan. 2018.
- [18] Z. Gao et al., “Robust estimation of carotid artery wall motion using the elasticity-based state-space approach,” *Med. Image Anal.*, vol. 37, pp. 1–21, Apr. 2017.
- [19] C. Szegedy et al., “Going deeper with convolutions,” in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 1–9.

[20] Yuan, Y.; Meng, M. Deep learning for polyp recognition in wireless capsule endoscopy images. *Med. Phys.* 2017, 44, 1379–1389.

[21] Wan S, Liang Y, Zhang Y. Deep convolutional neural networks for diabetic retinopathy detection by image classification. *ComputElectrEng* 2018;72:274–82

[22] Hyunghun Cho, Yongjin Kim, Eunjung Lee, Daeyoung Choi, Yongjae Lee, and Wonjong Rhee, “Basic Enhancement Strategies When Using Bayesian Optimization for Hyperparameter Tuning of Deep Neural Networks”, *IEEE Access*, Special Section on Scalable Deep Learning for Big Data, Volume 8, 2020, pp. 52588 – 52608.

[23] Mendes, R.; Cortez, P.; Rocha, M.; Neves, J. Particle swarms for feed-forward neural network training. In *Proceedings of the 2002 International Joint Conference on Neural Networks*, Honolulu, HI, USA, 12–17 May 2002; pp. 1895–1899.

[24] Blum, C.; Socha, K. Training feed-forward neural networks with ant colony optimization: An application to pattern classification. In *Proceedings of the International Conference on Hybrid Intelligent System*, Rio de Janeiro, Brazil, 6–9 December 2005; pp. 233–238.

Figures

Figure 1

Block diagram of Breast cancer image preprocessing.

Figure 2

Preprocessing of breast cancer image for breast and muscle identification (a) Original image (b) binary image (c) identification of breast and muscles

Figure 3

Breast cancer image before and after preprocessing (a) background removed image (b) muscle removed and breast alone present (c) Weiner filter applied image (d) image after applying CLAHE filter

Figure 4

Wavelet Packet Decomposition of Image for Level 2

(a)

(b)

(c)

(d)

Figure 5

Breast cancer image decomposition using wavelet Packets using db3 (a) Image after applying CLAHE filter (b) Quaternary tree (c) Reconstructed breast cancer image after denoising using WPD (d) Wavelet Packet Coefficients

Figure 6

Proposed Methodology for Breast Cancer Recognition

Figure 7

GoogleNet Inception model

Figure 8

PSO/ACO-MLP structure

(a)

(b)

Figure 9

Performance Curves of GoogleNet during Training with Learning Rate 0.001 (a)accuracy(b) loss function

(a)

(b)

Figure 10

Performance Curves of AlexNet during Training with Learning Rate 0.001 (a)accuracy (b) loss function

Figure 11

The recognition rate of PSO-MLP and ACO-MLP

(a)

(b)

Figure 12

ROC Curves for Test data for 0.0001 learning rate (a) GoogleNet (b) AlexNet