



Classification of Breast Cancer Using a Hybrid and Enhanced Recurrent Residual Convolutional Neural Network (ERResCNN)

S. Prakash^{1*}, K. Sangeetha²

Abstract

Females are affected by BC (Breast Cancer) more than any other type of cancer. BC has caused more deaths than any other diseases such as tuberculosis or malaria according to WHO (World Health Organization). The mortality rates due to BC in women are high making it a candidate for early detection for prevention and cure. Diagnosing BC is a complex task as it is interleaved with normal breast tissues. Image processing methods have been proposed for detecting BC, yet better segmentation methods are required. Fuzzy based approaches provide optimal results in segmenting BC images. Hence, this work uses Fuzzy approach combined with ResCNN (Recurrent Residual Convolution Neural Network) which is the optimized by a modified GA (Genetic Algorithm). The proposed ERResCNN classifying results in detecting BC from images is accurate and efficient in comparison to other methods.

Key Words: Additive White Gaussian Noise (AWGN), Fuzzy Clustering by Local Approximation of Membership (FLAME), Enhanced Recurrent Residual Convolution Neural Network (ERResCNN), Hybrid Genetic Grey Wolf Algorithm (HGGWA), Mammographic Image Analysis Society (MIAS).

66

DOI Number: 10.14704/nq.2021.19.2.NQ21019

NeuroQuantology 2021; 19(2):66-76

Introduction

Cancer is highly fatal causing more deaths than any other diseases around the globe. WHO agencies for cancer research like IARC (International Agency for Cancer Research) and AMC (American Cancer Society) report 17.1 million new global cancer cases in 2018 [1] which is about 25% of all cancers diagnosed in women. At least 53% of the cases are from developing countries, which represent 82% of the world population [1]. WHO estimates that cancer incidences might increase upto 27.5 million by 2040, with an estimated 16.3 million deaths [1]. BC ranks 2nd in mortality rates and mostly affects women [6] amongst four types cancers (ie, lung, breast and bowel [including anus], stomach, and prostate cancers. Breast tissues get abnormally

divided in BC and these abnormal cells start forming lesions in the breast which then grow into a tumor [7]. Researchers have also found that hormonal, lifestyle, and environmental changes also contribute to increasing the risk of breast cancer [2]. BC more often starts with malfunctioning of milk-producing ducts (invasive ductal carcinoma), though it may also begin from glandular tissues called lobules or other cells or tissues within the breast. BC tumors are called malignant when the cancerous cells spread to surrounding tissues and eventually to other parts of the body. BC is identified by breast lumps, breast's shape changes, dimpled skin, oozing fluids from the nipple and scaling red patch on the skin.

Corresponding author: S. Prakash

Address: ¹CSE Department, Sri Shakthi Institute of Engineering and Technology, Coimbatore, India; ²CSE Department, SNS College of Technology, Coimbatore, India.

¹E-mail: prakashdharsan@gmail.com

²E-mail: sangithaparakash@gmail.com

Relevant conflicts of interest/financial disclosures: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 12 January 2021 **Accepted:** 18 February 2021



Manual screening and reading examined values twice creates a heavy workload for radiologists leading to lesser efficiency and even radiologist scarcity [4]. At least twenty five percent of visible breast disorders found in mammographic images are missed [5] making it necessary to minimize interpretational errors using other techniques. Imaging techniques use images from medical modalities like thermography, ultrasound and mammography to detect BC. Mammography (Low dosed x-rays) images have been found to be an effective solution in BC detections. Mammography screening has been helping physicians in reducing BC mortalities [3] but even then, it is difficult for the radiologist to detect tumour or cancerous masses easily making the use of CAD (Computer Aided Diagnostic) system necessary. CAD techniques have also identified cancers in the breast [9]. One major issue with Mammography is its inability to detect micro calcifications (calcium deposits) inside breast tissues of thick breasts. ML (Machine Learning) techniques have also contributed to BC predictions and its related prognosis. ML techniques for BC's classification generally work in three stages namely pre-processing, extraction of features and classification [11]. For increasing accuracy in identifying BC tumours, DL (Deep Learning) approaches have also been used in researches as they help identify or detect or classify BC in images due to their inherent style of functioning like the human brain. This work proposes a hybrid mixture of DL techniques called ERResCNN ResNet (Residual Network) and RCNN (Recurrent Convolution Neural Network). The organization of the paper is a review of literature in section 2 followed by ERResCNN's methodology in detailed section 3. Section 4 displays results of the proposed DL technique while section 5 concludes this research work.

Review of Related Literature

DCNN (Deep Convolutional Neural Network) and SVM (Support Vector Machine) was used by Ragab et al [12] in CAD system for classifying malignant and benign BC tumours from mammography images. They divided the image by manually identifying ROI (Region Of Interest) and using threshold values on the identified regions. The divided region's salient features were extracted using DCNN which was then classified the data into only two classes by AlexNet for further processing. SVM is then connected to the network's final layer (fc) for better accuracy. The scheme used the

datasets DDSM (Digital Database for Screening Mammography) and CBIS-DDSM (Curated Breast Imaging Subset of DDSM) for implementations. Their results of AUC (Area Under the Curve) achieved showed high accuracy in their proposed segmentation and subsequent use of classifiers.

Dual Thresholding was introduced by Badawy et al [13] to segment BC areas in Mammographic images. The proposal used an enhanced dual thresholding and applied applying morphological operations. In post-processing segmented image border contours were added to original samples for helping physicians diagnose BC affected areas easily. Their results was very encouraging as manual thresholding reduced costly processing time and storage needs.

Gao et al [14] used SD-CNN (Shallow-Deep Convolution Neural Networks) for BC diagnosis. This kind of network is a shallow CNN which combines images to obtain a virtual image and extracts novel features from the combined image followed by an ensemble model which classifies lesions as benign or cancer. Their proposal was evaluated on 49 CEDM cases of Mayo Clinic and they achieved an accuracy of 0.84 in their AUC curve. The study followed it up with the creation of a database with 89 FFDM virtual images from the INbreast public database.

ELM (Extreme Learning Machine) was also used in BC detections by Wang et al [15]. They used a median filter to reduce noise in Mammographic images followed by contrast enhancements. A wavelet modulus and maxima transform with morphological operations and regional growth segmented tumour edges. The study extracted five textural and morphological features. ELM technique classified images into BC affected images based on the extracted features. Their results showed their model had better accuracy than SVM in BC detection due to ELM's learning and generalization capability.

Hidden Markov Trees were used by Hu et al [16] where in the model DTCWT (Dual-Tree Complex Wavelet Transform) was also used for identifying micro-calcifications of breast tissues in Mammography images. Their scheme identifies correlations between wavelet coefficients which are then modelled as statistical dependencies or non-Gaussian statistics of real signals for characterizing micro-calcifications in breast tissues. These DTCWT-HMT generated features are optimized by GA and ELM to classify samples with micro-calcifications as benign or malignant. Their



evaluations on DDSM, MIAS and Nijmegen, datasets showed the scheme’s effectiveness of classifying Mammographic images with micro-calcification in their AUC and ROC (Receiver Operating Characteristic) curve values.

DL technique CNN was used by Tan et al [17] in their study for detect BC in Mammographic images. The system BCDCNN (Breast Cancer Detection using Convolution Neural Networks) classified mammographic images into three classes namely non-cancerous, cancerous and normal. Their aim was to quicken BC diagnosis in classification of BC. Visual Images are converted into Digital image sequences in pre-processing where they are studied for getting a suitable parameter for CNN’s training. CNN then recognizes Mammographic images by comparing both types of images based on the parameter. BCDCNN’s Mammogram Classifications when evaluated was found to improve the accuracy in BC classifications from Mammographic images.

From the review, it can be said that heterogeneous breast densities create a challenge in detecting

masses in the breast. Traditional ML techniques are specific to particular density type or dataset, while DL techniques show promises in improving BC diagnosis.

Proposed Methodology

The proposed ERResCNN-HGGWA follows three steps namely Pre-processing, Segmentation and Classification. First is the pre-processing stage where AWGN (Additive white Gaussian Noise) removal technique removes noises from the image followed by a Gaussian blur while blurs the image. For segmentation, the second stage, FLAME (Fuzzy Local Approximation of Membership) groups image parts that can be segmented. ERResCNN, extended from CNN and HGGWA (Hybrid Genetic Grey Wolf Algorithm) which optimizes parameters classify BC affected Mammographic images. The architecture of ERResCNN was designed for Kaggle’s MIAS which has Mammographic BC images Figure 1 depicts the architecture of ERResCNN.

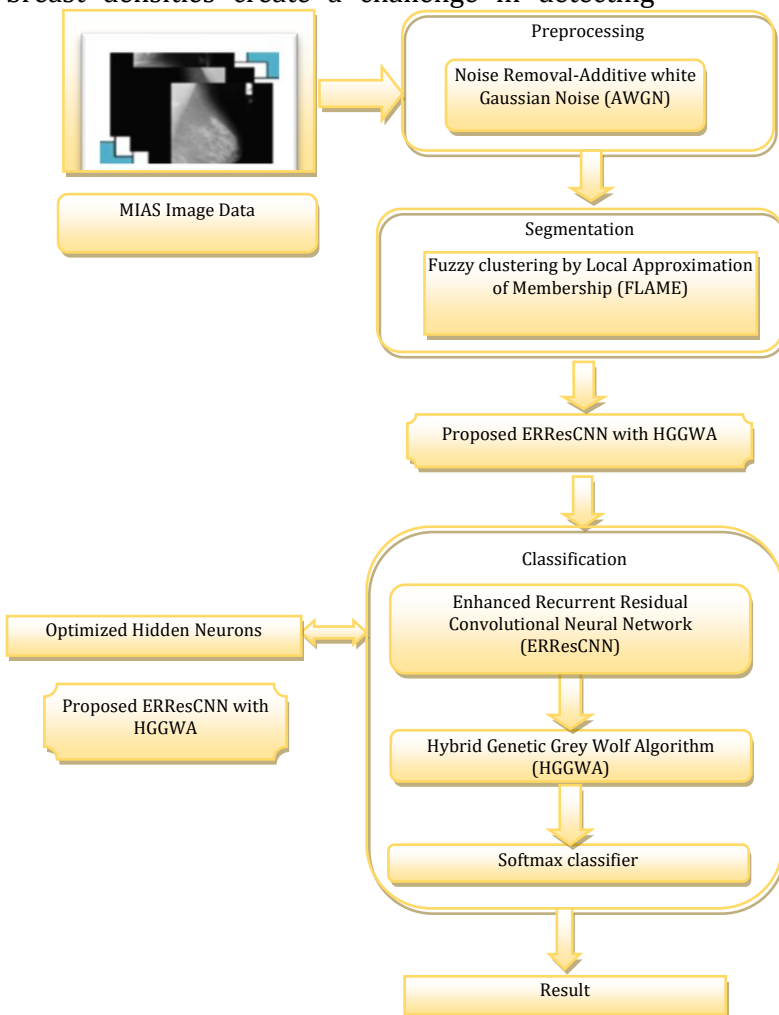


Figure 1. Architecture of ERResCNN



Preprocessing

Digital image is composed of picture elements (Pixels) which are finite, discrete numerical representation of its intensity or gray level and can be output as two dimensional spatial coordinates denoted as x, y. Noises in images reflect different intensity values in place of true pixel values. AWGN can mimic Information theory’s randomizations [18]. It is basically a noise model which uses modifiers denoting specific characteristics. It is Additive as more noise is added to any information that is intrinsic. Gaussian noise is statistical where the PDF (Probability Density Function) equals a normal distribution. The Gaussian random variable is given in Equation (1)

$$P(x) = \frac{1}{(\sigma\sqrt{2\pi}) * \frac{e^{-(x-\mu)^2}}{2\sigma^2 - \infty < 0 < \infty}} \tag{1}$$

here: P(x) – Image’s Gaussian distribution noise; μ / σ – mean/SD (Standard Deviation).

Gaussian noise can be reduced using conventional spatial filtering techniques which include mean filtering, median filtering and Gaussian smoothing.

Mean Filtering: The filtering is done using a sliding window that replaces the centers with mean of the pixels within the window. The filter removes noises like Gaussian noise but blurs the image. This smoothening may have undesirable results on high quality image pixels [19]. Assume I_{xy} as co-ordinates of the window of size $m \times n$, with center at (x, y) . The mean of the corrupted image is $g(x, y)$ in I_{xy} restoration of a point (x, y) is the mean computed using the pixels of I and given in Equation (2)

$$\hat{f}(x, y) = \frac{1}{mn} \sum_{(i,t) \in I_{xy}} g(i, t) \tag{2}$$

Median Filtering: This filter is statistical nonlinear filter on the image $f(x, y)$ outputting image $g(x, y) = \text{med}\{f(x - i, y - j), i, j \in W\}$ where W -two-dimensional mask, $n \times n$ -: mask size ($n = 1, 3, 5, \dots$). An image having zero mean noise and normal distribution, median filtering noise variances can be approximated as Equation (3)

$$\sigma_{med}^2 = \frac{1}{4nf^2(n)} \approx \frac{\sigma_i^2}{n + \frac{\pi}{2} - 1} \cdot \frac{\pi}{2} \tag{3}$$

where σ_i^2 - input power of the noise, n - filter size, $f(\bar{n})$ - noise density function where averaging noise variances are depicted in Equation (4)

$$\sigma_0^2 = \frac{1}{n} \sigma_i^2 \tag{4}$$

Median filters perform better than average filter in reducing random noises and specifically when narrow pulses are apart the filter’s $\frac{n}{2}$ pulse width is effective.

Gaussian smoothing or blurring by a Gaussian function is used to reduce noise in graphics software. The function transforms each pixel and can be depicted as equations (5) and (6) for unidimensional and two dimensional blurs.

$$G(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} \tag{5}$$

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{6}$$

where x -Horizontal distance from origin, y - Vertical distance from the origin, σ - SD of Gaussian distribution.

Segmentation

Image segmentation divides an image into smaller segments or objects with the aim of simplifying image representations into meaningful information for analysis. These segments together form the image where contours can be extracted. Pixels with similar characteristics like intensity, Colour and texture can be computed for identifying regions of similarity. This study uses a Fuzzy based clustering or soft clustering technique for its image segmentation. The objects in such a process can belong to more than one cluster as a Fuzzy process can assign membership which is then used for assigning objects to clusters. FCM (Fuzzy C-Means) in a popular algorithm [20] where memberships are assigned to pixels based on its distance from the cluster center in a cluster. The pseudo code is listed below.

Pseudo Code

Let D is image dataset, $D = \{x_i\}$, where $i = 1, 2 \dots n$; n is the size of D and k is number of clusters.

1. Randomly select ‘ c ’ images as cluster centers.

Repeat

2. Calculate the fuzzy membership ‘ μ_{ij} ’ using:

$$u_{ij} = \frac{1}{\sum_{l=1}^c \left(\frac{d_{ij}^2}{d_{il}^2}\right)^{\frac{1}{m-1}}}$$

3. Update the fuzzy centers ‘ V_j ’ using:

$$V_j = \frac{\sum_{i=1}^n \left((u_{ij})^m x_i\right)}{\sum_{i=1}^n (u_{ij})^m}, \text{ where } j = 1, 2, \dots, c$$

Until V_j estimate stabilize.

where $1 \leq i \leq n$ and $1 \leq j \leq C$. n is the number of images and c is the number of clusters, u_{ij} is the membership of i th image in the j th cluster. d_{ij} is the



distance between the i th image and j th cluster center. m is the fuzzification parameter and it should be more than one. If $m = 1$, then the problem is a crisp clustering. $m \in [1, \infty]$ and usually m is set to 2. V_j is the j th cluster center and c is the number of cluster.

• **FLAME Segmentation**

In FLAME clusters are defined in denser parts where assignment of cluster members is based on an object’s relationships with its neighbors. It constructs a kNN graph to locate outliers and cluster centers. Images with high local density or CSO (Cluster Supporting Objects) are assigned full membership for being cluster centers. Values lower than a threshold value are treated as outliers. They are also assigned memberships to create an group. Images outside the group get varying degrees of memberships for clustering supporting objects. In FLAME number of clusters and outliers are automatically determined based on a given k-NN number and threshold value. The pseudo code is listed below.

• **Pseudo Code**

1. Extracting Information structures from the dataset
 1. Create a KNN graph
 2. Compute object densities for all types (CSO, Outlier, Others) based on nearness in KNN
2. Approximations of Fuzzy Local memberships
 1. Fuzzy Initialization of membership
 - Assign CSOs a fixed membership to be in one cluster

- Assign Outliers a fixed membership to form an outlier group
 - Assign Others with equal memberships
2. Update all three objects iteratively with linear combination of its nearest neighbors fuzzy memberships (Local/Neighborhood Fuzzy Membership Approximation).
 3. Cluster construction
 1. Assign objects to clusters with highest membership (One-to-one object-assignment)
 2. Assign objects to clusters with membership values higher than the threshold value (One-to-Multiple object-assignment)

Classification

Classification predicts data point classes called categories or labels. It is a part of supervised learning where targets are provided along with input data. DL techniques have been successful in labeled data and have been used healthcare and computer vision [21]. RResCNN (Recurrent Residual Convolution Neural Network) is an improved DCNN architecture based residual networks and RCNN architectures [22]. Its advantage lies in better recognitions with fewer network parameters when compared to residual networks and RCNN. The networks have a concatenated inception unit and recurrent convolution layers with residual units (summation of inception unit and input features). The proposed model uses a combination of RResU stacks and transition units. Fig. 2 depicts a RResU

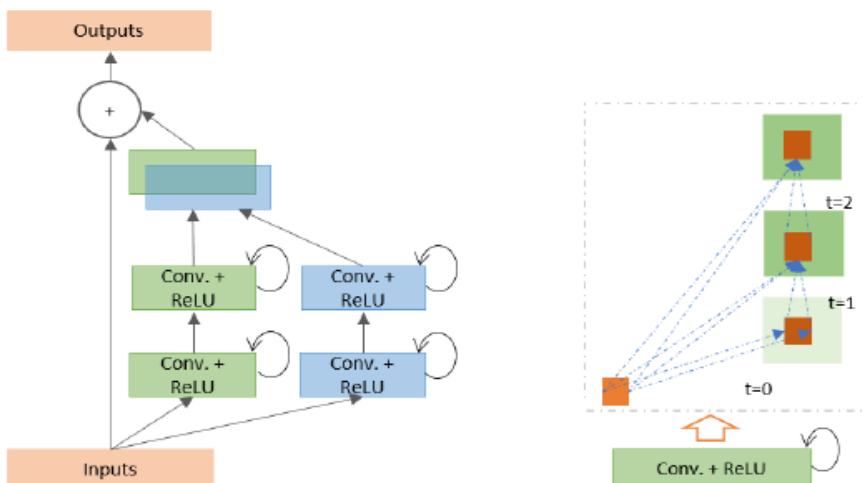


Figure 2. Displaying the Recurrent Residual Unit (RResU)



RResU which has RCLs (Recurrent Convolution Layers) with a residual layer is the most important part of RResCNN [23]. Recurrent convolutions on the residual unit are performed with different sized kernels. The outputs of the execution's time step are added to outputs of the previous time step. In Figure 2, $t = 2$ (0~2) implies one feed forward convolution along with 2 RCLs and the corresponding RCLs w.r.t different time steps are ($t = 2$ (0~2)) and ($t = 3$ (0~3)). RResU accumulates feature maps w.r.t time steps for ensuring better feature representations. RCLs work with discrete time steps expressed based on RResCNN. Assume x_l is the input sample in the l^{th} layer of RResCNN block, (i, j) is an input sample in the k^{th} feature map in RCL and $O_{ijk}^l(t)$ is output at time step t the the output can be expressed an an Equation (7) given below.

$$O_{ijk}^l = (w_k^f)^T * x_l^{f(i,j)}(t) + (w_k^r)^T * x_l^{r(i,j)}(t-1) + b_k \quad (7)$$

Where $x_l^{f(i,j)}(t)$ and $x_l^{r(i,j)}(t-1)$ - inputs for standard convolution layers and l^{th} RCL, w_k^f and w_k^r - weights for standard convolution layer and RCL of the k^{th} feature, b_k - bias. ReLU (Rectified Linear Unit) activation function is depicted in Equation (8)

$$y = f(O_{ijk}^l(t)) = \max(0, O_{ijk}^l(t)) \quad (8)$$

Where f - standard. Outputs of kernels and average pooling layer can be defined as $y_{1x1}(x)$, $y_{3x3}(x)$, and $y_{1x1}^p(x)$ respectively and $\mathcal{F}(x_l, w_l)$, the final output of RCNN is Equation (9)

$$\mathcal{F}(x_l, w_l) = y_{1x1}(x) \odot y(x) \odot y_{1x1}^p(x) \quad (9)$$

Where \odot - concatenation operation of a feature map axis or channel. RCNN outputs are added to RResCNN inputs where the residual operation is depicted as Equation (10)

$$x_{l+1} = x_l + \mathcal{F}(x_l, w_l) \quad (10)$$

Where x_{l+1} - inputs for next transition block, x_l - inputs of RResCNN block, w_l - kernel weight of the l^{th} block, $\mathcal{F}(x_l, w_l)$ - of l^{th} layer output from of RCNN unit. However, feature maps number equals residual unit feature maps. This RResU outputs become the inputs of the immediate next transition unit.

Transition unit operations include pooling, convolution and dropouts based on placement of the model. Non-overlapping max-pooling operation impact on model regularization negatively, hence

overlapped max-pooling is used for regularizing the network. Moreover, late use of a pooling layer increases non-linearity in features resulting in higher dimensional feature maps in the network's convolution layers while keeping the parameters to a minimum. Addition of a 1×1 filter increases the decision function's non-linearity without impacting convolution layers. This proposed scheme uses HGGWA for optimizing network parameters.

a. GWO (Grey Wolf Optimization) Algorithm

GWO imitates grey wolf hunting which move in packs with a four level hierarchy and number between 5 and 12 in a pack. GWO considers four levels namely alpha (α), beta (β), delta (δ), and omega (ω) where Male and Female Leaders are alpha and decide on pack activities like hunting. Beta assists in decisions by feedbacks. Delta is a scout and caretaker, while omega wolves obey other levels [24]. Thus in GWO these four parameters are used. The encircling behavior of the wolves can be depicted as equation (11)

$$\vec{X}(t+1) = \vec{X}_p(t) + \vec{A} \cdot \vec{D} \quad (11)$$

where \vec{A}, \vec{C} - coefficient vectors, \vec{X}_p - prey position vector, X - position of wolves in d -dimensional (d is the no of variables), (t) - no of iterations. \vec{D} is denoted by Equation (12)

$$\vec{D} = |\vec{C} \cdot \vec{X}_p(t) - \vec{X}(t)| \quad (12)$$

where $\vec{A} = 2\vec{a} \cdot \vec{r}_1 - \vec{a}$ and $\vec{C} = 2 \cdot \vec{r}_2$ and \vec{r}_1, \vec{r}_2 - random vectors in the interval $[0,1]$. \vec{a} - Linear vector decrease from 2 to 0 in iterations. During hunting alpha is the optimal solution, delta and beta are know about prey's possible position. These solutions are in use which forces omega to modify positions and reach the target. The updates are listed as Equation (13)

$$\vec{X}(t+1) = \frac{x_1 + x_2 + x_3}{3} \quad (13)$$

where \vec{x}_1, \vec{x}_2 and \vec{x}_3 - Best possible solution at an iteration t and $\vec{x}_1 = \vec{X}_\alpha - A_1 \cdot (\vec{D}_\alpha)$, $\vec{x}_2 = \vec{X}_\beta - A_2 \cdot (\vec{D}_\beta)$, $\vec{x}_3 = \vec{X}_\delta - A_3 \cdot (\vec{D}_\delta)$ and A_1, A_2 and A_3 are computed from the Equations. $\vec{D}_\alpha, \vec{D}_\beta, \vec{D}_\delta$ are calculated using $\vec{D}_\alpha = |\vec{C}_1 \cdot \vec{X}_\alpha - \vec{X}|$, $\vec{D}_\beta = |\vec{C}_2 \cdot \vec{X}_\beta - \vec{X}|$, $\vec{D}_\delta = |\vec{C}_3 \cdot \vec{X}_\delta - \vec{X}|$ and $\vec{C}_1, \vec{C}_2, \vec{C}_3$ are computed based on the equations.

\vec{a} is linear while \vec{r}_1 and \vec{r}_2 are randomized. GWO has a drawback in selecting random variables in deciding optimum positions. Hence, this work introduces GA in GWO to cross this hurdle.



b. Proposed HGGWA

The main motivation of proposing HGGWA is to cross the problem of random variable selection in GWO with a genetic algorithm. GWO uses static values for \vec{r}_1 and \vec{r}_2 which may lead to the problem of local minima. The proposed technique’s GA crosses and mutations help in selecting optimum control parameters.

HGGWA based on GA generates a population of $n1$ no of positions for GWO and $n2$ no of initial population for GA. In the the hunting process, grey wolves circle the prey for it to stop moving. HGGWA uses two control parameters (\vec{r}_1 and \vec{r}_2) using genetic operators. HGGWA’s dynamic crossover ratio and mutation ratio obtains a new population using: ranking the population based on fitness; average of fitness values; fixed as threshold value; discarding least fitness based on the threshold value [25].

GWO hunts follow a social hierarchy where optimum position is selected based on the fitness where, $X\alpha =$ the first search agent, $X\beta =$ the second search agent and, $X\gamma =$ the third search agent. The proposed system uses a softmax classifier as detailed below. In an input sample x , weight vector W and K distinct linear functions as softmax operation for the i^{th} class can be defined as equation (14)

$$P(y = i|x) = \frac{e^{x^T w_i}}{\sum_{k=1}^K e^{x^T w_k}} \tag{14}$$

RResCNN model was evaluated with many convolution layers in blocks where the no of layers were determined w.r.t the time step t. The proposed BC recognition model uses: two convolution layers; four RCNN blocks; transition blocks; fully connected/hidden layer; softmax layer. The last hidden state is used to classify data based on Equations (15) and (16)

$$y = softmax(w_k^f h_T + b_k) \tag{15}$$

$$y = softmax(w_k^r h_T + b_k) \tag{16}$$

where y - predicted gait type, w_k^f and w_k^r - output weights, b_k - output bias, respectively. The softmax classifier recognizes image patterns using its trained features. Every patch of a BC image x_i is transformed into a fixed-length trained feature vector \hat{x} . Hidden layer is used by softmax classifier to classify BC as malignant or benign. It estimates the probability of each class with which the data is classified where the total probability of all classes equals to 1. The softmax function uses normalization and exponentiation in finding class probabilities.

Results and Discussion

MIAS (Mini Mammographic) BC data set was used to test the proposed system. The MIAS dataset with benign and malignant tumors is available online and contains important risk factors which are used to diagnose BC in labs. and give reliable results. The tested set contained 322 breast images. The proposed framework of ERResCNN is compared with existing methods like Recurrent Neural Network (RNN) and Deep Neural Network (DNN) [26]. Performance metrics are like recall, f-measure, precision, and accuracy were used for analyzing ERResCNN-HGGWA framework’s performance. The metrics used is dependent a confusion matrix as it allows visualization of the performance of an algorithm [27] and is tabulated in table 1. Additionally, recall, precision and f-measure were used to analyze correct and incorrect decisions of the classifier. TP (True Positive) implies no of true detections of BC, FP (False Positives) is the no of non-mitosis mis-classified as BC. FN (False Negatives) is number of objects that were not detected and TN (True Negatives) implies no of detected non-breast cancer cases.

Table 1. Confusion Matrix

		Predicted	
		Positive	Negative
Actual	Positive	Tp	Fp
	Negative	Fn	Tn

Performance Measures used

Precision: Precision is finding how precise/accurate a model is from predicted positive to how many of them are actually positive. Precision is a good measure to determine, when the costs of False Positive is high.

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

Recall: Recall calculates how many of the Actual Positives a model capture through labeling it as Positive (True Positive). Recall shall be the model metric to select the best model when there is a high cost associated with False Negative.

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} = \frac{True\ Positive}{Total\ Actual\ Positive}$$

F1 Score: F-measure is a measure of a test’s accuracy



$$F_1 - score = \frac{1}{\frac{1}{recall} + \frac{1}{precision}}$$

Accuracy: It is the percentage of the test tuples that are classified properly by any algorithm.

$$Accuracy = \frac{\# of true positives + \# of true negatives}{\# of true positives + false negatives + false positives + true negatives}$$

The comparative performances of classification methods is listed in table 2.

Table 2. Comparative Performances of Classification Methods

Methods	Metrics			
	Precision (%)	Recall (%)	F-measure (%)	Accuracy (%)
DNN	85.6	85.6	84.8	91.2
RNN	87.2	89.8	88.6	93.2
ERResCNN	92.5	91.4	92.5	94.6

Figure 3 depicts comparative performances of classification methods in terms of precision with the proposed ERResCNN and existing methods like DNN and RNN. It can be seen that the proposed method has a higher precision rate of 92.5% and prediction.

can detect BC better. Comparatively DNN and RNN precision rates are 85.6% and 87.2%, lower than the proposed method. Hence, the proposed model is effective and meaningful for short-term disease

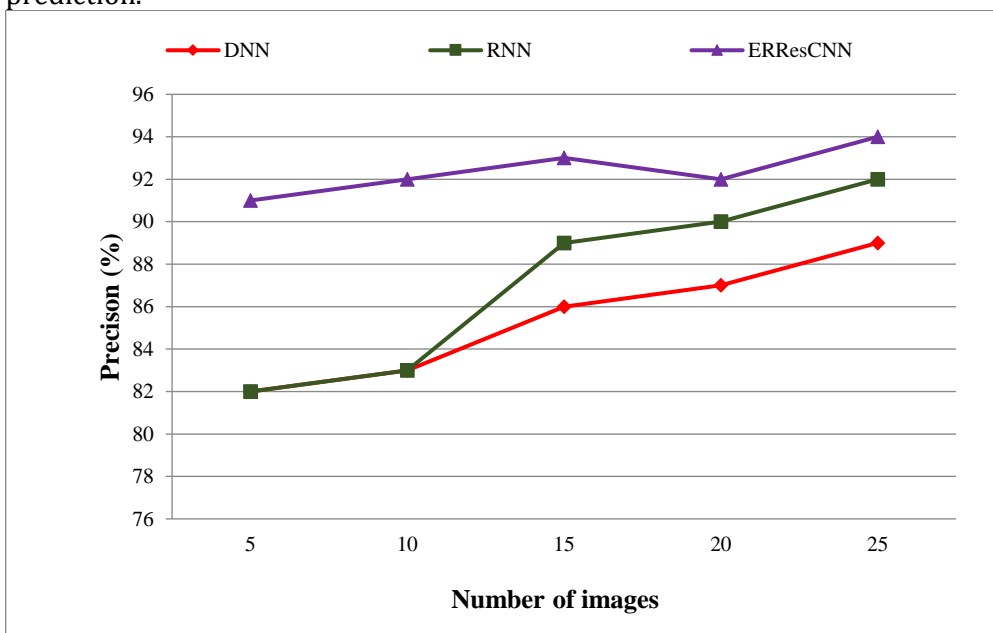


Figure 3. Comparative Performances of Classification Methods in Terms of Precision

The correlation comparison in various classification methods between proposed ERResCNN and existing methods such as DNN and RNN are shown in figure 3. From the above figure 3, proposed method can obtain high precision rate when compared to other existing methods. It is best way of getting the breast cancer exactly with the high precision rate of 92.5%. When comparing the

precision among the existing methods such as DNN and RNN are providing good precision rates of 85.6% and 87.2%, however which is lower than the proposed method. Hence, the proposed model is effective and meaningful for short-term disease prediction. Figure 4 depicts comparative performances of classification methods in terms of their recall values.



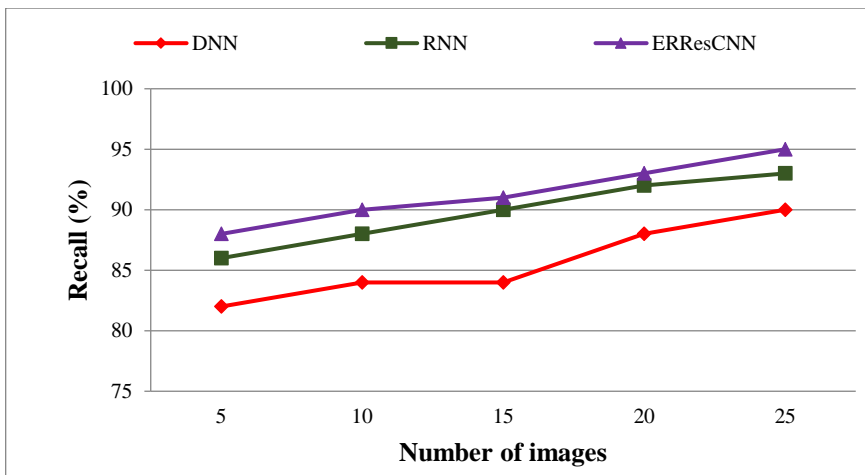


Figure 4. Comparative Performances of Classification Methods in Terms of Recall

The recall values from Figure 4 show that ERResCNN-HGGWA value of 91.4% in comparison with other classification methods shows its good detection rate. The proposed scheme’s effective image enhancement and HGGWA optimization puts

its performance higher than DNN and RNN which have recall values of 85.6% and 89.8%. Figure 5 depicts Comparative performances of classification methods in terms of F-Measure.

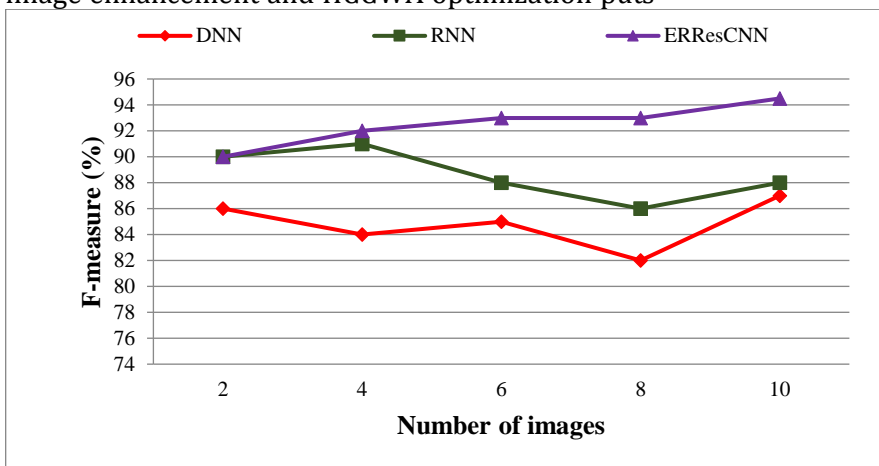


Figure 5. Comparative Performances of Classification Methods in Terms of F-Measure

The proposed ERResCNN-HGGWA in Figure 5 shows higher F-measure (above 90%) in comparison to other existing methods DNN and RNN which are less than (90%). Thus the proposed

system indicates it is suitable for BC detection from images. Figure 6 depicts the accuracy of the proposed system in comparison with other methods.

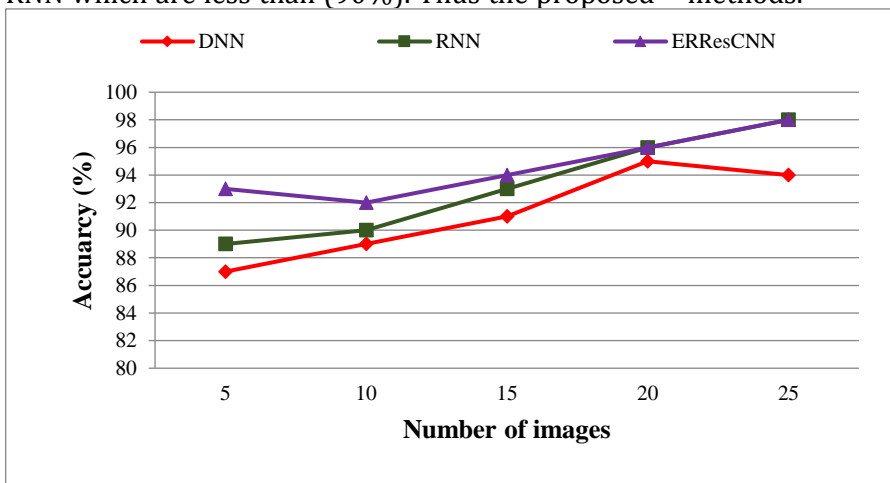


Figure 6. Comparative Accuracies of Classifiers



Figure 6 clearly indicates that the proposed ERResCNN-HGGWA is effective in identifying BC with an accuracy of 94.6%. Comparatively RNN and DNN score 93.2% and 91.2% which is lesser than the proposed system's accuracy.

Conclusion and Future Work

BC has high mortality rates and specifically in women. There is an imperative need to reduce this dreaded disease by early prognosis. Many systems have been proposed for identifying BC from Mammographic images. This work continues the same further by proposing a DL techniques for better efficiency and accuracy in detecting BC cases. Image Datasets have to be pre-processed, segmented and classification have issues when it comes to automating them. The proposed scheme of this work attempts to overcome many issues in healthcare disease diagnosis of BC from image datasets. The novelty in this work is in the techniques used in each and every stage, but end with optimizing network layer outputs. The results of the proposed ERResCNN-HGGWA model's evaluations also show considerable improvement than other existing methods. It is an accurate and computationally efficient method to detect BC from Image datasets as it scores highest in terms of accuracy with 94.6%. It can be concluded that the proposed scheme with its demonstrated results is a viable and implementable system for CAD systems. Future work can be in combining RCNN and auto encoders, using Hybrid optimization methods on large datasets using Grey Wolf Algorithm (GWA) with Cuckoo Search (CS) for improved classification accuracy results.

References

- American Cancer Society. Global Cancer: Facts & Figures, 4th edition
<http://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/global-cancer-facts-and-figures/global-cancer-facts-and-figures-4th-edition.pdf>, 2018.
- Lbachir IA, Es-Salhi R, Daoudi I, Tallal S. A new mammogram preprocessing method for Computer-Aided Diagnosis systems. In *IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA) 2017*; 166-171.
- Rimmer A. Radiologist shortage leaves patient care at risk, warns Royal College. *BMJ: British Medical Journal* 2017; 359.
- Weber RJ, Van Bommel RM, Louwman MW, Nederend J, Voogd AC, Jansen FH, Duijm, LE. Characteristics and prognosis of interval cancers after biennial screen-film or full-field digital screening mammography. *Breast cancer research and treatment* 2016; 158(3): 471-483.
- Lehman CD, Wellman RD, Buist DS, Kerlikowske K, Tosteson AN, Miglioretti DL. Breast Cancer Surveillance Consortium. Diagnostic accuracy of digital screening mammography with and without computer-aided detection. *JAMA internal medicine* 2015; 175(11): 1828-1837.
- Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Bray F. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *International journal of cancer* 2015; 136(5): E359-E386.
- Mehdy MM, Ng PY, Shair EF, Saleh NI, Gomes, C. Artificial neural networks in image processing for early detection of breast cancer. *Computational and mathematical methods in medicine* 2017.
- Ghoncheh M, Pournamdar Z, Salehiniya H. Incidence and mortality and epidemiology of breast cancer in the world. *Asian Pacific Journal of Cancer Prevention* 2016; 17(sup3): 43-46.
- Kim JH, Cha JH, Kim N, Chang, Y, Ko, MS, Choi YW, Kim HH. Computer-aided detection system for masses in automated whole breast ultrasonography: development and evaluation of the effectiveness. *Ultrasonography* 2014; 33(2): 105-115.
- Bickel H, Pinker K, Polanec S, Magometschnigg H, Wengert G, Spic C, Baltzer P. Diffusion-weighted imaging of breast lesions: region-of-interest placement and different ADC parameters influence apparent diffusion coefficient values. *European radiology* 2017; 27(5): 1883-1892.
- Zheng B, Yoon SW, Lam SS. Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms. *Expert Systems with Applications* 2014; 41(4): 1476-1482.
- Ragab DA, Sharkas M, Marshall S, Ren J. Breast cancer detection using deep convolutional neural networks and support vector machines. *PeerJ* 2019; 7: 1-23,
- Badawy SM, Hefnawy AA, Zidan HE, GadAllah MT. Breast cancer detection with mammogram segmentation: a qualitative study. *International Journal of Advanced Computer Science and Application* 2017; 8(10): 17-120.
- Gao F, Wu T, Li J, Zheng B, Ruan L, Shang D, Patel B. SD-CNN: A shallow-deep CNN for improved breast cancer diagnosis. *Computerized Medical Imaging and Graphics* 2018; 70: 53-62.
- Wang Z, Yu G, Kang Y, Zhao Y, Qu Q. Breast tumor detection in digital mammography based on extreme learning machine. *Neurocomputing* 2014; 128: 175-184.
- Hu K, Yang W, Gao X. Microcalcification diagnosis in digital mammography using extreme learning machine based on hidden Markov tree model of dual-tree complex wavelet transform. *Expert Systems with Applications* 2017; 86: 135-144.
- Tan YJ, Sim KS, Ting FF. Breast cancer detection using convolutional neural networks for mammogram imaging system. In *IEEE International Conference on Robotics, Automation and Sciences (ICORAS) 2017*; 1-5.
- Makandar A, Halalli B. Pre-processing of mammography image for early detection of breast cancer. *International Journal of Computer Applications* 2016; 144(3): 0975-8887.



Turajlić E, Karahodžić V. An adaptive scheme for X-ray medical image denoising using artificial neural networks and additive white Gaussian noise level estimation in SVD domain. *In CMBEBIH* 2017; 36-40.

Jecheva V, Nikolova E. Some clustering-based methodology applications to anomaly intrusion detection systems. *International Journal of Security and Its Applications* 2016; 10(1): 215-228.

Ball JE, Anderson DT, Chan Sr CS. Comprehensive survey of deep learning in remote sensing: theories, tools, and challenges for the community. *Journal of Applied Remote Sensing* 2017; 11(4): 042609.

Alom MZ Hasan M, Yakopcic C, Taha TM, Asari VK. Improved inception-residual convolutional neural network for object recognition. *Neural Computing and Applications* 2018: 1-15.

Wu S, Zhong S, Liu Y. Deep residual learning for image steganalysis. *Multimedia tools and applications* 2018; 77(9): 10437-10453.

Al-Tashi Q, Kadir, SJA, Rais HM, Mirjalili S, Alhussian H. Binary optimization using hybrid grey wolf optimization for feature selection. *IEEE Access* 2019; 7: 39496-39508.

Daniel E, Anitha J. Optimum green plane masking for the contrast enhancement of retinal images using enhanced genetic algorithm. *Optik* 2015; 126(18): 1726-1730.

Botchkarev A. Performance metrics (error measures) in machine learning regression, forecasting and prognostics: Properties and typology, arXiv preprint arXiv:1809.03006, 2018.

Ohsaki M, Wang P, Matsuda K, Katagiri S, Watanabe H, Ralescu A. Confusion-matrix-based kernel logistic regression for imbalanced data classification. *IEEE Transactions on Knowledge and Data Engineering* 2017; 29(9): 1806-1819.

Di Biase F. Quantum entanglement of consciousness and space-time a unified field of consciousness. *NeuroQuantology* 2019; 17(3): 80-85.

Köklü K. Is relativity theory also valid in biogenetics and mathematics?. *NeuroQuantology* 2019; 17(3): 53-58.



K. Sangeetha received her master's degree in Computer Science and Engineering and Ph.D. degree in Information and Communication Engineering from Anna University, Chennai. Currently she is working in the Department of Computer Science and Engineering, SNS College of Technology. Her current research interests include the field of Artificial Intelligence, machine learning, and medical data analysis. She has published numerous papers in Journals and Conferences.

Authors Profile



S. Prakash received his master's degree in Computer Science and Engineering and Ph.D. degree in Information and Communication Engineering from Anna University, Chennai. Currently he is working in the Department of Computer Science and

Engineering, Sri Shakthi Institute of Engineering and Technology. His current research interests include the field of medical data analysis, data mining and data structures. He has published numerous papers in international journals and conferences.