Academic Research Journal of Science and Technology (ARJST)
eISSN: 3048-9822 (online)
International, Double Blind-Peer Reviewed & Refereed, Monthly Open Access Journal
Volume 1 Issue 09 (May) 2025 Available online at: https://publications.ngmc.ac.in/journal/

BREAST CANCER CLASSIFICATION: A DEEP LEARNING BASED OUTLIER DETECTION INCORPORATING EGRET SWARM OPTIMIZATION ALGORITHM

Mrs. S. Maria Sylviaa[1], Dr. N. Sudha[2]

1.Research scholar, Bishop Appasamy College of Arts and Science, Coimbatore
2.Associate Professor, Bishop Appasamy College of Arts and Science, Coimbatore

Received 20/05/2025
Accepted 30/05/2025
Published 01/06/2025

ABSTRACT

Breast cancer (BC) continues to pose a major global health challenge, characterized by elevated incidence and mortality rates. Timely detection and precise classification are essential for successful treatment. Conventional BC therapies encompass surgery, radiation, and pharmacological interventions aimed at eradicating microscopic tumors. Recent advancements in machine learning (ML) and deep learning (DL) have demonstrated potential in improving the accuracy of BC diagnosis and classification. This study employed an innovative accurate classification model for BC utilizing a Deep Neural Network-Convolutional neural network Egret Swarm Optimization framework. The methodology comprises three primary stages: data pre-processing through Enhanced Linear Discriminant Analysis (ELDA), outlier identification via a Deep Neural Network (DNN), Ultimately, Convolutional Neural Network with Egret Swarm Optimization (CNN-ESO) algorithm categorizes the BC data as either benign or malignant. The efficacy of this method was confirmed using the Wisconsin Breast Cancer (WBC) and Wisconsin Diagnosis Breast Cancer (WDBC) datasets.

Keywords: BC, WBC, WDBC, DL, ML

1. INTRODUCTION:

Breast cancer (BC) treatment can be very effective, particularly when the illness is identified at an early stage. The majority of BC therapies encompass surgery, radiation, and medication, which aim at eradicating microscopic malignancies that have dispersed from a breast tumor into the bloodstream. This strategy has the prospective to save lives and stop the progression and dissemination of cancer. By the year 2020, there were 2.3 million newly diagnosed cases, and 685,000 people lost their lives to the disease. According to WHO, breast cancer (BC) is expected to become the most prevalent cancer worldwide **(Mridha et al., 2021)**. In the last five years, 7.8 million individuals have received a diagnosis of BC, following lung cancer. The most frequently occurring types of BC, which can be identified through microscopic examination, are invasive ductal carcinoma (IDC) and ductal carcinoma in situ (DCIS). DCIS has a slower growth rate and less effect on a patient's daily life.

Approximately 80% of BC diagnoses are identified as IDC, which is more lethal as it involves the entirety of the breast tissue, while DCIS represents 20-53% of cases. IDC impacts around 80% of individuals with BC. Compared to all other types of malignancy combined, BC leads to the highest number of lost disability-adjusted life years (DALYs) **(Mao et al., 2019)**. A lump or a painless lumpiness in the breast can both indicate breast cancer (BC). Women should seek medical attention immediately upon discovering any lump in their breasts, regardless of the discomfort it may cause **(Wang et al., 2018)**. There are many reasons for breast lumps, most of which are not harmful. Approximately 90% of breast swellings are benign and do not pose any health threats. Non-malignant conditions include breast infections and benign growths like cysts and fibroadenomas. A thorough medical assessment is essential **(Valvano et al., 2019)**. In recent decades, machine learning (ML) procedures have significantly progressed in the diagnosis and grouping of breast cancer (BC). Certain machine learning algorithms, like support vector machines and K-nearest neighbors , may not perform as well as anticipated when extraneous features are present **(Shaban et al., 2020)**. Prioritizing the removal of specific features before deploying machine learning algorithms can greatly improve the precision of classification systems **(Haq et al., 2021)**. are to avoid overfitting, enhance accuracy, boost learning performance, and lower computational expenses **(Haq et al., 2021)**. Instances of swarm intelligence methods include ant colony optimization (ACO), particle swarm optimization (PSO), grey wolf optimization (GWO), and the bat algorithm (BA), among others **(Nguyen et al., 2020)**. Nevertheless, solely relying on feature selection does not guarantee optimal classification performance, particularly when datasets are contaminated with defective or noisy information **(Rabie et al., 2020)**.

To enhance the effectiveness of classification methods, it is essential to eliminate such noisy data **(Rabie et al., 2020)**. The progression of outrebound data rejection, which involves removing or discarding data that significantly deviates from typical patterns, is vital prior to classification. A variety of machine learning algorithms interpret outliers as noise that should be eliminated to enhance the predictive power of the system.

2. LITERATURE REVIEW

Amarasinghe et al., 2018 introduced a novel method called DeepStreamCE, which leverages deep neural network activations in a streaming context to detect concept evolution.

Munir et al. 2018 established a framework for anomaly detection based on DNNs, which provides explanations for any anomalies identified. The creators of DeepAnT designed a DL-based approach for recognizing anomalies

Gao et al., 2020 The technique consists of two phases: a time-series prediction module and an anomaly detection module. DeepAnT employs convolutional neural networks (CNN) on unfiltered raw data, incorporating a max-pooling layer followed by two convolutional layers. **Gómez-Flores and Hernández-López et al., 2020**, a CNN was proposed for utilizing disruptions to enhance data augmentation in the time series detection of abnormalities in the amplitude and phase spectra within the frequency domain. However, this method is characterized by slower convergence rates.

Liu et al., 2020 also proposed a computer-aided diagnosis (CAD) method for classifying breast cancer (BC). They analyzed 39 morphological parameters to define breast tumor forms and assist in identifying BC types. The research utilized 892 mammography images from the BC Digital Repository (BCDR) and 2054 ultrasound images from the National Cancer Institute (INCa) in Rio de Janeiro, Brazil, for training purposes. Morphological traits, such as elongatedness, convexity, eccentricity, circularity, and area variations, were extracted using border and region descriptors. The area under the curve (AUC) for both image databases was found to be 82.0%. However, this method encountered challenges with misclassification rates.

Irfan et al., 2021proposed a computer-aided identification (CAD) method for classifying breast tumors based on corner feature mining. The area of interest (ROI) was analyzed to extract morphological characteristics such as roundness, aspect ratio, ellipticity, regularity, and roughness. An SVM classifier was then employed to determine whether the images represented malignant or benign cuts. The suggested method utilized 192 ultrasound images, which included 71 cancerous and 121 benign cases. This technique achieved an exactness of 67.31%, with a sensitivity of 47.62%, a specificity of 80.65%, and a positive predictive value (PPV) of 62.50%.

Lahoura et al., 2021implemented deep learning (DL) techniques for segmenting ultrasonic images of breast lesions using morphological erosions along with dilated semantic segmentation networks (Di-CNN). The image segments were introduced into DenseNet201 utilizing assignment knowledge for feature extraction. This study involved 780 breast tumors. Combining CNN-activated feature vectors with the SVM classifier yielded a correctness of 90.11%, a precision of 98.45%, and a remarkably high precision of 98.9%. However, the method encountered issues related to overfitting.

Adebiyi et al., 2022 A machine learning model based on extreme learning machine (ELM) was proposed for categorizing breast cancer in a cloud environment. The extreme model was combined with techniques such as AdaBoost, Naive Bayes, SVM, perceptron, and k-NN. The dataset used in this study was the Wisconsin Breast Cancer Diagnostic data (WBCD), which includes 32 attributes and 569 samples. The findings revealed that this method attained an accuracy of 98.68%, a precision of 90.54%, a recall of 91.30%, and an F1-score of 81.29%.

3. METHODS

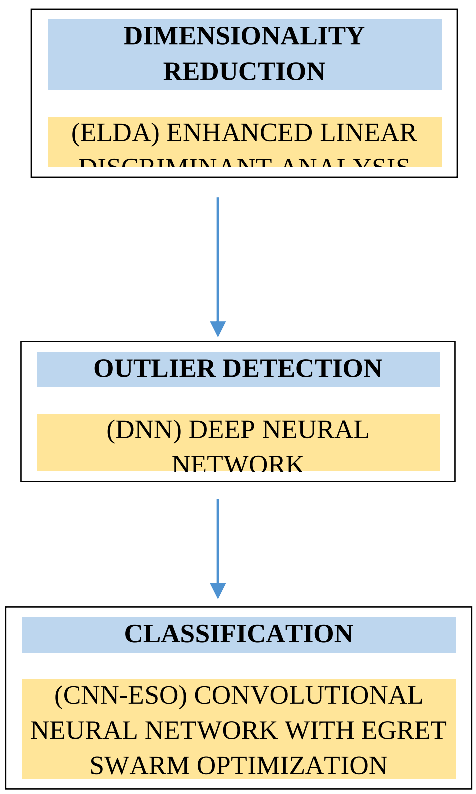


FIG 1: OVERALL FLOW

Here, this work contains three different phases for diagnosing breast cancer. In the first phase the dimensions are reduced from the WDBC and WBC Dataset after reducing the dimensions the output is given as input to the second phase which is outlier detection. After detecting outliers, they are removed from the dataset and the remaining data is moved to classification phase. In, which the data is classified as benign and malignant.

3.1. DATASET USED:

Here, two different dataset WBC (Wisconsin Breast Cancer) and WDBC (Wisconsin Diagnostic Breast Cancer) dataset collected from UCI Repository. Where, WDBC dataset has 569 instances with 30 features. WBC dataset has 699 instances with 9 features.

3.2. ELDA FOR DIMENSIONALITY REDUCTION

In supervised machine learning, a method called linear discriminant analysis can be employed to enhance categorization tasks. It is utilized for modelling differences among groups because it requires distinguishing between two or more categories. This technique allows for the transformation of a feature from a single dimension into a minimal-dimensional space. It works by maximizing the variance between classes while minimizing the variance within classes, reducing aspects from high-dimensional spaces to lower ones. Linear discriminant analysis (LDA) is extensively applied in ML classification tasks. It reduces feature dimensions by effectively limiting the variance between classes relative to the variance within the classes. Therefore, an enhancement to LDA (Algorithm 1) is demonstrated by calculating the class means and prior probabilities simply by supplying the independent variables for the dataset. The average vector dimensions for each group in the dataset are computed. To define the scattering matrix, the k Eigenvectors with the highest Eigenvalues can be selected, which are then ordered in descending sequence to form a dk matrix W, consisting of Eigenvectors ($e1$, $e2$, $e3$, $e4$, …, ed) alongside the corresponding Eigenvalues (1, 2, …, d). Employing the previously constructed eigenvector matrix W, the data is recreated in a new subspace. If $Y = X$W, its matrix representation is straight forward. The matrix of covariance for every group is calculated along with a pooled covariance matrix. This process involves labeling classes and determining the LDA discriminant function.

1. $En =, xU\, j \mid yj = cn, j = 1, \cdots, m - 1, n = 1, 2$ //class-specific groupings
2. $\mu i = mean(Ei), i = 1, 2$//class means
3. $C = (\mu 1 - \mu 2)(\mu 1 - \mu 2)U$//between scattering mediums of class
4. $Zn = En - 1mn\mu U\, i, n = 1, 2$//standards for the middle class
5. $Tn = ZU\, n\, Zn, n = 1, 2$// standards for the class scatter
6. $T = T1 + T2$//internal scattering
7. $\lambda 1, x = eigen(T - 1C)$ //estimate central eigenvector m

ALGORITHM 1: ENHANCED LINEAR DISCRIMINANT ANALYSIS

3.3 DNN FOR OUTLIER DETECTION

A decision-making algorithm utilizing artificial neural networks is referred to as DNN **(Adege et al. 2018) (Sun et al. 2017)**. Neurons, which are components of neural networks, are part of the DNN framework. As the name suggests, A deep neural network is characterized by having several hidden layers positioned between the input and output layers. The construction of a deep neural network during the DNN training phase involves a single input, output, and three hidden layers to enhance abstraction capabilities for improved performance. The DNN used in this study is based on the feeding-forward concept of artificial neural networks. Specifically, input training data is employed when attributes in the training data coincide with neurons in the input layer nodes **(Yusuf et al. 2021)**. To connect the nodes of the output layer, the weights that link to the input layer (l) are initially generated and assigned to the hidden layer nodes through an activation function.

The ReLU function depends on the results produced by each prior layer of the network, as shown in Equation (2), which further computes the bias (b) from the weighted inputs for the next layer in the network by applying the activation function across all three concealed layers.

$$\boldsymbol{ReLU(x) = \max(0, x)} \quad \textbf{(1)}$$

$$\boldsymbol{x^l = f(z^l)} \quad \textbf{(2)}$$

$$\boldsymbol{z^l = w^l x^{l-1} + b^l} \quad \textbf{(3)}$$

Formally, **Equation (2)** and **Equation (3)** determine the outputs of the hidden layers within the networks: The final activation function, $f(z^l)$, as stated in **Equation (2)**, allows the ReLU function to produce l, the current layers, along with w and b, which are the weights and biases of the hidden layers. The layers, which are fully integrated, function as the final output decision layer and are suitable with sigmoid activations as defined by **Equation (4)**. Sigmoid classification is employed as it assesses the neuron representations as either outliers in normal data.

$$\textbf{sigmoid}(z) = 1 / 1 + e^{-z} \quad \textbf{(4)}$$

The data instances $x = (x1, x2, \ldots, xn)$ are transformed into an outlier mark, which reflects the likelihood of being classified as an outlier, by utilizing the sigmoid activation function. Statistically, normal instances are more frequently observed, making them appear more likely, while outlier cases occur less commonly since they represent unusual events **(Goni et al. 2020)**. Given that the probability rate range from (0, 1), this information facilitates the differentiation between outlierness and normalcy when categorizing data instances as outliers or normal. Specifically, a higher probability suggests a stronger indication of being an outbound data. In **Equation (5)**, the function $\varepsilon(x|\theta)$ accounts for the outlier score.

$$\varepsilon(x|\theta) = \sum^{k}{}_{i,j=1} \quad \textbf{(5)}$$

The influence of the interaction is represented by the trainable parameter w_{ij}, where z_{ij} consolidates the i^{th} feature values of x within the representative space Z in lower dimensions. DNN training phases yield models that identify outliers and standard training data.

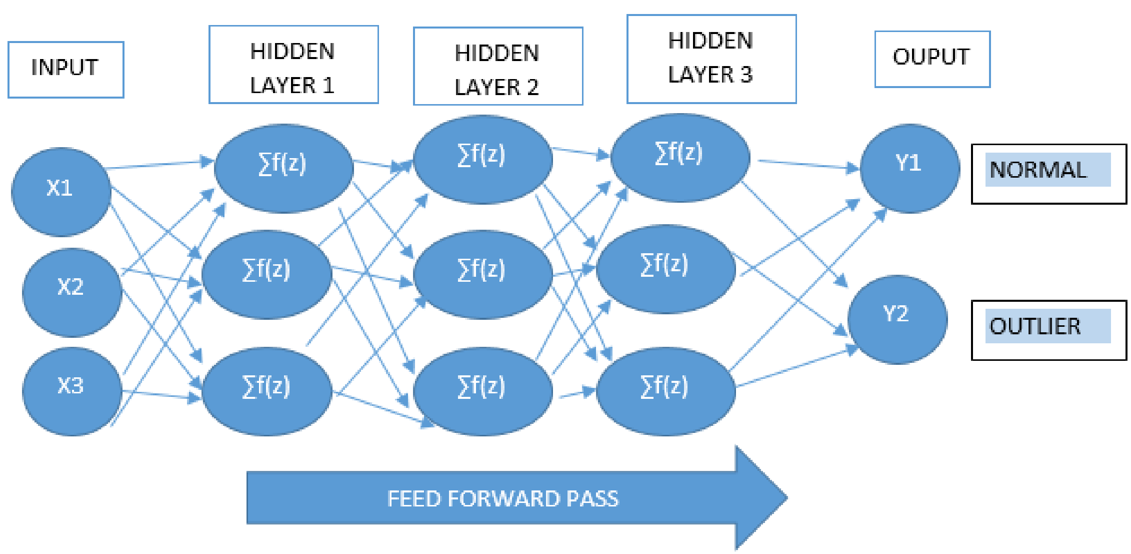


FIG 2: BLOCK DIAGRAM OF DNN

3.4 CNN-ESO FOR CLASSIFICATION

ESO has integrated the advantages of both methods and created a corresponding mathematical framework to measure the behaviors. The combative nature of the Great Egret and the ambush technique of the Snowy Egret influenced the strategies **(Chen et al. 2022)**, the three primary elements of ESO include the aggressive approach, the discerning scenario, and the ambush strategy. ESO operates as a parallel algorithm. A team is made up of three egrets; egret A takes the lead while egrets B and C unpredictably surround their target. Each component is detailed below. While Egret A evaluates a C, it selectively navigates based on the positions of the other egrets and conducts a search aligned with the gradient of the parameters. Meanwhile, Egret B participates in global random exploration. By combining historical information and randomness in its gradient estimation, ESO is more capable of balancing its exploration and exploitation efforts, thus enabling swift searches for viable solutions.

i. Sit-and-Wait Strategy

Considering that the location of the i[th] egret squad is $x_i \in R^n, n$ is the element of problem, $A(*)$ is the Snowy Egret's evaluation method of the potential existence of target in its present surroundings. $\hat{y}$is the evaluation of the prey in the region at the moment,

$$\hat{y}_i = A(x_i) \tag{6}$$

The evaluation technique modified as,

$$\hat{y}_i = w_i . x_i \tag{7}$$

where the $w_i \in R^n$ is the evaluation techniques weight. The error $e_i$defined as,

$$e_i = \|\hat{y}_i - y_i\|^2/2 \tag{8}$$

Also, $\hat{g}_i \in R^n$, the practical slope of ω_i, can be recreated using the incomplete derivative of w_i for the error Equation (8), and its track is $\hat{d}_i$.

$$\hat{g}_i = \frac{\partial \hat{e}_i}{\partial w_i} \tag{9}$$

$$= \frac{\partial \|\hat{y}_i - y_i\|^2/2}{\partial w_i} = (\hat{y}_i - y_i). x_i$$

$$\hat{d}_i = \frac{\hat{g}_i}{\|\hat{g}_i\|} \tag{10}$$

The directional modification of the squad's best position is given by $d_{h,i} \in R^n$, whereas the optimal location for all squads is given by $d_{g,i} \in R^n$.

$$d_{h,i} = \frac{x_{ibest} - x_i}{|x_{ibest} - x_i|} . \frac{f_{ibest} - f_i}{|x_{ibest} - x_i|} + d_{ibest} \tag{11}$$

$$d_{g,i} = \frac{x_{gbest} - x_i}{|x_{gbest} - x_i|} . \frac{f_{gbest} - f_i}{|x_{gbest} - x_i|} + d_{gbest} \tag{12}$$

The integrated gradient$g_i \in R^n$signified as below, and $r_h \in [0,0.5), r_g \in [0,0.5)$:

$$g_i = (1 - r_h - r_g). \hat{d}_i + r_h . d_{h,i} + r_g . d_{g,i} \tag{13}$$

An adaptive weight adjustment technique is used, β_1 is 0.9 and β_2 is 0.99

$$m_i = \beta_1 . m_i + (1 - \beta_1). g_i \tag{14}$$

$$v_i = \beta_1 . v_i + (1 - \beta_1). g_i^2 \tag{15}$$

$$w_i = w_i - m_i / \sqrt{v_i} \tag{16}$$

Based on Egret A's assessment of the existing condition, the site for the following sampling $x_{a,i}$defined as

$$x_{a,i} = x_i + step_a . \exp\left(-\frac{t}{0.1 t_{max}}\right) . hop . g_i \tag{17}$$

$$y_{a,i} = f(x_{a,i}) \quad (18)$$

here t and t_{max} is the difference between the present iteration and upper limit iteration time, whereas hop is the distance between the solution space's low bound and up bound. $step_a \in (0,1]$ is indicated by Egret A's step size factor. $y_{a,i}$ is the fitness of $x_{a,i}$ i .

ii. Aggressive Strategy

The following behavior of Egret B, which has a propensity to constantly look for prey

$$x_{b,i} = x_i + step_b . \tan(r_{b,i}) . hop/(1+t) \quad (19)$$

$$y_{b,i} = f(x_{b,i}) \quad (20)$$

here $r_{b,i}$ is a arbitary number in $(-\pi/2, \pi/2)$, $x_{b,i}$ is Egret B's anticipated position and $y_{b,i}$ is the fitness.

The encompassing mechanism is used by Egret C as a revision method for its location since it chooses to follow its prey actively:

$$D_h = x_{ibest} - x_i \quad (21)$$

$$D_g = x_{gbest} - x_i \quad (22)$$

$$x_{c,i} = (1 - r_i - r_g) . x_i + r_h . D_h + r_g . D_g \quad (23)$$

$$y_{c,i} = f(x_{c,i}) \quad (24)$$

D_h is the break matrix among the group of egrets' present and its ideal position while D_g compares to all Egret squads' optimal location. $x_{c,i}$ is the position of Egret C is anticipated. $step_b \in (0,$ 1] is Egret B's step size factor. r_h and r_g are random numbers in [0, 0.5).

iii. Discriminant Condition

The Egret squad decides on its strategy once each bird has made a decision, then chooses the best course of action and executes it collectively. $x_{s,i}$ is the key matrix of i-th Egret squad

$$x_{s,i} = [x_{a,i}\, x_{b,i} x_{c,i}], \quad (25)$$

$$y_{s,i} = [y_{a,i}\, y_{b,i} y_{c,i}], \quad (26)$$

$$c_i = argmin(y_{s,i}), \quad (27)$$

$$x_i = \begin{cases} x_{s,i|c_i} & if \quad y_{s,i|c_i} < y_i \quad or\, r < 0.3, \\ x_i & else \end{cases} \quad (28)$$

The Egret squad chooses the option if the negligible value of $y_{s,i}$ is higher than existing fitness y_i. Alternatively, if the arbitrary number $r \in (0,1)$ is 30% of the time a worse plan will be accepted if the value is less than 0.3.

This approach, unlike gradient descent, reduces the likelihood of getting stuck at the pivot of the optimization challenge. CNNs are a distinct form of ANN that utilize supervised learning and mimic the processing of layers found in the human visual cortex. This methodology enables the identification of distinct patterns within the input data, facilitating object recognition through a hierarchical arrangement of specialized hidden layers. The initial layers can detect basic elements like curves and lines, and as one delves deeper into the layers, the system can recognize increasingly complex shapes such as silhouettes or human faces. The architecture of a CNN comprises four types of layers: convolutional layers, subsampling layers, fully connected layers, and output layers. In the convolution layer, input features undergo convolution with a kernel (filter). Typically, a kernel within the convolution matrix is known as a screen, while the resulting features from this convolution process are called feature maps of size i*i.

A CNN can incorporate several convolutional layers, where the inputs and outputs of subsequent layers are derived from the feature vector. Each convolution layer contains a set of n filters, which is applied to the input, and the depth of the resulting feature maps (n*) corresponds to the total number of filters used in the convolution. The importance is to note that each filter map represents a specific feature located at a certain position in the input. The output from the l-th convolution layer is denoted as C_i^((l)), consisting of these feature maps. It is computed using

$$C_i^{(l)} = B_i^{(l)} + \sum_{j=1}^{a_i^{(l-1)}} K_{i,j}^{(l-1)} * C_j^{(l)}$$

Where, $B_i^{(l)}$is the bias matrix and $K_{i,j}^{(l-1)}$is convolution filter or kernel of size a*a that connects the jth feature map in layer (l − 1) with the i-th feature map in the same layer **J. Schmidhuber et.al. [106]**

POOLING LAYER

The primary purpose of this layer is to minimize the dimensionality of the feature chart obtained from the preceding convolution layer in a spatial manner. The subsampling process occurs between the mask and the feature maps. Various subsampling techniques have been suggested, including average pooling, sum pooling, and max pooling.

CLASSIFICATION LAYER

The classification layer serves to make the final decision and is typically the last layer in the network. It defines how the training of the network penalizes the variation between the forecasted labels and the actual labels. Generally, the softmax function is employed in the classification layer to generate predictions for cancer. The weight values are optimized using an Egret swarm optimization (ESO).

The most effective data is forecasted, aiding in the adjustment of CNN hyper parameters for cancer detection, facilitated by ESO. In this incident, the fitness value of ESO is defined by a mix of various attributes within the cancer dataset. As a output, the cancer data with the highest fitness among all available options is selected. The suggested algorithm is employed to determine whether the cancer is malignant or benign

Input: Set the pre-defined building blocks of CNN, Population Size (N), Maximum iteration, classification of breast cancer, No. of epochs, weight value, bias value.

Output: The tuned structure of CNN to classify breast cancer data.

STEP 1: Initialization of Population

STEP 2: Compute fitness function evaluation f for all particles using ESO function

STEP 3: Initialize bounds of hyperparameter of CNN

STEP 4: Assume s=stating point solution

STEP 5: For iter=1 to Max iter do

STEP 6: For i=1 to N do

STEP 7: While $f(k) \Leftarrow f(s) do$

STEP 8: For all k ℵ-neighbours(s) do

Generate an $k \Leftarrow$ ℵ-neighbours(s)

if fitness(k)>fitness(s) then

Replace k with s;

End if

End for

End while

End for

End for

ALGORITHM 2: CNN-ESO PSEUDOCODE

ALGORITHM 2: CONVOLUTIONAL NEURAL NETWORK WITH EGRET SWARM OPTIMIZATION

3.5 RESULT:

This research was implemented in **MATLAB R 2018a**. The proposed method undergoes a comprehensive evaluation using various datasets to determine its robustness. To analyze and assess the data, evaluation metrics such as recall, precision, F-Measure, accuracy, and the kappa statistic are employed.

Accuracy refers to how well objects are correctly identified.

$$Accuracy = \frac{TP+TN}{Total\ no.of\ Samples}$$

Precision (P) relates to the degree of exactitude of the classifier.

$$\text{Precision (P)} = \frac{TP}{TP+FP}$$

Recall (R) is used to judge the sensitivity or effectiveness of the classifier

$$\text{Recall (R)} = \frac{TP}{TP+FN}$$

F-Measure serves as the harmonic mean of precision and recall.

$$F - Measure = 2 * \left[\frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}\right]$$

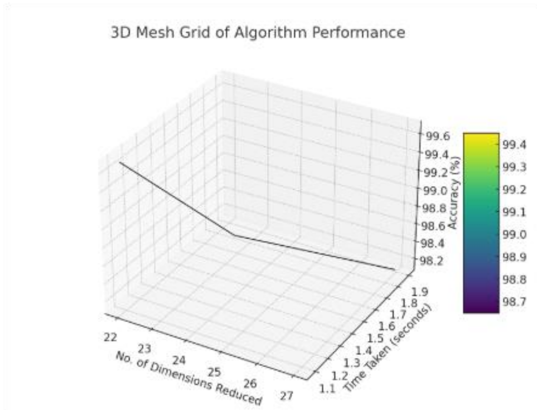
Kappa statistics is the distinction between observed and anticipated agreement serves as an alternative evaluation metric. This value ranges from 0 to 1, where a value of 1 signifies complete agreement.

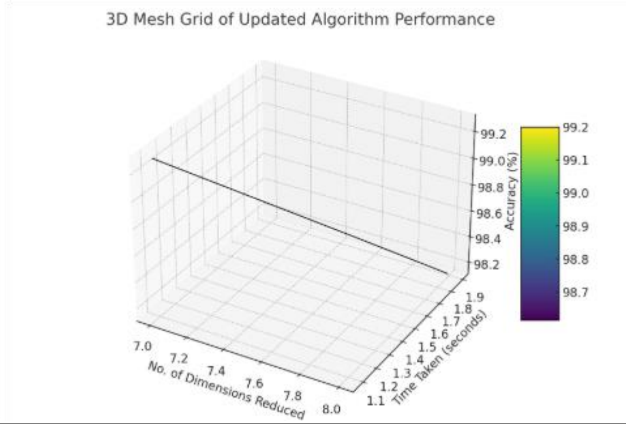
$$K = \frac{p_0 - p_e}{1 - p_e} = 1 - \frac{1 - p_o}{1 - p_e}$$

<u>DIMENSIONALITY REDUCTION</u>

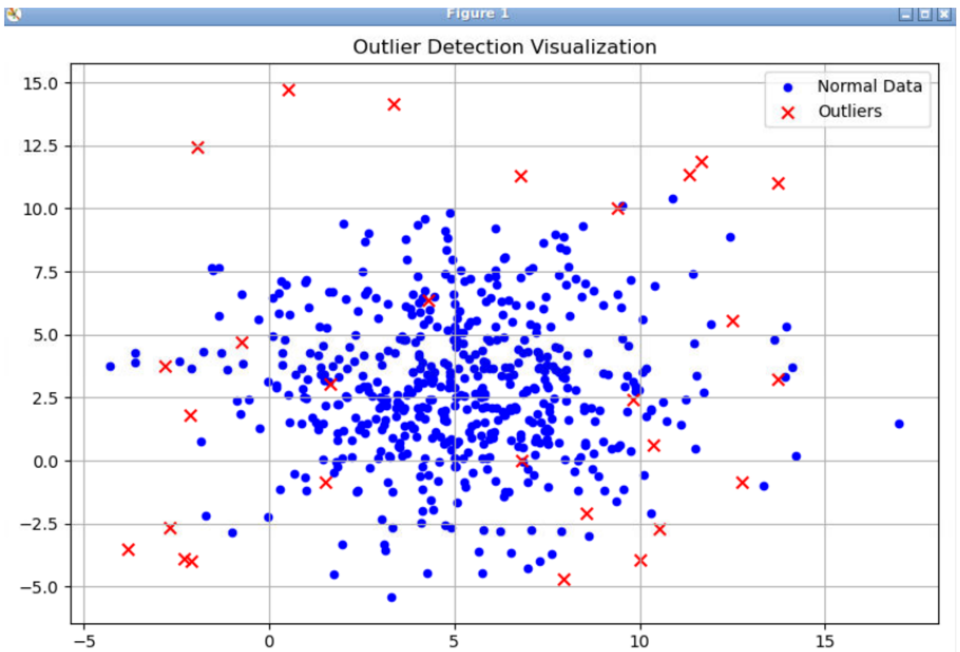
S.NO	ALGORITHMS	TIME TAKEN (WDBC)	ACCURACY(%) (WDBC)	TIME TAKEN (WBC)	ACCURACY(%) (WBC)
1	PCA	1.8930 s	98.10	1.8891 s	98.13
2	ICA	1.14 s	99.2	1.134 s	99.1

3	ELDA	1.093 s	99.7	1.087 s	99.3

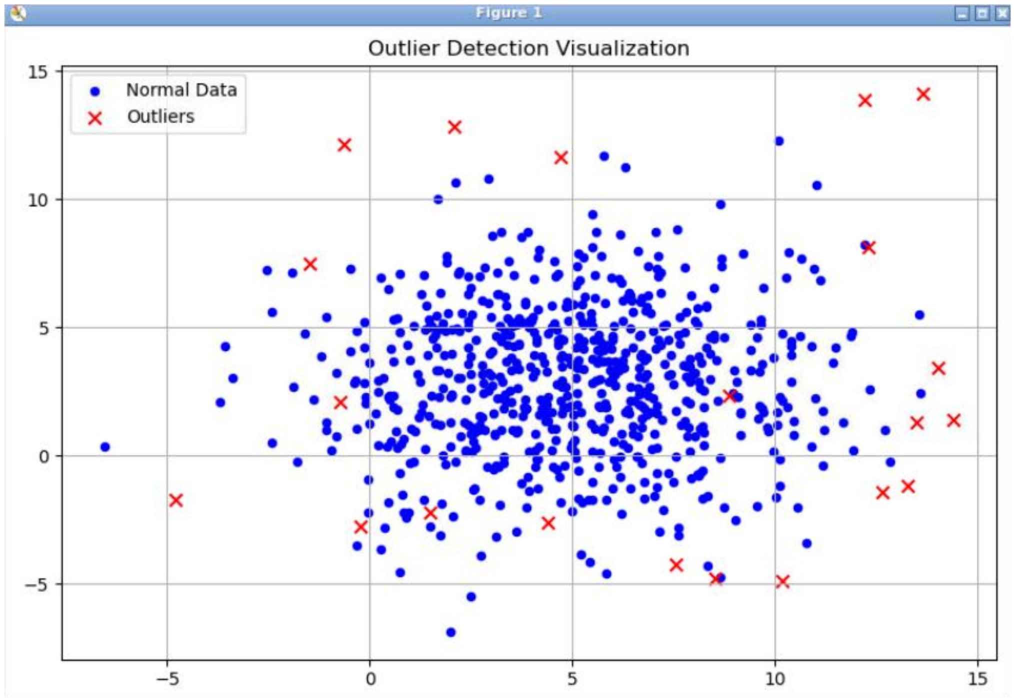




OUTLIER DETECTION -WDBC Dataset

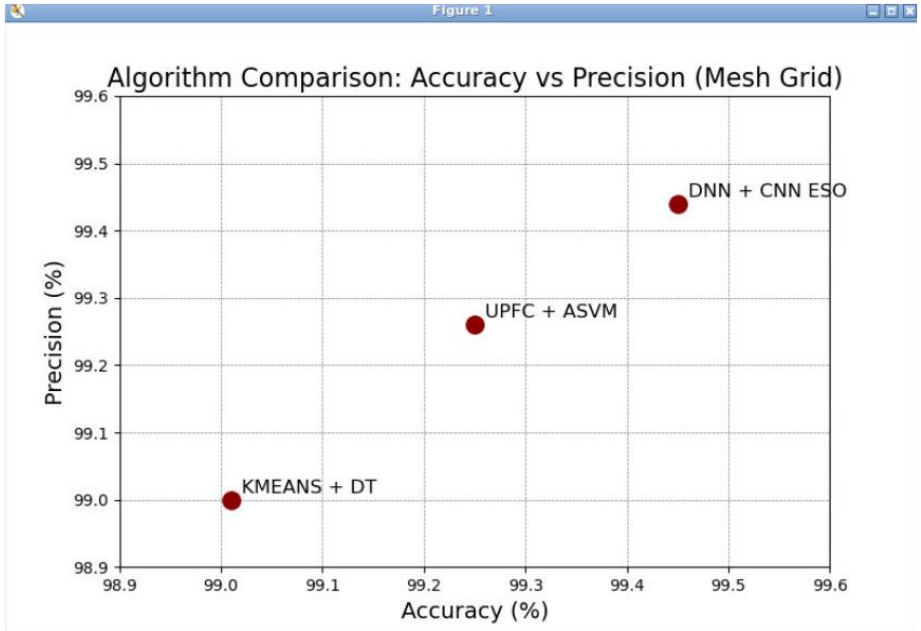


OUTLIER DETECTION-WBC DATASET



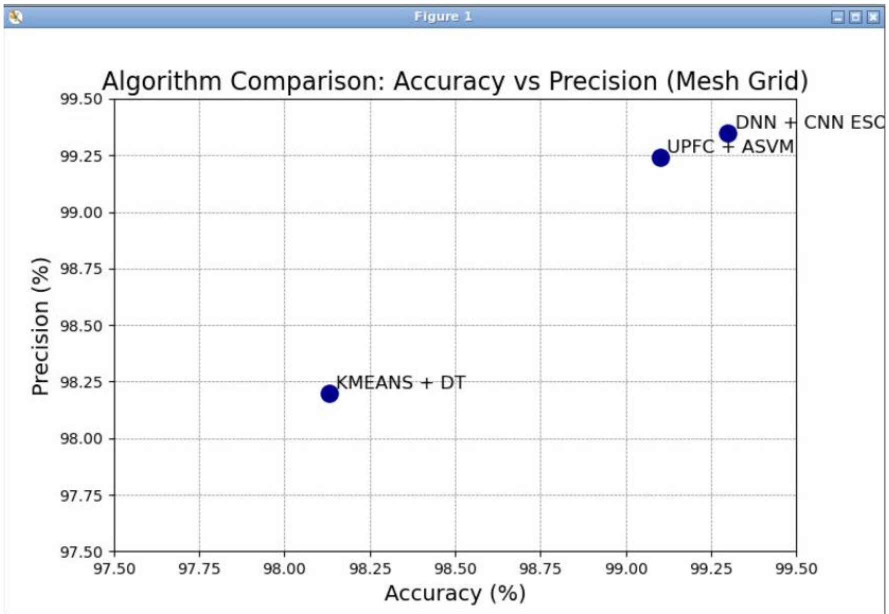
CLASSIFICATION ACCURACY-WDBC Dataset

SNO	ALGORITHM	ACCURACY	PRECISION	F MEASURE	RECALL	KAPPA STATISTICS
1	KMEANS with DT	99.01	99.0	99.1	99.0	97.87
2	UPFC+ASVM	99.25	99.26	99.29	99.33	99.37
3	DNN+CNN ESO	99.45	99.44	99.50	99.53	98.55



CLASSIFICATION ACCURACY-WBC Dataset

SNO	ALGORITHM	ACCURACY	PRECISION	F MEASURE	RECALL	KAPPA STATISTICS
1	KMEANS with DT	98.13	98.2	98.4	98.1	96.2
2	UPFC+ASVM	99.10	99.24	99.12	99.01	98.0
3	DNN+CNN ESO	99.30	99.35	99.36	99.42	98.48



3.6 CONCLUSION

In summary, the early identification and precise categorization of breast cancer (BC) are vital for successful treatment. The proposed ELDA-DNN-CNN ESO algorithm incorporates three essential processes: Dimensionality Reduction, outlier detection, and classification, showing outstanding performance metrics. The model achieves remarkable classification accuracies of 99.30% and 99.45% for the Wisconsin BC (WBC) and Wisconsin Diagnostic BC (WDBC) datasets, respectively. When compared to current methods, the ELDA-DNN-CNN ESO algorithm excels in accuracy, precision, F-measure, recall, and kappa statistics, as demonstrated through 10-fold cross-validation. These results highlight the efficacy of merging DL with evolutionary strategies for BC diagnosis, providing a promising resource for early identification and precise categorization. This progress has considerable potential to enhance patient outcomes and streamline treatment approaches through accurate and timely diagnosis.

3.7 REFERENCES

[1]. Mridha, M. F., Hamid, M. A., Monowar, M. M., Keya, A. J., Ohi, A. Q., Islam, M. R., & Kim, J. M. (2021). A comprehensive survey on deep-learning-based breast cancer diagnosis. Cancers, 13(23), 6116.
[2]. Mao, N., Yin, P., Wang, Q., Liu, M., Dong, J., Zhang, X., ... & Hong, N. (2019). Added value of radiomics on mammography for breast cancer diagnosis: a feasibility study. JACR, 16(4), 485-491.
[3]. Wang, H., Feng, J., Bu, Q., Liu, F., Zhang, M., Ren, Y., & Lv, Y. (2018). Breast mass detection in digital mammogram based on gestalt psychology. J. Healthc. Eng., 2018
[4]. Valvano, G., Santini, G., Martini, N., Ripoli, A., Iacconi, C., Chiappino, D., & Della Latta, D. (2019). Convolutional neural networks for the segmentation of microcalcification in mammography imaging. J. Healthc. Eng., 2019.
[5]. Shaban, W. M., Rabie, A. H., Saleh, A. I., & Abo-Elsoud, M. A. (2020). A new COVID-19 Patients Detection Strategy (CPDS) based on hybrid feature selection and enhanced KNN classifier. Knowledge-Based Systems, 205, 106270.
[6]. Haq, A. U., Zeb, A., Lei, Z., & Zhang, D. (2021). Forecasting daily stock trend using multi filter feature selection and deep learning. Expert Systems with Applications, 168, 114444.
[7]. Nguyen, B. H., Xue, B., & Zhang, M. (2020). A survey on swarm intelligence approaches to feature selection in data mining. Swarm and Evolutionary Computation, 54, 100663.
[8]. Rabie, A. H., Ali, S. H., Saleh, A. I., & Ali, H. A. (2020). A new outlier rejection methodology for supporting load forecasting in smart grids based on big data. Cluster Computing, 23, 509-535.

[9]. Amarasingh, K., Kenney, K., & Manic, M. (2018, July). Toward explainable deep neural network-based anomaly detection. In 2018 11th international conference on human system interaction (HSI) (pp. 311-317). IEEE.
[10]. Munir, M., Siddiqui, S. A., Dengel, A., & Ahmed, S. (2018). DeepAnT: A deep learning approach for unsupervised anomaly detection in time series. Ieee Access, 7, 1991-2005.
[11]. Gao, J., Song, X., Wen, Q., Wang, P., Sun, L., & Xu, H. (2020). Robusttad: Robust time series anomaly detection via decomposition and convolutional neural networks. arXiv preprint arXiv:2002.09545.
[12]. Gómez-Flores, W., & Hernández-López, J. (2020). Assessment of the invariance and discriminant power of morphological features under geometric transformations for breast tumor classification. Computer methods and programs in biomedicine, 185, 105173.
[13]. Lahoura, V., Singh, H., Aggarwal, A., Sharma, B., Mohammed, M. A., Damaševičius, R., ... & Cengiz, K. (2021). Cloud computing-based framework for breast cancer diagnosis using extreme learning machine. Diagnostics, 11(2), 241.
[14]. Liu, Y., Ren, L., Cao, X., & Tong, Y. (2020). Breast tumors recognition based on edge feature extraction using support vector machine. Biomedical Signal Processing and Control, 58, 101825.
[15]. Irfan, R., Almazroi, A.A., Rauf, H.T., Damaševičius, R., Nasr, E.A. and Abdelgawad, A.E., (2021). Dilated semantic segmentation for breast ultrasonic lesion detection using parallel feature fusion. Diagnostics, 11(7), p.1212.
[16]. Adebiyi, M. O., Arowolo, M. O., Mshelia, M. D., & Olugbara, O. O. (2022). A linear discriminant analysis and classification model for breast cancer diagnosis. Applied Sciences, 12(22), 11455.
[17]. Sun, W., Tseng, T. L. B., Zhang, J., & Qian, W. (2017). Enhancing deep convolutional neural network scheme for breast cancer diagnosis with unlabeled data. Computerized Medical Imaging and Graphics, 57, 4-9.
[18]. Adege, A. B., Lin, H. P., Tarekegn, G. B., & Jeng, S. S. (2018). Applying deep neural network (DNN) for robust indoor localization in multi-building environment. Applied Sciences, 8(7), 1062.
[19]. Chen, Z., Francis, A., Li, S., Liao, B., Xiao, D., Ha, T. T., ... & Cao, X. (2022). Egret swarm optimization algorithm: an evolutionary computation approach for model free optimization. Biomimetics, 7(4), 144.
[20]. Goni, M. O. F., Hasnain, F. M. S., Siddique, M. A. I., Jyoti, O., & Rahaman, M. H. (2020, December). Breast cancer detection using deep neural network. In 2020 23rd International Conference on Computer and Information Technology (ICCIT) (pp. 1-5). IEEE.
[21]. Yusuf, A. B., Dima, R. M., & Aina, S. K. (2021). Optimized breast cancer classification using feature selection and outliers detection. JNSPS, 298-307.
[22]. J. Schmidhuber, "Deep Learning in Neural Networks: An Overview," Neural Networks, vol. 61, pp. 85–117, Jan. 2015. https://doi.org/10.1016/j.neunet.2014.09.003