



# Enhancing Breast Cancer Classification: A Deep Learning Approach with Outlier Detection with Egret Swarm Optimization Algorithm and Feature Selection Integration

S. Maria Sylvania<sup>1\*</sup>, N. Sudha<sup>1</sup>

## Abstract

**Background:** Breast cancer (BC) remains a significant global health concern, with high incidence and mortality rates. Early detection and accurate classification are crucial for effective treatment. Traditional BC treatments include surgery, radiation, and medication aimed at eliminating microscopic malignancies. Advances in machine learning (ML) and deep learning (DL) have shown promise in enhancing BC diagnosis and classification accuracy. **Method:** This study used a novel classification model for BC using a Deep Neural Network-Genetic Algorithm-Evolutionary Strategy Optimization (DNN-GA-ESO) approach. The methodology involves three key phases: data preprocessing using Improved Linear Discriminant Analysis (ILDA), outlier detection via a Deep Neural Network (DNN), and feature selection using a Genetic Algorithm (GA). Finally, the Evolutionary Strategy Optimization (ESO) algorithm classifies the BC data as benign or malignant. The effectiveness of this approach was validated using the Wisconsin Breast Cancer (WBC) and Wisconsin Diagnosis Breast Cancer (WDBC) datasets.

**Significance** | Early Breast Cancer diagnosis via data mining is crucial. Proposed methods achieve high accuracy, improving treatment outcomes for patients significantly.

\*Correspondence. S. Maria Sylvania, Department of Computer Science, Bishop Appasamy College of Arts and Science, Coimbatore.  
E-mail: mariasylvaa1991@gmail.com

Editor Deepak Kumar Sahu, And accepted by the Editorial Board Mar 03, 2024 (received for review Jan 07, 2024)

**Results:** The proposed DNN-GA-ESO method demonstrated superior performance in classifying BC. For the WBC dataset, the model achieved an accuracy of 99.30%, precision of 99.35%, F-measure of 0.9936, recall of 99.42%, and kappa statistic of 99.48%. For the WDBC dataset, it achieved an accuracy of 99.45%. These results significantly outperformed existing methods such as K-means with Decision Tree, UPFC with ASVM, and other standard ML algorithms. **Conclusion:** The DNN-GA-ESO approach enhances BC classification accuracy through efficient outlier detection and feature selection. This method surpasses traditional and current techniques, providing a more reliable and precise diagnostic tool for early BC detection. The integrated meta-algorithm offers a promising solution for medical diagnostics, potentially improving patient outcomes through early and accurate detection of BC.

**Keywords:** Deep learning, Parkinson's disease, VGG19, MRI images.

## Introduction

Breast cancer (BC) treatment can be highly effective, especially when the disease is detected early. Most BC treatments involve surgery, radiation, and medication, targeting microscopic malignancies that have entered the circulation from a breast tumor. This approach can save lives and halt the growth and spread of cancer. By 2020, there were 2.3 million new cases diagnosed, and 685,000 individuals died from the disease. The World Health

### Author Affiliation.

<sup>1</sup> Department of Computer Science, Bishop Appasamy College of Arts and Science, Coimbatore, India.

### Please cite this article.

S. Maria Sylvania, N. Sudha (2024). Enhancing Breast Cancer Classification: A Deep Learning Approach with Outlier Detection with Egret Swarm Optimization Algorithm and Feature Selection Integration, *Journal of Angiotherapy*, 8(3), 1-13, 9520

2207-8843/© 2024 ANGIOTHERAPY, a publication of Eman Research, USA.  
This is an open access article under the CC BY-NC-ND license.  
(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).  
(<https://publishing.emanresearch.org>).

Organization (WHO) predicts that BC will be the most common cancer globally (Mridha et al., 2021). Following lung cancer, 7.8 million people have been diagnosed with BC in the past five years. The most common types of BC, which can be identified using a microscope, are invasive ductal carcinoma (IDC) and ductal carcinoma in situ (DCIS). DCIS grows more slowly and has less impact on the patient's daily activities. Up to 80% of BC cases are diagnosed as IDC, which is more deadly as it involves the entire breast tissue, whereas DCIS accounts for 20-53% of cases. IDC affects around 80% of people with BC. BC causes the most disability-adjusted life years (DALYs) lost compared to all other cancer types combined (Mao et al., 2019).

Any woman after puberty, in any region of the world, can develop BC at any time. The number of women who passed away from BC between 1930 and 1970 remained essentially steady. However, in the 1980s, life expectancy started to increase in countries that implemented various treatments to eliminate invasive illnesses and early detection systems. Unlike other cancers, BC is not communicable. While there is a clear association between HPV infection and cervical cancer, there is no established correlation between viral or bacterial infections and BC. Approximately 50% of BC cases occur in women who are 40 years of age or older. BC is more common in women who take hormone therapy after menopause, consume alcohol excessively, or have a family history of the illness.

A lump or a painless thickening of the breast are both signs of BC. Regardless of how uncomfortable a lump is, women should seek medical assistance as soon as they detect one in their breasts (Wang et al., 2018). There are numerous causes of breast lumps, the majority of which are harmless. About 90% of breast lumps are benign and do not pose any health risks. Non-malignant abnormalities include breast infections and benign tumors such as cysts and fibroadenomas. A comprehensive medical evaluation is necessary (Valvano et al., 2019).

Breast imaging is often used to determine whether a tumor is malignant or benign, sometimes involving the removal of a tissue sample for further analysis. Women who experience chronic symptoms lasting more than one month should undergo tests such as breast imaging and tissue sampling (biopsy) (Devi & Devi, 2016). Machine Learning (ML) algorithms have made significant advancements in diagnosing and classifying breast cancer (BC) over the past few decades. These techniques involve three main procedures: preprocessing, feature extraction, and classification. Preprocessing mammography films to enhance their peripheral and intensity distributions can improve the evaluation and interpretation of images. Various methods have been developed for this purpose, many of which have proven effective (Khan et al., 2019). The application of ML is expanding and may soon be offered as a standard service. However, ML remains a complex field that

typically requires specialized knowledge and skills. A reliable ML approach demands a broad range of competencies, including feature engineering, preprocessing, and classification techniques (Huang et al., 2019).

Some ML algorithms, such as support vector machines (SVM) and K-nearest neighbors (KNN), often fail to perform as expected due to the presence of unnecessary features (Shaban et al., 2020). Eliminating certain characteristics before applying ML algorithms can significantly enhance the accuracy of classification systems (Haq et al., 2021). Feature selection involves identifying the essential qualities from the original set. The primary goals of the feature selection process are to prevent overfitting, increase accuracy, improve learning performance, and reduce computational costs (Haq et al., 2021). Feature selection procedures are categorized into two main types: filter and wrapper techniques (Abualigah & Dulaimi, 2021). Filtering techniques identify important features without initially classifying the data, while wrapper techniques focus on the most critical attributes using classification algorithms. Wrapper techniques are generally preferred over filter techniques for classification accuracy (Chaudhuri & Sahu, 2021).

The use of swarm intelligence algorithms for feature selection has significantly increased. Swarm intelligence (SI) is a technique for modeling the collective behavior of biological groups and consists of multi-agent frameworks inspired by nature's swarm behaviors (Sun et al., 2020). It mimics the behaviors of animal herds competing for territory. SI has been applied to various complex problems, including guiding artificial intelligence, predicting social behaviors, and enhancing networking and telecommunications (Xue & Shen, 2020). The feature selection community finds SI appealing due to its accessibility and capability to perform global searches.

Examples of swarm intelligence techniques include ant colony optimization (ACO), particle swarm optimization (PSO), grey wolf optimization (GWO), and the bat algorithm (BA), among others (Nguyen et al., 2020). However, feature selection alone does not always yield the best classification performance, especially when datasets contain defective or noisy data (Rabie et al., 2020). To optimize the efficacy of classification techniques, it is crucial to exclude such noisy data (Rabie et al., 2020).

Outlier rejection, the practice of removing or discarding data that significantly deviates from the norm, is essential before classification. Many machine learning algorithms regard outliers as noise, which must be removed to improve the system's predictive capabilities. There are two types of outlier strategies: the standard outlier approach and the spatial outlier technique (Chambers et al., 2020).

The main aim of this research is to enhance breast cancer (BC) classification using a deep learning (DL) based outlier detection algorithm. Despite numerous research efforts and methodologies, misclassification and lower accuracy persist in BC datasets. To address these issues, this study proposes the DNN-GA-ESO approach, which aims to improve overall BC classification performance. The primary contributions of this study include data pre-processing, feature selection, outlier identification, and the classification procedure.

The ESO classification is employed for the BC dataset to more accurately determine whether an output class is benign or malignant. By utilizing efficient algorithms, the proposed technique yields more precise results for the specified dataset.

## 1. Literature Review

A considerable number of current outlier identification methods are based on machine learning (ML). However, some modern outlier detection methods based on deep learning (DL) have also been introduced. Chambers et al. (Amarasinghe et al., 2018) proposed a novel DNN-based method called DeepStreamCE, which utilizes deep neural network activations in a streaming setting to identify concept evolution. Another notable study by Munir et al. (2018) provides a framework for DNN-based anomaly identification, offering explanations for any detected anomalies.

The developers of DeepAnT propose a DL-based solution for anomaly recognition (Gao et al., 2020). This method involves two stages: a time-series predictor module and an anomaly detection module. DeepAnT trains convolutional neural networks (CNN) on raw data without removing anomalies, using a max-pooling layer followed by two convolutional layers.

In a recent study by Gómez-Flores and Hernández-López (2020), a CNN is proposed as a technique for using disruptions for data augmentation in the time series detection of abnormalities in the phase and amplitude spectra of the frequency domain. However, this approach has issues with lower convergence rates.

Gómez-Flores and Hernández-López (Liu et al., 2020) suggested a CAD-assisted categorization of breast cancer (BC). They utilized 39 morphological parameters to define breast tumor forms and help specify BC types. The study employed 892 mammography images from the BC Digital Repository (BCDR) and 2054 ultrasound images from the National Cancer Institute (INCa) of Rio de Janeiro, Brazil, for training. Morphological features, including elongatedness, convexity, eccentricity, circularity, and area variations, were extracted based on border and region descriptors. The AUC for both databases was determined to be 82.0%. However, the approach faced issues with misclassification error rates.

Liu et al. (Irfan et al., 2021) introduced an edge feature extraction-based CAD for breast tumor classification. The region of interest

(ROI) was used to extract morphological properties such as roundness, aspect ratio, ellipticity, regularity, and roughness. The SVM classifier was used to classify images as either malignant or benign lesions. The proposed approach utilized 192 ultrasound images, including 71 cancerous and 121 benign cases. The technique achieved an accuracy of 67.31%, sensitivity of 47.62%, specificity of 80.65%, and PPV of 62.50%.

Irfan et al. (Lahoura et al., 2021) suggested deep learning (DL) segmentations of ultrasonic breast lesion images using morphological erosions and dilated semantic segmentation networks (Di-CNN). The segments of the images were loaded into DenseNet201 with transfer learning for feature extraction. This research utilized 780 breast tumors. Using CNN-activated feature vectors combined with the SVM classifier, the method achieved an accuracy of 90.11%, precision of 98.45%, and a high precision of 98.9%. However, it faced issues with overfitting.

Lahoura et al. (Adebisi et al., 2022) suggested an extreme learning machine (ELM)-based ML model for cloud-based BC categorization. The ELM model was applied using the AdaBoost, Naive Bayesian, SVM, perceptron, and k-NN methods. The dataset used was the Wisconsin BC Diagnosis (WBCD), comprising 32 attributes and 569 items. The findings showed that the approach achieved an accuracy of 98.68%, precision of 90.54%, recall of 91.30%, and an F1-score of 81.29%.

## 2. Materials and Methods

We have used a technology with a three-phase procedure for diagnosing cancer, as illustrated in Figure 1. In the first phase, data dimensionality is reduced using Hybrid Independent Component Analysis (HICA). This step simplifies the dataset while retaining essential information. In the second phase, a Deep Neural Network (DNN) is employed for outlier detection within the cancer dataset, identifying any data points that deviate significantly from the norm. Following this, feature selection is performed using a Genetic Algorithm (GA). GAs are frequently used for search and optimization tasks due to their efficiency in locating the optimal or near-optimal features in large and complex search spaces. This makes them particularly suitable for diagnosing cancer by developing models to interpret test results, especially those from non-invasive procedures. In the final phase, the Evolutionary Strategy Optimization (ESO) algorithm is utilized to determine whether the identified malignancy is benign or malignant. The effectiveness of this proposed technique has been validated using the Breast Cancer (BC) dataset.

### 2.1. Input Cancer Dataset

The measures of BC cases have been recorded using the BC dataset. BC in Wisconsin (WBC) and Wisconsin Diagnosis BC data are utilized to make an evaluation. It comes from the UCI data warehouse. There are 699 occurrences of WBC, two benign and

malignant classes, and 9 integer-valued properties. For the next step, a WDBC with 568 instances, benign and malignant classes and 32 integer-valued properties is employed.

**2.2. Preprocessing using Improved Linear Discriminant Analysis (ILDA) Algorithm**

In supervised ML, a technique known as linear discriminant analysis can be used to streamline categorization issues. It is used for modelling set differences because it needs discriminating among two or more groups. A feature can be transferred from A single dimension into a space with lower dimensions using this tool. To increase between-class variance to within-class variation, aspects are downscaled from high-dimensional to low-dimensional regions (24). Linear discriminant analysis (LDA) is widely used in ML classifications (Ricciardi et al. 2020). Features are reduced in dimension by its capacity to limit between-class variation to within-class variance. Thus, an improvement to LDA (Algorithm 1) is shown by calculating the class means and prior probability by merely providing the independent variables for the sample (Zhao et al. 2018). Each group in the dataset has its average vector dimensions calculated. To determine the scattering matrix, the k Eigenvectors with the highest Eigenvalues may be selected, and the Eigenvectors can be arranged in descending order to create a dk matrix W, involving Eigenvectors  $(e_1, e_2, \dots, e_d)$  and the associated Eigenvalues  $(1, 2, \dots, d)$ . Using the eigenvector matrix W created before, the data is then recreated in a new subspace. If  $Y = XW$ , its matrix representation is simple. Each group's covariance matrix is computed together with the pooled one. Labeling classes and finding the LDA discriminant function follow.

**Algorithm 1: Improved Linear Discriminant Analysis**

1.  $En = xUj | yj = cn, j = 1, \dots, m - 1, n = 1, 2$   
//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.  $Zn = En - mn\mu_i, n = 1, 2$ //standards for the middle class
5.  $Tn = ZUnZn, n = 1, 2$ // standards for the class scatter
6.  $T = T1 + T2$ //internal scattering
7.  $\lambda_1, x = eigen(T - 1C)$  //estimate central eigenvector m

**2.3. Outlier detection for removing the outliers using Deep Neural Network (DNN)**

A decision-making algorithm based on artificial neural networks is known as DNN (Adege et al. 2018) (Sun et al. 2017)). Neurons, a type of neural network, comprise the DNN. A deep neural network has numerous hidden layers among the output and input layers, as implied by the name. A single input, output, and three hidden layers make up a deep neural network, is constructed during the DNN

training phase to improve the abstraction characteristics for greater capacity. The research's DNN relies on the idea of a feed-forwards of artificial neural networks. Input training data is used particularly when characteristics in the training data are equal to neurons in input layer nodes (Yusuf et al. 2021). To link the output layer nodes to the link weights that connect the input layer (l) to the hidden layer nodes, weights (w) are first created using an activation function. Two neurons are present in the output layer nodes in Figure 2 for categorizing events as normal or outliers.

Based on the results of each previous layer of the network, the ReLU function is used in this case as Eq (2) calculates bias (b) from weighted inputs for the next network layer using the activation function for all three hidden levels.

$$Relu(x) = max(0, x) \tag{1}$$

Formally, equations (3) and (4) ascertain outputs of hidden layers inside networks:

$$x^l = f(z^l) \tag{2}$$

$$z^l = w^l x^{l-1} + b^l \tag{3}$$

The output activation function,  $f(z^l)$ , as in Eq. (2), the Relu function generates l, current layers, and w and b, hidden layers' weights and biases.

Completely integrated layers act as final output decision layer and are appropriate with sigmoid activations determined by Eq. (5). Sigmoid classification are is used as it evaluates representations of neurons as outliers in normal data.

$$sigmoid(z) = \frac{1}{1 + e^{-z}} \tag{4}$$

The data instances  $x = (x_1, x_2, \dots, x_N)$  are mapped to an outlier score, which indicates the probability of being an outlier, using the sigmoid activation function.

According to statistics, normal instances are more likely to occur because they are measured more frequently, but outlier examples are more likely to occur since they are unusual events (Goni et al. 2020). Since the values of probability range among (0, 1), this knowledge allows to distinguish among outlierness and normalcy when classifying data examples as outliers or normal. Particularly, a larger probability indicates an outlier to a greater degree.

In Equation (6), the  $\epsilon(x|\theta)$  function takes the outliers score into account.

$$\epsilon(x|\theta) = \sum_{i,j=1}^k w_{ij} z_i z_j \tag{5}$$

The weight of the interaction is indicated by the trainable parameter  $w_{ij}$ , where  $z_i$  integrates x's ith feature values in representative space Z in low dimensions. DNN training steps produce models that find outliers and normal training data.

The suggested framework's detection phase is its last step. The unknown information is classified as either outliers or normal instances using the trained deep neural network model. The trained DNN model is then tested on the test data, and it gives an unseen instance an outlieriness score. The final outlieriness score is calculated using the sigmoid response vectors' cumulative probability.

**2.4. Feature Selection using Genetic Algorithm (GA)**

In this work, feature selection is done by using GA. As a type of inductive learning approach, GA offers several advantages for effective feature selection. It can quickly and efficiently find near-optimal solutions in nonlinear and complicated search spaces. GA is a stochastic generic search technique that can efficiently comb across broad search spaces, which is typically necessary for selecting attributes. Furthermore, GAs do a global search as opposed to many other search algorithms' limited, greedy searches. Reproduction, crossover, and mutation are the three operators that make up GA (Rupali et al. 2021) (Alzubaidi et al. 2016). Crossover joins excellent strings to try to generate better children; mutation modifies a string locally to try to build a better string. Reproduction chooses a good string (subset of input characteristics). Binary bits make up the string: 1 denotes the attribute selection, while 0 denotes dropping the attribute. The population is assessed and checked for algorithm termination in each generation. The population is operated upon by the three GA operators and then reevaluated if the termination condition is not met. After a predetermined number of generations, this process is repeated. The GA is built on the ideas of "Genetics and Natural Selection" is utilized to determine input values using this output values obtained. The fittest individuals are chosen for reproduction using a genetic algorithm, which is a form of natural selection that produces children for the following generation. It simply serves as a method of problem-solving to offer an ideal response (Mishra et al. 2020).

Figure 3 shows the nature of GA which are exchanges of bits by selected parents to produce offsprings. Bit counts  $b$  selected from parents  $P_n$  with parameters:  $0 < k < 1$

**GA algorithm pseudo code:**

Input: BC datasets

Output: Best features

Start

using a random selection process to create the first population from the BC dataset

Do for the designated number of generations

Given the population's size (features),

Using uniform probability, choose two people to be parent1 and parent2 (the characteristics of the BC patient).

Transition to create a new person

alteration in a kid

Determine the distance  $d1$  between the child and parent1, and  $d2$  between the child and parent 2.

Determine the fitness of the following: kid, parent1, parent2, and  $f$ , correspondingly.

*if*( $d1 < d2$ )*and* ( $f > f1$ )*then*

Replace parent1 with child

Else

*f*( $d2 \leq d1$ )*and* ( $f > f2$ ) *then*

Replace parent2 with child.

Feature modified step:

Repeat

Population  $\leftarrow$  feature subset  $F_m$

Generation =0;

Loop for  $i$  from 1 to size population (do) using  $b = K * P_n$

S1  $\leftarrow$  selection (population, fitness)

S2  $\leftarrow$  selection (population, fitness)

Child  $\leftarrow$  crossover (s1, s2) check feasibility of  $n$  element

Child  $\leftarrow$  mutate (child) check feasibility of  $n$  element

Fitness (more relevant and important features)

Generation = generation +1

Until generation < max\_generation

M=m+1

End if

End if

End for

End for

Obtain best individual solution as optimal results

**2.5. ESO for Breast cancer data classification**

ESO has incorporated the benefits of both techniques and developed a corresponding mathematical model to quantify the behaviours. The Great Egret's aggressive style and the Snowy Egret's sit-and-wait strategy drove the strategies (Chen et al. 2022). As seen in Figure 4, the three main components of ESO are the violent approach, the discriminating scenario, and the sit-and-wait technique. ESO is a parallel algorithm. Three egrets make up one squad; egret A leads the way, while egrets B and C move erratically and encircle. Each component is described in depth below.

Figure 4 depicts the various characters and search choices of the Egret Squad. While Egret C selectively investigates according to the location of better egrets, On the basis of the parameters' gradient, Egret A will determine a descent plane and conduct a search. Egret B will engage in global random wandering. ESO will be better able to balance its exploration and exploitation efforts and conduct quick searches for workable solutions as a result. Because ESO uses both historical data and stochasticity in the gradient estimate, unlike gradient descent, it is less likely to reach the saddle point of the optimisation problem. Unlike other meta-heuristic methods,



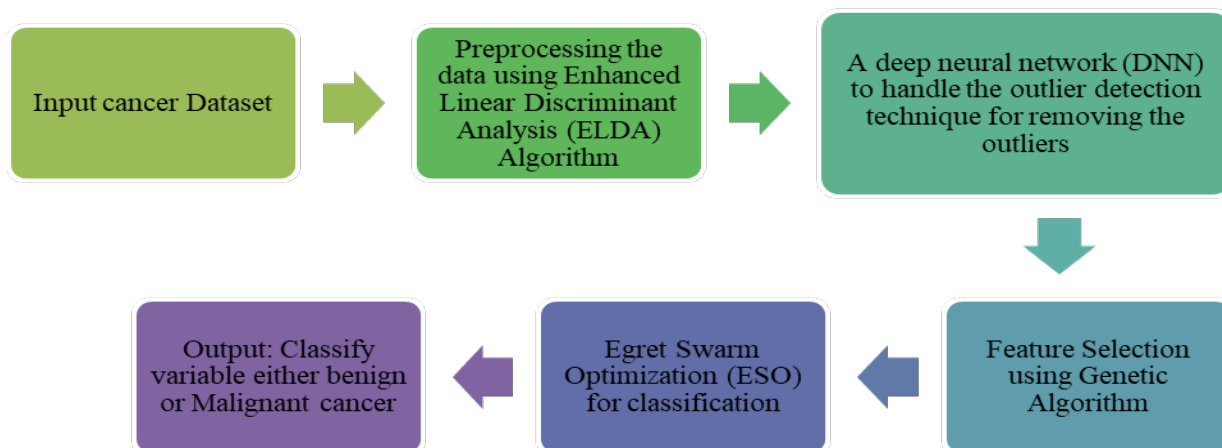


Figure 1. Overall Flow of the Proposed Review Methodology

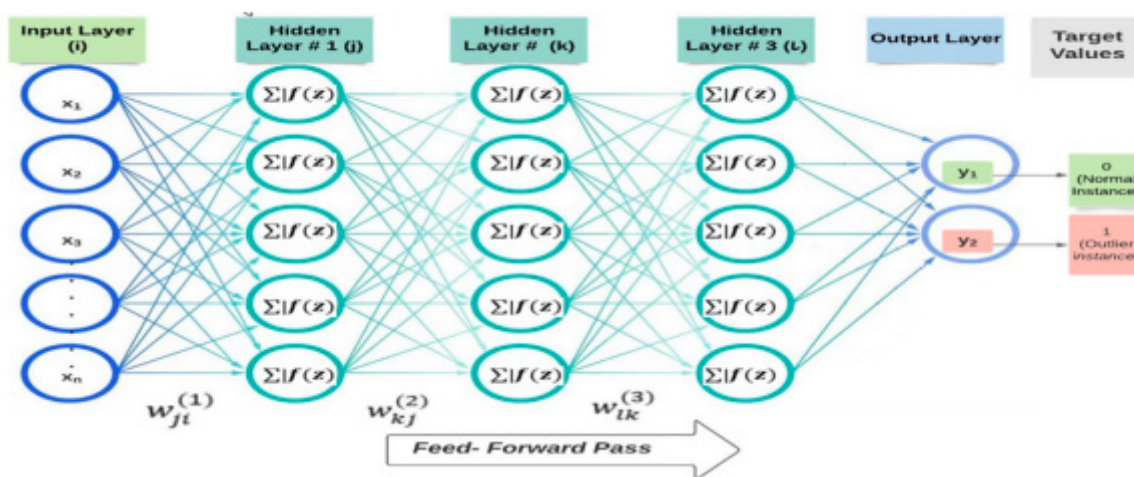


Figure 2. DNN Block for outlier detection

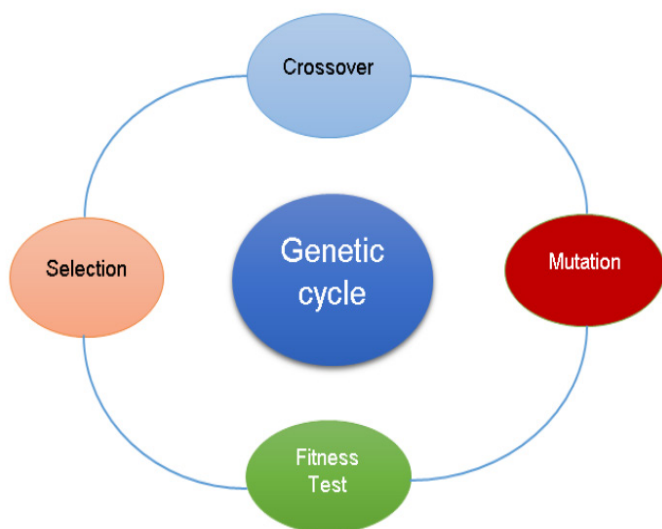


Figure 3. Nature of genetic algorithm

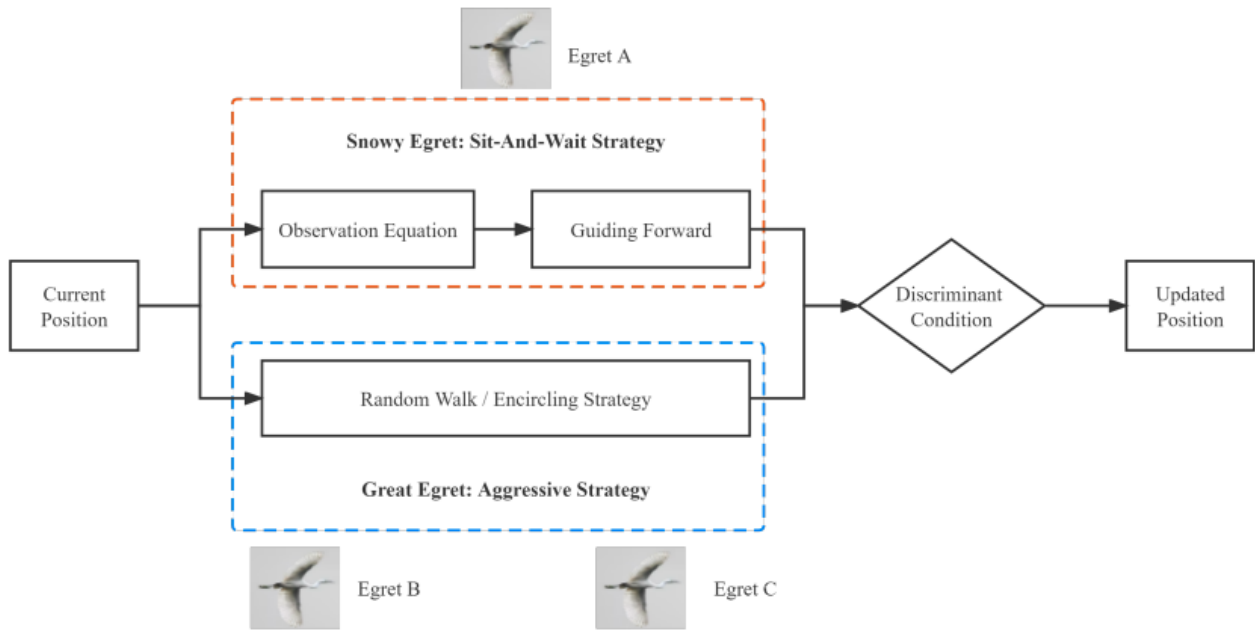


Figure 4. The Egret Swarm Optimization Algorithm's Framework.

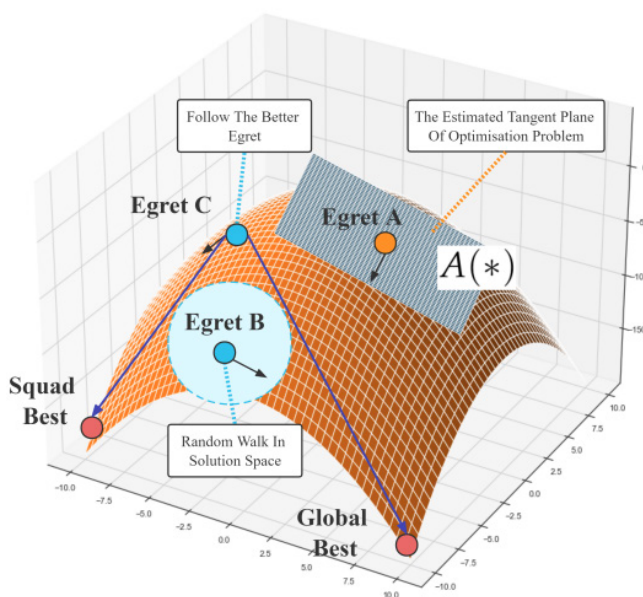


Figure 5. The Detailed ESOA Search Behavior.

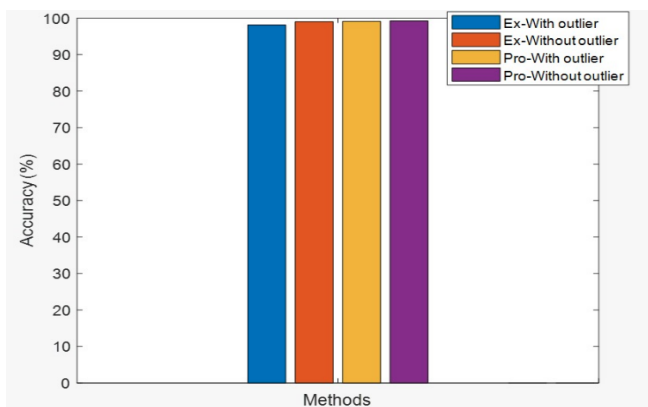


Figure 6. Accuracy Comparison based on Outliers.

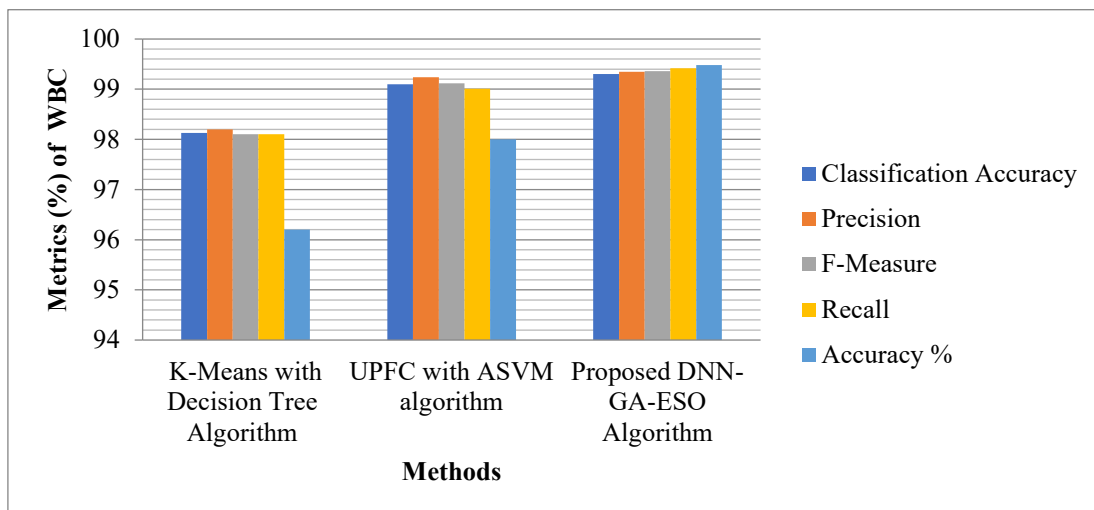


Figure 7. Comparison of approaches on WBC dataset using various metrics

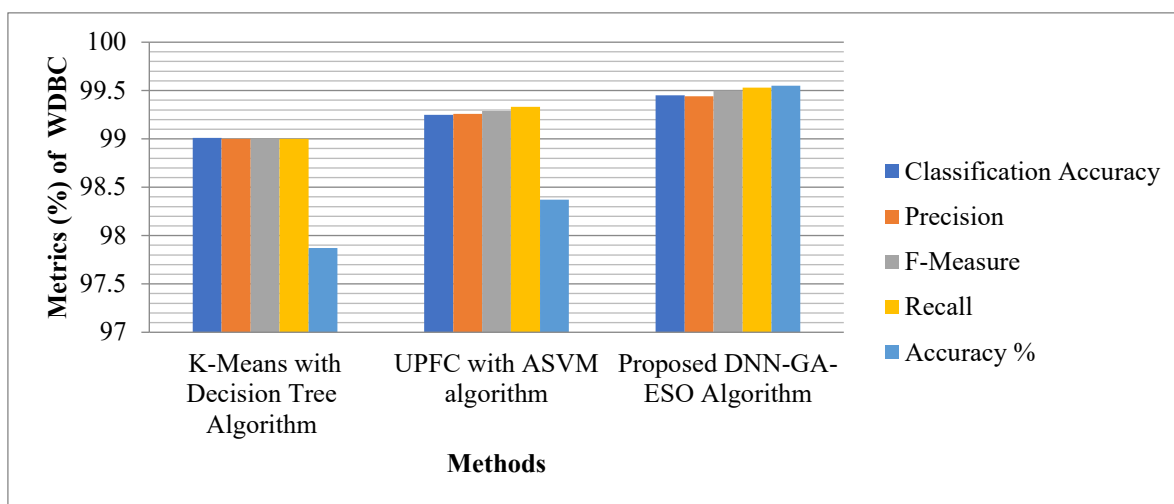


Figure 8. Comparison of approaches on WDBC dataset using various metrics

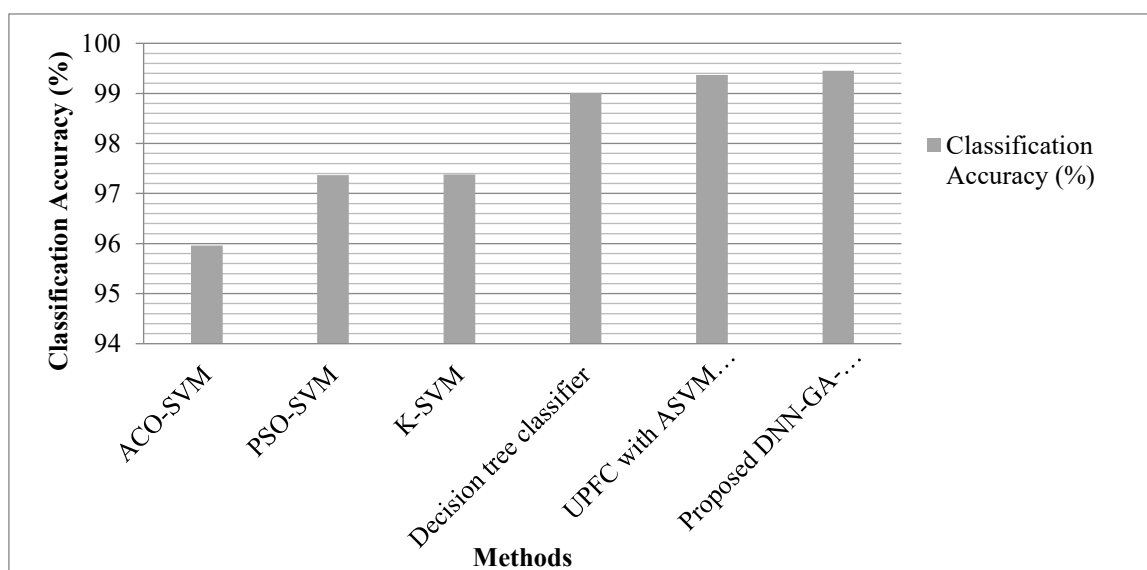


Figure 9. Comparison of classification Accuracy using various existing and proposed method



**Table 1.** Detailed statistics for WBCD dataset(N=699)

Attribute	Min Value	Max Value	Mean	Std.Dev
Clump Thickness	1	10	4.418	2.816
Uniformity of Cell Size	1	10	3.134	3.051
Uniformity of Shape	1	10	3.207	2.972
Marginal Adhension	1	10	2.807	2.855
Single Epithelial Cell Size	1	10	3.216	2.214
Bare Nuclei	1	10	3.545	3.644
Bland Chromatin	1	10	3.438	2.438
Normnal Nucleoli	1	10	2.867	3.054
Mitoses	1	10	1.589	1.715
Class	1	10	Benign (458)	Malignant (241)

**Table 2.** Detailed Statistics for BC dataset After removing Outliers (N=578)

Attribute	Min Value	Max Value	Mean	Median	median of the absolute deviation (MAD)	Std.Dev
Clump Thickness	1	10	3.855	4.012	4.550	2.529
Uniformity of Cell Size	1	10	2.451	2.651	2.989	2.589
Uniformity of Shape	1	10	2.581	2.761	3.012	2.552
Marginal Adhension	1	10	2.220	2.440	2.854	2.318
Single Epithelial Cell Size	1	10	2.756	2.890	3.121	1.809
Bare Nuclei	1	10	2.77	2.910	3.212	3.186
Bland Chromatin	1	10	3.003	3.025	3.065	2.201
Normnal Nucleoli	1	10	2.187	2.301	2.345	2.493
Mitoses	1	10	1.257	1.281	1.541	2.01
Class	1	10	Benign (444)			Malignant (134)

**Table 3.** The Assessment Metric Values attained in the WBC and WDBC Data Sets

Evaluation Metrics	Existing K-Means with Decision Tree Algorithm		UPFC with ASVM algorithm		Proposed DNN-GA-ESO Algorithm	
	WBC Value	WDBC Value	WBC Value	WDBC Value	WBC Value	WDBC Value
Classification accuracy (%)	98.13	99.01	99.10	99.25	99.30	99.45
Precision	98.2	99.0	99.24	99.26	99.35	99.44
F-Measure	98.1	99.0	99.12	99.29	99.36	99.50
Recall	98.1	99.0	99.01	99.33	99.42	99.53
Kappa statistics	96.2	97.87	98	98.37	98.48	98.55

**Table 4.** The suggested system and other BC classifiers achieved classification accuracy

Existing Author	Method	Classification Accuracy (%)
Prasad et al (2010)	ACO-SVM	95.96
Prasad et al (2010)	PSO-SVM	97.37
BichenZheng et al (2014)	K-SVM	97.38
Priya&Karthikeyan (2019)	Decision tree classifier	99.01
Existing Approach	UPFC with ASVM algorithm	99.37
Proposed Approach	Proposed DNN-GA-ESO Algorithm	99.45

ESO predicts the tangent plane of the optimisation problem, allowing for a rapid drop to the current finest point.

i. Sit-and-Wait Strategy

The Snowy Egret uses  $A(*)$  to assess the possibility of prey's presence in its current surroundings, given the positions of i-th egret squads are  $x_i \in R^n$ ,  $n$ , where  $n$  implies problems' dimensions.  $y$  represents evaluations of prey in regions (Wei et al. 2023).

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

The evaluation technique modified as,

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

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

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

Also,  $\hat{g}_i \in R^n$ , Equation (3) allows us to reproduce the practical gradient of  $w_i$ , whose direction is  $\hat{d}_i$ , using the partial derivative of  $w_i$ .

$$\begin{aligned} \hat{g}_i &= \frac{\partial e_i}{\partial w_i} \\ &= \frac{\partial \|\hat{y}_i - y_i\|^2 / 2}{\partial w_i} \end{aligned} \tag{9}$$

$$= (\hat{y}_i - y_i) \cdot x_i$$

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

Egrets denote to superior egrets when preying, use their expertise to predict the behavior of prey and adding their own ideas, as shown in Figure 5. The directional correction 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  $d_{g,i} \in R^n$ .

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

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

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

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

The method of adaptive weight updating is applied(23),  $\beta_1$  is 0.9 and  $\beta_2$  is 0.99

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

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

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

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

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

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

here  $t$  and  $t_{max}$  is the difference between the current and maximum iteration time, while  $hop$  is the solution space's lowbound and upbound. The step size factor of Egret A is  $step_a \in (0, 1)$ . The fitness of  $x_{a,i}$  is denoted by  $y_{a,i}$ .

2.2.2. Aggressive Strategy

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

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

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

here  $r_{b,i}$  represents random numbers in  $(-\pi/2, \pi/2)$ ,  $x_{b,i}$  implies Egret B's anticipated positions and  $y_{b,i}$  represents fitness values.

Encircling mechanisms are used by Egrets C as revisions for their locations as they choose to follow their preys actively:

$$D_h = x_{ibest} - x_i \tag{21}$$

$$D_g = x_{gbest} - x_i \tag{22}$$

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

$$y_{c,i} = f(x_{c,i}) \tag{29}$$

$D_h$  is the gap 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.  $step_b \in (0, 1)$  implies Egret B's step sizes.  $r_h$  and  $r_g$  represents random numbers in the interval  $(0, 0.5)$ .

2.2.3. 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.  $x_{s,i}$  represents solution matrices of i-th Egret squads:

$$x_{s,i} = [x_{a,i} \ x_{b,i} \ x_{c,i}], \tag{24}$$

$$y_{s,i} = [y_{a,i} \ y_{b,i} \ y_{c,i}], \tag{25}$$

$$c_i = argmin(y_{s,i}), \tag{26}$$

$$x_i = \begin{cases} x_{s,i|c_i} & \text{if } y_{s,i|c_i} < y_i \text{ or } r < 0.3, \\ x_i & \text{else} \end{cases} \tag{27}$$

Egrets select if minimum  $y_{s,i}$  is greater than fitness  $y_i$ . The worst plan will be approved if random numbers  $r \in (0, 1)$  are 30% with less than  $0.3 \cdot r \in (0, 1)$

Pseudo Code

The Algorithm 1 used to build the ESO pseudo-code has two main objectives: obtaining the predicted location matrix for the Egret squad, and to choose an improved scheme using a discriminant condition. ESO needs an initial matrix  $x^0 \in R^{P \times N}$  of the P size Egret Swarm location as input, while returning the optimal position  $x_{best}$  and fitness  $y_{best}$ .

**Algorithm 1: ESO algorithm for BC classification**

**Input:** BC dataset:  $x^0$ : the P size Egret Swarm position  $\in R^P \times N$  ;

**Objective function:** Classifier accuracy

**Output:**  $x_{best}$ : Optimal;  $y_{best}$ : Optimal fitness;

Begin

function SITANDWAIT(x)

Update the integrated gradient  $g$  via Eq. (9 to 13)

Update the weight of observation method  $\omega$  by Eq. (15)

Get the expected position  $x_a$  of Egret A by Eq.(17)

return  $x_a$

end function

function AGGRESSIVE( $x_a$ )

Get the expected feature position  $x_b$  of Egret B by Eq. (19)

```

Get the expected feature position  $x_c$  of Egret C by Eq. (21)
and (22)
return  $x_b, x_c$ 
end function
while  $t < t_{max}$  do
 $x_a^t \leftarrow \text{SITANDWAIT}(x^t)$ 
 $x_b^t, x_c^t \leftarrow \text{AGGRESSIVE}(x^t)$ 
Get next position ( $x^{t+1}$ )
end while
return  $x_{best}, y_{best}$  optimal features
Obtain higher BC classifier results
End
    
```

The best data for cancer detection is carried out by ESO. Here, the fitness value of ESO is set by a combination of different attributes of cancer data. Hence, the cancer data with greatest fitness among all possible data is chosen. Two data sets have been used with the ESO classification algorithm. The proposed algorithm is used for diagnose the cancer is either malignant or benign.

### 3. Results

Our research leveraged the MATLAB tool for early breast cancer (BC) diagnosis and performance evaluation.

#### A. Dataset Description

The BC dataset encompasses measures of BC cases recorded from the Wisconsin Diagnosis and Wisconsin BC (WBC) datasets, originating from the UCI ML data repository. The WBC dataset comprises 9 integer-valued attributes, 699 instances, and 2 classifications (malignant and benign). On the other hand, the WDBC dataset consists of 32 integer-valued attributes, 568 instances, and the same 2 classifications. Training and test data are prepared using 10-fold cross-validation (8).

#### Wisconsin BC Dataset

Table 1 presents a detailed description of the attributes included in the original WBC dataset, with its dimensionality reduced using Independent Component Analysis (ICA). Furthermore, a novel dataset was created by removing 120 extreme values (all malignant) from the original dataset, resulting in 578 instances. The statistics of this new dataset post-outlier removal are provided in Table 2.

#### B. Wisconsin Diagnosis BC Dataset (WDBC)

The WDBC dataset includes data for each cell nucleus, comprising an ID number, diagnosis (B = benign, M = malignant), and 10 real-valued features such as compactness, area, concave points, smoothness, symmetry, concavity, and fractal dimension. The original WBC dataset employed Unified Power Flow Controller (UPFC) for outlier identification and ICA for dimensionality reduction. The resultant new dataset comprised 424 instances after eliminating 56 outliers and 88 extreme malignant cases from the original dataset.

Figure 6 illustrates the Accuracy Comparison based on Outliers, presenting a comparison graph of outlier detection accuracy between existing and proposed methods.

#### C. Performance Evaluation

The performance of the proposed approach is thoroughly evaluated using datasets to ascertain its stability. Assessment metrics like recall, precision, F-Measure, accuracy, and kappa statistic are utilized to identify and evaluate the data.

Accuracy is defined as the accuracy of correctly determining objects

$$Accuracy = \frac{TP+TN}{Total\ no.\ of\ Samples} \quad (28)$$

Precision (P) is defined as determine exactness of the Classifier

$$Precision\ (P) = \frac{TP}{TP+FP} \quad (29)$$

Recall (R) is defined as to calculate the sensitivity or the classifiers' efficiency

$$Recall\ (R) = \frac{TP}{TP+FN} \quad (30)$$

F-Measure is a potential harmonic mean of recall and precision.

$$F - Measure = 2 * \left[ \frac{Precision * Recall}{Precision + Recall} \right] \quad (31)$$

Where TP - True Positive, FP - False Positive, FN - False Negative, TN - True Negative.

Kappa Statistic: The difference among observed and expected agreement is a different evaluation metric. The value is between 0 and 1, and a value of 1 indicates full agreement. To determine the kappa statistic value, use equation 1.

$$K = \frac{p_o - p_e}{1 - p_e} = 1 - \frac{1 - p_o}{1 - p_e} \quad (32)$$

where,  $p_o$  is the 10 related agreements that were recorded, and  $p_e$  is agreement probability.

#### A. Result Analysis

A robust evaluation methodology employing 10-fold cross-validation is employed in our study. This method entails dividing the dataset into 10 equal-sized sections through random sampling. While one section is reserved for testing, the remaining segments are utilized for training the base learner. This process is iterated 10 times, ensuring that each partition is tested exactly once. Subsequently, individual data are aggregated based on their mean accuracy. The experiments were conducted using MATLAB, implementing the ASVM classification algorithm. The results depicted in Table 3 and Figure 7 underscore the efficacy of the categorization model in diagnosing BC. Notably, the proposed model achieves a classification accuracy of 99.10% and 99.35% with WBC and WDBC data, respectively, showcasing a highly promising performance compared to existing techniques.

Table 4 provides a comprehensive comparison of various existing classifiers from the literature alongside the classification accuracies achieved using the proposed approach.

Figure 7 juxtaposes the performance of K-Means with Decision Tree Algorithm, UPFC with ASVM algorithm, and the Proposed

DNN-GA-ESO Algorithm, utilizing diverse metrics on the WBC dataset such as Classification accuracy (%), Precision, F-Measure, Recall, and Kappa statistics. The proposed ESO method outperforms existing approaches, boasting a classification accuracy of 99.30%, Precision of 99.35, F-Measure of 0.9936, Recall of 99.42, and Kappa statistics of 99.48. Conversely, existing methods yield lower performance metrics, underscoring the superiority of the proposed DNN-GA-ESO algorithm in enhancing BC classifier performance through efficient outlier detection.

Similarly, Figure 8 presents a comparison of the same methodologies on the WDBC dataset. The proposed DNN-GA-ESO algorithm again outperforms existing methods, showcasing higher accuracy, precision, F-Measure, recall, and Kappa statistics. The utilization of the ILDA algorithm for preprocessing further enhances the classifier accuracy, solidifying the conclusion that the proposed approach significantly improves BC classifier performance.

Finally, Figure 9 highlights the comparison of classification accuracy between the recommended DNN-GA-ESO algorithm and existing methodologies. The proposed algorithm achieves an impressive accuracy of 99.45%, outperforming other algorithms such as UPFC with ASVM (99.37%), Decision tree classifier (99.01%), K-SVM (97.38%), PSO-SVM (97.37%), and ACO-SVM (95.96%). This reinforces the superiority of the suggested approach in BC classification.

#### 4. Conclusion

In conclusion, early detection and accurate classification of breast cancer (BC) are critical for effective treatment. This study introduces a robust classification model leveraging deep learning (DL) and genetic algorithms (GA) for enhanced BC diagnosis. The proposed DNN-GA-ESO algorithm integrates three core processes: outlier detection, feature selection, and classification, demonstrating superior performance metrics. The model achieves impressive classification accuracies of 99.30% and 99.45% for the Wisconsin BC (WBC) and Wisconsin Diagnosis BC (WDBC) datasets, respectively. Compared to existing methodologies, the DNN-GA-ESO algorithm outperforms in terms of accuracy, precision, F-measure, recall, and kappa statistics, as validated by 10-fold cross-validation. These findings underscore the effectiveness of combining DL with evolutionary strategies for BC diagnosis, offering a promising tool for early detection and accurate classification. This advancement holds significant potential for improving patient outcomes and optimizing treatment strategies through precise and early BC diagnosis.

#### Author contributions

S. M. S., N. S. conducted the methodology, collected and analyzed the data, and wrote the original draft, conceptualized the project,

supervised the work, reviewed and edited the writing, and managed the project. T. M. Praneeth Naidu developed the methodology, interpreted the data, and reviewed and edited the writing.

#### Acknowledgment

The authors were grateful to the Department.

#### Competing financial interests

The authors have no conflict of interest.

#### References

- Abualigah, L., & Dulaimi, A. J. (2021). A novel feature selection method for data mining tasks using hybrid sine cosine algorithm and genetic algorithm. *Cluster Computing*, 24, 2161-2176.
- 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..
- 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.
- Alzubaidi, A., Cosma, G., Brown, D., & Pockley, A. G. (2016, October). Breast cancer diagnosis using a hybrid genetic algorithm for feature selection based on mutual information. In *2016 International Conference on Interactive Technologies and Games (ITAG)* (pp. 70-76). IEEE.
- Amarasinghe, 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.
- Chambers, L., Gaber, M. M., & Abdallah, Z. S. (2020). DeepStreamCE: A Streaming Approach to Concept Evolution Detection in Deep Neural Networks. *arXiv preprint arXiv:2004.04116*.
- Chaudhuri, A., & Sahu, T. P. (2021). A hybrid feature selection method based on Binary Jaya algorithm for micro-array data classification. *Computers & Electrical Engineering*, 90, 106963.
- 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.
- Devi, R. D., & Devi, M. I. (2016). Outlier detection algorithm combined with decision tree classifier for early diagnosis of breast cancer. *Int. J. Adv. Eng. Technol*, 12, 93-98.
- Gao, J., Song, X., Wen, Q., Wang, P., Sun, L., & Xu, H. (2020). Robuststad: Robust time series anomaly detection via decomposition and convolutional neural networks. *arXiv preprint arXiv:2002.09545*.
- 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.
- 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 (ICCI) (pp. 1-5). IEEE.
- 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.
- Huang, Q., Chen, Y., Liu, L., Tao, D., & Li, X. (2019). On combining biclustering mining and AdaBoost for breast tumor classification. *IEEE Transactions on Knowledge and Data Engineering*, 32(4), 728-738.
- 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.
- Khan, S., Islam, N., Jan, Z., Din, I. U., & Rodrigues, J. J. C. (2019). A novel deep learning-based framework for the detection and classification of breast cancer using transfer learning. *Pattern Recognition Letters*, 125, 1-6.
- 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.
- 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.
- 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.
- Mishra, A. K., Roy, P., & Bandyopadhyay, S. (2020). Genetic algorithm-based selection of appropriate biomarkers for improved breast cancer prediction. In *Intelligent Systems and Applications: Proceedings of the 2019 Intelligent Systems Conference (IntelliSys) Volume 2* (pp. 724-732). Springer International Publishing.
- 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.
- 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.
- 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.
- 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.
- Ricciardi, C., Valente, A. S., Edmund, K., Cantoni, V., Green, R., Fiorillo, A., ... & Cesarelli, M. (2020). Linear discriminant analysis and principal component analysis to predict coronary artery disease. *J. Health Inform.*, 26(3), 2181-2192.
- Rupali, Verma, R., Handa, R., & Puri, V. (2021). Feature Selection Using Genetic Algorithm for Cancer Prediction System. In *Advances in Communication and Computational Technology: Select Proceedings of ICACCT 2019* (pp. 1197-1212). Springer Singapore.
- 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.
- Sun, W., Tang, M., Zhang, L., Huo, Z., & Shu, L. (2020). A survey of using swarm intelligence algorithms in IoT. *Sensors*, 20(5), 1420.
- 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.
- 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.
- 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.
- Wei, P., Shi, X., & Zhou, J. (2023, December). ESOA Algorithm Based on learning rate optimization in Convolutional neural networks. In *2023 IEEE International Conference on Data Mining Workshops (ICDMW)* (pp. 435-439). IEEE.
- Xue, J., & Shen, B. (2020). A novel swarm intelligence optimization approach: sparrow search algorithm. *Systems science & control engineering*, 8(1), 22-34.
- Yusuf, A. B., Dima, R. M., & Aina, S. K. (2021). Optimized breast cancer classification using feature selection and outliers detection. *JNSPS*, 298-307.
- Zhao, H., Wang, Z., & Nie, F. (2018). A new formulation of linear discriminant analysis for robust dimensionality reduction. *IEEE Transactions on Knowledge and data engineering*, 31(4), 629-640.