Breast Cancer Detection Using Novel Weight Based Convolutional Neural Network With Simulated Annealing (Nwcnn-Sa) Based Outlier Detection Algorithm Combined With Weighted Deep Autoencoder (Wdae) - Chimp Optimization Algorithm (Choa) For Classification

S. Maria Sylviaa[1]*, N. Sudha[2]

[1]Research Scholar,Department of Computer Science,Bishop Appasamy College of Arts and Science, Coimbatore, Email: mariasylviaa1991@gmail.com
[2]Associate Professor, Department of Computer Science, Bishop Appasamy College of Arts and Science, Coimbatore, Email: sudhanatarajan105@gmail.com
*Corresponding Author

Received: 10.04.2024 Revised : 12.05.2024 Accepted: 22.05.2024

ABSTRACT
One of the main causes of mortality worldwide, which has been rising in recent years, is cancer. There are two categories for this illness: benign and malignant. Breast cancer, which manifests in breast tissue despite a person's gender, is a single of the initial and largest types of cancer in the human body. The detection of cancer of the breast has benefited greatly from the application of deep learning. This study applies the selection of features and modification to the Wisconsin Diagnostic Breast Cancer dataset. Finding the most pertinent traits is meant to help categorize a diagnosis as benign or malignant. In the beginning of this study project, the data's dimensionality is reduced using the Non-Negative Matrix Factorization (NMF) Algorithm. After that, the outliers from the cancer dataset are found using the Novel weight-based Convolutional neural network with Simulated Annealing (NWCNN-SA) detection approach. Then, Remora Optimization Algorithm (ROA) is used to pick the attributesfrom the data with the outliers eliminated. After that, all of the altered images are given to the Weighted Deep AutoEncoder (WDAE) with Chimp Optimization Algorithm (ChOA), also known as the WADE-ChOAclassifier, for the training process. This helps to maximize the detection accuracy by classifying incoming clinical images of breast cancer as benign or malignant without requiring any prior knowledge about the appearance of cancer. Utilizing the breast cancer database, evaluations the efficacy of the suggested technique. A dataset of breast cancer cases, prior to as well as following the elimination of outliers, was used for the experiments. The results show that the proposed strategy is more accurate and performs better. The findings of the research on breast cancer will help in patient diagnosis and care. The suggested method's effectiveness has been evaluated using the breast cancer database. The trials employed a dataset of instances of breast cancer, both prior to and following the outliers were removed. The outcomes demonstrate how much more precise and effective the suggested method is. The results of this breast cancer study might help in detection and supervision of patients.

Keywords: Breast Cancer, Clustering Algorithm, Data Mining, machine learning, Optimization algorithm, Classification Algorithm, Outlier Detection.

INTRODUCTION

Healthcare technology encompasses several aspects including maintaining along with retrieving digital medical information and associated equipment for patients. Diagnosing and treating haematological problems has long been complicated by the identification of malignancy. These days, an astonishing number of individuals suffer from one or more illnesses. Medical research has advanced tremendously in the last several years. Notwithstanding these developments, the general population still knows a little regarding illnesses and their symptoms. It seems that a significant proportion of the populace is coping with health problems, certain of them may be harmful [1]. It is possible to lower the demand for caretakers and overall health care expenditures while simultaneously increasing the accuracy of early diagnosis of fatal illnesses by using sensible, safe practices and utilizing contemporary technologies.

Intelligent decision-making processes and technology breakthroughs have the potential to save many lives [2-4].

Cancer is characterized by rapid and abnormal cell proliferation brought on by a confluence of certain genetic and epigenetic abnormalities. Unchecked cell division promotes the development of tumorsIf a tumor spreads quickly to other physiological systems and organs, cancer may already be incurable at the stage of diagnostic [5]. At some point in their lives, one in 8 female gets a breast cancer diagnosis; fewer than 1% of instances impact non-female individuals [6]. An estimated 2.1 million female identified a breast cancer diagnosis each year, with the age range of 40 to 70 having the greatest death rates [7]. So, early detection is crucial for a good prognosis in cases of breast cancer. An early identification greatly improves the chances of survival, even in cases when the illness's initial symptoms are not extremely severe [1]. Breast cancer screening methods include mammography, fine needle aspiration cytology (FNAC), and ultrasound-guided clinical biopsies. The low percentage of cancer detected by mammography in cases with large breasts varies from 10% to 30% [8–9]. Accurately identifying cancer cells is essential to reducing mortality rates. To increase the proportion of cancer patients who survive overall, effective preliminary detection and therapy are necessary [10–13].

For this reason, a favorable prognosis in patients of breast cancer depends on early identification. Although the early signs of the illness may not be exceedingly severe, early identification significantly increases the odds of survival [1]. Fine needle aspiration cytology (FNAC), ultrasound-guided clinical biopsy, and mammography are some of the techniques used in breast cancer screening. Mammography's poor cancer detection rate in instances with big breasts ranges from 10% to 30% of cases not reported [8–9]. It is necessary to precisely identify cancer cells in order to lower the death rate from the disease, and it is necessary to provide effective early diagnosis and therapy in order to raise the percentage of cancer patients who cure overall [10–13]. It entails managing databases, statistical methods, and artificial intelligence via the discovery of novel patterns in vast volumes of data. To find the relevant patterns, data mining techniques combine theoretical explanations with rules for association, categorization, estimation, organization, and pattern identification [16]. Historical data is being used more and more by medical researchers to find and take advantage of patterns and correlations within a wide range of factors in an effort to forecast how a disease will develop. The primary problem with data mining is the preliminary information collecting, or preprocessing. In data warehouses, two strategies are employed: cleaning and integration.Prior to the application of data mining tools, data is filtered and altered. Noise-filled data is not helpful. Corrupt data and noisy data can occasionally be identical. But now, its definition has expanded to encompass any type of data that is difficult for computers to comprehend and assess accurately, including unstructured text [16]. In this study, outlier identification is utilized to improve the accuracy of the breast tumor diagnosis.

Incomplete data is a common feature of large databases. Patterns that deviate significantly from the typical pattern that most of the data exhibits are called outliers, commonly known as noise in the data. As such, this data should not be used in traditional data mining techniques. Thus, in order to improve the quality of the data and provide an accurate data mining result, outlier identification—also known as anomaly detection—is required to locate the outliers. An observation is considered an outlier if it differs sufficiently from the rest that it is difficult to determine if it is the result of some other factor. Even though they are typically the result of errors in measurement, execution, etc., Outliers may be viewed as noise in general or their relevance can be reduced based on the revised outliers estimate in an effort to analyze the data set. In the current study, anomalies throughout the breast tumor data set were identified using the application of the outlier detection algorithm (ODA) [17]. Following the identification of outliers, the preprocessed data is split into two groups: benign and malignant cancer, using a classification technique. Data can be categorized and assigned to a predetermined set or group. The categories were formed before to the assessment data analysis procedure using the classification approach, which is also referred to as supervised learning. Any classification technique must start with class definition, which is a prerequisite that is frequently determined by the characteristics of the acknowledged data category—which is determined by the data found in an attribute's value.

SI has been applied in real-world settings to tackle a wide range of difficult problems, including managing AI, predicting social behaviors, assisting networks and telecommuting, etc. [18]. The features community is really interested in SI since it is so approachable and has global search capabilities. Among the numerous swarm intelligence approaches that are currently available are the bat algorithm (BA), particle swarm optimization (PSO), ant colony optimization (ACO), grey wolf optimization (GWO), and many more [19]. Relying exclusively on the attribute choice approach will not yield the greatest classification results, as there are relatively few cases in the dataset collection that reflect noisy or low-quality data [20]. Therefore, rushed data should be ignored in order to increase the classification systems' performance [20]. Outlier rejection is the process of eliminating or undervaluing noisy data that deviates significantly

from the rest. Therefore, before applying the classification algorithm to the datasets, the outliers need to be removed. Algorithms used in machine learning frequently regard outliers as noise. It must be removed since it makes it more difficult for the algorithm to forecast future events. The classic outlier technique and the spatial outlier strategy are the two types of outlier methods [21].
The WDBC and WBC data sets are used by the hybrid deep learning-based outlier detection (OD) method in this study to locate outliers. Once an outlier has been found, the smaller sample's suggestiveness for benign or malignant malignancy is determined using the decision tree categorization approach. During the categorization process, data elements are organized based on previously established groups or groupings. A more accurate diagnosis of a breast cancer dataset is provided by the Weighted Deep AutoEncoder (WDAE) with Chimp Optimization Algorithm (ChOA) classifier, which determines whether an output class is benign or malignant in comparison to previous studies.

RELATED WORKS

Sadhukhan et al. [22] utilized a model that was based on elements found in the cell nucleus and developed using machine learning (ML). SVM and K-NN algorithms were employed in this paper. They ascertained and examined the classifiers' performance. They used a dataset that contained the values of the attributes they had taken out of the cell slide photos using FNAC in order to evaluate and contrast the work with the use of a Bayesian Network. They found that every effort was put into creating an algorithm which might enable one to determine the nature of a tumor—whether benign or malignant. The image displayed different lighting in many areas, with differences depending on light intensity. Gray, White, and Black were the three classification levels for the picture. The efficiency of the model was remarkably high, at 97.49 percent. The experts concluded that an early breast cancer detection technique may be developed in the future. This approach would employ computers to identify the key characteristics on slide images of the cells that were obtained using FNAC. The slide images would then be instantly loaded into the preprocessing of the machine learning model.
Osmanović et al. [23] information compiled among 699 instances in the machine learning repository at UCI. Nine neurons in the artificial neural network (ANN) used in the research represented the number of features, while one output neuron indicated the type of lump and whether it was malignant or benign. Furthermore, a confusion matrix unique to the backpropagation network was supplied. According to the research, there is a 99% chance that an artificial neural network (ANN) built with care can identify an ailment. This demonstrated even more how likely it was that they would receive a negative grade—97.6%. The ANN still has promise for further advancements. In addition to monitoring the patient's status, it might assist doctors in diagnosing patients more quickly. The authors suggested creating the instrument featuring a graphical user interface (GUI) to assist medical professionals who are not used to using these kinds of instruments without a suitable interface.
Negi et al. [24] proposed using the following four machine learning (ML) model-based techniques to identify breast cancer: random forests, kNN, SVM, and Bayesian networks. The authors set out to develop a systematic strategy for the detection and categorization of breast cancer. When comparing the Random Forest method with the Support Vector Machine (SVM) method, the Random Forest approach has the best chance of correctly diagnosing cancers. It also succeeded in achieving perfect precision and uniqueness. By using characteristics such the rolled-up features measure, prediction accuracy, recall, operating characteristic area of the receiver curve, and training time, it is possible to compare data from small and large datasets accurately. Their results suggest that further study is still needed in this field.
Irfan et al. [25] methods such as the dilated semantic segmentation network (Di-CNN) and suggested morphological erosion for the deep learning segmentation of ultrasound breast lesion images. Feature extraction was performed on the image segments that were fed into DenseNet201 via transfer learning. 780 instances of breast cancer were subjected to this research. DenseNet201-activated feature vectors increased the accuracy and precision to 98.45% and 98.9%, respectively, when paired alongside an SVM classifier. Using CNN-activated feature vectors, the method's accuracy in recognizing the breast tumor was 90.11 percent.
Lahoura et al. [26] Created a cloud-based machine learning model for BC categorization using an extreme learning machine (ELM). Following that, the ELM model was implemented using AdaBoost, k-NN, SVM, Naive Bayesian, and perceptron approaches. The data were obtained from the Wisconsin Breast Cancer Diagnosis (WBCD) database. The dataset consisted of 569 items with 32 attributes. The approach achieved an F1-score of 81.29%, recall of 91.30%, precision of 90.54%, and accuracy of 98.68%.
Magna et al. [27] Breast cancer cases were classified using deep learning, machine learning, and word insertion techniques based on the patients' medical histories. They attempted to provide a system of recommendations that backs the doctor's assessment.

Reddy et al. [28] used the support-value method of deep neural networks (DNNs) to detect breast cancer. The experimental results show that the recommended DNNs surpass the most sophisticated approaches.
Saxena and Gyanchandani [29] inspected machine learning techniques based on histology for computer-aided breast cancer screening. After a detailed analysis of various methods, it was shown that deep learning constituted the majority of machine learning research on breast cancer.
Kayikci and Khoshgoftaar [30]this breast cancer dataset was examined by us earlier. For managing data from many sources, they had already created a multimodal framework. To look at this model more, we used a variety of machine learning techniques. With respect to decision trees, they achieved an accuracy of 82%; for random forests, it was 90%; and for support vector machines, 88%.

PROPOSED METHODOLOGY

For the purpose of lowering the dimensionality of the data, this work is the first to use the Non-Negative Matrix Factorization (NMF) approach. Next, we employ NWCNN-SA, a novel weight-based convolutional neural network with simulated annealingoutlier identification approach to the cancer dataset in order to identify the outliers. The characteristics that remain after the outliers have been eliminated are then selected using the Remora Optimization Algorithm (ROA). In order to determine the maliciousness of the data, the WADE-ChOA Chimp Optimization Algorithm is utilized to feed all of the modified pictures into the Weighted Deep AutoEncoder (WDAE). The proposed procedure diagram may be as seen in Figure 1.

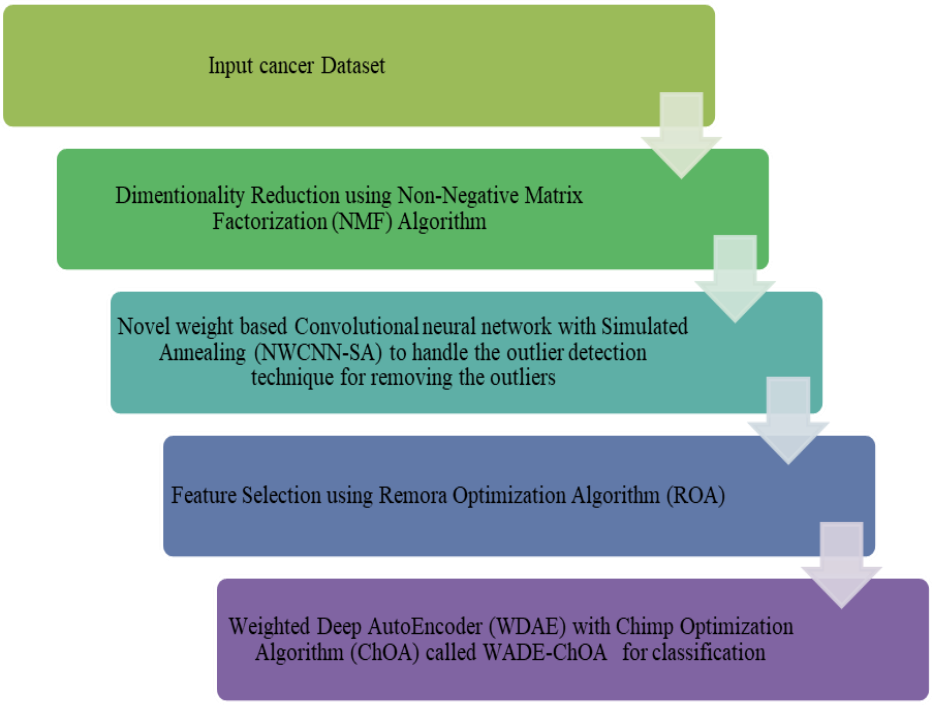


Figure 1. Overall Flow of the Proposed Research Methodology

Input Dataset

The Wisconsin Diagnosis Breast Cancer Dataset (WDBC) provides information on fine needle aspiration (FNA) of breast cancer based on characteristics found in a digital photograph. The findings recognized characteristics of the imaged cellular nuclei. The sample has 569 data points, with 357 classed as benign and 212 as malignant. The dataset was divided into ten groups based on the following features: radius, texture, perimeter, field, smoothness, compactness, concavity, concave points, symmetry, and fractal dimension.For every function, the mean, standard error, and the "worst" or worst (mean of the three largest values) approximations have been made. Therefore, thirty functions would be included in the dataset.
The Wisconsin Breast Cancer Dataset, or WBC, is made up of 699 samples that were taken from the UCI collection.. In this category, there are 241 malignant samples and 458 normal ones. Ten traits and one class are also included in the dataset. Benign and malignant are the two distinct classifications at the class level. In addition, the dataset has one missing attribute. One of the properties is an id, or example code number. 2. The cluster has a thickness of one to ten. 3. Cell size consistency: 1–10. 4. Homogeneity of cell shape: 1–10. 5. Attachment to the periphery: 1–10. One epithelial cell has a diameter between one and ten.7. Nuclei without shells: 1–10. 8. Chromatin diffuse (1–10) 9. Regular Nucleoli (Worksheets 1–10).

There are four types of malignant mitoses and two types of benign mitoses. There are 699 entries in the WOBC dataset, and each one has nine properties besides class and id number. A scale of 1 to 10 is used to grade these nine features, with 10 being the most severe state.

Dimensionality Reduction using Non-Negative Matrix Factorization (NMF) Algorithm

A matrix may be approximated by multiplying two matrices, W and H, which is what matrix factorization is defined as. There are two types of matrix factorization algorithms: those that employ restrictions on the components W and H, and those that indicate the approximation's correctness via an error function. That W and H elements be nonnegative is required by the family of nonnegative factorizations. Only positive values can be found in the data matrix.

When factorizing a non-negative matrix factorization (NMF), the goal is frequently to find the Euclidean distance, or Frobenius standard, between the components of A and the principal components of WH. This metric is commonly used in research and is well-known. On the other hand, different metrics of distance (or similarity) can be applied and frequently lead to different factorizations. I Loss functions are usually modified to correspond with certain application areas.

The Kullback-Leibler divergence was employed as a distance metric in [31]. $D_s(u_j * w_i)$ is a symmetric divergence of u_j with regard to w_i given by [2]:

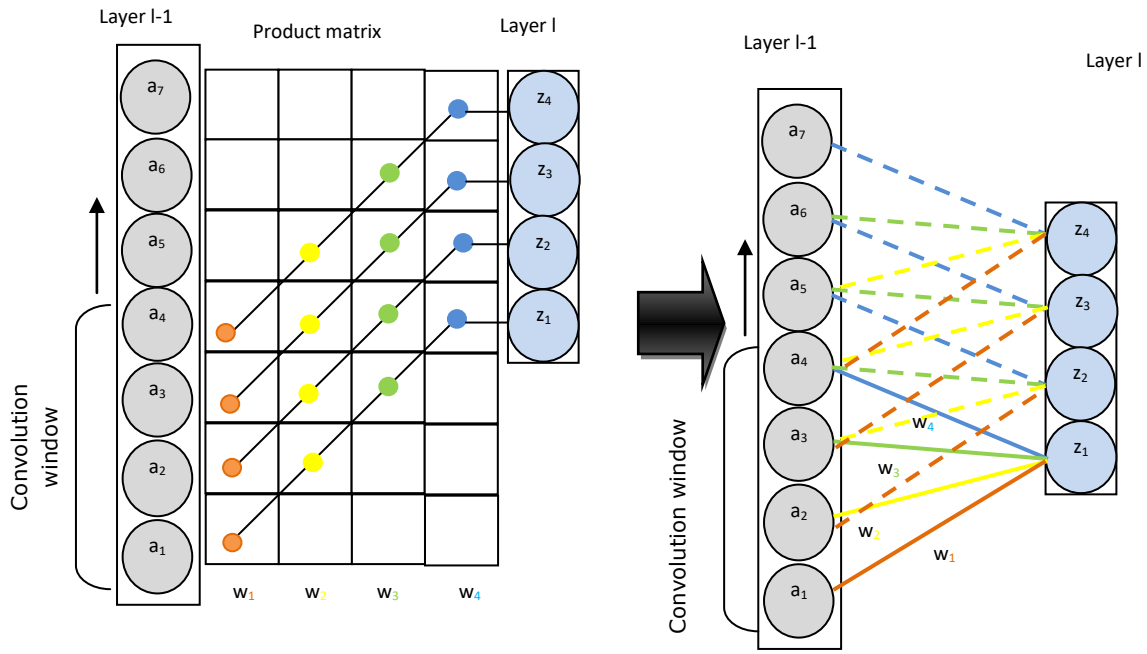
$$D_s(u_j * w_i) = D(u_j \| w_i) + D(u_j \| w_i) \quad (1)$$

Where

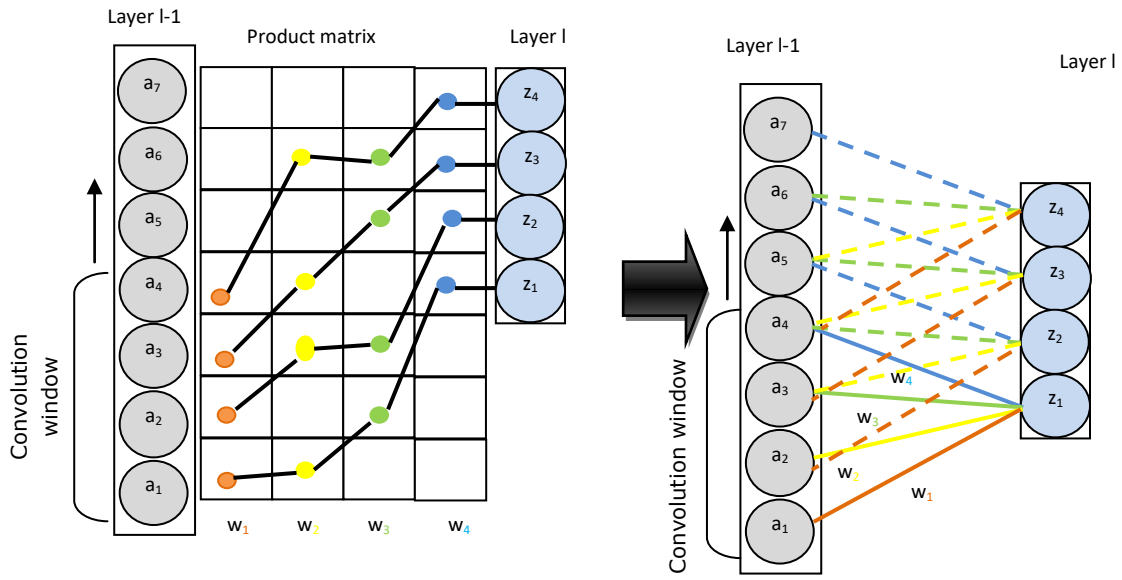
$$D(x\|z) = \sum_l \frac{x(l)}{\|x\|_1} \log\left(\frac{x(l\|z\|_1)}{z(l\|x\|_1)}\right) \quad (2)$$

Novel weight based Convolutional neural network with Simulated Annealing (NWCNN-SA) to handle the outlier detection technique for removing the outliers

Currently, diagnostic samples are distributed throughout a multidimensional region. The recommended method finds the optimal weights for CNN convolutional layers using a dynamic process. By employing simulated annealing (SA), the weight parameters of the classifier are constructed. After determining which features are most important to the positive class, SA is calculated for each attribute. Instead of using a convolution's usual linear inner product, the convolutional filter weights are dynamically matched to the input window values to lessen volatility. The input values and weights have the same characteristics. In Figure 2, a conventional convolutional layer with linear weight alignment is contrasted with a suggested NWCNN with dynamic weight alignment.



(a) Convolution with linear weight alignment



(b) Dynamic weight alignment in convolution

Figure 2. Comparison of Classical Linear Convolutions

Figure 2(b) shows both of the 1D convolutions at stride 1 with four weights (w1,..., w4), whereas Figure 2(a) shows the suggested convolution with dynamic weight alignment. Layers l through 1 indicate the current layer, which contains entities a1,..., and a7. Convolution with entities z1,..., and z4 produces a feature map represented by layer l. The blue circle represents the total of all the products, whereas each weight and input product is represented by a dot. An artificial neural network, or CNN, is made up of one or more convolutional layers. Convolutional layers' fundamental characteristics are sparse connectivity and common parameters. The weights of a convolutional layer, specifically, share the local receptive field of each matched output element. This reduces convolution operations to forward computations of convolutional layers, with the shared weights serving as the filtering process and the result being a feature map. The official expression for the convolutional layer's feature map $z_j^{(l)}$ is:

$$z_j^{(l)} = \sum_{i=0}^{I-1} w_i^{(l)} a_{j+(I-1)}^{(l-1)} + b^{(l)} \quad (3)$$

With l standing for the convolutional layer, l - 1 for the preceding layer, istands for the filter index, and I for the window size, for every element j the representations of the shared weights, the bias, and the prior layer activations are as follows:by $w_i^{(l)}, a_{i+j}^{(l-1)}$, and $b^{(l)}$. To put it another way, , $z_j^{(l)}$provides the inner product of each window of the preceding layer, $a_j^{(l-1)}, \ldots, a_{j+(l-1)}^{(l-1)}$and the shared weights w^l.The weights of the inputs inside the window match this inner product linearly. However, there could be circumstances in which some weights must match very optimal inputs; these scenarios could involve feature translation, noisy entities, and scale disparities inside the filter. The standard interior combination of a convolution is the same similarity function. The main concept is to align the weights so that they may be forcefully activated since the input windows are identical but just slightly mismatched.

In [32], a Simulated Annealing (SA) suggestion was proposed. This process is dependent upon the annealing process, which produces the basic state of matter, or the lowest energy of a solid state. At these low temperatures, it is not feasible to grow a single crystal inside the melt and reach a stable state of matter. Annealing processes aim to melt a material and then progressively lower the temperature; this may entail extended periods of time at almost freezing temperatures. If this isn't done, the material can be out of balance and form very faulty crystals. Since the framework in the case of glass is locally optimal due to the condition's metastability, crystals are unable to form a crystalline order. The Boltzmann distribution is one of the quantitative elements of the SA method. The formula below gives the probability of any genuine condition of x.

$$p(x) = e^{\frac{-\Delta f(x)}{kT}} \quad (4)$$

where T is the temperature, k is Boltzmann's constant, and f(x) is the energy configuration. The following stages provide the usual SA optimization approach in the event of an optimization problem:

1. To create the initial solution vector, first compute the goal function and then randomly choose x0, the system's first solution.

2. Establish the beginning temperature: The temperature T is an essential component for the efficient application of SA. When the number is too large, more reduction is needed for the data to converge. The process might not be optimal and the global optimum might be exceeded if the search parameters are very restrictive.
3. Select a novel strategy that is comparable to the current one: A new solution x0+'x is identified as a recent new solution based on T. The objective functions for x0 and x0+'x are reflected in the expressions f(x0+'x) and f(x0), respectively.
4. If f(x0+'x) = f(x0), then x0+'x = x0 and we update the current optimal solution. Otherwise, we move on to step 6. Conversely, x0+'x can also be accepted with a probability based on equation (4) if f(x0+'x) > f(x0).
5. Regularly lower the temperature: As temperature T decreases throughout the search, there is initially a higher chance of accepting deteriorating movements, but this chance gradually decreases.
6. Until a pausing criterion is satisfied, steps 2–6 must be repeated: The computation is finished when the termination requirements are satisfied. If not, go through steps 2 through 6 again.
According to statistics, outlier instances are more likely to occur because they are unusual events, while normal occurrences are more likely to occur because they are evaluated more frequently. When classifying data examples as outliers or normal, this information helps differentiate between outlier Ness and normalcy since probability values range from [0, 1]. Specifically, a greater chance denotes a more extreme outlier.
The outlierness score is considered by $\varepsilon(x|\theta)$ function in Eq. (5)

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

In the representation space Z, the weight of the interaction is represented by the trainable parameter w_{ij}, and the ith feature value for x is low-dimensionally integrated as z_i. A trained model with the ability to discriminate between normal and anomalous training samples is produced by the FCNN-SA training phases.
According to the suggested architecture, the detection phase comes last. We classify the unknown data as either typical occurrences or outliers using the well-established deep neural network model. When an unknown occurrence is assigned an outlierness score by the FCNN-SA model, test data are utilized to assess the model after training. The final score for outlierness is obtained by utilizing the cumulative probability of the sigmoid response vectors.

Feature Selection using Remora Optimization Algorithm (ROA)
The retrieved characteristics are chosen using the ROA method. ROA is a newly suggested swarm intelligent optimization method inspired by the remora's adsorbing host in the ocean. It is significantly simpler for the remora to get nourishment by depending on the host than on the the remora itself. The principal hosts in the ROA are sailfish and whales. As a result, ROA adjusts the locations using various formulae from the SFO algorithm and WOA, taking global exploration and local exploitation into account. Furthermore, ROA will determine whether the remora requires a host transition via an experience assault. If the host is able to locate food, the remora will search for food remnants near the host, a process known as host feeding [33]. The detailed process of the remora absorbing host, switching hosts, and hunting for food is depicted in Figure 3.

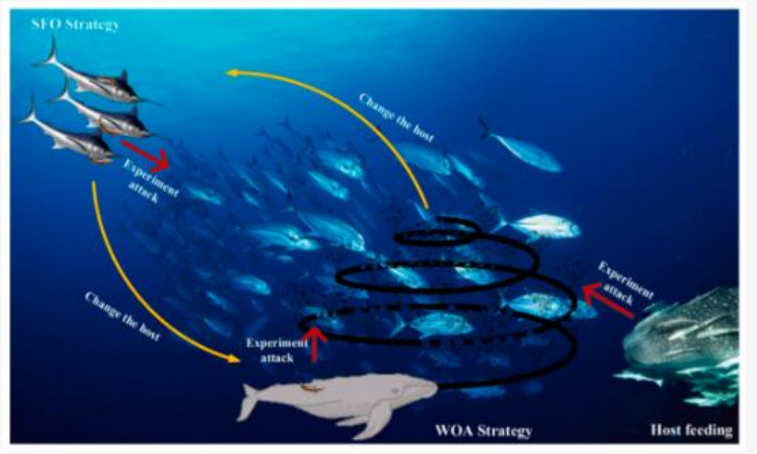


Figure 3. The schematic diagram of different stages of ROA

i. Initialization

Equation (6) initializes every outcome in the space of search in each dimension; the initialization is a set of randomly selected solutions.

$P_i = lb_j + rand \times (ub_j - lb_j)$ (6)

Between intervals [0, 1], random numbers can be entered into the variable rand . The formula indicates the jth dimension in the search space's upper and lower bounds as $lb_j$and ub_j, respectively, and the index value of the answer as i.

ii. Free Travel (Exploration)

The strategy of Sailfish Optimization (SFO): Because sailfish can swim quickly in the ocean, ROA selects sailfish as the primary host for the exploration phase. The SFO algorithm's elite approach is used to update the remora's location. The following is the revised Formula:

$$P_i^{t+1} = P_{best}^t - (rand \times \left(\frac{P_{best}^t + P_{rand}^t}{2}\right) - P_{rand}^t \quad (7)$$

Here, t indicates the number of iterations that are currently in progress, P_{best}^t indicates the current ideal placement, and P_{rand}^t indicates the random individual's position within the population. P_{best}^t represents the location of the ith remora following a relocation.

Experience Attack: The host serves as the remora's main source of food. If Remora cannot get food from their present host, they will choose to travel to a new one. As a result, the remora will only make a small range of exploratory movements around the host. This process is known as an experience assault. This is how the specific formula looks:

$$P_i^{att} = P_i^t + (P_i^t - P_i^{pre}) \times randn \quad (8)$$

randn can produces a random integer in the interval [0, 1] that follows the normal distribution. P_i^{att} denotes the location of the remora following the exploratory movement, P_i^t the ith remora's current position, and P_i^{pre} representsto the ith remora's position in the preceding generation.

The fitness values of P_i^{att} and $P_i^t$are represented by the variables $f(P_i^{att})$ and $f(P_i^t)$, respectively, sincef() is utilized to assess the fitness value that corresponds to the present position. The remora will decide to move hosts if $P_i^{att'}$s fitness value is greater than P_i^t's. The following is the formula for changing hosts:

$S(i) = round(rand)$ (9)

The host of the ith remora will be found via S(i).A whale will be chosen as the host and absorbed by the remora if S(i)= 1. In the event that not, the sailfish will be chosen as the host. Rounding is the purpose of the round.

iii. Eat Thoughtfully (Exploitation)

The technique of the Whale Optimization Algorithm (WOA) involves simulating the behavior of remora adsorbing on whales using a portion of its formulae. The populace now uses the following formula to update their position:

$$P_i^{t+1} = D \times e^a \times \cos(2\pi h) + P_i^t \quad (10)$$

$$D = |P_{best}^t - P_i^t| \quad (11)$$

$$a = round \times (h - 1) + 1 \quad (12)$$

$$h = -(1 + \frac{t}{T}) \quad (13)$$

Where D represents the difference between the actual and ideal selves, a is a random integer between [-1, 1], h declines gradually from 1 to -2, and e represents an infinite noncyclic fraction having an integer value of (2.71828).

Host feeding: This process, which occurs when the host locates food, is what we refer to as the remora searching for food nearby. The following is the actual mathematical formula:

$$P_i^{t+1} = P_i^t + A \quad (14)$$

$$A = B \times (P_i^t - C \times P_{best}) \quad (15)$$

$$B = 2 \times V \times rand - V \quad (16)$$

$$V = 2 \times (1 - t/T) \quad (17)$$

The distance that the remora will travel is represented by A in the formula, the remora factor, having a value of 0.1, is represented by C, and the host and remora's volumes are simulated by B and V.

The following is the ROA pseudocode for Algorithm 2:

Algorithm 2: The ROA pseudocode

Set the parameters to their initial values (N for population size, T for total iteration, and C for remora factor)
Use Formula (20) for initializing the population.
While (t < T)
Return search agents to the border from outside the search space
Evaluating the fitness value of each search agent and updating Pos_{Best}
For i in 1 to N
If S(i) == 0, then
Select a sailfish to serve as the host for the Formula (21) position update
Elseif S(i) == 1
Select a whale to serve as the host for the Formulas (24)– (27) position update.
End if
Move to the host's immediate vicinity using Formula (22).
If $f(Pos_i^{att}) < f(Pos_i^t)$
Update Posit and switch hosts using Formula (23)
Else
Formulas (28)–(31): Host feeding
End if
End for
t = t + 1
End While
Return Pos_{Best}^t

Chimp Optimization Algorithm (ChOA) in conjunction with Weighted Deep AutoEncoder (WDAE) for classification

The chosen attributes were categorized using WADE-ChOA. Auto Encoder is among the more effective techniques for reducing the dimensionality of data. AE-based network embedding techniques are frequently used to modify the input vectors or loss function. Structural deep network embedding (SDNE) was used to develop an improved semi-supervised deep model that improved the input vectors while maintaining the network framework. First- and second-order proximity were both considered via SDNE. The first-order proximity was added to the loss function as supervisory information in the model to better reflect the local network architecture. Under deep network representations, the loss function was improved in adversarial regularized autoencoders as both the locality preservation loss and the AE loss decreased simultaneously. Smoothly regularized vertex representations that accurately reflect the network topology were generated by the model by taking into consideration both locality-preserving constraints and global reconstruction with simultaneous impact.

Weighted Deep Auto Encoder (WDAE)

The step-by-step instructions for the WDAE algorithm are given here. Using deep autoencoders, the topological properties of the network are retrieved. First, the idea of autoencoders is presented. The network's topology is represented by the low-dimensional matrix H, which is created after creating a deep autoencoder and eliminating the features from the similarity matrix X. An unsupervised deep learning tool called the autoencoder can match input values with output values by using a back-propagation mechanism. The input is compressed by an autoencoder into a latent spatial illustration before the output is rebuilt. The autoencoder consists of two encoders, one for each. By using a latent spatial representation created by the encoder, the decoder may rebuild the input from its compressed form. To produce a low-dimensional vector h_i, take a vector $x_i \in R^{n*1}$ in the similarity matrix $X = \{x_1, x_2, \ldots, x_n\}$ obtained as an input to the autoencoder. Subsequently, the decoder receives this low-dimensional vector h_i and returns a vector x_i' that has the same dimensions as x_i. The low-dimensional vector h_i can accurately capture the original vector information if there is strong similarity between the input vector x_i and x_i'. To decrease the reconstruction error, we train the network using back-propagation and modify the encoder and decoder parameters. We can treat the output vector x_i' in this instance as being identical to the input vector x_i. Lastly, examine vector h_i, the intermediate layer, obtained from the x_i feature result. The autoencoder employs the particular training procedure listed below.

The autoencoder takes its input from the similarity matrix X, which is generated during data preparation. The correlation between each node in the network and every other node is represented by each column vector $x_i \in R^{n*1}$ of X. Formula (18) is used to acquire the coding output $h_i \in R^{k*1}$ for each coding layer after feeding x_i become a k-neuron encoder.

$h_i = s(Wx_i + b)$ (18)

where is the coding layer activation function denoted by s. The sigmoid function found in formula (19) will be utilized in this inquiry. The weight matrix $W \in R^{k*n}$, while the offset vector $b \in R^{k*1}$.

A. $S(x) = \frac{1}{1+e^{-x}}$(19)

Determine the result of the coding layer, h_i such that $h_i \in R^{k*1}$, where h_i is the low-dimensional vector associated with location i. Formula (20) is then utilized in training to get the decoding layer's output data, $x_i' \in R^{n*1}$. This is done after passing h via the decoding layer.

B. $x_i' = s(W'h_i + b')$ (20)

The decoding layer's weighted matrix $W' \in R^{k*n}$, while its offset vector $b' \in R^{k*1}$. During training, the autoencoder adjusts the encoding layer weight matrix $W \in R^{k*n}$, encoding layer bias vector $b \in R^{k*1}$, decoding layer weight matrix $W' \in R^{k*n}$, and decoding layer bias vector $b' \in R^{k*1}$ using the back-propagation algorithm to reduce the reconstruction error of x_i and x_i'. Thus, the value produced by formula (21) is reduced.

C. $L(x, x') = \|x - x'\|^2$ (21)

D. $= \|x - s(W'h_i + b')\|^2$ (22)

E. $= \|x - s(W's(Wx_i + b) + b')\|^2$ (23)

For the output to precisely correspond with the input, sparsity criteria are imposed to the hidden layer units. Neurons triggered by sigmoid activation functions exhibit activity when their output value is 1 and inactivity in other situations. Another interpretation of sparsity is that the neuron is not activated for the great majority of the time. Apply KL divergence as a sparsity constraint to the autoencoder and utilize the neuron's output as the activation degree.

F. $\sum_{j=1}^{k} KL(\rho | \left|\frac{1}{n}\sum_{i=1}^{n} h_i\right.)$ (24)

Using a mean of ρ and ρj, the average activation degree of the hidden layer for n samples is ρj= 1 n _ ni=1 h i, where KL _ ρ_ ρj _denotes the relative entropy of the two variables. The following formula may be used to compute KL divergence:

G. $\sum_{j=1}^{k} KL(\rho | \left|\hat{\rho}_j\right.) = \sum_{j=1}^{k} \rho \log \frac{\rho}{\hat{\rho}_j} + (1-\rho)\log \frac{1-\rho}{1-\hat{\rho}_j}$ (25)

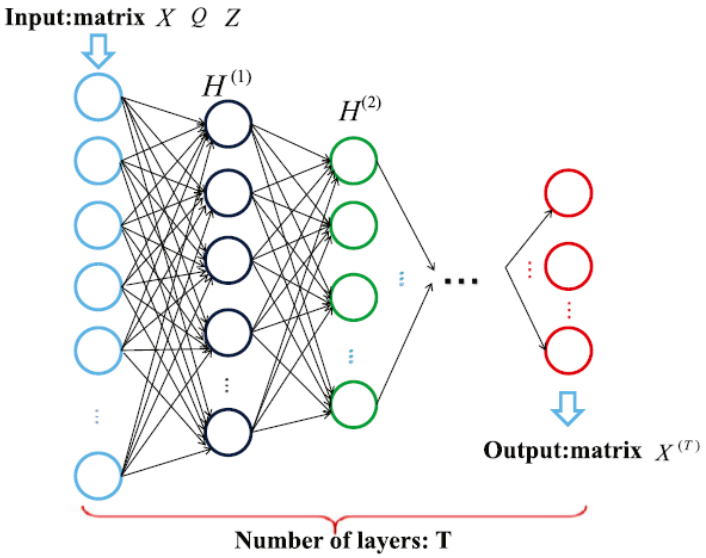


Figure 4. The structure of the deep autoencoder

Formula (26) summarizes the recreated error caused by the proposed autoencoder.

H. $L(x, x') = \|x - s(W's(Wx_i + b) + b')\|^2 + \partial \sum_{j=1}^{k} KL(\rho || \frac{1}{n}\sum_{i=1}^{n} s(Wx_i + b))$ (26)

The autoencoder significantly decreases the number of parameters that must be taught, makes training easier, and avoids the local minimal values and problems with overfitting that previous autoencoders

have. A deep autoencoder is made up of many autoencoders layered together. A deep autoencoder's architecture is shown in Figure 3.A deep autoencoder, unlike a regular autoencoder, has numerous hidden layers. Each layer receives and delivers new information, building on the characteristics of the previous layer. The preceding part used a single autoencoder to extract the low-dimensional feature vector h_i and train a three-layer network $x_i \rightarrow h_i \rightarrow x_i'$. The preceding autoencoder's middle-layer vector $h_i$is used as the source vector for the subsequent autoencoder, following a layer-by-layer training procedure. The training procedure continues until the desired number of layers is obtained. The Chimp Optimization Algorithm is employed to solve this problem.

Chimp Optimization Algorithm (ChOA)

The ChOAexperimental optimization method is based on the hunting behavior of chimpanzees. Chimpanzees in the "Attacker," "Barrier," "Chaser," and "Driver" subpopulations are distinguished from one another by the varying degrees of cognition and ability they exhibit when hunting. All chimpanzee species possess the ability to think for themselves and may organize and carry out a search strategy to locate and anticipate food. Even in circumstances when they have obligations, people are socially motivated to obtain benefits and have sex in the latter stages of their search [11]. It's a unique time of year for hunting activities.The average ChOA is as follows: Assume that X is the location of the i-th chimpanzee likewisethat there are N chimps overall. The greatest answer is "Attacker," followed by "Chaser," "Barrier," and "Driver." The way chimps' approach and encircle their prey, as well as the functioning of their position update algorithms, are explained below:

$$D = \left|C.X_{prey}(t) - m.X_{chimp}(t)\right| \quad (27)$$

I. $$X_{chimp}(t+1) = X_{chimp}(t) - A.D \quad (28)$$

J. $$A = f.(2.r_1 - 1), C = 2.r_2 \quad (29)$$

K. $$m = chaotic - value \quad (30)$$

whereby the random vectors r1 and r2 have values between 0 and 1. The coefficient of nonlinear decay, f, decreases linearly from 2.5 to 0 in relation to the number of repetitions. As of right now, t repeats have occurred. The random vector A has a value between f and -f. Chimpanzee locations are impacted by material rewards, as depicted by the chaotic factor M. The position of the prey inside [0, 2] is influenced by a random variable, C, on a per-chimp basis. When C > 1 and C < 1, respectively, the degree of effect rises and falls. Figure 5.9 illustrates the position updates technique. Other chimpanzees in the community can be located based on their positions as the driver, chaser, barrier, and attacker. These locations are determined using the position update formulae shown below:

$$D_{Attacker} = |C_1 * X_{Attacker} - m_1 * X| \quad (31)$$

L. $$D_{Barrier} = |C_2 * X_{Barrier} - m_2 * X| \quad (32)$$

M. $$D_{Chaser} = |C_3 * X_{Chaser} - m_3 * X| \quad (33)$$

N. $$D_{Driver} = |C_4 * X_{Driver} - m_4 * X| \quad (34)$$

O. $$X_1 = X_{Attacker} - A_1 * D_{Attacker} \quad (35)$$

P. $$X_2 = X_{Barrier} - A_2 * D_{Barrier} \quad (36)$$

Q. $$X_3 = X_{Chaser} - A_3 * D_{Chaser} \quad (37)$$

R. $$X_4 = X_{Driver} - A_4 * D_{Driver} \quad (38)$$

S. $$X(t+1) = (X_1 + X_2 + X_3 + X_4)/4 \quad (39)$$

Where C1, C2, C3, and C4 are comparable to C; A1, A2, A3, and A4 are similar to A; And m1, m2, m3, and m4 are all comparable to m.

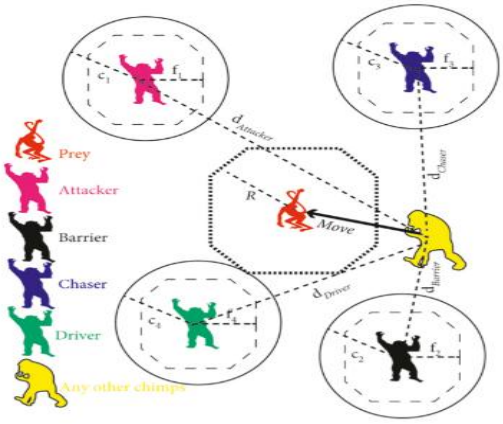


Figure 5. Procedure of position updations

In summary, generating some potential solutions (a stochastic chimp population) is the first step in the process of searching for prey in ChOA. After then, the group's members are split up into four self-regulating groups at random: driver, chaser, barrier, and attacker. To update its amounts, each clustermember applies the group technique. The chimps driving, pursuing, blocking, and attacking make estimates about potential prey locations during the repetition phase. Solutions update the distance of each group member from the target. To improve convergence and prevent local optima, the a and c vectors are adjusted.

RESULT AND DISCUSSION

Numerous data analysis tools have been developed. This study makes use of the MATLAB tool to diagnose early breast cancer and evaluate performance.

The proposed approach's performance is analyzed and tested utilizing datasets to establish its stability. Assessment measures like as recall, precision, F-Measure, accuracy, and the kappa statistic are used to identify data.

Accuracy is defined as the precision of accurately detecting items.

$$\text{Accuracy} = \frac{\text{TP}+\text{TN}}{\text{Total no .of Samples}} \qquad (40)$$

Determining the exactness of the classifier is the definition of precision (P).

$$\text{Precision (P)} = \frac{\text{TP}}{\text{TP}+\text{FP}} \qquad (41)$$

Computing the sensitivity or the effectiveness of the classifiers is what is meant by recall (R).

$$\text{Recall (R)} = \frac{\text{TP}}{\text{TP}+\text{FN}} \qquad (42)$$

Recall and accuracy have a possible harmonic mean called F-Measure.

$$\text{F} - \text{Measure} = 2 * \left[\frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}\right] \qquad (43)$$

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

The Kappa Statistic is an evaluative tool that compares the actual and predicted levels of agreement. One represents complete agreement. The number ranges from 0 to 1. Utilizing equation 1, get the kappa statistic value.

$$K = \frac{p_0 - p_e}{1 - p_e} = 1 - \frac{1 - p_o}{1 - p_e} \qquad (44)$$

This denotes the agreement probability p_e and the ten associated agreements that were documented p_o.

```
Attributes reduced

datas1 =

  5×1 cell array

    {'Samplecode'           }
    {'ClumpThickness'       }
    {'UniformityofCellSize' }
    {'UniformityofCellShape'}
    {'MarginalAdhesion'     }
```

Figure 6. Attributes Reduction

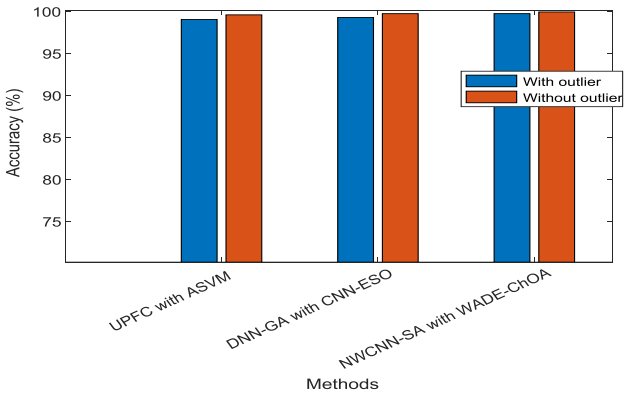


Figure 7. Accuracy Comparison based on Outliers

A comparison of accuracy depending on outliers is presented in Figure 7. An accuracy comparison graph for outlier identification is provided by both the current and the planned study.

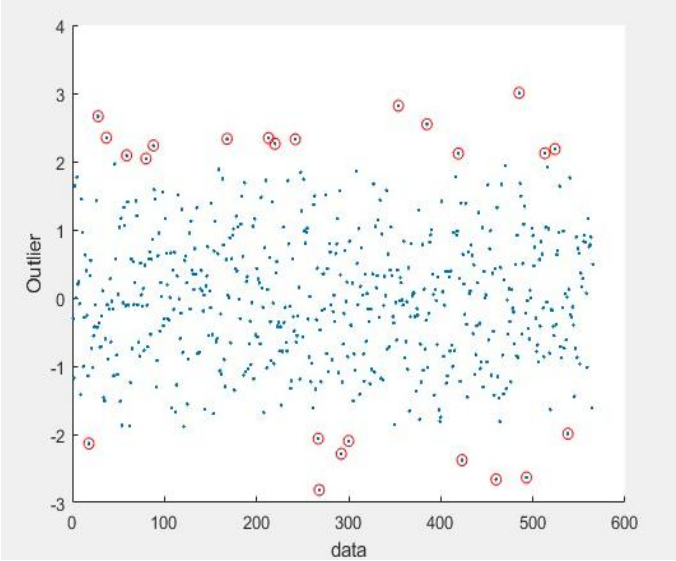


Figure 8. Outlier Graph

Figure 8 shows the outlier detection graph.

Result Analysis

The study uses 10-fold cross validation in this instance. Ten equal-sized pieces were randomly selected from the data set. The other sections are used to train the beginning student, while one division is used for assessment. To test each partition only once, the previous approach is repeated ten times. After then, the individual statistics are aggregated according to their average correctness.The trials were carried out using MATLAB. The algorithm for ASVM classification is put into practice. Table 3 and Figure 5 data show how effective the classification method is for BC diagnosis. The proposed model achieves a very optimistic anticipated performance of 99.10% classification accuracy for WBC data and 99.35% classification accuracy for WDBC data, compared with the results from current approaches.

Table 3. The Values Obtained for Evaluation Metrics in WBC And WDBCData Sets

Evaluation Metrics	**Existing K-Means with Decision Tree Algorithm**		**UPFC with ASVM algorithm**		**DNN-GA-ESO Algorithm**		**Proposed NWCNN-SA with WADE-ChOA**	
	WBC Value	**WDBC Value**	**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	99.50	99.65
Precision	98.2	99.0	99.24	99.26	99.35	99.44	99.41	99.48
F-Measure	98.1	99.0	99.12	99.29	99.36	99.50	99.43	99.52
Recall	98.1	99.0	99.01	99.33	99.42	99.53	99.50	99.59
Kappa statistics	96.2	97.87	98	98.37	98.48	98.55	99.55	99.63

For comparison's sake, Table 4 presents the classification accuracies achieved using many popular classifiers from the literature as well as the methodology that is recommended.

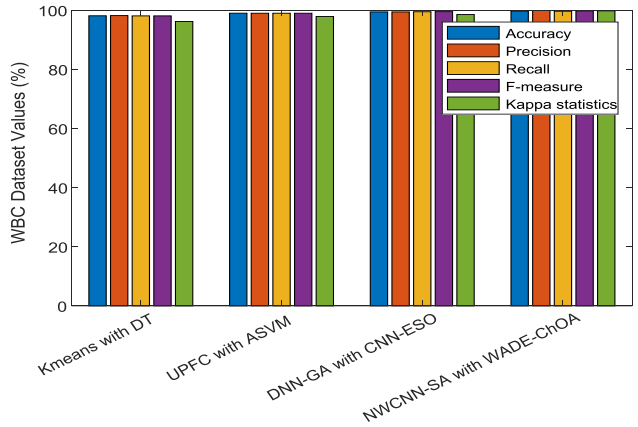


Figure 9. Comparison of approaches on WBC dataset using various metrics

Figure 9 presents a comparison of several metrics on the WBC dataset, including classification accuracy (%), precision, F-Measure, recall, and kappa statistics, between K-Means and Decision Tree method, UPFC and ASVM, DNN-GA-ESO method, and suggested NWCNN-SA with WADE-ChOA Algorithm. When compared to the current method, the suggested NWCNN-SA with WADE-ChOA algorithm performs better in terms of classification accuracy (%), precision (99.41%), F-Measure (99.43), recall (99.50), and kappa statistics (99.55).

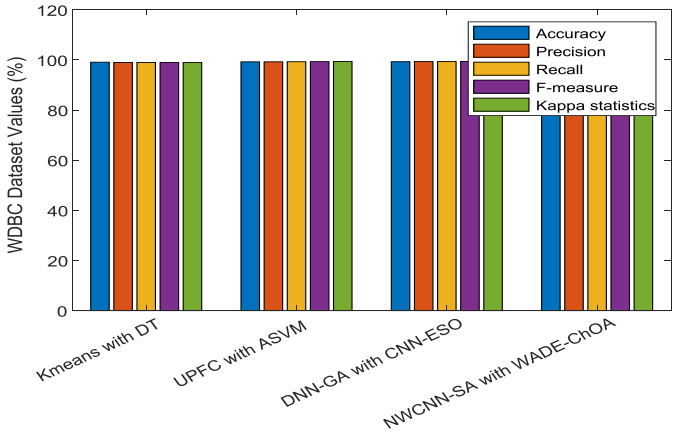


Figure 10. Comparison of approaches on WDBC dataset using various metrics

Figure 10 presents a comparison of many metrics on the WDBC dataset, including classification accuracy (%), precision, F-Measure, recall, and kappa statistics, between K-Means and Decision Tree method, UPFC and ASVM method, DNN-GA-ESO, and Proposed NWCNN-SA with WADE-ChOA Algorithm.

Table 4. Classification Accuracies achieved with proposed system andother existing classifiers inBC

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
Phase 1 Approach	UPFC with ASVM algorithm	99.37
Phase 2 Approach	DNN-GA-ESO Algorithm	99.45
Proposed Approach	Proposed NWCNN-SA with WADE-ChOA	99.65

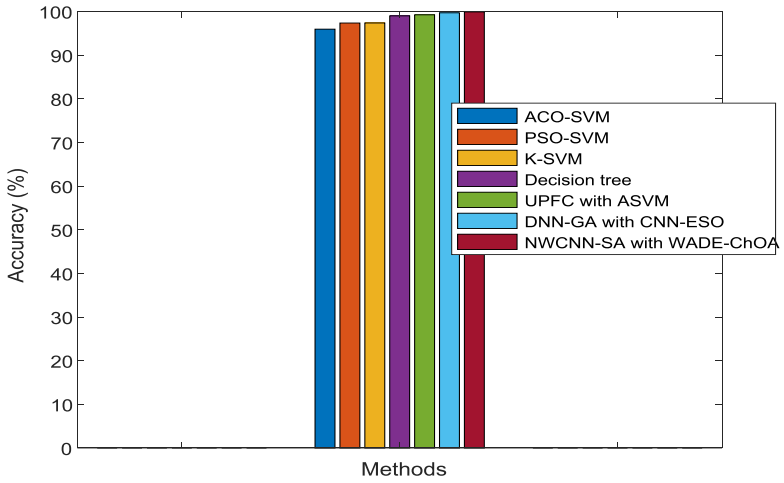


Figure 11. Comparison of classification Accuracy using various existing and proposed method

Figure 11 depicts a comparison of classification accuracy between recommended and existent approaches. While existing algorithms, such as the UPFC with ASVM algorithm, decision tree classifier, K-SVM, PSO-SVM, ACO-SVM, and DNN-GA-ESO obtain 99.37%, 99.01%, 97.38%, and 97.37%, respectively, and 99.45%, reach 99.65%, the suggested technique achieves 99.65%. As a consequence, the suggested suggested NWCNN-SA with WADE-ChOA algorithm outperforms the present approaches.

CONCLUSION

Early detection of breast cancer is critical for dealing with medical difficulties. The usage of data mining techniques improves accuracy. This study proposed a deep learning-based classification model and outlier detection technique. Three methodologies, such as outlier identification, feature selection, and classification processes, have been investigated throughout this study, and their efficacy has been measured using a range of measures. The 10-fold cross validation approach produces an accuracy of 99.30% for the WBC dataset with an outlier and 99.65% for the WDBC dataset without an outlier. According to the data, the Integrated Meta-algorithm beat other models in terms of accuracy. Overall, it was shown that integrated algorithms were more successful at identifying breast cancer than other procedures. Furthermore, the best data mining approach was applied to include the high-risk level and develop the suitable classification for spreading information for patients' treatment from ordinary people.

REFERENCES

[1] Kamboj A, Tanay P, Sinha A, et al. Breast cancer detection using supervised machine learning: a comparative analysis. Singapore: Springer;2021:263–9.
[2] Solanki P, Baldaniya D, Jogani D, et al. Artificial intelligence: new age of transformation in petroleum upstream. Pet Res 2021. doi:10.1016/j.ptlrs.2021.07.002.
[3] Fotouhi H, Čaušević A, Lundqvist K, et al. Proceedings - International Computer Software and Applications Conference, Atlanta, United States. IEEE Computer Society; 2016. doi:101109/COMPSAC20168.
[4] Hady AA, Ghubaish A, Salman T, et al. Intrusion detection system for healthcare systems using medical and network data: a comparison study. IEEE Access 2020;8:106576–84. doi:10.1109/ACCESS.2020.3000421.
[5] Singh S, Deep A, Mohanta G, et al. In next generation point-of-care biomedical sensors technologies for cancer diagnosis. Singapore: Springer;2017:253–78.
[6] Mihaylov I, Nisheva M, Vassilev D. Machine learning techniques for survival time prediction in breast cancer. Lecture Notes in Computer Science Springer, Cham. 2018: 186–94, doi:10.1007/978-3-319-99344-7_17.
[7] Benbrahim H, Hachimi H, Amine A. Comparative study of machine learning algorithms using the breast cancer dataset. AdvIntell Sys Comp 2020;1103:83–91. doi:10.1007/978-3-030-36664-3_10.
[8] Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. CA Cancer J Clin 2016;66(1):7–30. doi:10.3322/caac.21332.

[9] Kaushal C, Singla A. Analysis of breast cancer for histological dataset based on different feature extraction and classification algorithms. Adv Intelligent Sys Comp 2021;1165:821–33. doi:10.1007/978-981-15-5113-0_69.
[10] Uhr JW. Cancer diagnostics: one-stop shop. Nature 2007;450:1168–9. doi:10.1038/4501168a.
[11] Geiger TR, Peeper DS. Metastasis mechanisms. BiochimBiophysActa 2009;1796:293–308. doi:10.1016/j.bbcan.2009.07.006.
[12] Liberko M, Kolostova K, Bobek V. Essentials of circulating tumor cells for clinical research and practice. Critical Rev Oncol 2013;88:338–56. doi:10.1016/j.critrevonc.2013.05.002.
[13] Wang K, He MQ, Zhai FH, et al. A novel electrochemical biosensor based on polyadenine modified aptamer for label-free and ultrasensitive detection of human breast cancer cells. Talanta 2017.
[14] Akay, M. F. (2009). Support vector machines combined with feature selection for breast diagnosis. Expert Systems with Applications, 36, 3240–3247
[15] Ubeyli, E. D. (2007). Implementing automated diagnostic systems for breast cancer detection. Expert Systems with Applications, 33, 1054–1062
[16] Karabatak, M., & Ince, M. C. (2009). An expert system for detection of breast cancer based on association rules and neural networks. Expert Systems with Applications, 36, 3465–3469
[17] R. DelshiHowsalya Devi, M.IndraDevi,A Novel Hybrid Algorithm for Outlier Detection Using Weka Interface. International Journal of Applied Engineering Research, ISSN 0973-4562 Vol. 10 No.55 (2015)
[18] Xue J, Shen B (2020) A novel swarm intelligence optimization approach: sparrow search algorithm. SystSci Control Eng 8(5):22–34. https://doi.org/10.1080/21642583.2019.1708830
[19] Nguyen B, Xue B, Zhang M (2020) A survey on swarm intelligence approaches to feature selection in data mining. Swarm EvolComput 54:100663. https://doi.org/10.1016/j.swevo.2020.100663
[20] Rabie A, Ali S, Saleh A, Ali H (2020) A new outlier rejection methodology for supporting load forecasting in smart grids based on big data. ClustComput Springer 23:509–535. https://doi.org/10.1007/s10586-019-02942-0
[21] Chambers L, Gaber MM, Abdallah ZS. DeepStreamCE: a streaming approach to concept evolution detection in deep neural networks, 2020;http://arxiv.org/abs/2004.04116.
[22] Sadhukhan S, Upadhyay N, Chakraborty P. Breast cancer diagnosis using image processing and machine learning. Adv Intel Sys Comp 2020;937:113–27. doi:10.1007/978-981-13-7403-6_12.
[23] Osmanović A, Halilović S, Ilah LA, et al. Machine learning techniques for classification of breast cancer. IFMBE Proc 2019;68:197–200. doi:10.1007/978-981-10-9035-6_35.
[24] Negi R, Mathew R. Machine learning algorithms for diagnosis of breast cancer. lecture notes on data engineering and communications technologies. Springer Sci Bus Media Deutschl GmbH 2020;31:928–32.
[25] Irfan, R., Almazroi, A. A., Rauf, H. T., Damaševičius, R., Nasr, E. A., &Abdelgawad, A. E. (2021). Dilated semantic segmentation for breast ultrasonic lesion detection using parallel feature fusion. Diagnostics, 11(7), 1212.
[26] V. Lahoura, H. Singh, A. Aggarwal et al., "Cloud computing-based framework for breast cancer diagnosis using extreme learning machine," Diagnostics, vol. 11, no. 2, p. 241, 2021.
[27] Magna AAR, Allende-Cid H, Taramasco C, Becerra C, Figueroa RL. Application of machine learning and word embeddings in the classification of cancer diagnosis using patient anamnesis. IEEE Access. 2020;8:106198–213.
[28] Vaka AR, Soni B, Reddy S. Breast cancer detection by leveraging machine learning. ICT Exp. 2020;6(4):320–4.
[29] Saxena S, Gyanchandani M. Machine learning methods for computer-aided breast cancer diagnosis using histopathology: a narrative review. J Med Imaging Radiat Sci. 2020;51(1):182–93.
[30] Kayikci S, Khoshgoftaar T. A stack based multimodal machine learning model for breast cancer diagnosis. In: 2022 international congress on human–computer interaction, optimization and robotic applications (HORA). IEEE; 2022. p. 1–5.
[31] Berry, M. W., Browne M., Langville A. N., Pauca V. P., Plemmons R. J., "Algorithms and Applications for Approximate Nonnegative Matrix Factorization", Computational Statistics and Data Analysis, Vol. 52, No. 1, pp. 155-173, 2007.
[32] Amine, K. (2019). Multiobjective simulated annealing: Principles and algorithm variants. Advances in Operations Research, 2019.
[33] Jia, H., Peng, X., & Lang, C. (2021). Remora optimization algorithm. Expert Systems with Applications, 185, 115665.