

# Classification of Digital Mammogram Images using Wrapper based Chaotic Crow Search Optimization Algorithm

**R.Reenadevi<sup>1\*</sup>, T.Sathiya<sup>2</sup>, B.Sathiyabhama<sup>3</sup>**

<sup>1,2,3</sup>Department of Computer Science and Engineering, Sona College of Technology, Salem, India

\*reenadevi@sonatech.ac.in

## ABSTRACT

Breast disease is a major cause of death in both women and men around the world. An early diagnosis of breast cancer disease with the assistance of mammogram images is crucial for patients to be treated properly so that they can live a healthy life. While several early-stage strategies for diagnosing breast cancer disease have been developed, most of them are not effective. In this research, an optimized Wrapper-based Chaotic Crow Search Algorithm (WCCSA) is developed to improve the diagnosis of Breast cancer disease. The proposed WCCSA can be used with Probabilistic Neural Network (PNN) to identify the mammogram images as malignant, benign, and normal, support individuals to receive appropriate care treatment in earlier. The effectiveness of WCCSA with PNN is evaluated using a mini-Mammographic Image Analysis Society (MIAS) dataset of 322 images, and the performance were compared to those of other machine learning algorithms. An evaluation result shows that the proposed WCCSA with PNN method seeks an optimal subset of features while maintaining stability, and achieved an accuracy of 97%.

**Keywords:** Breast cancer, Mammogram, Wrapper, Chaotic Crow search, Probabilistic Neural Network, Classification.

## 1. Introduction

Breast cancer is an epidemic in the world that diagnoses millions of new cases annually. Aside from that, over 100,000 women die every year [1,22]. One in every 30 people is now predicted to die of breast cancer. The radiologist's early detection of breast cancer lowers the global death rate. There are several approaches used for detecting breast cancers among which digital mammography is the most popular and reliable technique used by radiologists. This test also ensures the identification of other pathologies that indicate the existence of cancer as benign, malignant, or normal. The most significant advancement in breast imaging is digital mammography. The early detection of breast irregularities can be done through a computer-aided detection (CAD) system with machine learning [2] technologies on digital mammogram images. The CAD will help radiologists make accurate and reliable medical diagnoses. CAD will also help to ensure that visual fatigue and the effort needed during the diagnosis process are avoided as a result of humane investigative errors.

Several researchers have recently investigated and suggested a CAD method for detecting and identifying breast abnormalities in mammography images. A CAD method consists of several stages such as preprocessing, enhancing, segmenting, extraction of features, selection of features, and classification. In this, Image enhancement is a preprocessing phase that is used to enhance digital images and improve image quality. Image segmentation [3,21] is one of the essential steps in the diagnostic process for the removal of breast cancer muscle and segmenting tumors from images of the mammograms. From the segmented images, essential features are extracted for further processing. Some of the important feature extraction [4,23] techniques used in the diagnosis of breast image such as texture, detector, morphological, model-based, and statistical. Statistical features [5] are the spatial distribution of intensity values of the pixels that provides information about a particular pixel. Selection of features [6,20] is an essential additional activity to be performed before model formation. It has been proved that effectively eradicate redundant and irrelevant features. Also, it can boost classifier efficiency, reduce computer costs and reduce storage requirements.

In recent decades, evolutionary computation has received a lot of attention for solving optimization problems. Taking inspiration from nature, several algorithms have been suggested. One of the meta-heuristic approaches influenced by crows' intelligent actions while hiding and stealing foods is the Crow Search Algorithm (CSA) [7]. Crows are thought to be among the world's most intelligent animals. These evolutionary algorithms can be modified to be used for feature selection. A new CAD method based on modified wrapper-based chaotic CSA is

proposed in this paper to distinguish abnormal and normal breast masses from mammogram images. The following are the main contributions of this work:

1. For efficient feature selection, a natural-based novel optimization method called WCCSA has been implemented.
2. WCCSA has been experimented with in mini-MIAS datasets for the recognition of breast cancer disease and has a 97 % accuracy rate.
3. The PNN algorithm is used for classification, and it is validated on 322 images of MIAS datasets.
4. The proposed algorithm's efficiency is compared to that of machine learning algorithms. The results show that the WCCSA outperforms the compared models, and WCCSA's performance significantly improved.

## 2. Literature Review

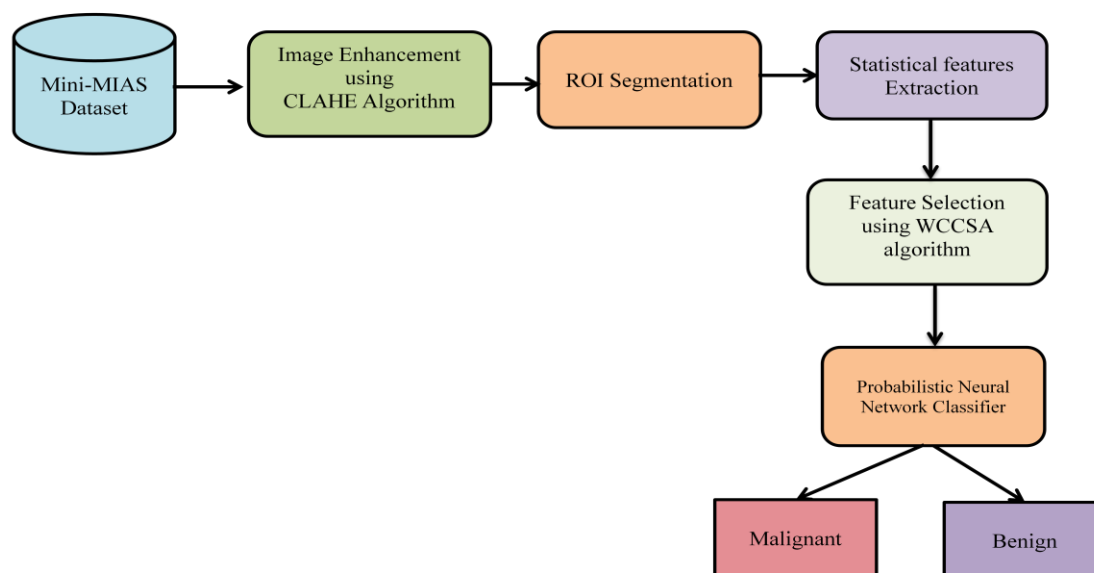
Alqudah et al. [8] created a CAD method for automatic breast mass segmentation and two-stage classification. In the first level, the masses are classified into seven groups, which are achieved using a PNN classifier. In the second stage, the masses are classified into two groups as benign and malignant using a support vector machine (SVM) algorithm. They have employed two different breast cancer datasets such as mammographic image analysis society database (MIAS-DB) of 322 images, breast cancer digital repository (BCDR) with 936 images for their analysis. They have achieved an accuracy of 97.08%, 99.18% respectively. For identifying the regions of interest, Parvathavarthini et al. [9] proposed a fuzzy-based crow search algorithm for finding the region of interest. They have used the mini-MIAS database for their analysis.

Arafa et al. [10] developed a CAD system for breast cancer recognition using a Gaussian Mixture Model (GMM) for mammogram image segmentation and SVM for classification. The mini-MIAS dataset used for their analysis and achieved an accuracy of 92.5%. Elgin et al. in [11] developed a framework for clinical diagnosis of the Wisconsin Diagnostic Breast Cancer (WDBC) dataset. They have used a wrapper-based method by integrating three different bio-inspired algorithms, known as, Glowworm Swarm Optimization algorithm, lion optimization algorithm, and Differential Evolution. An AdaBoost algorithm with SVM was used for the selection of features, gradient descent backpropagation neural network algorithm were used for classification and obtained 98.4% of accuracy.

Guha et al. [5] proposed an chaotic based whale survival algorithm, which employs the wrapper method for classification and a filter-based method to choose the subset of features. They have used four types of chaotic maps such as circular, logistics, piecewise, and tent to direct the choice of the kind of movement pursued by whales when looking for hunt. Arora et al. [12] implemented a hybrid based algorithm by combining Crow Search Algorithm with Grey Wolf Optimization to solve the optimization problems. Ahmed et al. [19] implemented a hybrid based optimization algorithm for selecting the optimized features in the applications of medical analysis. They have integrated the chaos theory with crow search algorithm and fuzzy c-means (FCM) technique. To avoid the sensitivity of local optimization the CSA is applied to achieve global optimization. In this, FCM objective function is employed as a cost function for the chaotic CSA.

## 3. Materials and Methods

The mini-MIAS database dataset [13] is used in this work for the diagnosis of breast cancer. The database consists of 322 mammogram images with annotated information. All images are available in 1024 x 1024 resolutions. The proposed CAD system is illustrated in **Figure 1**. It consists of phases such as enhancement, segmentation, feature extraction, selection, and classification. The contrast-limited adaptive histogram equalization (CLAHE) technique is applied to improve the input mammogram images [14]. The regions of interest (ROI) segmentation technique is used to removing the tumor region and suppress any objects in the mammograms. The GLCM statistical features were obtained from the segmented images and are used as an input of feature selection.

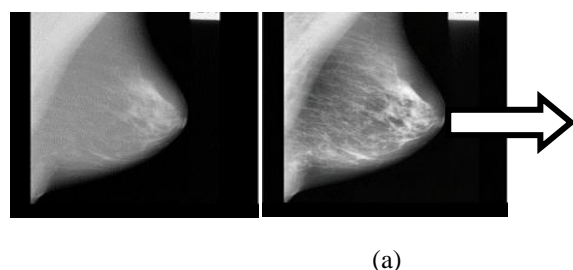


**Figure 1. Proposed CAD system flow**

Finally, an optimized wrapper-based chaotic crow search feature selection algorithm with a probabilistic neural network is used to choose that feature that maximizes the accuracy of the classifier.

### 3.1 Image Enhancement

Enhancement involves the processing of images to improve contrast and remove noise to aid radiologists in detecting the abnormalities. CLAHE (Contrast Limited Adaptive Histogram Equalization) is an image quality enhancement technique that has been shown to work well on medical images. CLAHE is superior to histogram techniques in medical imaging [14]. The following **Figure 2** shows the enhancement of the image using a threshold value of 0.0005.

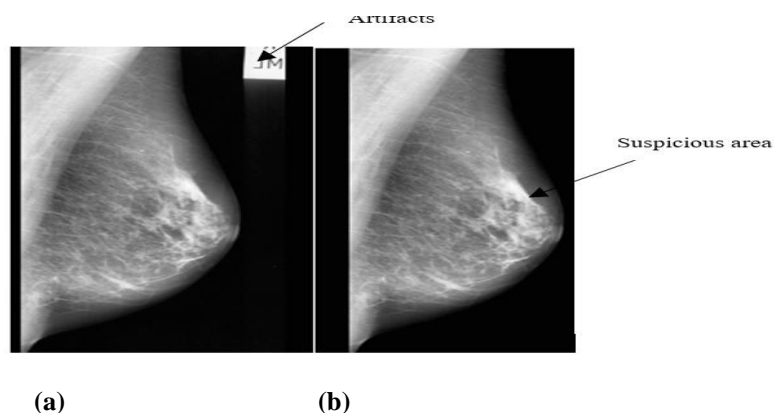


**Figure 2. (a) Abnormal mass of an original image and (b) Enhanced image using CLAHE method**

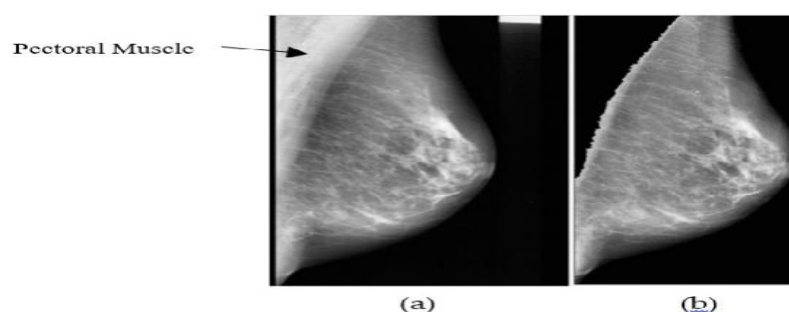
The intensity value of the histogram is directly related to contrast enhancement. This means both the slope and the height of the histogram are minimized by regulating the improvement of contrast.

### 3.2. Image Segmentation

Segmentation divides an image into segments with identical characteristics and properties. It is primarily intended to make simpler the image to enable easy analysis. The region of interest (ROI) [15] was derived from the enhanced mammograms as shown in Figure 3 and Figure 4, except for the pectoral muscle and all other objects by eliminating the whole breast. The triangle-shaped region on one side of the mammogram's MLO view is the pectoral muscles.



**Figure 3. (a) Enhanced image of an Original abnormal mass and (b) Suspicious image artifacts**



**Figure 4. (a) Enhanced image of an Original abnormal mass and (b) Pectoral muscle removal.**

The pectoral muscles have a comparable density to the dense tissues of the mammograms. Consequently, pectoral muscle extraction is an important factor in the accurate findings of cancer cells.

### 3.3. Feature Extraction

In an initial process of extraction, each image is separated into  $16 \times 16$  blocks. Following that, from each block of an image within the spatial domain, the statistic features [16] are determined. The mean of these statistical values such as root means square (RMS), variance, entropy, maximum, minimum, mean, standard deviation, and range is determined. Then all these functions were being merged into a single function vector to effectively extract the features.

### 3.4 Feature selection

The selection of features is achieved to reduce the irrelevant features and to improve the entire system performance. The main objective of the proposed algorithm is to determine a subset that predicts the target effectively while reducing the cost of the computations. In this work, proposed algorithm is extensively applied to achieve the benefits of reducing the complexity of the model, training faster, and improving the accuracy.

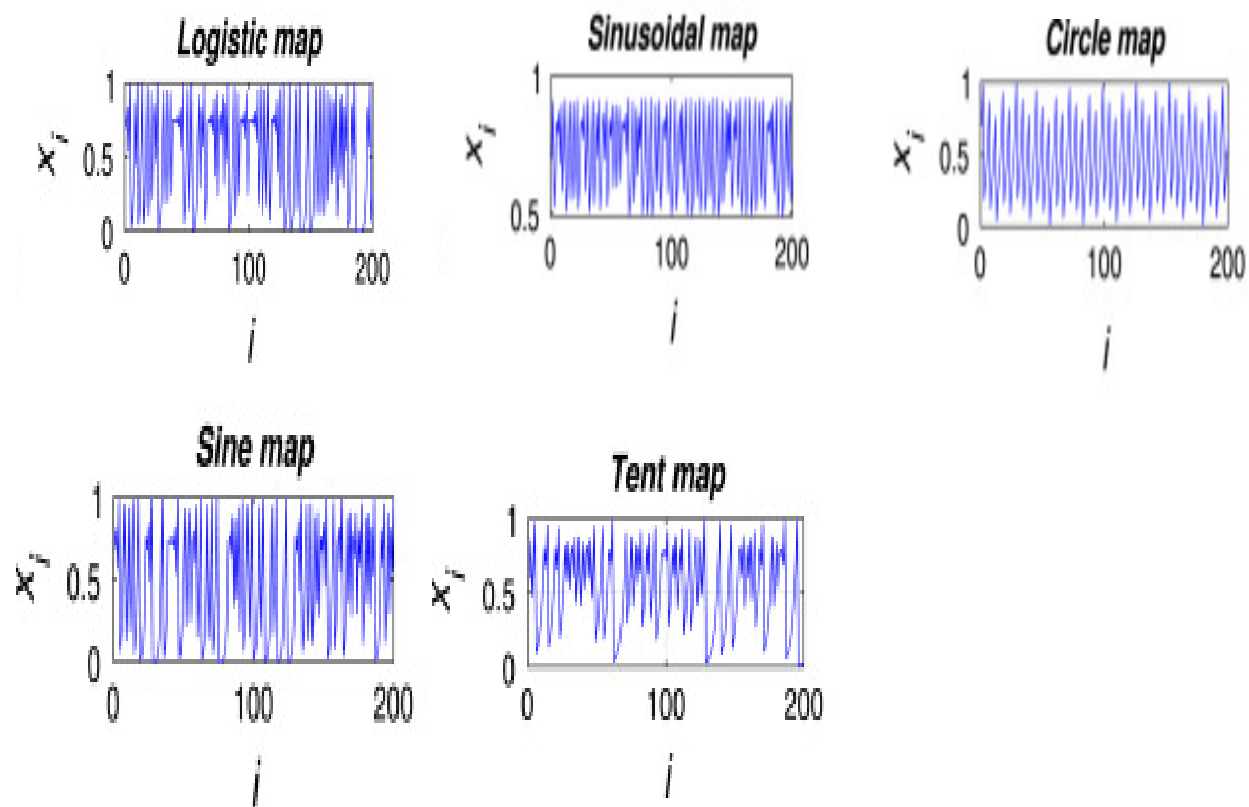
#### 3.4.1 The proposed algorithm

In this proposed work, an optimized wrapper-based chaotic CSA is specifically applied to select the best features from the segmented features of ROI. The proposed algorithm is presented in Algorithm 1. In this, chaotic maps are used for updating the random variables. The purpose of the chaotic sequence is to update the crow position which influences the achievement of optimal results and rate of convergence. In this, there are five types of chaotic maps such as logistic, sine, circular, sinusoidal, and tent are used. These maps are defined mathematically as illustrated in Table 1.

**Table1. Mathematical definitions of Chaotic maps**

Name	Definition	Range
Circle	$p_{q+1} = \text{mod}(p_q + d - (c/2\pi) \sin(2\pi p_q), 1)$ $d = 0.2, c = 0.5$	(0,1)
Logistic	$p_{q+1} = (1 - p_q)cp_q, c = 4$	(0,1)
Sine	$p_{q+1} = \sin(\pi p_q) c/4, c = 4$	(0,1)
Sinusoidal	$p_{q+1} = cp_q \sin(\pi p_q), c = 2.3$	(0,1)
Tent	$p_{q+1} = p_q/0.7, p_q < 0.7$ $10/3(1 - p_q), p_q \geq 0.7$	(0,1)

The graphical representations of the chaotic maps employed in WCCSA are shown in **Figure 5**. The visualization demonstrates the random existence of the chaotic mappings. The chosen chaotic maps can be distinguished and the results of incorporating various chaotic maps in the proposed approach can also be properly seen.

**Figure 5. Visualization of Chaotic maps**

For each iteration, the following fitness function is used to test any crow spot.

$$Fnt = \text{Max} (Acc + w_f \times (1 - L_f L_t))(1)$$

Where  $Acc$  defines the overall classification accuracy.  $w_f$  indicates the weighted factor of the range [0, 1],  $L_t$  denotes the number of features used,  $L_f$  denotes the feature-length. The value of  $w_f$  is 0.8 and it is also employed to monitor the significance of the classifier for selected features. In this work, there are eight statistical features such as root mean square (RMS), variance, entropy, maximum, minimum, mean, standard deviation, and range were chosen.

These features were given as the input of the PNN classifier for classifying the breast images into benign and malignant. The PNN classifier is utilized to evaluate the usefulness of the features that are selected. The best solution is to optimize the precision in classification and reduce the chosen features.

---

**Algorithm 1** Chaotic crow search algorithm
 

---

```

1: Set the initial values of  $M$ ,  $AP$ ,  $fl$ , and  $tMax$ .
2: Initialize the crow position  $y$  randomly.
3: Evaluate the fitness function of each crow  $F_n(y)$ .
4: Initialize the memory of search crow  $N$ 
5: Set  $t := 1$ . {Counter initialization}.
6: repeat
7:   for ( $j = 1 : j \leq M$ ) do
8:     Get value of chaotic map  $C$ 
9:     if  $C_z \geq AP^{z,t}$  then
10:       $y^{j,t+1} = y^{j,t} + C_j \times fl^{j,t} \times (N^{z,t} - y^{j,t})$ 
11:     else
12:       $y^{j,t+1} = A$  random position of the search
13:    end if
14:     $y^{j,t+1} = \begin{cases} 1 & \text{if } (s(y^{j,t+1})) \geq rand() \\ 0 & \text{Otherwise} \end{cases}$ 
15:  end for
16:  Check the feasibility of  $y^{j,t+1}$ 
17:  Evaluate the new position of crow  $F_n(y^{j,t+1})$ 
18:  Update the crow's memory  $N^{j,t+1}$ 
19:  Set  $t = t + 1$ . {Iteration counter increasing}.
20: until ( $t < tMax$ ). {Termination criteria satisfied}.
21: Produce the best solution  $N$ .
  
```

---

### 3.4.2. Probabilistic Neural Network

PNN is a kind of feedforward neural network that is used as a classifier with an approximation of each layer's probability distribution function [18]. The approximation is achieved by the use of the parzen window and nonparametric function provided in the step of the Bayesian rule. It is divided into four sections which are completely interconnected: input, sequence, summation, and output. In contrast to other neural network types, the pattern part is triggered using an exponential function rather than a sigmoid. In this work the spread constants to be set as 0.2 after comparing the PNN output accuracy over different spread constant values. **Figure 6** shows a simple PNN classification structure.

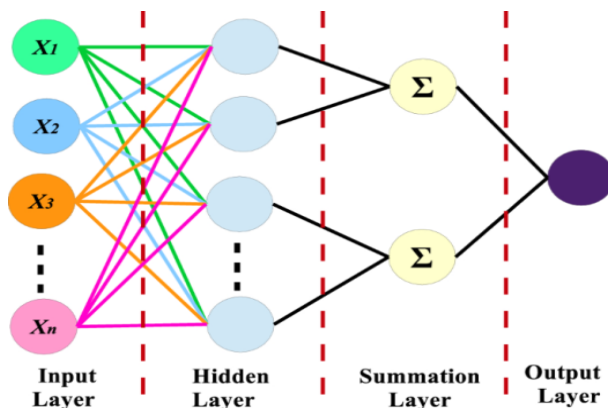


Figure 6. Structure of PNN

#### 4. Experimental Results and Discussion

The mammogram images to be examined are taken from the mini- MIAS [13] database of 322 images with a resolution of 1,024 x 1,024 pixels. The data are randomly divided into two separate parts: the training and testing of k-fold data sets. The value of k is set to 10 to ensure that the obtained results are stable. To evaluate the proposed system efficiency, a desktop computer with a 2.0 GHz Core 2 CPU and a 3GB of memory running Ubuntu is used. An algorithm is implemented by using a recent Python version. The parameters which are used in the WCCA algorithm are initialized as given in **Table 2**.

**Table 2. Initialization of parameters**

WCCA Parameters	Value
AP	0.1
M	30
Lower bound	0
f l	2
tMax	50
Upper bound	1
D	Total features

To increase accuracy by reducing features and computing costs, the proposed WCCSA algorithm is used. Instead of estimating the accuracy using some formula, the fitness function is modified by the fact it increases the measuring cost. PNN is used to forecast classification accuracy. To validate the efficiency of the proposed algorithm, it has been compared to other machine learning classifiers. The holdout approach is assessed by 75% of the training dataset and 25% of the testing dataset. Table 3 displays the statistical analysis performed and features extracted from the datasets to interpret the extracted.

**Table 3. Statistical analysis of feature subset**

Feature	Statistical analysis			
	Minimum	Maximum	Mean	Standard Deviation
Range	65.25	126.47	87.03	11.76
RMS	68.78	126.47	100.36	9.70
Standard Deviation	15.23	34.48	19.34	2.41
Variance	375.05	2078.91	656.39	176.06
Maximum	110.52	174.14	141.43	10.89
Entropy	3.27	5.13	4.32	0.33
Mean	62.11	122.29	96.73	10.08
Minimum	28.07	81.23	58.84	9.84

The effectiveness of the proposed system is analyzed by using the metrics of accuracy, sensitivity, specificity, and F1-score. The parameters are specified by Equations (2)–(5), and Table 4 demonstrates an assessment of the proposed algorithm and existing classification algorithms.

$$\text{Accuracy (\%)} = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \quad (2)$$

$$\text{Sensitivity (\%)} = \text{TPR} = \frac{TP}{TP + FN} \times 100 \quad (3)$$

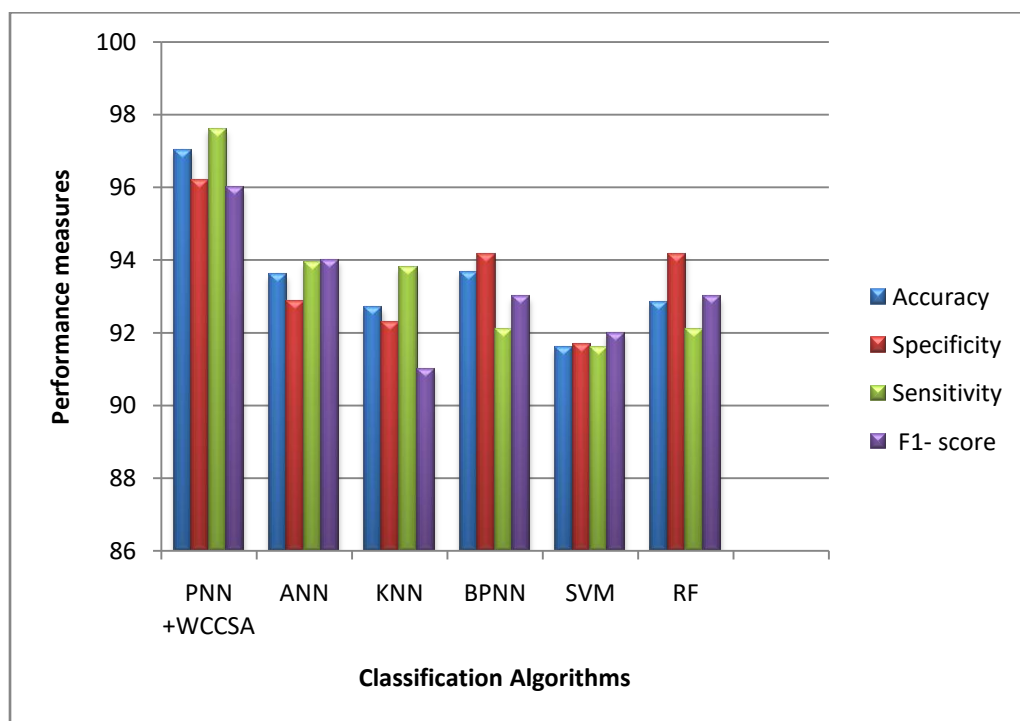
$$\text{Specificity (\%)} = \frac{TN}{TN + FP} \times 100 \quad (4)$$

$$F1 \text{ score } (\%) = \frac{\text{Specificity} \times \text{Sensitivity}}{\text{Specificity} + \text{Sensitivity}} \quad (5)$$

**Table 4. Comparison of proposed and existing classification algorithms using average results of 10-fold cross-validation measure**

Algorithm	Accuracy (%)	Specificity (%)	Sensitivity (%)	F1 score (%)
<b>PNN +WCCSA</b>	<b>97</b>	<b>96.2</b>	<b>97.6</b>	<b>96</b>
ANN	93.6	92.86	93.94	94
KNN	92.7	92.3	93.8	91
BPNN	93.67	94.17	92.10	93
SVM	91.6	91.67	91.59	92
RF	92.85	94.17	92.10	93

The following **Figure 7** shows the comparison of various classification algorithms such as ANN, SVM, KNN, RF, and BPNN with the proposed WCCSA algorithm and PNN.



**Figure 7. Comparison of the classification algorithms**

According to the above findings, the proposed WCCSA method outperforms the other existing algorithms in terms of accuracy (97%), sensitivity (96.2%), specificity (97.6%), and F1-score (96 %).

## 5. Conclusion

This work implemented a new CAD system for the diagnosis of breast cancer using a novel optimized wrapper-based hybridization algorithm of WCCSA. This study employs five chaotic maps to improve CSA's efficiency and convergence speed. To solve the difficulties of feature selection, WCCSA is used. The proposed method has been validated on mini-MIAS datasets. This work employs four different assessment parameters, including precision, sensitivity, specificity, and F1-score. WCCSA's performance is compared to that of popular and recent classification algorithms such as ANN, SVM, KNN, RF, and BPNN. According to the experimental findings, WCCSA outperforms the other classification algorithms. Furthermore, the results illustrated that WCCSA combined with PNN can improve the overall performance of the CAD system. In the future, chaotic maps will be incorporated with other meta-heuristic algorithms.

## References

1. Momenimovahed, Z., & Salehiniya, H. (2019). Epidemiological characteristics of and risk factors for breast cancer in the world. *Breast Cancer: Targets and Therapy*, 11, 151.
2. Aswathy, M. A., & Mohan, J. (2020). Analysis of Machine Learning Algorithms for Breast Cancer Detection. In *Handbook of Research on Applications and Implementations of Machine Learning Techniques* (pp. 1-20). IGI Global.
3. Chanda, P. B., & Sarkar, S. K. (2020). Detection and Classification of Breast Cancer in Mammographic Images Using Efficient Image Segmentation Technique. In *Advances in Control, Signal Processing and Energy Systems* (pp. 107-117). Springer, Singapore.
4. Dhal, K. G., Ray, S., Das, A., & Das, S. (2019). A survey on nature-inspired optimization algorithms and their application in image enhancement domain. *Archives of Computational Methods in Engineering*, 26(5), 1607-1638.
5. Guha, R., Ghosh, M., Mutsuddi, S., Sarkar, R., & Mirjalili, S. (2020). Embedded chaotic whale survival algorithm for filter-wrapper feature selection. *arXiv preprint arXiv:2005.04593*.
6. Sayed, G. I., Hassanien, A. E., & Azar, A. T. (2019). Feature selection via a novel chaotic crow search algorithm. *Neural computing and applications*, 31(1), 171-188.
7. Anter, A. M., & Ali, M. (2020). Feature selection strategy based on hybrid crow search optimization algorithm integrated with chaos theory and fuzzy c-means algorithm for medical diagnosis problems. *Soft Computing*, 24(3), 1565-1584.
8. Alqudah, A. M., Algharib, H. M., Algharib, A. M., & Algharib, H. M. (2019). Computer-aided diagnosis system for automatic two stages classification of breast mass in digital mammogram images. *Biomedical Engineering: Applications, Basis and Communications*, 31(01), 1950007.
9. Parvathavarthini, S., Karthikeyani Visalakshi, N., & Shanthi, S. (2019). Breast cancer detection using crow search optimization based Intuitionistic fuzzy clustering with neighborhood attraction. *Asian Pacific journal of cancer prevention: APJCP*, 20(1), 157.
10. Arafa, A. A. A., El-Sokary, N., Asad, A., & Hefny, H. (2019). Computer-aided detection system for breast cancer based on GMM and SVM. *Arab Journal of Nuclear Sciences and Applications*, 52(2), 142-150.
11. Elgin Christo, V. R., Khanna Nehemiah, H., Minu, B., & Kannan, A. (2019). Correlation-Based Ensemble Feature Selection Using Bioinspired Algorithms and Classification Using Backpropagation Neural Network. *Computational and mathematical methods in medicine*, 2019.
12. Arora, S., Singh, H., Sharma, M., Sharma, S., & Anand, P. (2019). A new hybrid

- algorithm based on grey wolf optimization and crow search algorithm for unconstrained function optimization and feature selection. *IEEE Access*, 7, 26343-26361.
13. sSuckling J, Parker J, Dance D, Astley S, Hutt I, Boggis C, Ricketts I et al. (2015). Mammographic Image Analysis Society (MIAS) database v1.21 [Dataset].
14. Pisano, E. D., Zong, S., Hemminger, B. M., DeLuca, M., Johnston, R. E., Muller, K., & Pizer, S. M. (1998). Contrast limited adaptive histogram equalization image processing to improve the detection of simulated spiculations in dense mammograms. *Journal of Digital imaging*, 11(4), 193.
15. YengecTasdemir, S. B., Tasdemir, K., & Aydin, Z. (2020). A review of mammographic region of interest classification. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, e1357.
16. Novitasari, D. C. R., Lubab, A., Sawiji, A., &Asyhar, A. H. (2019). Application of Feature Extraction for Breast Cancer using One Order Statistic, GLCM, GLRLM, and GLDM. *Advances in Science, Technology and Engineering Systems Journal*, 4(4), 11 120.
17. Mohamed, B. A., &Afify, H. M. (2017). Mammogram compression techniques using haar wavelet and quadtree decomposition-based image enhancement. *Biomedical Engineering: Applications, Basis and Communications*, 29(05), 1750038.
18. Qader, K., & Adda, M. (2014, September). Fault classification system for computer networks using fuzzy probabilistic neural network classifier (fpnnc). In *International Conference on Engineering Applications of Neural Networks* (pp. 217-226). Springer, Cham.
19. Anter, A. M., & Ali, M. (2020). Feature selection strategy based on hybrid crow search optimization algorithm integrated with chaos theory and fuzzy c-means algorithm for medical diagnosis problems. *Soft Computing*, 24(3), 1565-1584.
20. Sathiya, T., &Sathiyabhama, B. (2019). Fuzzy relevance vector machine based classification of lung nodules in computed tomography images. *International Journal of Imaging Systems and Technology*, 29(3), 360-373.
21. B. Sathiyabhama, J. Jayanthi, T. Sathiya, A. K. Ilavarasi, S. Udhaya Kumar, V. Yuvarajan, K. Gopikrishna, “A novel Feature Selection Framework based on Grey Wolf Optimizer for Mammogram Image Analysis”, *Journal of Neural Computing and Applications*, 2020.
22. B. Sathiyabhama, J. Jayanthi, T. Sathiya, A. K. Ilavarasi, S. Udhaya Kumar, V. Yuvarajan, “A grey wolf optimization for feature subset selection in the classification of breast cancer data”, *Journal of Soft Computing*, 2020.
23. Rajeswari, C., Sathiyabhama, B., Devendiran, S., &Manivannan, K. (2014). Bearing fault diagnosis using wavelet packet transform, hybrid PSO and support vector machine. *Procedia Engineering*, 97, 1772-1783.
24. Asraf Yasmin, B., Latha, R., & Manikandan, R. (2019). Implementation of Affective Knowledge for any Geo Location Based on Emotional Intelligence using GPS. *International Journal of Innovative Technology and Exploring Engineering*, 8(11S), 764–769. <https://doi.org/10.35940/ijitee.k1134.09811s19>
25. Manikandan, R., Latha, R., & Ambethraj, C. (1). An Analysis of Map Matching Algorithm for Recent Intelligent Transport System. *Asian Journal of Applied Sciences*, 5(1). Retrieved from <https://www.ajouronline.com/index.php/AJAS/article/view/4642>