



# METAHEURISTIC ENSEMBLE FEATURE SELECTION AND WEIGHTED AVERAGE LSTM NETWORK FOR RECURRENCE BREAST CANCER PREDICTION

662

<sup>1</sup>Preetha G, <sup>2</sup>Suban Ravichandran

<sup>1</sup>Research Scholar, Department of Computer & Information Science, Annamalai University, India

<sup>2</sup>Associate Professor, Department of Information Technology, Annamalai University, India

## ABSTRACT:

Now days, Breast cancer is among the extreme perilous kind of cancer in women kind throughout the world. The sole method to enhance treatment choices, which reduces mortality rates and enhances patients' chances of survival is early analysis. However, with the numerous characteristics in the dataset, it is a complex process. As a result, early and precise detection and categorization of breast cancer are major study areas. The major objective of this paper was to select important features using Metaheuristic Ensemble Feature Selection (MHEFS) before submitting these essential features to a classification procedure. MHEFS approach is introduced to combine informative features which are obtained using Opposition Colony Predation Algorithm (OCPA), Monarch Butterfly Optimization (MBO), and Linear Function based Animal Migration Optimization (LFAMO). Features in the MHEFS subset are those features which are present in features sets of individual methods. OCPA has been introduced in order to solve an algorithm's acceleration and it is able to find a global feature selection solution. MBO has been constructed to expedite and lionize monarch butterfly migration. The animal migratory patterns in all major animal clusters served as the inspiration for the LFAMO algorithm. MHEFS approach could improve feature selection led to the most robust result, and results are combined via the majority voting. Weighted Average Long Short Term Memory Network (WALSTM) had been introduced for breast cancer recurrence forecast. The results of the proposed classifier and existing classifiers have been experimented using MATrix LABoratory (MATLAB). The classifier was evaluated, and experimented using the public Wisconsin Breast Cancer Dataset (WBCD). Experimental results it concludes that the proposed system accomplishes healthier than the current system in terms of specificity, recall, precision, accuracy, and f-measure.

**INDEX TERMS:** Breast cancer, Metaheuristic Ensemble Feature Selection (MHEFS), Linear Function based Animal Migration Optimization (LFAMO), Opposition Colony Predation Algorithm (OCPA), Monarch Butterfly Optimization (MBO), and Weighted Average Long Short Term Memory Networks (WALSTMs).

DOI Number: 10.48047/NQ.2022.20.20.NQ109067

NeuroQuantology2022;20(20): 662-677

## 1. INTRODUCTION

Breast cancer is a prevalent malignancy in females as well as the second greatest reason of fatality in women. Breast cancer has become more common over the world, with more instances recorded each year. It is more prevalent in women than in males, in equated to other types of cancer. If this disease is not diagnosed and treated quickly, it can be fatal. The return of breast cancer in a person for whom the prior disease had gone into remission is referred to as breast cancer pathology. Lessening is the intended effect of chemotherapy and oncologists' ongoing treatment [1,2,3]. The repetition of breast cancer or any other disease is one of the most terrifying thoughts that a cancer patient faces. As a result, it turns out to be one of the issues affecting their quality of life. Nevertheless, of its significance, it is occasionally stated across maximum breast cancer data sources, rendering study into its forecast more difficult [4].

Early identification and diagnosis are critical for illness prevention and survival, and pathology diagnosis represents the most trustworthy benchmark of all approaches. In recent times, supercomputing diagnostic tools and artificial intelligence approaches for medical diagnosis have provided automated procedures for realistic judgements by utilizing quantitative metrics and machine learning techniques [5,6]. These technologies have the potential to increase diagnostic accuracy. Medical records dimensions encompass multiple aspects, each of which comprises of a different variety of value.

One strategy that can deal with this difficulty and is significant in classification is feature selection. Wrapper, Filter, and Embedded are three ways to feature selection approaches. The goodness of a specified subsection of attributes is determined in the wrapper approach by training and assessing a classification using only



the attributes contained in the proposed subsection. It is also suggested that decreasing the amount of source features in a classifier is favorable to constructing an outstanding prognostic and less cost computational prototype. This variation is critical because a partial feature subgroup indicates reduced test and treatment costs in the field of clinical diagnostics [7,8]. Additionally, reducing multiplicity might eliminate unnecessary parameters, but diminishing noise can upshot in more strong learning methods owing to the connection of lesser features [8]. The much more classic feature selection approaches choose the greatest pertinent, non-redundant characters while disregarding the features' inherent interdependence layout. Because of the single method, feature selection has a frequent drawback of lower classification accuracy.

Ensemble Feature Selection (EFS) is a technique that combines the outcome of numerous feature selectors to optimize performance and free the user from the need to select a specific strategy [9,10]. Following the completion of the EFS, classification is carried out using ensemble learning [11]. Ensemble Learning procedures are a cluster of algorithms castoff to improve a classification model's predicted accuracy. It improves final classification performance by integrating the benefits of numerous classifiers into a single autonomous unit [12]. Several hypotheses are trained using ensemble approaches, sometimes referred to as committee-based training or learning multi - classifier systems [12]. However, because the dataset contains a huge amount of elements, the principle of ensemble learning does not solve the high dimensionality problem. Consequently, it is common to use a pretreating phase to eliminate unimportant elements and lower the issue's dimensionality [12]. A right feature selection can lead to an enhancement in the inductive learner's generalizations capacity, learning speed or the simplicity of the inferred model.

It is necessary to enhance efficiency in this task. Utilizing the Breast Cancer dataset of Wisconsin Diagnostic, ensemble feature selection-based enhancement is used to discover the far more significant elements for the categorization of benign type or malignant cancer diagnosis. Numerous real-world applications frequently include optimization difficulties. Practical activities and difficulties that must be solved may usually be stated as an optimization problem. Furthermore, Monarch Butterfly Optimization (MBO) is opted to optimize the sensitivity and specificity measurements at the same time. Finally, the performance of the approaches utilized is compared.

Traditional classifier-based ensemble learning, on the other hand, takes a long time to improve their suggested models and attain the desired performance [13]. Deep Learning (DL) algorithms have become widely employed in cancer in breast detection. Ensemble Deep Learning technology (EDL) technique is a market leader in determining whether tumours are malignant or benign. It can acquire hierarchies of important features automatically from the raw data provided. As a result, the EDL encoder might be a potentially better classifier than the usual Machine Learning type ensemble technique. When compared to typical classification methods, the EDL model performs better. As a result, the purpose of this study is to provide ensemble learning in classification and feature selection. The research work's key contribution is arranged as follows. Firstly, Wisconsin Breast Cancer Dataset (WBCD) is collected from benchmark repository. Then input data are preprocessed to eliminate irrelevant and/or missing value data by z-score normalisation. Following that, Metaheuristic Ensemble Feature Selection (MHEFS) are performed by combining three methods such as Opposition Colony Predation Algorithm (OCPA), MBO, and Linear Function based Animal Migration Optimization (LFAMO). The results of individual feature selection methods are combined via the majority voting. Classification is carried out utilizing the WALSTM from the selected attributes, and their results are evaluated between the classifiers.

## 2. LITERATURE REVIEW

Alikovi and Subasi [14] used several data mining approaches to diagnose breast cancer. In this research, two distinct WBCD were employed to test the system. The system is divided into two phases. In the first stage, the Genetic Algorithm (GA) is used to eliminate pointless traits and extract significant and valuable attributes. The second phase involves using a variety of data mining techniques to choose between two distinct groups of persons who have breast cancer or do not have it. In the second step, several individual and multiple classifier methods were employed to build an accurate system for Breast cancer categorization. The outcomes of this study may lead to the development of novel methods for the early recognition of breast cancer.

Punitha et al. [15] designed an Integrated Artificial Immune system and Artificial Bee Colony based breast cancer diagnosis system (IAIS-ABC-CDS) in an Artificial Neural Network (ANN) for parallel computing of efficient feature extraction and attribute tuning. The standard diagnosis strategies of IAIS-ABC-CDS with Resilient Back-Propagation Techniques (RBPT) and GA rooted ANN with Multilayer Perceptron (GA-ANN-MLP) are compared to the IAIS-ABC-CDS with Momentum-based Gradient Descent Backpropagation (MBGD) that improves the localized search strategy by using the profits of simulated annealing (SA). Under the Wisconsin dataset, the IAIS-ABC-CDS produces a mean cataloguing percentage of 99.34% and 99.11% in ANN.

Alfian et al. [16] used a Support Vector Machine (SVM) in conjunction by such an exceptionally randomized trees classifier (extra-trees) to deliver a BC risk indicator-based early analysis. The extra-trees attribute



was applied to eliminate unnecessary elements, whereas SVM was employed to determine the state of breast cancer. Machine learning models used a BC dataset of 116 people to forecast BC, while segmented ten-fold cross validation was used to evaluate the system. The approach is predicted to progress the precision of breast cancer detection according to the risk factors. Furthermore, the prediction model might be used for computer-based breast cancer forecast. The approach is suggested to enhance diagnosis policymaking schemes by properly envisaging BC illness.

Antunaset al [17] proposed using feature selection and multi-objective optimization on the WBCD dataset. Its goal is to discover the most important traits that will help determine whether the conclusion is benign or malignant. In the feature selection job, two classifiers will be utilized; one based on NN and the other on SVM. The optimization process's goal functions are to maximize sensitivity and specificity at the same time. When the methodologies were compared, it was discovered that NNs performed better.

To diagnosis breast cancer, Huanget al. [18] designed an improved machine learning system. The basic concept behind this approach is to employ the Fruit Fly Optimization Algorithm (FOA) augmented with the Levy Flight (LF) method to improve 2 important SVM features and create an LFOA-based SVM (LFOA-SVM) for identifying breast cancer. For the very first time, breast cancer is diagnosed using the high-level traits that were extracted from the volunteers.

An automated breast cancer diagnostic method utilizing a GA for simultaneous feature selection and ANN parameter enhancement was developed by Ahmad et al. [19]. The three alternative backpropagation techniques—Levenberg-Marquardt (GAANN LM), resilient back-propagation (GAANN RP), and gradient descent with momentum (GAANN GD)—and their results are contrasted. These techniques are used to regulate the load of ANNs. By combining the parasitic worker bee stage of the ABC by the bubble net attacking method of whale optimization, Stephan et al. [20] created a mixture artificial bee colony with whale optimization algorithm (HAW). Employee bees utilize humpback whales to locate better locations for food sources during the attacking phase. The suggested mutative initialization step, which makes up the exploratory phase of the HAW procedure, is opted to strengthen the poor exploration of traditional ABC.

In order to examine the efficacy of some current data mining techniques in foretelling the recurrence of breast cancer, Sakri et al [21] devised their method. In order to improve the prediction model's accuracy, it incorporates a Particle Swarm Optimization (PSO) as a feature selection method into 3 recognized classifiers: K-nearest neighbour (KNN), Naive Bayes (NBs), and rapid decision tree learner.

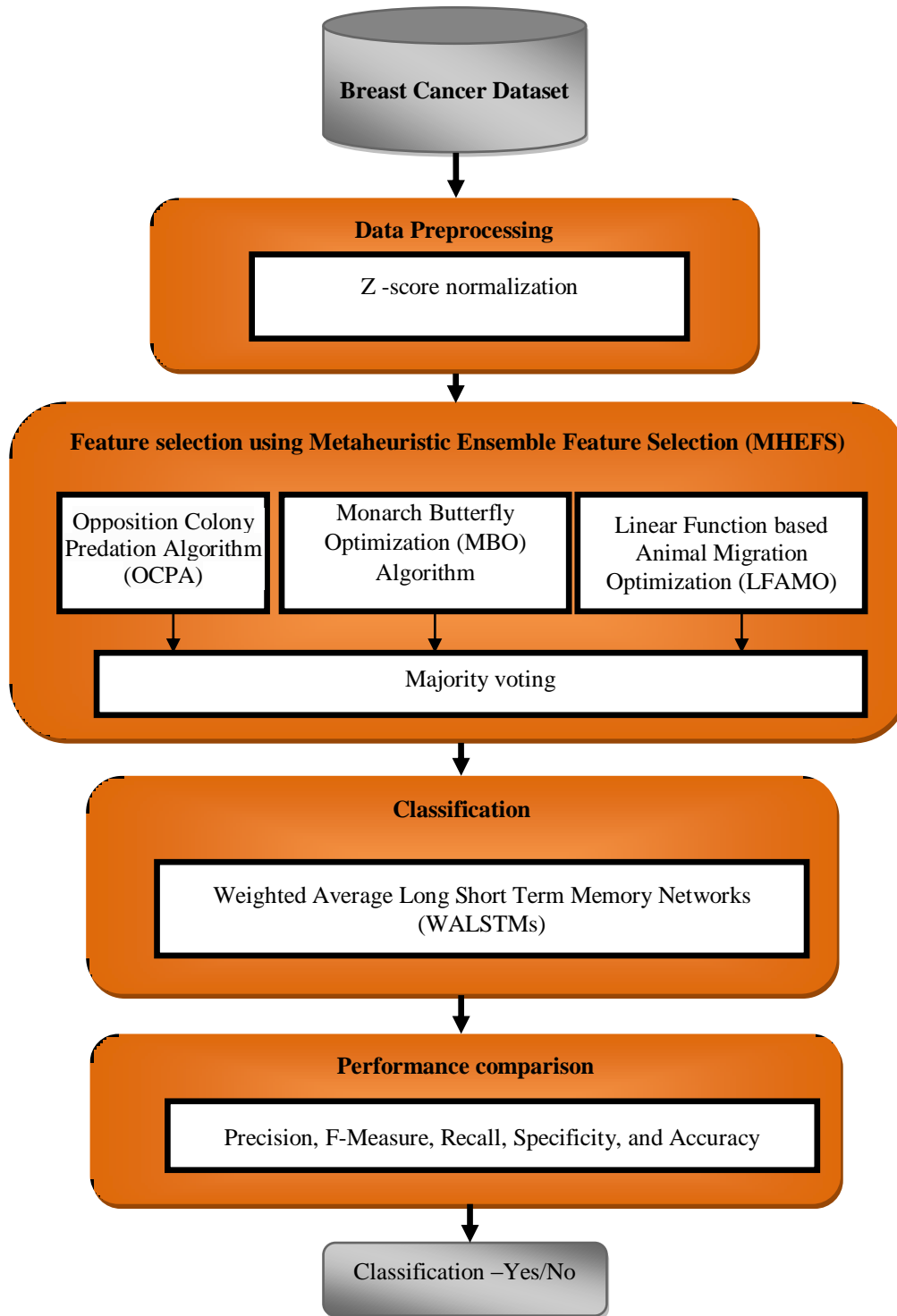
Jumanto et al [22] implemented to build a classification system for detecting breasts using the Backpropagation method, by adding a method of forward selection for feature selection from the many features that exist in the BC dataset, because not all features can be used in the NB, SVM classification process. The results of combining the Backpropagation method and the method of forward selection can increase the detection accuracy of breast cancer patients by 98.3%.

Extensible Breast Cancer Prognosis Framework (XBPF) was suggested by Aavula and Bhramaramba[23] for the prognosis of breast cancer and covers vulnerability or risk evaluation, recurrence or revitalization of the cancer after steadfastness, and survivorship. In order to increase prognosis efficiency, the representative feature subset selection (RFSS) approach is used with SVM. To conduct experimentations, the SEER (Surveillance, Epidemiology and End Results) dataset is used. To show that the concept is sound, a prototype is constructed. The empirical investigation showed that the framework, rather than concentrating on a specific feature like susceptibility, survival, and recurrence separately, is beneficial in predicting the prognosis of breast cancer. Comparing SVM-RFSS to cutting-edge prognostic techniques, it has demonstrated a considerable performance increase.

### 3. PROPOSED METHODOLOGY

Firstly, Wisconsin Breast Cancer Dataset (WBCD) is collected from benchmark repository. Then input data are preprocessed to eliminate irrelevant and/or missing value data by z-score normalization. The persistence of Metaheuristic Ensemble Feature Selection (MHEFS) is to discard irrelevant or superfluous features from a given feature vector. MHEFS is done based on Opposition Colony Predation Algorithm (OCPA), Monarch Butterfly Optimization (MBO), and Linear Function based Animal Migration Optimization (LFAMO) algorithm. MHEFS are initially presented to cope with the steadiness of feature selection method. The results of individual feature selection methods are combined via the majority voting. Classification is carried out utilizing the Weighted Average Long Short-Term Memory Network (WALSTM) according to the selected attributes. Figure 1 displays the designed flowchart of proposed system.





**FIGURE 1. METAHEURISTIC ENSEMBLE FEATURE SELECTION (MHEFS) AND DEEP LEARNING BASED PROPOSED SYSTEM**

### 3.1. Data Normalization

As an input, the WBCD is collected from kaggle. The input data are preprocessed to eliminate irrelevant and/or missing value data. Zero mean normalization is another name for Z score normalization. The data in this instance has been homogenous using the standard deviation and mean. Then the equation (1) is described as below,

$$d' = \frac{d - \text{mean}(p)}{\text{std}(p)} \quad (1)$$



where Mean (p) -is the average of altogether p's element values, Std(p)-is the standard deviation of all p's value, d be the dataset, d' be the normalized dataset.

### 3.2. Metaheuristic Ensemble Feature Selection (MHEFS)

Metaheuristic Ensemble Feature Selection (MHEFS) is done based on Opposition Colony Predation Algorithm (OCPA), Monarch Butterfly Optimization (MBO), and Linear Function based Animal Migration Optimization (LFAMO) procedure. In order to construct the best possible subset, the MHEFS procedure combines the OCPA, MBO, and LFAMO. It improves the feature space and lowers the possibility of selecting an imbalanced subset. MHEFS extracted from individual methods are combined based on majority voting. It tries to combine them to find a better common feature subset for all major types of breast cancer.

#### 3.2.1. Opposition Colony Predation Algorithm (OCPA)

In attempt to pick the top features for animal hunting parties, the colony predation algorithm (CPA) is developed. This mapping method comprises spreading prey, encircling prey, aiding the hunter who has the best chance of success, and pursuing additional targets [24]. The following equation (2) represents individual cooperative communication and food-seeking behavior for optimal selection,

$$\vec{X}_j^i(t+1) = r \cdot \vec{X}_j^i(t) + (1-r) \cdot \left( \frac{\vec{X}_1(t) + \vec{X}_2(t)}{2} \right) \quad (2)$$

there r is a value chosen randomly in [0, 1],  $\vec{X}_j^i(t)$  is the target prey, and  $\vec{X}_1$  and  $\vec{X}_2$  are two features of WBCD dataset is closed to the  $\vec{X}_j^i(t)$ ,  $j \in 1, 2, \dots, \text{dim}$ . The newest update spot denoted by  $\vec{X}_j^i(t+1)$ . The forecast approach is comprehended as  $\vec{X}(t+1) = \vec{X}_{\text{best}} - S \cdot (r_1(\text{ub} - \text{lb}) + \text{lb})$ , here  $\vec{X}(t+1)$  is the novel feature at the (t + 1)<sup>th</sup> iterations,  $\vec{X}_{\text{best}}$  is the feature location of the food. S is the power of prey and its complete value diminutions with the amount of valuations from 0,  $r_1$  is the  $[R_1, R_2, R_3, \dots, R_j]$ ,  $j = \text{dim}$  signifies the feature dimension of the populace, *ub* refers the higher bound and *lb* refers the lower bounds values of feature from the WBCD dataset. S is intended by  $S = 2 \cdot S_0 \cdot r_2 - S_0$  and  $S_0 = a - t \cdot \left( \frac{a}{N} \right)$ , where N is the no. of samples with parameters,  $S_0$  reduces in value from a to 0, and t represents the present number of iterations for optimum selection of features from the WBCD dataset and  $r_2$  is a value in [0, 1] in random. The hunting group will circle the classification of disease diagnosis and proceed to explore it. The following is a mathematical description of this step,

$$\vec{X}(t+1) = \vec{X}_{\text{best}} - 2 \cdot S \cdot D \cdot e^1 \cdot \tan\left(1 \cdot \frac{\pi}{4}\right) \quad (3)$$

D refers the gap among the present discrete and the prey. The probabilities of using these two predatory ploys are represented by the following mathematical expressions,

$$\vec{X}(t+1) = \begin{cases} \vec{X}_{\text{best}} - S \cdot (r_1 \cdot (\text{ub} - \text{lb}) + \text{lb}) & r_2 \geq 0.5 \\ \vec{X}_{\text{best}} - 2 \cdot S \cdot D \cdot e^1 \cdot \tan\left(1 \cdot \frac{\pi}{4}\right) & r_2 < 0.5 \end{cases} \quad (4)$$

It could be articulated as  $\vec{X}(t+1) = \vec{P}_{\text{nearest}}$ , here the nearby predator(feature) in the support cluster (dataset) is denoted by  $\vec{P}_{\text{nearest}}$ .  $\vec{P}$  refers the predator nearby the target. The pursuit for the food (optimal selection of features) can be described as follows,

$$D_1 = \text{abs}(2 \cdot r_4 \cdot \vec{X}_{\text{rand}} - \vec{X}(t)) \vec{X}(t+1) = \vec{X}_{\text{rand}} - S \cdot D_1 \quad (5)$$

where  $D_1$  is the distance a random group traveled,  $r_4$  is produced randomly in [0, 1], and  $\vec{X}_{\text{rand}}$  is a new feature which is also random. To help expedite the divergence of an algorithm, Opposition-based learning may be used, which is frequently used to find a global selection of features from the dataset. The contradictory value of a number  $X \in [L_b, l_u]$  is calculated by  $\bar{X} = \text{ub} + \text{lb} - X$ , and *lb* and *ub* are the lowest and higher limits of the feature in the WBCD search space. For multidimensional decision space,  $X_i = (X_{i1}, X_{i2}, X_{i3}, \dots, X_{ij})$  with  $X_{ij} \in [l_{bj}, l_{uj}]$ ,  $j = 1, 2, \dots, n$ , the opposite point can be calculated as mentioned here:  $\bar{X}_i = (\bar{X}_{i1}, \bar{X}_{i2}, \bar{X}_{i3}, \dots, \bar{X}_{ij})$ ,  $\bar{X}_{ij} = l_{bj} + l_{uj} - x_{ij}$ . It was detected that when  $f(\bar{X}_i)$  is optimal to number of features in the dataset via  $f(X_i)$ , the opposing point  $X_i$  may be substituted by the comparative arrangement  $\bar{X}_i$ . There are three elements to the proposed OCPA. Setting up the search population is the first step in ensuring the best possible selection of dataset features. Finishing the original CPA algorithm operators comes second, and finishing opposition-based learning comes third. Algorithm 1 shows





the proposed OCPA algorithm's whole process. The real CPA operation is carried out first, and then opposition-based training follows.

**ALGORITHM 1. PSEUDO CODE OF OCPA**

**INPUT:**No. of Populace Size N, T is the measure of maximum repetitions, g=0,

**OUTPUT:** Best position  $X_b$ , Best fitness value  $f_{value}$ ;

1. Randomly Prepare population  $X_i (i = 1, 2, \dots, N)$ ;
2. **Begin**
3. **While**  $g < T$  **do**
4.     **for**  $(i = 1:N)$  **do**
5.         Make  $X_i$  inside the search range;
6.         Compute the fitness of  $X_i$ ;
7.         Modify the  $X_b, f_{value}$ ;
8.         Modify the  $S, S_0, \alpha, r_t$
9.     **For**  $(j=1;dim)$  **Do**
10.         Update the  $X_1, X_2, i_1, r$ ;
11.         If  $rand < g/T$   $X_i = X_b$ ;
12.         Calculate the  $X_j^i$  by equation(1)
13.     **For** $(i=1: N)$  **Do**
14.         Update  $S, l$ ;
15.     **If** $abs(S) < 2 \times a/3$
16.         Compute the  $X_i$  by formula (4);
16.     **Else**
17.         Compute the  $X_i$  by equation (),
17.         
$$\vec{X}(t + 1) = \begin{cases} \vec{P}_{nearest} abs(r_6) < 1 \\ \vec{X}_{rand} - S.D_1 abs(r_6) > 1 \end{cases}$$
17.         \*/Opposition-based learning \*/
17.         **for**  $(i = 1 : N)$  **do**
18.         Compute the opposition  $\vec{X}_i$  of  $X_i$ ;  $q$
19.         **if**  $f(\vec{X}_i) < f(X_i)$  **then**
20.             Replace  $X_i$  by  $\vec{X}_i$  ;
21.         **Else**
22.             Keep  $X_i$  ;
23.     Update  $X_b$ ;
24.      $g = g + 1$
25. **End-Loop**;

**3.2.2. MBO - Monarch Butterfly Optimization Algorithm**

Monarch Butterfly Optimization simulates how monarch butterflies migrate in the wild to solve the feature selection issue in WBCD. As a result, here are 2 approaches to appraise the monarch butterfly placements. The migration operator creates the offsprings first (updating the feature position). Then, the butterfly correcting operator adjusts the feature placements of the other butterflies. The migration operator and butterfly adjustment operator play a significant role in determining the hunt vector of the monarch butterfly species in the MBO procedure. The MBO approach is excellent for feature selection and is perfectly suited for balancing intensification and diversification. It is a crucial phenomenon in the issue of feature selection.

**Migration Operator:** The amount of butterflies on  $land_1$  and  $land_2$  can be intended as  $ceil(p * NP)(NP_1, subpopulation 1, SP_1)$  and  $NP - NP_1 (NP_2, subpopulation 2, SP_2)$  which is has been used for feature selection. Where,  $ceil(x)$  rounds off  $x$  to the closest feature not lesser than  $x$ . Consequently, when  $r \leq p$ , then  $x_{i,k}^{t+1}$  is produced by the equation below [25],

$$x_{i,k}^{t+1} = x_{r_1,k}^t \tag{6}$$

where  $x_{i,k}^{t+1}$  is the  $k^{th}$  feature of  $x_i$ , and  $x_{r_1,k}^t$  is the  $k^{th}$  feature of  $x_{r_1}$ . Butterfly  $r_1$  is selected from  $SP_1$  in a random manner.  $r$  is given in the following form in the Equation (7),

$$r = rand * peri \tag{7}$$



here, the migration period is  $peri$  [25]. In assessment, when  $r > p$ , then  $x_{r_1,k}^t$  is specified by equation (8),

$$x_{i,k}^{t+1} = x_{r_2,k}^t \quad (8)$$

where  $x_{r_2,k}^t$  is the  $k^{th}$  feature of  $x_{r_2}$ , and butterfly $_2$  is chosen from  $SP_2$  in a random manner.

## ALGORITHM 2. MIGRATION OPERATOR

668

1. **Begin**
2. **for**  $i= 1$  to  $NP_1$  (for every monarch butterflies in  $SP_1$  (Number of samples and features in the WBCD dataset) **do**
3.     **for**  $k=1$  to  $D$  (all the features in  $i^{th}$  monarch butterfly) **do**
4.     **Generate** a random number  $rand$  by uniform dispersal via equation (7)
5.     **if**  $r \leq p$  **then**
6.         Select a monarch butterfly randomly in  $SP_1$  (say  $r_1$ );
7.         Engender the  $k^{th}$  feature of the  $x_i^{t+1}$  as equation (6).
8.     **else**
9.         In Random manner, select a monarch butterfly in  $SP_2$  (say  $r_2$ )
10.         Generate the  $k^{th}$  feature of the  $x_i^{t+1}$  as equation (8).
11.     **end if**
12.     **end for k**
13. **end for i**
14. **End**

**Butterfly Adjusting Operator:** For the butterfly  $j$ , if  $rand$  is no more than  $p$ , the  $k^{th}$  feature is given as [25],

$$x_{j,k}^{t+1} = x_{best,k}^t \quad (9)$$

where  $x_{j,k}^{t+1}$  is the  $k^{th}$  feature of  $x_j$ . Similarly,  $x_{best,k}^t$  is the  $k^{th}$  feature of the dataset  $x_{best}$  that is the most excellent monarch butterfly in Land 1 and Land 2.  $t$  is present generation number. It could be expressed as follows,

$$x_{j,k}^{t+1} = x_{r_3,k}^t \quad (10)$$

where  $x_{r_3,k}^t$  is the  $k^{th}$  feature of  $x_{r_3}$  is randomly nominated in Land 2,  $r_3 \in \{1, 2, \dots, NP_2\}$ . Under this circumstance, if  $rand > BAR$ , it could be additionally restructured as such [25],

$$x_{j,k}^{t+1} = x_{j,k}^{t+1} + \alpha \times (dx_k - 0.5) \quad (11)$$

there  $dx$  is the pace step of butterfly  $j$ ,  $BAR$  indicates butterfly adjusting rate.  $dx$  is can be computed by carrying out Levy flight.  $\alpha$  is the weighting factor that is given as equation (13).

$$dx = Levy(x_j^t) \quad (12)$$

$$\alpha = S_{max}/t^2 \quad (13)$$

Where  $t$  is the current generation and  $S_{max}$  is max walk phase which a monarch butterfly discrete can travel in one stride. The larger  $\alpha$ , signifying lengthy step of exploration, rises the effect of  $dx$  on  $x_{j,k}^{t+1}$  and cheers the procedure of search, whereas the smaller  $\alpha$ , indicating short step of examination, reductions the impact of  $dx$  on  $x_{j,k}^{t+1}$  and boosts the procedure of exploitation. Accordingly, the description of butterfly regulating operator might be provided in the below algorithm 3.

## ALGORITHM 3. BUTTERFLY ADJUSTING OPERATOR

1. **Begin**
2. **for**  $j= 1$  to  $NP_2$  (for all monarch butterflies in  $SP_2$ ) **do**
3.     Compute the walk phase  $dx$  by equation (12);
4.     Compute the weighting factor  $\alpha$  by equation (13);
5.     **for**  $k=1$  to  $D$  (all the features in  $i^{th}$  monarch butterfly) **do**
6.         In random manner **generate** a no.  $rand$  by uniform distribution
7.     **if**  $rand \leq p$  **then**



8. Randomly choose a monarch butterfly in  $SP_1$  (say  $r_1$ );
9. Generate the  $k^{th}$  feature of the  $x_j^{t+1}$  as equation (9).
10. **else**
11. Randomly chose a monarch butterfly in  $SP_2$  (say  $r_3$ )
12. Generate the  $k^{th}$  feature of the  $x_i^{t+1}$  as equation (10).
13. **if** rand > BAR **then**
14.  $x_{j,k}^{i+1} = x_{j,k}^{j+1} + \alpha \times (dx_i - 0.5)$
15. **end if**
16. **end if**
17. **end for k**
18. **end for j**
19. **End**

#### ALGORITHM 4. MONARCH BUTTERFLY OPTIMIZATION (MBO) ALGORITHM

##### Begin

**Step 1: Initialization :** Make the generation counter  $t = 1$   
 initialize the populacePof NP monarch butterfly entities in random  
 set the maximum generation MaxGen, monarch butterfly number  $NP_1 \& NP_1$ , max step  $S_{Max}$ , migration period peri, and the migration ratio p.

##### Step 2: Fitness evaluation

Appraise each monarch butterfly conferring to its feature position by accuracy.

##### Step 3: While the best solution is not found or $t < MaxGendo$

Sort out all the monarch butterfly individuals related to their accuracy.

Divide monarch butterfly individuals into two subpopulations ( $SP_1 \& SP_2$ );

**For**  $i = 1$  to  $NP_1$  ((for all **monarch** butterflies in  $SP_1$ ) **do**

Produce new  $SP_1$  according to Algorithm 2.

**End for**  $i$

**For**  $j = 1$  to  $NP_2$  (for each monarch butterflies in  $SP_2$ ) **do**

Generate new  $SP_2$  corresponding to Algorithm 3.

**End for**  $j$

Syndicate the equally newly-generated features addicted to particular whole dataset;

Calculate the dataset(population) according to the newly selected features;

$t = t + 1$ ;

**Step 4:** end while

**Step 5:** Output the best result

##### End

#### 3.2.3. Linear Function based Animal Migration Optimization (LFAMO) algorithm

The algorithm, Animal Migration Optimization (AMO) was created to handle global optimization issues and was motivated by the animal migration patterns that could be observed in all foremost animal clusters, including fish and birds. The fundamental operation of the algorithm is described in terms of two idealized assumptions: (1) The herd's leader animal will be determined by the animal with the highest quality, and the leader animal will indeed be preserved for future generations. (2) The herd has a set number of animals, and each one will be substituted with a novel animal with likelihood  $P_a$ . In this situation, they will quit the group and later be replaced by a new animal [26].

**Migration process:** The migration process and the population update method make up the AMO algorithm. The animal's transition from one location to another is covered by the migration process. The following list of animal herds makes up the migrating animal population. Equation (14) displays this migrating population,

$$\text{Population} = \{X_1, X_2, \dots, X_{NP}\} \quad (14)$$

where  $X_{NP}$  and  $X_i$ , respectively, stand for the dataset's population and feature sizes. The procedure starts with a populace that is initialized randomly and it is expressed as follows,





$$X_i = X_{\min} + \text{rand} * (X_{\max} - X_{\min}) \quad (15)$$

where  $X_i$ ,  $X_{\min}$  and  $X_{\max}$  stand for a population feature, the minimum and maximum boundaries of the search space, and a feature in the population, respectively. A "rand" is a number between 0 and 1 generated randomly that has a uniform distribution.

**Animal Relocation:** An animal should stick to three rules when migrating: (1) stay clear of your neighbours; (2) travel in the same direction as your neighbours; and (3) stay near to your neighbours.

**Population Updating Process:** Even throughout this update process, the computer duplicates how some creatures quit the group and a few migrate to the new community. Classification accuracy is taken into account as a fitness function in the suggested study. While the probability value for the population's most compatible member is 1, the probability value for its most mismatched member is  $1/NP$ .

### 3.2.4. Majority voting

Voting is the ensemble approach, which is employed to syndicate the results of methods. The majority voting process is utilized to integrate results from each sub-model. It is a meta-classifier that uses a majority vote to identify FS methods that are conceptually similar or dissimilar. Majority voting is used to forecast the eventual class label, which is the class tag that FS models most usually predict. It is described by equation (16),

$$y = \text{mode}(FS_1(x), FS_2(x), FS_3(x)) \quad (16)$$

Where  $y$  = diagnosis class label and  $FS_1(x), FS_2(x), \dots, FS_m(x)$  = FS models.

### 3.3. Classification using Weighted Average Long Short Term Memory Network (WALSTM)

Based on the specific selected features, classification is performed by utilizing WALSTM. The LSTM is an RNN framework designed to replicate temporal sequences and their distant relationships more accurately. Input, forget, and output gates and a cell activation element make up a typical LSTM cell [28]. Such devices gather activation signals from a variety of causes and regulate cell activity by using certain multipliers. The LSTM gates might avert the rest of the network from altering the contents of memory cells for a large number of time steps. LSTM input gate is described as follows,

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \quad (17)$$

The definition of forget gate as follows,

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f) \quad (18)$$

The definition of cell gate as follows,

$$c_t = f_t c_{t-1} + i_t \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \quad (19)$$

The output gate is described as follows,

$$o_t = \sigma(W_{x0}x_t + W_{h0}h_{t-1} + W_{c0}c_t + b_0) \quad (20)$$

Lastly, hidden state is computed as follows,

$$h_t = o_t \tanh(c_t) \quad (21)$$

$\tanh$  - hyperbolic tangent activation function,  $x_t$  - input at  $t$  time,  $W$  and  $b$  - network parameters (Weights, Biases). The weight value is updated based on the weighted average scheme. When this is the case the system can apply a weighting to each value. The weighted mean can be calculated with the subsequent equation (22),

$$W = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i} \quad (22)$$

Where,  $x_i$  is the features,  $w_i$  is a Corresponding weight for each feature,  $\sigma$  is signified by logistic sigmoid function,  $i$ ,  $f$ ,  $o$  and  $c$  are input, forget, output, cell state, congruently. Weight matrices for peephole acquaintances are indicated by letters  $W_{ci}$ ,  $W_{cf}$  and  $W_{co}$ . The input gate chooses the input ratio. This ratio has an impact on the equation when defining the cell state (17). The forget gate confirms if prior memory  $h_{t-1}$  is passed or not. It is determined in equation (18) and cell gate is determined in the equation (19). The output gate determines whether or



not the memory cell's outcome is passed. This may be seen in the equation (20). In addition to the three gates, they can use LSTM to solve the disappearing and bursting gradient difficulties. In the LSTM-Recurrent Neural Network (RNN) architecture, the recurrent hidden layer is substituted by an LSTM cell. The output gate classifies the dataset into two categories such as cancer or normal.

#### 4. EXPERIMENTAL RESULTS

The efficiency of the proposed WALSTM classifier, and existing classifiers such as Weighted Minkowski Radial Basis Function based Support Vector Machine (WMRBF-SVM), NB and fast decision tree learner (Reduced Error Pruning Tree (REPTree)) are evaluated using MATLAB. The dataset is composed from <https://www.kaggle.com/uciml/breast-cancer-wisconsin-data>. The proposed scheme is compared with the previous NB, REPTree, and WMRBF-SVM approaches via accuracy, recall, f-measure, specificity, and precision.

**Precision:** The ratio of correctly anticipated positive annotations to all correctly foretold positive observations determines precision, which establishes the significance of the results. It is calculated by equation (23),

$$\text{Precision} = \frac{TP}{TP+FP} \quad (23)$$

**Recall:** This is the ratio of properly estimated positive observations to the all observations in actual class - yes. It is calculated by equation (24),

$$\text{Recall} = \frac{TP}{TP+FN} \quad (24)$$

**F-measure:** The weighted mean of Precision and Recall serves as a measure of experiment accuracy. It is calculated by equation (25),

$$\text{F-measure} = 2 * \frac{(\text{Recall} * \text{Precision})}{(\text{Recall} + \text{Precision})} \quad (25)$$

**Specificity:** The ratio of how many data points were correctly identified as negative to how many data points were truly negative is the classifier's specificity. It is calculated by equation (26),

$$\text{Specificity} = \frac{TN}{FP+TN} \quad (26)$$

**Accuracy:** Accuracy is the most correct performance estimation that accurately distinguishes an event and is simply a prediction of the right predictions in all observations. It is calculated by equation (27),

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN} \quad (27)$$

where TP - True Positive, FN - False Negative, FP - False Positive, and TN- True Negative. Table 1 represents the overall performance comparison of classifiers with respect to several FS methods.

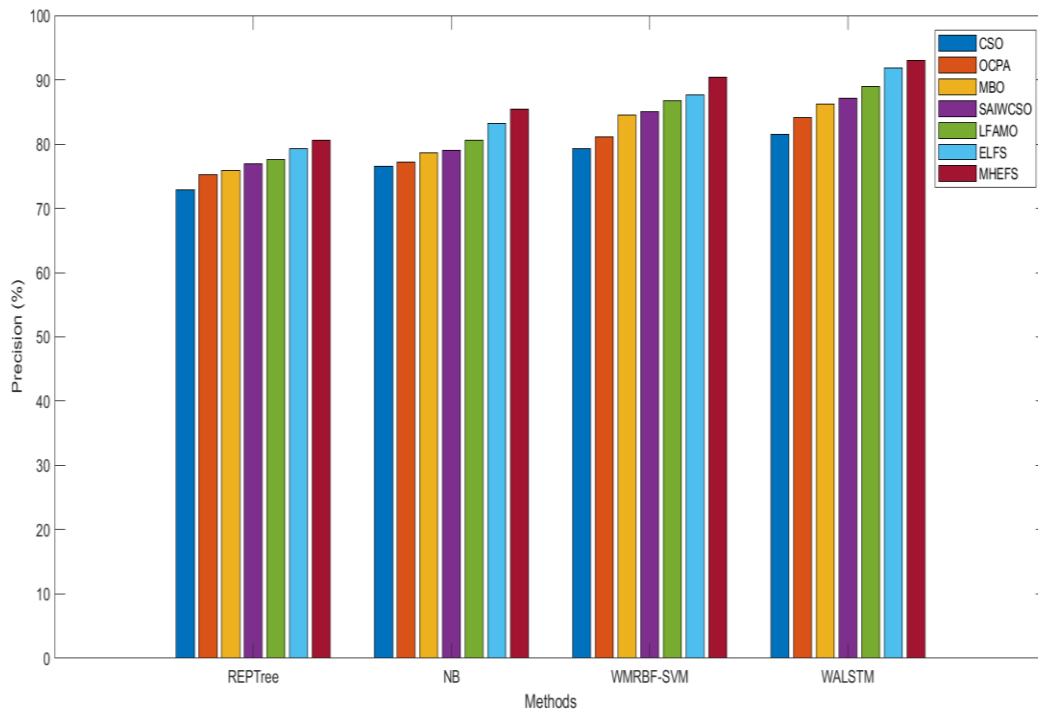
**TABLE 1. METRICS RESULTS COMPARISON OF CLASSIFIERS VS. FS METHODS**

Precision (%)							
Methods	CSO	OCPA	MBO	SAIWCSO	LFAMO	ELFS	MHEFS
REPTree	72.8368	75.2948	75.9580	77.0037	78.6439	79.3043	80.6685
NB	76.6001	77.1686	78.7060	79.0558	80.5712	83.2440	85.4997
WMRBF-SVM	79.3043	81.0838	84.5237	85.0231	86.7354	87.6125	90.3841
WALSTM	81.4725	84.1040	86.2436	87.2078	89.0021	91.7983	93.0756
Recall (%)							



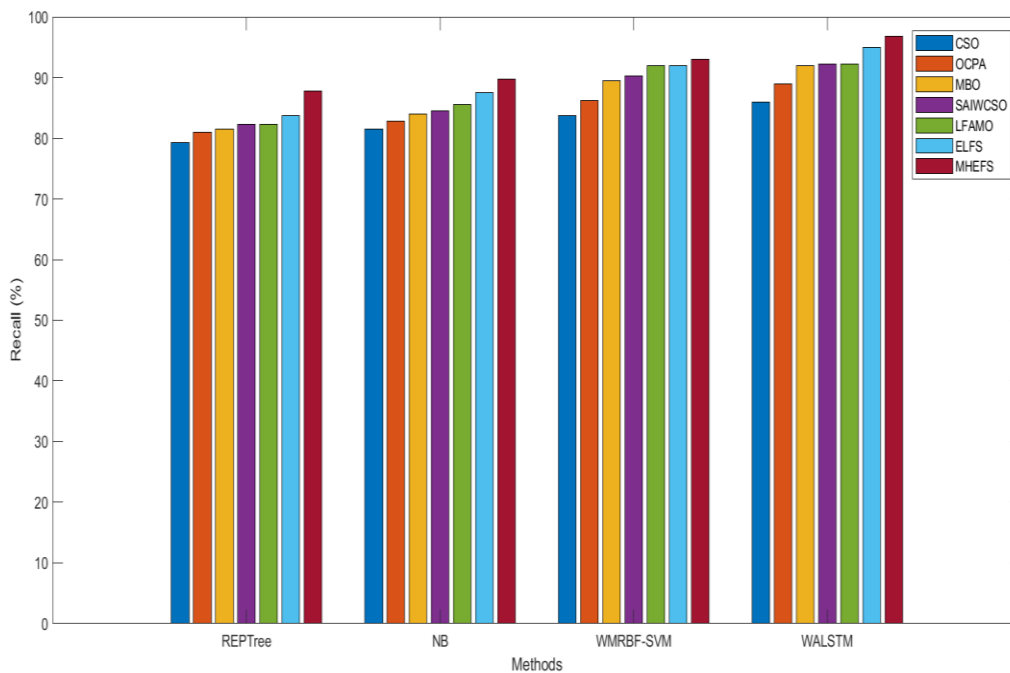
Methods	CSO	OCPA	MBO	SAIWCSO	LFAMO	ELFS	MHEFS
<b>REPTree</b>	79.3514	81.0569	81.5495	82.2697	82.9932	83.7662	87.8377
<b>NB</b>	81.5122	82.8183	83.9939	84.5425	85.5837	87.5168	89.7709
<b>WMRBF-SVM</b>	83.7662	86.2853	89.5246	90.3008	91.9876	92.0436	93.0475
<b>WALSTM</b>	86.0203	89.0319	91.0063	92.1899	93.3086	95.0179	96.7794
<b>F-Measure (%)</b>							
Methods	CSO	OCPA	MBO	SAIWCSO	LFAMO	ELFS	MHEFS
<b>REPTree</b>	76.0941	78.1758	78.7537	79.6367	79.8899	81.5352	84.2531
<b>NB</b>	79.0561	79.9934	81.3499	81.7991	83.0774	85.3804	87.6353
<b>WMRBF-SVM</b>	81.5352	83.6845	87.0241	87.6619	89.3615	89.8280	91.7158
<b>WALSTM</b>	83.7464	86.5679	88.6249	89.6988	91.1553	93.4081	94.9275
<b>Specificity (%)</b>							
Methods	CSO	OCPA	MBO	SAIWCSO	LFAMO	ELFS	MHEFS
<b>REPTree</b>	80.9755	82.3624	83.4020	84.3232	86.2312	87.7377	90.5501
<b>NB</b>	83.8057	85.4416	85.9011	87.0306	87.4849	88.9613	92.5865
<b>WMRBF-SVM</b>	84.8771	90.1443	92.2231	93.1148	93.3855	94.2816	94.8188
<b>WALSTM</b>	88.6716	91.5510	93.8203	94.5113	95.4184	96.1124	97.8841
<b>Accuracy (%)</b>							
Methods	CSO	OCPA	MBO	SAIWCSO	LFAMO	ELFS	MHEFS
<b>REPTree</b>	77.3234	80.6691	81.4126	82.5279	83.2714	84.7584	85.5019
<b>NB</b>	81.7844	82.8996	83.6431	84.3866	85.8736	88.1041	89.9628
<b>WMRBF-SVM</b>	84.7584	86.2454	89.2193	89.5911	91.0781	91.8216	93.3086
<b>WALSTM</b>	86.6171	88.8476	90.7063	91.4498	92.5651	94.7955	95.9108





**FIGURE 2. PRECISION RESULTS COMPARISON OF CLASSIFIERS VS. FS METHODS**

Figure 2 depicts the precision results comparison of classifiers methods with respect to several FS methods. The results are varied between four different classifiers (REPTree, NB, WMRBF-SVM, and proposed WALSTM) along FS methods. The results are varied between each method, it shows that the proposed system has higher results of 93.0756%, whereas other methods such as REPTree, NB, and WMRBF-SVM also gives improved results of 80.6685%, 85.4997%, and 90.3841% for proposed MHEFS algorithm. The WALSTM algorithm gives highest results of 81.4725%, 84.1040%, 86.2436%, 87.2078%, 89.0021%, and 91.7983% for CSO, OCPA, MBO, SAIWCOS, LFAMO, and ELFS algorithm.

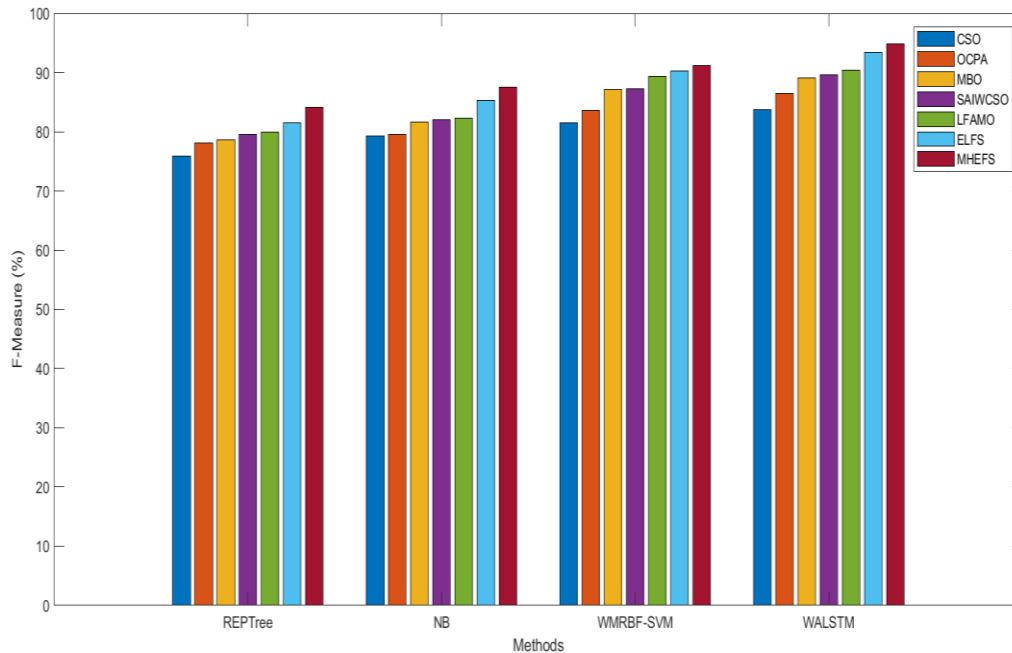


**FIGURE 3. RECALL RESULTS COMPARISON OF CLASSIFIERS VS. FS METHODS**

Recall results are varied between four different classifiers (REPTree, NB, WMRBF-SVM, and proposed WALSTM) along FS methods are illustrated in figure 3. The results are varied between each methods, it shows that

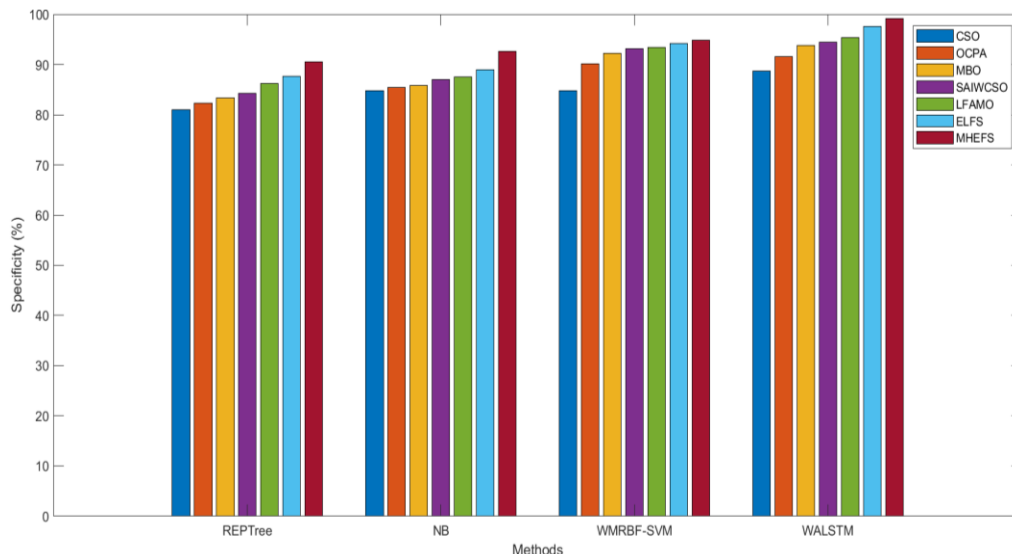


the proposed system has higher results of 96.7794%, whereas other methods such as REPTree, NB, and WMRBF-SVM also gives improved results of 87.8377%, 89.7709%, and 93.0475% for proposed MHEFS algorithm. The WALSTM algorithm gives highest results of 86.0203%, 89.0319%, 91.0063%, 92.1899%, 93.3086%, and 95.0179% for CSO, OCPA, MBO, SAIWCOS, LFAMO, and ELFS algorithm.



**FIGURE 4. F-MEASURE RESULTS COMPARISON OF CLASSIFIERS VS. FS METHODS**

Four different classifiers (REPTree, NB, WMRBF-SVM, and proposed WALSTM) along FS methods via F-measure comparison is illustrated in figure 4. The results are varied between each methods, it shows that the proposed system has higher results of 94.9275%, whereas other methods such as REPTree, NB, and WMRBF-SVM also gives improved results of 84.2531%, 87.6353%, and 91.7158% for proposed MHEFS algorithm. The WALSTM algorithm gives highest results of 83.7464%, 86.5679%, 88.6249%, 89.6988%, 91.1553%, and 93.4081% for CSO, OCPA, MBO, SAIWCOS, LFAMO, and ELFS algorithm.

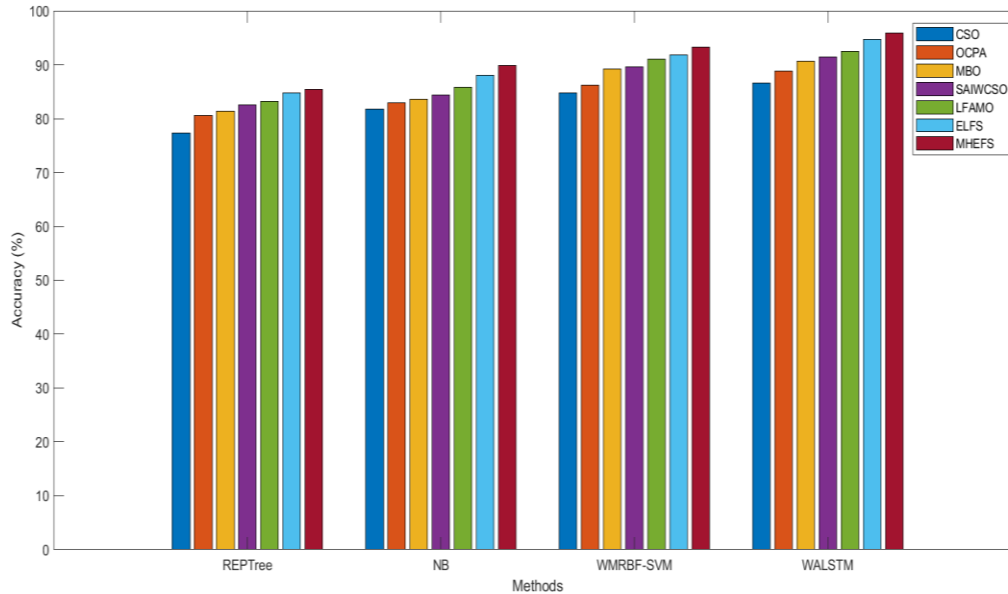


**FIGURE 5. SPECIFICITY RESULTS COMPARISON OF CLASSIFIERS VS. FS METHODS**

Classifiers such as REPTree, NB, WMRBF-SVM, and proposed WALSTM along FS methods by specificity comparison are illustrated in figure 5. The results are varied between each methods, it shows that the proposed system has higher results of 97.8841%, whereas other methods such as REPTree, NB, and WMRBF-SVM also gives improved results of 90.5501%, 92.5865%, and 94.8188% for proposed MHEFS algorithm. The



WALSTM algorithm gives highest results of 88.6716%, 91.5510%, 93.8203%, 94.5113%, 95.4184%, and 96.1124% for CSO, OCPA, MBO, SAIWCSO, LFAMO, and ELFS algorithm.



**FIGURE 6. ACCURACY RESULTS COMPARISON OF CLASSIFIERS VS. FS METHODS**

Accuracy results comparison among several classifiers with FS methods is illustrated in figure 6. It shows that the proposed system has higher results of 95.9108%, whereas other methods such as REPTree, NB, and WMRBF-SVM also gives improved results of 85.5019%, 89.9628%, and 93.3086% for proposed MHEFS algorithm. The WALSTM algorithm gives highest results of 86.6171%, 88.8476%, 90.7063%, 91.4498%, 92.5651%, and 94.7955% for CSO, OCPA, MBO, SAIWCSO, LFAMO, and ELFS algorithm.

**5. CONCLUSION AND FUTURE WORK**

In this paper, Metaheuristic Ensemble Feature Selection (MHEFS) and Weighted Average Long Short-Term Memory Network (WALSTM) have been presented for breast cancer recurrence prediction. Initially, Z score normalization has been introduced for data normalization. Then, the MHEFS algorithm is performed depending on Opposition Colony Predation Algorithm (OCPA), Monarch Butterfly Optimization (MBO), and Linear Function based Animal Migration Optimization (LFAMO). MHEFS algorithm is introduced to integrate many different feature selection approaches (OCPA, MBO, and LFAMO), and create an optimal best subset via majority voting. OCPA is introduced as a mapping strategy for selection of optimal features with animal hunting parties by learning. MBO, monarch butterflies are rationalized in 2ways (Butterfly adjusting operator and Migration operator). LFAMO, migration process and the population update has been introduced for feature selection. From this majority voting process is used to integrate results from each algorithm. Finally, the WALSTM classifier has been working depending on input gate, forget gate, output gate with cell activation element for recurrence prediction. Results it concludes that the proposed classifier gives better results via precision (93.0756%), recall (96.7794%), F-measure(94.9275%), specificity(97.8841%), and accuracy(95.9108%).Performance of MHEFS method is verified using gene expression, and real time dataset of WBCD. The convergence of the MHEFS technique will be scrutinized theoretically by Markov chains and dynamic systems. The system had been implemented to analyze the hazard factors of breast cancer via the real time dataset.

**REFERENCES**

1. Akram, M., Iqbal, M., Daniyal, M. and Khan, A.U., 2017. Awareness and current knowledge of breast cancer. *Biological research*, 50(1), pp.1-23.
2. McDonald, E.S., Clark, A.S., Tchou, J., Zhang, P. and Freedman, G.M., 2016. Clinical diagnosis and management of breast cancer. *Journal of Nuclear Medicine*, 57(Supplement 1), pp.9S-16S.
3. Farr, R. Wuerstlein, A. Heiduschka, C. F. Singer, and N. Harbeck, "Modern risk assessment for individualizing treatment concepts in early-stage breast cancer," *Rev. Obstetrics Gynecol.*, vol. 6, nos. 3\_4, pp. 165\_173, 2013.
4. DeSantis, C.E., Ma, J., Gaudet, M.M., Newman, L.A., Miller, K.D., Goding Sauer, A., Jemal, A. and Siegel, R.L., 2019. Breast cancer statistics, 2019. *CA: a cancer journal for clinicians*, 69(6), pp.438-451.





5. Li Q, Chen H, Huang H, Zhao X, Cai Z, Tong C, Liu W, Tian X. An enhanced Grey wolf optimization based feature selection wrapped kernel extreme learning machine for medical diagnosis. *Comput Math Methods Med.* 2017; 2017:9512741.
6. Zhao X, Zhang X, Cai Z, Tian X, Wang X, Huang Y, Chen H, Hu L. Chaos enhanced grey wolf optimization wrapped ELM for diagnosis of paraquatpoisoned patients. *ComputBiol Chem.* 2018. <https://doi.org/10.1016/j.compbiolchem.2018.11.017>.
7. K. Kourou, T. P. Exarchos, K. P. Exarchos, M. V. Karamouzis, and D. I. Fotiadis, "Machine learning applications in cancer prognosis and prediction," *Comput. Struct. Biotechnol. J.*, vol. 13, pp. 8-17, Dec. 2015.
8. Nguyen, Q.H.; Do, T.T.; Wang, Y.; Heng, S.S.; Chen, K.; Ang, W.H.M.; Philip, C.E.; Singh, M.; Pham, H.N.; Nguyen, B.P.; et al. Breast Cancer Prediction using Feature Selection and Ensemble Voting. In *Proceedings of the 2019 International Conference on System Science and Engineering (ICSSE)*, Dong Hoi City, Vietnam, 19–21 July 2019; pp. 250–254.
9. Agarap, A.F.M., 2018, On breast cancer detection: an application of machine learning algorithms on the wisconsin diagnostic dataset. In *Proceedings of the 2nd international conference on machine learning and soft computing*, pp.5-9.
10. Hoque, N., Singh, M. and Bhattacharyya, D.K., 2018. EFS-MI: an ensemble feature selection method for classification. *Complex & Intelligent Systems*, 4(2), pp.105-118.
11. Chen, C.W., Tsai, Y.H., Chang, F.R. and Lin, W.C., 2020. Ensemble feature selection in medical datasets: Combining filter, wrapper, and embedded feature selection results. *Expert Systems*, 37(5), pp.1-10.
12. Seijo-Pardo, B., Porto-Díaz, I., Bolón-Canedo, V., & Alonso-Betanzos, A. (2017). Ensemble feature selection: homogeneous and heterogeneous approaches. *Knowledge-Based Systems*, 118, pp.124-139.
13. Das, A., Mohanty, M. N., Mallick, P. K., Tiwari, P., Muhammad, K., & Zhu, H. (2021). Breast cancer detection using an ensemble deep learning method. *Biomedical Signal Processing and Control*, 70, pp.1-11.
14. Aličković, E. and Subasi, A., 2017. Breast cancer diagnosis using GA feature selection and Rotation Forest. *Neural Computing and applications*, 28(4), pp.753-763.
15. Punitha, S., Al-Turjman, F. and Stephan, T., 2021. An automated breast cancer diagnosis using feature selection and parameter optimization in ANN. *Computers & Electrical Engineering*, 90, p.106958.
16. Alfian, G., Syafrudin, M., Fahrurrozi, I., Fitriyani, N.L., Atmaji, F.T.D., Widodo, T., Bahiyah, N., Benes, F. and Rhee, J., 2022. Predicting Breast Cancer from Risk Factors Using SVM and Extra-Trees-Based Feature Selection Method. *Computers*, 11(9), p.136.
17. Antunes, A.R., Matos, M.A., Costa, L.A., Rocha, A.M.A. and Braga, A.C., 2021, July. Feature Selection Optimization for Breast Cancer Diagnosis. In *International Conference on Optimization, Learning Algorithms and Applications* (pp. 492-506). Springer, Cham.
18. Huang, H., Feng, X.A., Zhou, S., Jiang, J., Chen, H., Li, Y. and Li, C., 2019. A new fruit fly optimization algorithm enhanced support vector machine for diagnosis of breast cancer based on high-level features. *BMC bioinformatics*, 20(8), pp.1-14.
19. Ahmad, F., Mat Isa, N.A., Hussain, Z., Osman, M.K. and Sulaiman, S.N., 2015. A GA-based feature selection and parameter optimization of an ANN in diagnosing breast cancer. *Pattern Analysis and Applications*, 18(4), pp.861-870.
20. Stephan, P., Stephan, T., Kannan, R. and Abraham, A., 2021. A hybrid artificial bee colony with whale optimization algorithm for improved breast cancer diagnosis. *Neural Computing and Applications*, 33(20), pp.13667-13691.
21. Sakri, S.B., Rashid, N.B.A. and Zain, Z.M., 2018. Particle swarm optimization feature selection for breast cancer recurrence prediction. *IEEE Access*, 6, pp.29637-29647.
22. Jumanto, J., Mardiansyah, M.F., Pratama, R.N., Al Hakim, M.F. and Rawat, B., 2022. Optimization of breast cancer classification using feature selection on neural network. *Journal of Soft Computing Exploration*, 3(2), pp.105-110.
23. Aavula, R.; Bhramaramba, R. XBPF: An Extensible Breast Cancer Prognosis Framework for Predicting Susceptibility, Recurrence and Survivability. *Int. J. Eng. Adv. Technol.* 2019, 2249–8958.
24. Tu, J.; Chen, H.; Wang, M.; Gandomi, A.H. The Colony Predation Algorithm. *J. Bionic Eng.* 2021, 18, 674–710.



25. Wang, G.G., Deb, S. and Cui, Z., 2019. Monarch butterfly optimization. *Neural computing and applications*, 31(7), pp.1995-2014.
26. Lai, Z., Feng, X., & Yu, H. (2019). An Improved Animal Migration Optimization Algorithm Based on Interactive Learning Behavior for High Dimensional Optimization Problem. In *2019 International Conference on High Performance Big Data and Intelligent Systems (HPBD&IS)*, pp. 110-115.
27. Kiziloz, H.E., 2021. Classifier ensemble methods in feature selection. *Neurocomputing*, 419, pp.97-107.
28. Zhang, Y., Wang, D., Chen, Y., Shang, H. and Tian, Q., 2017, Credit risk assessment based on long short-term memory model. In *International conference on intelligent computing* (pp. 700-712). Springer, Cham.

