A Novel Adaptive Mutation Enhanced Elephant Herding Optimization (Ameho) Based Feature Selection And Kernel Extreme Learning Machine (Kelm) Classifier For Breast Cancer Diagnosis

R.S.Padma Priya[a], Dr.P.Senthil Vadivu[b]

[a]Assistant Professor, Dr.N.G.P.Arts and Science College, Research Scholar, Hindusthan College of Arts and Science, Coimbatore, Tamilnadu, India
[b]Professor& Head of Computer Applications (UG), Hindusthan College of Arts and Science, Coimbatore, Tamilnadu, India, priypadma@gmail.com, sowjupavan@gmail.com

ABSTRACT: Breast cancer is one among the deadly diseases in women that have the high death percentage in the world. An accurate and initial recognition of breast cancer via dataset is still a difficult task. Due to the huge dispersion of the disease, automatic recognition schemes can benefit physicians to categorize the tumors as benign or malignant. Though, carrying out an automatic recognition is time intense and produces reduced accuracy. Meanwhile numerous statistics of features are existing in the dataset. Over the centuries, meta-heuristic optimization methods have remained and useful for Feature Selection (FS), as these are able to overwhelmed the limits of old optimization methods. Data mining techniques can sustenance doctors in analysis decision-making method. This paper reports feature selection and classification method for breast cancer analysis. The proposed system has dual steps. In the first step, in order to remove trivial features, wrapper method using Adaptive Mutation Enhanced Elephant Herding Optimization (AMEHO)built on FS for variety of useful and important features. In AMEHO algorithm, clan bring up-to-date operator for sorting according to appropriateness and three degrees of freedom (α, β, γ). Local optima problematic is resolved by announcing adaptive mutation operator. This method decreases the computational difficulty and rapidity of data mining method. This FS algorithm is used to improve the accuracy of analysis (benign and malignant). In the following stage, Kernel Extreme Learning Machine (KELM) classifier is active to choose for two diverse categories of subjects with or without breast cancer. To assess efficiency of proposed process, three different classifiers such as K Nearest Neighbour (KNN), Naïve Bayes (NB), Multi-Layer Perceptron (MLP) and Support Vector Machine (SVM) on Wisconsin Diagnosis Breast Cancer (WDBC) and Wisconsin Original Breast Cancer (WOBC) using University of California, Irvine (UCI) repository. Performance metrics such as Precision, Recall, F-measure, Accuracy, Area Under Curve (AUC), statistical measure via k-fold cross proof is taken for linking the proposed system with the present works.

KEYWORDS: Breast cancer, Data mining, Feature Selection (FS), Adaptive Mutation Enhanced Elephant Herding Optimization (AMEHO), Kernel Extreme Learning Machine (KELM) classifier, Wisconsin Breast Cancer Dataset (WBC), Wisconsin Diagnosis Breast Cancer (WDBC), and University of California, Irvine (UCI).

1. INTRODUCTION

Breast cancer is one of the most common cancers in women, and it is a leading cause of death globally [1,2]. In developed countries, it is the second leading cause of death for women. According to global cancer estimates, the number of confirmed cases in 2018 was expected to reach 18,078,957, with 9,555,027 deaths (52.85%) [3]. Breast cancer affects 2,088,849 people (11.55% of the population) and claims the lives of 626,679 people per year (6.56 percent). 60 percent of all deaths occur in low-income developing countries, like Ethiopia [3, 4, 5]. The final diagnosis can be difficult to achieve, particularly for a medical specialist, due to the large number of specifics. Improvements in laboratories have allowed the collection of large medical databases, which necessitates the discovery of secret associations in documents.

To fix these issues, data mining techniques are widely used in the medical field [6,7]. One of the applications of database analysis is automated diagnostic systems. This service will help doctors make better decisions. Another application is to look for ways to enhance patient outcomes, reduce costs, and improve clinical trials. Furthermore, in the case of serious diseases such as cancer, where early detection improves long-term survival and lowers costs, the need for automated diagnosis has become particularly pressing. Early detection has been shown to reduce the number of deaths among breast cancer patients. As a result, more effective breast cancer detection methods are needed. These methods

may aid in the classification of patients into either a "benign" group that may or may not have breast cancer, or a "malignant" group that has strong evidence of breast cancer.

2. SIGNIFICANCE OF THE STUDY

The dangers of malignant tumors are greater than those of benign tumors. As previously noted, early detection of breast cancer increases the likelihood of a successful cure. Doctors will need high-precision, high-reliability diagnostic systems to help them distinguish between benign and malignant breast tumors in order to achieve this goal. One of the problems is the complexity of diagnostic device features. The classification algorithm becomes more unstable as these features become more irrelevant and repetitive, reducing learning precision. One method for dealing with this problem is feature selection (FS), which is needed in classification [8]. FS is a pre-processing technique for data mining that is commonly used in mathematics, pattern analysis, and medicine. It aids in the elimination of all such ineffective and redundant features, thus reducing processing time and storage space. Therefore, the subsequent machine learning or data mining algorithms' overall classification (or prediction) output increases [9].

Based on feature evaluation criteria, FS strategies are divided into three classes [10]: The filter solution employs such techniques to evaluate the selected subset while omitting the classifier algorithm. Features are evaluated in the filter system using pre-defined parameters like Information Gain, ReliefF, Chi-square, Fisher Score, Laplacian Score, and so on, and the most important features are chosen based on those criteria. Wrapper scheme, on the other hand, uses a learning algorithm to evaluate function subsets and then chooses the right one for the task [10]. The wrapper approach evaluates the consistency of a subset of features by learning and evaluating a classifier that only uses the variables in the proposed subset. In other words, the intrinsic properties of the data are used to determine the utility of a subset of features [11]. In the embedded process, the best subset of features is chosen during the model development step. Filter methods are faster than wrapper methods because they do not require learning algorithms, but wrapper methods generally achieve higher accuracy [12].

SI algorithms are increasingly being used in FS [13], as SI has been shown to solve NP-hard computational problems, one of which is finding an ideal function subset. In recent years, SI algorithms have become more general, and Basir and Ahmad [14] compared swarm algorithms for feature selection and reduction. The only bio-inspired swarm algorithms found were Particle Swarm Optimization (PSO), Ant Colony Optimization Algorithms (ACO), Artificial Fish Swarm Algorithms (AFSA), Artificial Bees Colony Algorithms (ABC), Firefly Algorithms (FA), and Bat Algorithms (BA). FS Evolutionary Computation (EC) techniques were thoroughly examined by Xue et al [15]. Evolutionary Algorithms (EAs), Simulation Algorithms (SIs), and other EC paradigms were classified. Several optimization algorithms have been influenced by swarm intelligence (SI). Traditional algorithms have several advantages over SI-based algorithms.

SI-based algorithms with meta-heuristic algorithms have been very popular in solving various optimization problems in the last decade [12] because of their ability to resolve local optima, non-derivative mechanism, and stability. Two key characteristics of a meta-heuristic algorithm are exploration or diversification [10], which is the ability to search the entire solution space for the right solution in each iteration while ignoring local optima, and exploitation or intensification, which means finding a better solution about the obtained solution, leading to faster convergence. A effective meta-heuristic algorithm balances exploration and manipulation. The Elephant Herding Optimization (EHO) algorithm simulates elephant herding behaviour to solve global optimization problems [16, 17]. The EHO algorithm, on the other hand, is prone to being trapped within local optima due to limited exploration or an incorrect exploration-exploitation balance.

In previous studies, the EHO algorithm was shown to solve feature selection issues. Ismaeel et al [18] implemented the Enhanced Elephant Herding Optimization (EEHO) algorithm to reduce the vulnerability of the initial EHO to local minima. In this paper, the adaptive mutation operator is used to solve the local minima problem of the worst response. In this article, an improved version of EHO is used as a wrapper feature selection method to find the perfect function subsets. This study proposes a wrapper feature selection approach based on Adaptive Mutation Enhanced Elephant Herding Optimization (AMEHO). To improve the algorithm's discovery and exploitation capabilities, the enhanced model of EHO includes an inertia weight parameter and an adaptive mutation operator. Previous AMEHO feature selection research used a fitness function that was designed to enhance classification accuracy. The fitness function achieves a balance between the number of features selected and classification precision. The Kernel Extreme

Learning Machine (KELM) classifier is used to perform the chosen classification task and compute fitness function. The idea was tested using data from Wisconsin's breast cancer databases.

3. LITERATURE REVIEW

Ghosh et al [19] proposed a two-stage model for feature selection in microarray datasets. The gene rankings for different filter methods vary a lot, and the efficacy of the rankings is determined by the dataset. An Ensemble of Filter (EOF) method is created by combining and intersecting the top-n features of Relief, chi-square, and symmetrical uncertainty. Designers can use this ensemble to combine all three ranks' information into a single subset. In the next step, the Genetic Algorithm (GA) is applied to the union and intersection to produce fine-tuned results, with the former outperforming the latter. The proposed model has been shown to be classifier stable using three classifiers: The Multi-Layer Perceptron (MLP), Support Vector Machine (SVM), and K Nearest Neighbour (KNN). The definition was tested using datasets from the colon, liver, leukaemia, Small Round Blue Cell Tumours (SRBCTs), and prostate cancer. Experiment results demonstrate the model's dominance over country approaches.

Jain et al. [20] proposed a two-phase hybrid cancer classification model with Correlation-based Feature Selection (CFS) and improved-Binary Particle Swarm Optimization (iBPSO). This iBPSO model selects a low dimensional set of prognostic genes to classify biological samples of binary and multiclass cancers using the Naive–Bayes (NBs) classifier with stratified 10-fold cross-validation. The proposed iBPSO algorithm also solves the problem of early convergence to the local optimum in standard BPSO. The proposed model was tested on 11 benchmark microarray datasets with different cancer types. The results of the experiments were comparable to those of seven other well-known methods, and our model outperformed them in most respects in terms of classification accuracy and number of selected genes. Classification accuracy is checked for seven out of eleven datasets with a very small prognostic gene subset on all eleven datasets (up to 1.5 percent).

Jeyasingh and Veluchamy [21] proposed the Modified Bat Algorithm (MBA) for feature selection to delete irrelevant features from an initial dataset. The Bat algorithm was modified to use simple random sampling to distinguish random instances from the database. The most common features in the dataset were identified by using the global best features. The attributes are used to train a Random Forest (RF) classification algorithm. The MBA feature selection algorithm enhanced the RF classification's efficiency in detecting breast cancer. The Wisconsin Diagnosis Breast Cancer Dataset (WDBC) was used to validate the output of the proposed MBA feature selection algorithm. The proposed algorithm outperformed the competition in terms of the Kappa statistic, Mathew's Correlation Coefficient, Precision, F-measure, Recall, Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Relative Absolute Error (RAE), and Root Relative Squared Error (RRSE).

Mafarja and Mirjalili [22] developed two hybridization models based on the Whale Optimization Algorithm (WOA). The Simulated Annealing (SA) algorithm is used to develop the best solution found after each iteration of the WOA algorithm in the first model, while it is used to develop the best solution found after each iteration of the WOA algorithm in the second model. In this case, the aim of using SA is to boost exploitation by locating the most promising areas found by the WOA algorithm. On 18 standard benchmark datasets from the University of California, Irvine (UCI) repository, the proposed approaches are comparable to three well-known wrapper feature selection strategies in the literature. The experimental results show that the proposed methods are more efficient at optimizing classification precision than other wrapper-based algorithms, suggesting that the WOA algorithm can search the feature space and choose the most descriptive attributes for classification tasks.

To increase population diversity in the search domain, Tubishat et al. [23] used Opposition Based Learning (OBL) during the initialization phase of the Salp Swarm Algorithm (SSA). The second improvement is the introduction and use of a new Local Search Algorithm with SSA to improve its exploitation. To validate and check the performance of the proposed enhanced SSA, ISSA was applied to 18 datasets from the UCI repository (ISSA). ISSA was also related to four well-known optimization algorithms, including the Genetic Algorithm (GA), Particle Swarm Optimization (PSO), Grasshopper Optimization Algorithm (GOA), and Ant Lion Optimizer (ALO). The research shows that ISSA outperforms all baseline algorithms in terms of fitness values, precision, convergence curves, and feature reduction in many of the datasets used. As shown by results obtained on a variety of datasets, the wrapper FS model can be used in a variety of expert and intelligent device applications.

Al-Tashi et al [24] proposed a binary version of the hybrid Grey Wolf Optimization and Particle Swarm Optimization to solve function selection (BGWOPSO). PSOGWO is a modern hybrid optimization algorithm that utilizes the strengths of both GWO and PSO. Despite its superior performance, the original hybrid solution is ideal for problems with a continuous search space. To find the best solutions, the wrapper-based method K-Nearest Neighbors (KNNs) classifier with Euclidean separation metric is used. The findings show that BGWOPSO outperforms the binary GWO (BGWO), binary PSO, Binary Genetic Algorithm (BGA), and Whale Optimization Algorithm (WOA) with Simulated Annealing while using multiple efficiency metrics such as precision, choosing the best accurate result, and computational time (SA).

To select the best function subset for classification, Emary et al. [25] proposed a new binary version of the Grey Wolf Optimization (GWO). In the first method, individual steps against the first three best choices are binarized, and in the second method, stochastic crossover is used to find the changed binary grey wolf position of the three basic movements. The second approach uses a sigmoidal function to squash the continuous changed position, after which the values are stochastically thresholder to find the new binary grey wolf position. The binary Grey Wolf Optimization (bGWO) methods are used in the FS domain to classify feature subsets that optimize classification accuracy while reducing the quantity of selected features. The proposed binary versions were linked to two common optimizers in this domain using a particle swarm optimizer and genetic algorithms. The results show that the proposed bGWO would search the feature space for optimal feature combinations regardless of the initialization or stochastic operators used.

Ghosh et al [26] proposed the Binary Sailfish (BSF) optimizer to deal with FS problems, which is a binary variant of the recently proposed Sailfish Optimizer (SFO). The SFO's continuous search space is converted to a binary one using the sigmoid transformation function. Combine the BSF optimizer with another recently proposed meta-heuristic algorithm, adaptive-hill-climbing (AHC), to increase the BSF optimizer's exploitation potential. On 18 standard UCI datasets, the suggested BSF and ABSF algorithms are equivalent to 10 advanced meta-heuristic FS methods. BSF and ABSF algorithms are superior in solving FS problems, according to the results.

Hegazy et al [27] improved the structure of simple SSA to improve process accuracy, performance, and convergence time. A new control parameter, inertia weight, has been added to adjust the current best answer. The Improved Salp Swarm Algorithm (ISSA) is a cutting-edge method that is currently being tested in a feature selection process. For feature collection, the ISSA algorithm is combined with the KNNs classifier, and the accuracy of the ISSA algorithm is assessed using twenty-three UCI datasets. The ISSA is linked to the basic SSA and four other swarm strategies. The results show that the suggested solution outperformed the other optimizers in terms of classification accuracy and feature reduction. Hegazy et al. [28] suggested a new chaotic SSA algorithm (CSSA) to assess the usability of the SSA algorithm, which employs chaotic maps to increase the algorithm's accuracy. The CSSA algorithm is combined with the KNN classifier to solve the FS problem, and twenty-seven datasets are used to assess the CSSA algorithm's accuracy. The proposed chaotic SSA (especially Tent map) outperformed standard SSA and other optimization algorithms.

4. PROPOSED METHODOLOGY:

The proposed methodology in this study is built on feature selection and classification techniques. This description is typically broken down into four distinct measures. The first development is to generate a dataset list, which is obtained from the UCI ML repository. Adaptive Mutation Enhanced Elephant Herding Optimization (AMEHO) based on wrapper FS algorithm is implemented for selection of perceptive and important features in the second phase. The third stage is classification, and for enhancing the generalization efficiency of the ELM learning algorithm, a Kernel Based Extreme Learning Machine (KELM) classifier is recommended based on the kernel function. The KELM classification algorithm learns from a training dataset with class labelling to build the hypothesis. The performance phase is the preceding stage. The validation dataset is used to measure the model's classification accuracy. For designing and examination of the various classifiers, the research was accepted out in MATrixLABoratory R2014a (MATLAB 2014a). The architecture diagram of the proposed design is shown in Figure 1.
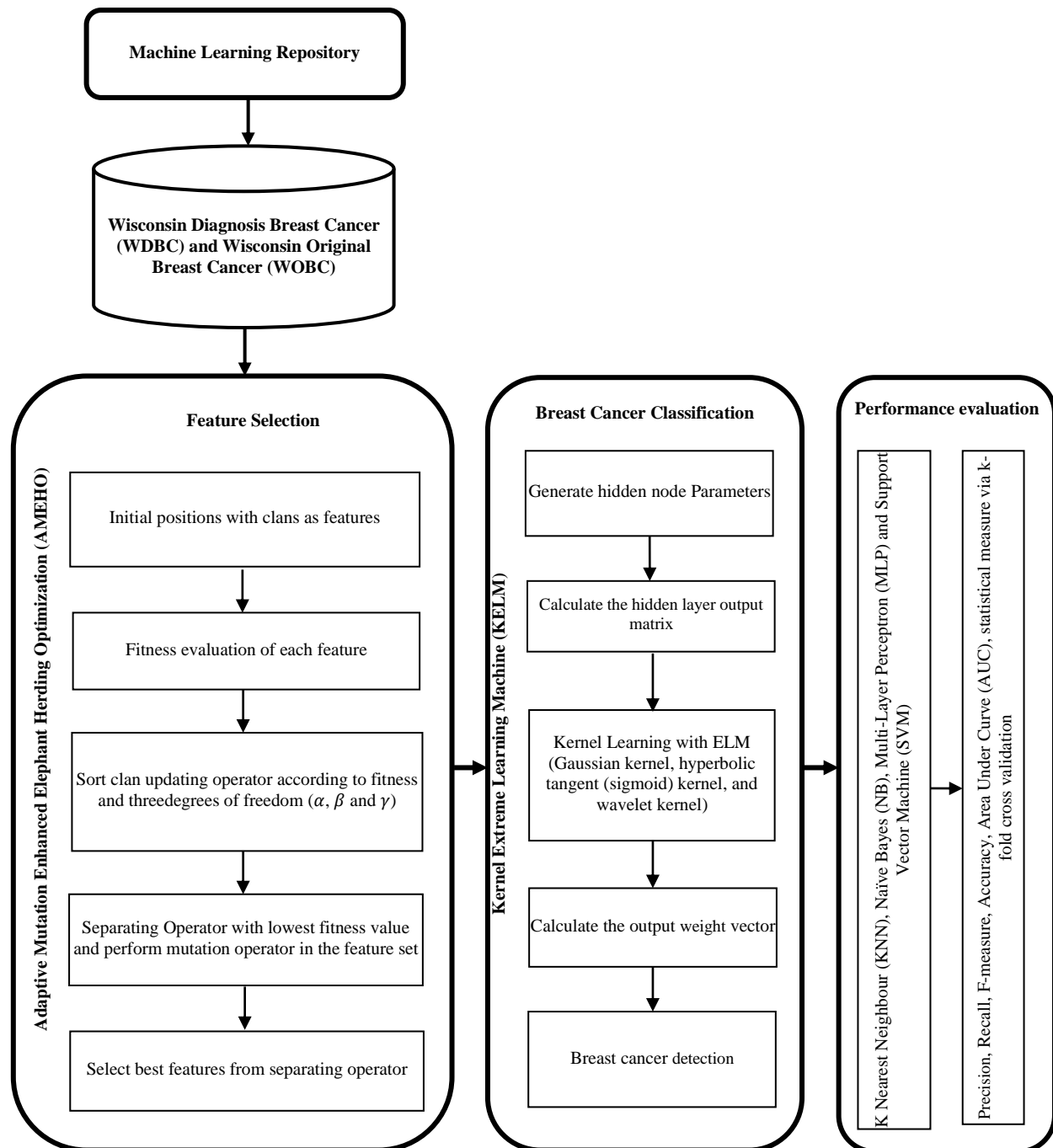
FIGURE 1. Overall architecture of proposed Breast Cancer Diagnosis Model

4.1. Dataset Collection:

Data collection necessitates collecting information. The data for this study made from the Machine Learning (ML) repository at the University of California, Irvine (UCI). It's an open-source project. This source assembles widely accessible BC datasets such as WDBC and WOBC.

4.2. Feature Selection:

Feature Selection (FS) is a technique for reduction the quantity of features in breast cancer datasets and selecting a subset of them. FS is frequently used in data pre-processing to classify significant features that were formerly unknown and to delete redundant or redundant features that are not relevant to the classification process. The aim of FS is to rise classification accuracy. Wrapper approach estimates feature subsets using a learning algorithm and then chooses the best feature subset for the task. Meta-heuristic algorithms for selecting features from a dataset have been applied in the last decade.

The Adaptive Mutation Enhanced Elephant Herding Optimization (AMEHO) algorithm, which is based on the FS algorithm, is proposed for choosing breast cancer features. It resolves the difficulties of quick, unjustified convergence towards the unique EHO's source, the EHO's local optima issues, and the exploration-exploitation conversation offs of control [29]. The clan informing operator and the separating operator with transmutation operator are used in AMEHO to accomplish discovery and exploitation. In every assembly, the suitable individual elephant (Accuracy of the classifier for BC recognition (the one with the extreme fitness)) in a clan c_i is selected as the matriarch (m_i) at time t as follows by equation (1),

$$m_i^t = \underset{x \in x_i}{\operatorname{argmax}} F(x) \tag{1}$$

Where x_i is the set of individual elephants in clan i(number of examples in the BC dataset).

4.3. Clan updating operator

Respectively distinct elephant j(features of BC dataset) in clan i has an old position $x_{i,j}^t$. Its novel position $x_{i,j}^{t+1}$ is affected by the clan matriarch m_i^t according to equation (2),

$$x_{i,j}^{t+1} = x_{i,j}^t + \alpha\left(m_i^t - x_{i,j}^{t+1}\right) + \beta\left(c_i^t - x_{i,j}^{t+1}\right) + \gamma * r \tag{2}$$

Where $\alpha \in [0,1]$ is a parameter for determining the impact of the clan matriarch on the elephant (features of BC) at new position ,$\beta \in [0,1]$ is a parameter that determines the likelihood of elephants to move towards the clan center, $\gamma \in [0,1]$ is a parameter for determining the likelihood of elephants to wander (walk in a random direction),

$$r = (2 * rand - 1)(x_{max} - x_{min})(x_{amut}) \tag{3}$$

$x_{mutation}$ One of the most sensitive parameters is the mutation rate parameter. It has been shown that reducing the time to find the optimum by utilizing a mutation rate variance scheme that adapts the mutation rate parameter during the algorithm's run reduces the time to find the optimum. A mutation rate adaptation scheme is proposed in this report, which adjusts the mutation rate separately for each elephant position on the clan based on input from the performance and failure rates of the current clan's features.

The location of the elephant is written into the chromosomes and undergoes mutation in the self-adaptive method. The main premise is that better elephant (Feature) values contribute to better range, which improves classification accuracy, and that these feature values can remain in the society since they are among the surviving features. In their paper, they propose a self-adaptation mechanism for a single mutation rate per organism. By equation, the mutation of this mutation rate value achieves the current mutation rate (4). In this equation, γ is the learning rate for defining the likelihood of elephants to wander (walk in a random direction) and controls the variation speed. It is stated by equation (4),

$$x_{amut} = \left(1 + \frac{f_{max} - f_{avg}}{f} . \exp\left(\gamma . N(0,1)\right)\right)^{-1} \tag{4}$$

f signifies the fitness value of the feature, f_{max} signifies the best fitness value of the existing generation, and f_{avg} represents the regular fitness value of the present generation. r is a random vector drawn from a even distribution, x_{min} and x_{max} correspondingly are lower and upper limits of an separate elephant position (feature value of the BC dataset), proposed work in addition reflect approximately the innovative elephant position $x_{mutation}$ by means of merging the bit position of x_{min} and x_{max} . c_i^t is the midpoint of the clan and is predictable according to equation (5),

$$c_i^t = \frac{1}{n_i} \times \sum_j x_{i,j}^{t+1} \tag{5}$$

According to EHO, the original matriarch position at time $t + 1$ is continuously anywhere between the clan center and the source irrespective of its earlier position at time t. Inappropriately, this outcomes in an unfounded merging in the direction of the origin as exposed in Figure 2.
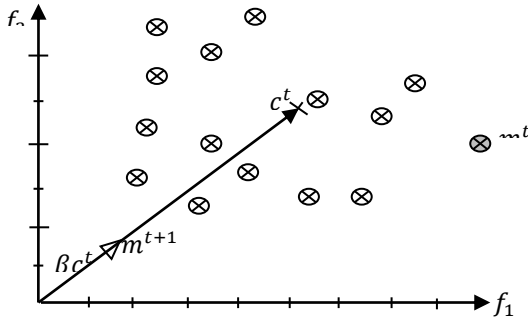
FIGURE 2. AMEHO biased Matriarch Updating Operator

Noticeably, the new position of the matriarch is not exaggerated with its old position (feature position) at all, which counterspontaneous. In addition, for small values of, the matriarch is unexpectedly (and unjustifiably) interpreted after the source. For huge values of β, the matriarch is rapidly (and unjustifiably) interpreted following to the clan center(samples). To fix the matriarch updating operator by equation (6), must be used in its place.

$$m^{t+1} = \beta(c^t - m^{t+1}) \tag{6}$$

The origin has unknown to do with this new informing operator. So, the matriarch's new position (feature position) is an affinal grouping of its old position (feature position) and the clan center dependent on a controller parameter β. According to EEHO, the matriarch position at time $t+1$ is always on the conventional line linking the clan center and the matriarch's position at time t. EHO, a on its own parameter (α) concurrently controls the impression of the matriarch position as fine as the random walk. Therefore, it is not likely to switch the exploration-exploitation trade-off. Similarly, according to EEHO, a single updating operator is used for equally the matriarch and the rest of the clan relatively than two. The new updating operator workers three dissimilar switch parameters (α, β , and γ) relatively than two only (as in EHO) for autonomously controlling the impression of the clan center, the effect of the matriarch, and the impact of distinct random walk correspondingly as in Figure 3.
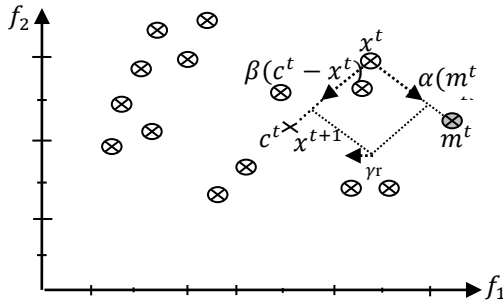

FIGURE 3. AMEHOgeneralized Elephant Updating operator with three degrees of freedom (α, β , and γ)

4.4. Separating operator

Equation (7) of an individual elephant can be used to measure the separating operator (feature position),

$$x_{i,worst}^{t+1} = x_{min} + (x_{max} - x_{min}) * rand * x_{amut} \tag{7}$$

Where x_{min} and x_{max} are correspondingly the lower and upper bounds of an individual elephant position (feature position) and $x_{i,worst}^{t+1}$ is the individual elephant with the minimum fitness in clan c_i (dataset samples), x_{amut} is the adaptive mutation parameter figured via the fitness function (See equation (4)). For the separating operator, let's jump with rand function, which is a distinctive implementation of Pseudo Random Number Generator (PRNG) that produces a uniformly scattered random number in the interval [0,1]. The Probability Density Function (PDF) must be continuously equal to 1 on [0,1] and zero otherwise. To make a real random quantity consistently spread in the range $(x_{min}, x_{max}, x_{amut})$, the output of the function rand should be scaled then explained. To make a random integer quantity regularly spread in a assumed range, floor function must be used as illustrated in Figure 4. Obviously, $floor([x_{min}, x_{amut}, x_{max}]) \in [x_{min}. x_{max-1}. x_{amut}]$; consequently, to make a discrete uniform distribution in the range $[x_{min}, x_{amut}, x_{max}]$, a constant uniform distribution in the range $[x_{min}, x_{max+1}]$ must be used. The PMF (the probability

mass function) must be continuously equal to $1/(x_{max} * x_{amut} - x_{min} + 1)$ on $[x_{min}, x_{amut}, x_{max}]$ and zero otherwise. The clan updating operator, separating operator, and the full EEHO algorithm is showed in Algorithm 1. Furthermore, Figure 5 demonstrations the whole flow diagram of EEHO, as well as equally the clan updating and separating operators:
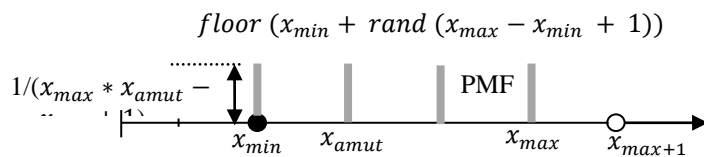
$$floor\ (x_{min} + rand\ (x_{max} - x_{min} + 1))$$

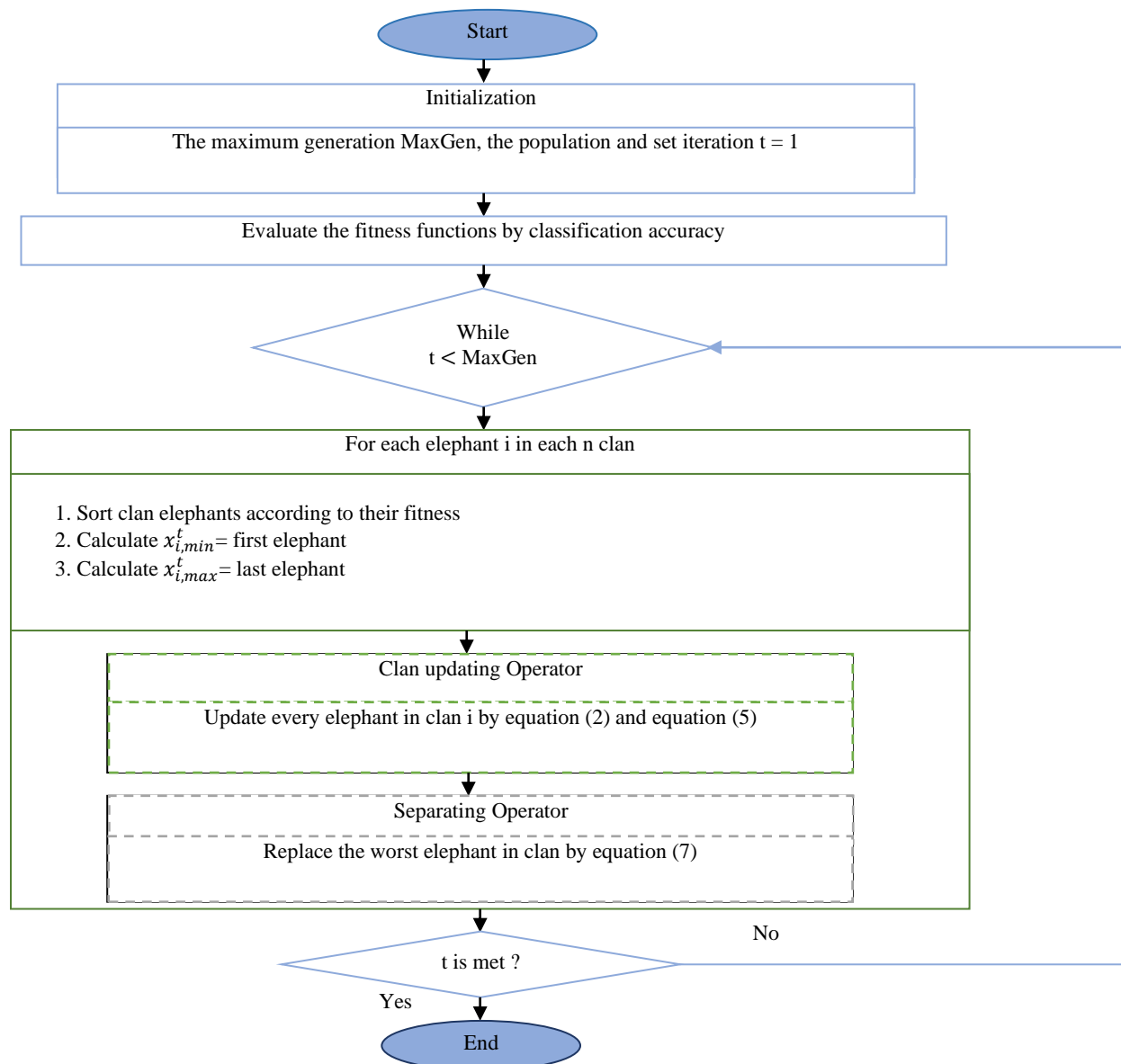**FIGURE 4. A discrete uniform distribution in the range** $[x_{min}. x_{max}. x_{amut}]$

FIGURE 5. Flowchart Of Adaptive Mutation Enhanced Elephant Herding Optimization (AMEHO)

ALGORITHM 1. ADAPTIVE MUTATION ENHANCED ELEPHANT HERDING OPTIMIZATION (AMEHO)

Initialization: Set the current generation t = 1 and the max number of generations maxgen

While t ≤ maxgen do

for i = 1 to n Clan do

Sort the population of C_i in a non-increasing order of fitness

Set $x^t_{i,best}$ = the first individual elephant

Set $x^t_{i,worst}$ = the last individual elephant

for i = 1 to n Clan do

Sort the population of C_i in a non-increasing order of fitness

Set $x^t_{i,best}$ = the first individual elephant

Set $x^t_{i,worst}$ = the last individual elephant

Set the matriarch m^t_i to x^t_{ibest}

Replace the worst individual elephant $x^{t+1}_{i,worst}$ in C_i [equation (7)]

for j = 1 to n_{Ci} do

Update every individual elephant $x^{t+1}_{i,j}$ in C_i [equation (2)and (5)]

end for

t = t + 1

end for

end while

4.5. Classification :

Breast cancer screening is important for early detection. Kernel Based Extreme Learning Machine (KELM) classifier was used for selected features in the UC Irvine Machine Learning Repository database, based on Adaptive Mutation Enhanced Elephant Herding Optimization (AMEHO) in the WDBC and WOBC. The KELM classifier [30, 31] is suggested as having a high learning speed and good results. In KELM, the hidden layer's initial parameters do not need to be modified, and almost every nonlinear piecewise continuous function can be used as a hidden neuron. Consequently, for N arbitrary separate breast cancer examples $\{(x_i, t_i)|x_i \in \mathbb{R}^n, t_i \in \mathbb{R}^m, i = 1,\dots,N\}$, the output function in ELM with L hidden neurons by equation (8),

$$f_L(x) = \sum_{i=1}^{L} \beta_i h_i(x) = h(x)\beta \tag{8}$$

where $\beta = [\beta_1, \beta_2, \dots, \beta_L]$ is the vector of the output weights between the hidden layer of L nerve cell and the output nerve cell and $h(x) = [h_1(x), h_2(x), \dots, h_L(x)]$ is the output vector of the hidden layer with respect to the input x, which maps the numbers from input space to the KELM feature space [32,33]. For decreasing the training error and refining the simplification performance of neural networks, the exercise error and the output weights must be minimalized at the similar period by equation (9),

$$Minimize: ||H\beta - T||, ||\beta|| \tag{9}$$

The least squares solution of equation (9) depending on KKT conditions be able to be written by equation (10),

$$\beta = H^T \left(\frac{1}{C} + HH^T\right)^{-1} T \tag{10}$$

Where H is the hidden layer output matrix, C is the regulation coefficient, and T is the predictable output matrix of examples. Then, the output purpose of the KELM learning algorithm by equation (11),

$$f(x) = h(x)H^T \left(\frac{1}{C} + HH^T\right)^{-1} T \tag{11}$$

If the feature mapping $h(x)$ is unidentified and the kernel matrix of KELM built on Mercer's circumstances can be well-defined by equation (12),

$$M = HH^T : m_{ij} = h(x_i)h(x_j) = k(x_i, x_j) \tag{12}$$

Therefore, the output function $f(x)$ of the Kernel Based Extreme Learning Machine (KELM) can be written efficiently as by equation (13),

$$f(x) = [k(x, x_1), \dots k(x, x_N)] \left(\frac{1}{c} + M\right)^{-1} T \tag{13}$$

Where $M = HH^T$ and $k(x, y)$ is the kernel function of hidden neurons of on its own hidden layer feed-forward neural networks. There are numerous kernel functions substantial the Mercer disorder accessible from the current literature, such as linear kernel, polynomial kernel, Gaussian kernel, and exponentialkernel [34]. The subsequent three typical kernel purposes are used for imitation and performance investigation and the selected kernel roles are as follows.
(1) Gaussian kernel is well-defined by equation (14),

$$k(x, y) = \exp(-a||x - y||) \tag{14}$$

(2) Hyperbolic tangent (sigmoid) kernel by equation (15),

$$k(x, y) = \tanh(bx^T y + c) \tag{15}$$

(3) Wavelet kernel by equation (16),

$$k(x, y) = \cos\left(d\frac{||x - y||}{e}\right) \exp\left(-\frac{||x - y||^2}{f}\right) \tag{16}$$

Where Gaussian kernel function is a representative local kernel function and tangent kernel function is a distinctive global nuclear function, correspondingly [34]. Also, the composite wavelet kernel function is also used for challenging the performance of set of rules. In the directly above three kernel purposes, the adjustable parameters a, b, c, e, and f play a foremost role in the routine of KELM and must be changed wisely based on the transmutation parameter.

5. EXPERIMENTAL RESULTS:

The simulation consequences exposed that reducing the features of a dataset can advance the routine of a classifier. In applied via the use of MATrixLABoratory R 2014 a (MATLAB 2014a). The system description of Intel Core™ i7-11375H processor (12M Cache, up to 5.00 GHz processor) with 11th generation, 4.00 GB RAM windows 8.1 pro, 64-bit operating system, operation system, and 1 TB hard disk. K Nearest Neighbour (KNN), Naïve Bayes (NB), Multi-Layer Perceptron (MLP) and Support Vector Machine (SVM)**.** For individually dataset, 80% of the examples are used for exercise the perfect and rest 20% are used for testing. Functional the FS approaches on the train data, and determined which features are to be involved in the carefully chosen feature subsection. From test data, only those features are selected and then classification metrics are measured based on test data using classifiers.

5.1.Dataset Description

The WDBC and WOBC datasets from the (UCI) ML Repository have been used in this analysis (See Table 1). There are 569 cases, 357 of which are benign (not affected) and 212 of which are malignant (affected).

TABLE 1.DATASETS USED FOR EXPERIMENTATION

Datasets	Records	Attributes	Missing Values	Class Distribution

| WDBC | 569 | 32 | No-0% | 357 benign, 212 malignant |
| WOBC | 699 | 10 | Yes- 2.28% | 458 benign, 241 malignant |

WDBC: Dataset, which includes Fine Needle Aspiration (FNA) of a breast mass based on features computed from a digitized picture. The features also identified the characteristics of the cell nuclei that existed in the picture. There are 569 data points in the sample, with 212 belonging to Malignant and 357 to Benign. The dataset has been divided into ten categories based on the following characteristics: I radius, ii) texture, iii) perimeter, iv) field, v) smoothness, vi) compactness, vii) concavity, viii) concave points, ix) symmetry, and x) fractal dimension. Three important data, respectively Mean, Standard Error, and "worst"/largest (mean of the three largest values), have been estimated for each function. As a result, there would be a total of 30 dataset functions.

WOBC: The dataset contains 699 samples gathered from the UCI repository [28]. There are 458 benign samples and 241 malignant samples in that group. In addition, the dataset contains ten attributes and one class. The class level is divided into two categories: benign and malignant. In addition, the dataset contains a missing attribute. The characteristics include: 1. Sample code number: id number, 2. 2. Thickness of the clump: 1–10. 3. Cell Size Uniformity: 1–10. 4. Cell Shape Uniformity: 1–10. 5. Adhesion to the margins: 1–10. Size of a single epithelial cell: 1–10 7. Nuclei without a shell: 1–10. 8. Chromatin smudges: 1–10 9. Normal Nucleoli (numbers 1–10) Class: Mitoses 1–10 and Mitoses: 1–10 (2 for benign, 4 for malignant). There are 699 records in the WOBC dataset, each with nine attributes save for the id number and class. These nine characteristics are rated on a scale of 1–10, with 10 indicating the most pathological condition.

5.2. Performance Evaluation Metrics :

By using MATLAB environment, verify the robustness of classifiers using two benchmark datasets. Classification parameters such as precision, recall, f-measure, accuracy, and Area under Curve (AUC) the Receiver Operating Characteristic are used to evaluate these classifiers. True Positive (TP), False Positive (FP), True Negative (TN) and Positive Negative (PN) are four effective steps determined from the uncertainty matrix output from Table 2.

TABLE 2. CONFUSION MATRIX

Total population		**Predicted class**	
		Prediction Positive	**Prediction Negative**
Actual class	**Condition Positive**	**True Positive (T_{pos})**	**False Negative (F_{neg})**
	Condition Negative	**False Positive (F_{pos})**	**True Negative (T_{neg})**

The subsequent metrics are used to estimate the performance of this research work.

True Positive (TP) (T_{pos}): Malignant people correctly identified as malignant
True Negative (TN) (T_{neg}): Benign people correctly identified as benign
False Positive (FP)(F_{pos}): Benign people incorrectly identified as malignant
False Negative (FN)(F_{neg}): Malignant people incorrectly identified as benign

Precision

The fraction of correctly categorized instances or samples from those classified as positives is evaluated by precision. As a result, equation (17) provides the precision assessment technique,

$$\text{Precision} = \text{True Positive/ (True Positive + False Positive)} = T_{pos} / (T_{pos} + F_{pos}) \quad (17)$$

Recall

The model's capacity to accurately infer positives from actual positives is measured by recall. As a result, calculation is used to measure the recall is given in equation (18),

$$\text{Recall} = \text{True Positive/ (False Negative True Positive)} = T_{pos} / (F_{neg} + T_{pos}) \quad (18)$$

F-*Measure*

F-Measure is a method for combining precision and recall into a single measure that encompasses both. Equation (19) calculates the standard F-measure as follows,

$$\textbf{F-Measure} = 2* \text{(Precision * Recall) / (Precision + Recall)} \quad (19)$$

Accuracy

The model's capacity to accurately forecast both positives and negatives out of all the predictions is represented by the accuracy value. It defines the ratio of the number of true positive and true negatives out of all the predictions mathematically is shown in equation (20).

$$\textbf{Accuracy} = T_{pos} + T_{neg} / (T_{pos} + T_{neg} + F_{pos} + F_{neg}) \quad (20)$$

Area Under Curve (AUC)

Area Under Curve (AUC) where in the curve is the receiver operating characteristic (ROC-curve) curve. The ROC-curve is the graphical plot of the True Positive Rate (TPR) versus the False Positive Rate (FPR) for a binary classifier as its discrimination threshold is wide-ranging.

*Results comparison between classifiers and datasets*

This section shows the performance comparison results of the various classifiers with respect to metrics mentioned above. The classifiers results are shown between two different datasets such as WDBC and WOBC. After completing the dataset and all missing data was obtained, the following indicators are used for evaluating the experimentation results are precision, recall, f-measure, accuracy and AUC between BC datasets under classifiers (See Table 3).

TABLE 3. NUMERICAL RESULTS COMPARISON OF CLASSIFIERS FOR BREAST CANCER DATASETS VS. METRICS

| Classifiers	WDBC Results (%)				
	Precision	**Recall**	**F-Measure**	**Accuracy**	**AUC**
EOF+KNN	76.7494	78.2338	77.4845	77.8559	72.2338
EOF+NB	79.0087	80.1946	79.5973	80.3163	74.1946
EOF+MLP	83.9060	85.4130	84.6528	85.0615	79.4310
EOF+SVM	88.4017	90.0640	89.2251	89.4552	84.0640
AMEHO+KELM	91.1877	91.6416	91.4141	91.9156	85.6416
Classifiers	WOBC Results (%)				
	Precision	**Recall**	**F-Measure**	**Accuracy**	**AUC**
EOF+KNN	76.4646	76.8883	76.6759	77.0000	70.8883
EOF+NB	80.7272	80.0082	80.3661	81.0000	74.0082
EOF+MLP	85.1880	86.0222	85.6030	85.5000	80.0222
EOF+SVM	89.1816	89.3062	89.2439	89.5000	83.3062
AMEHO+KELM	91.1130	91.8514	91.4807	91.5000	85.8514
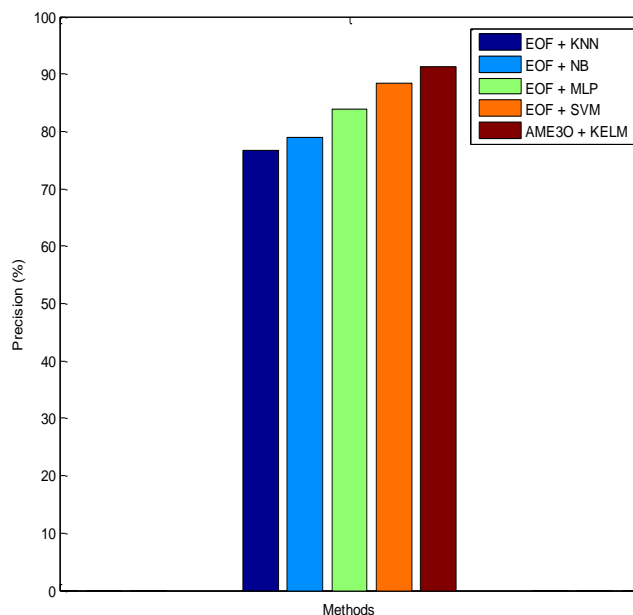
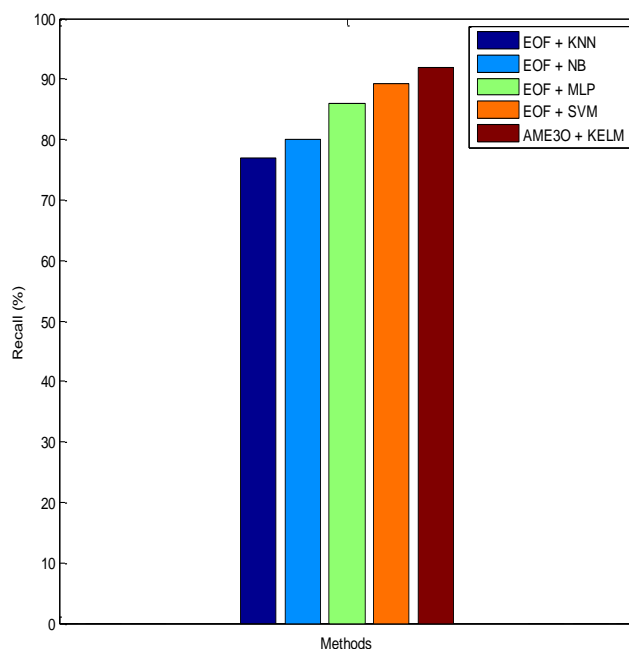**(a) Precision comparison of WDBC dataset vs. Classifiers**
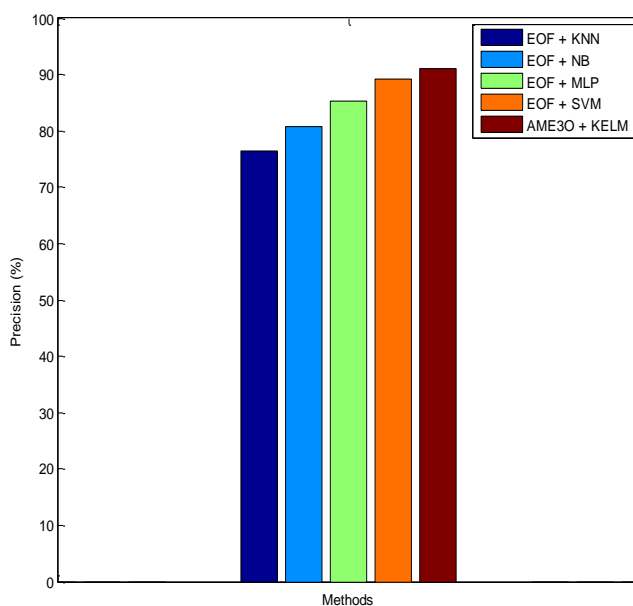
(b) Precision comparison of WOBC dataset vs. Classifiers

FIGURE 6. PRECISION ANALYSIS OF CLASSIFEIRS VS. BC DATASETS

Overall precision results comparison of the BC datasets under different classifiers such as EOF+KNN, EOF+NB, EOF+MLP, EOF+SVM and proposed AMEHO+KELM classifier is shown in the figure 6. Figure 6(a) shows the performance comparison results of the different classifiers under WDBC dataset and WOBC dataset in figure 6(b). Figure 6(a) and figure 6(b), it concludes that the proposed AMEHO+KELM classifier gives higher precision results of 91.1877% and 91.1130% for WDBC and WOBC dataset respectively. The methods like EOF+KNN, EOF+NB, EOF+MLP, and EOF+SVM produces precision value of 76.7494%, 79.0087%, 83.9060%, and 88.4017% for WDBC dataset (See Table 3).
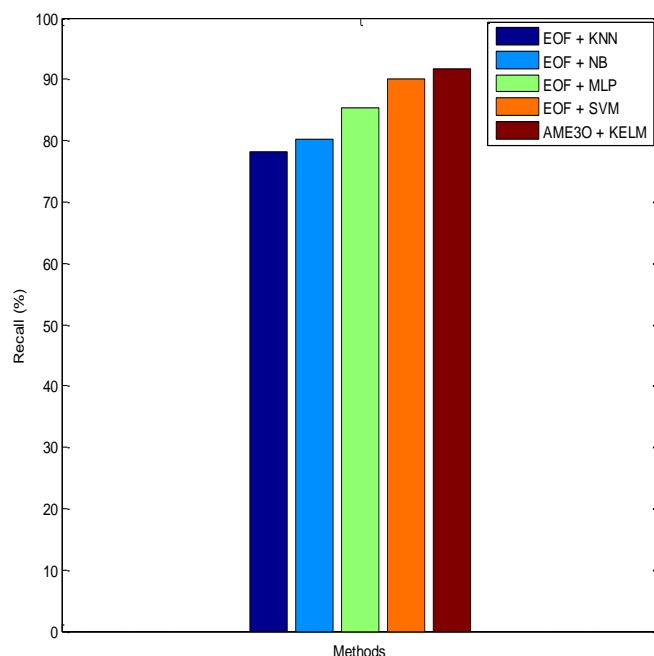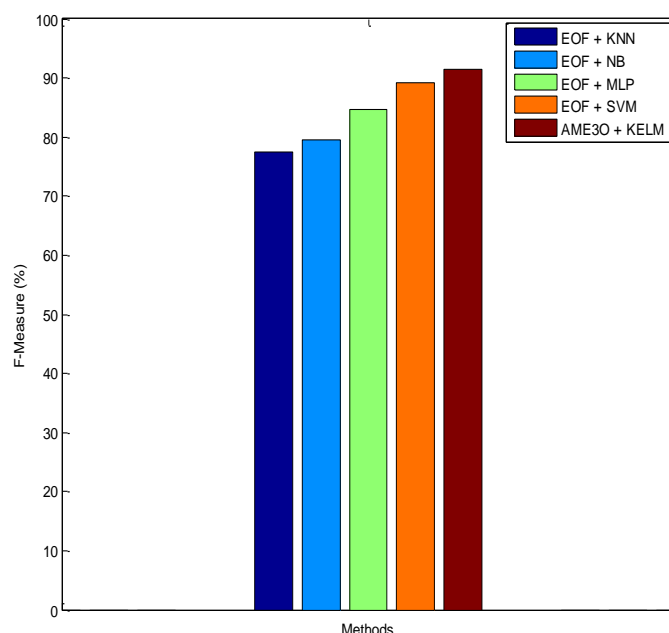
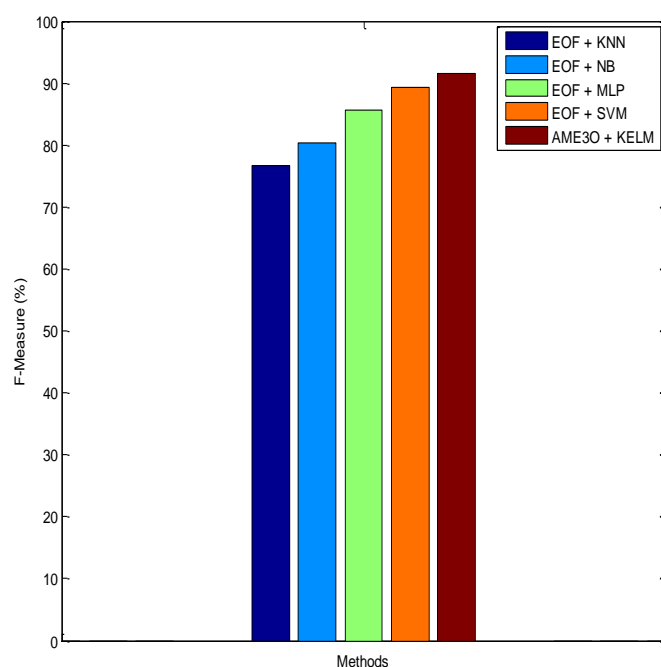**(a) Recall comparison of WDBC Dataset vs. Classifiers**

(b) Recall comparison of WOBC Dataset vs. Classifiers

FIGURE 7. RECALL ANALYSIS OF CLASSIFEIRS VS. BC DATASETS

Different classifiers such as EOF+KNN, EOF+NB, EOF+MLP, EOF+SVM and proposed AMEHO+KELM classifier with respect to recall results are shown in the figure 7. Figure 7(a) shows the recall results of different classifiers under WDBC dataset and WOBC dataset in figure 7(b). Figure 7(a) and figure 7(b), it concludes that the proposed AMEHO+KELM classifier gives higher recall results of 91.6416% and 91.8514% for WDBC and WOBC dataset respectively. The methods like EOF+KNN, EOF+NB, EOF+MLP, and EOF+SVM produces recall value of 78.2338%, 80.1946%, 85.4130%, and 90.0640% for WDBC dataset(See Table 3).
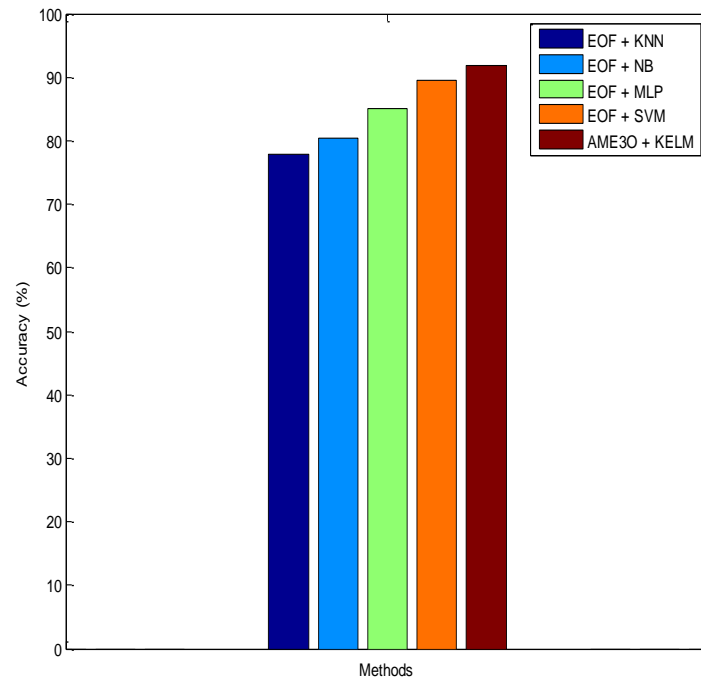
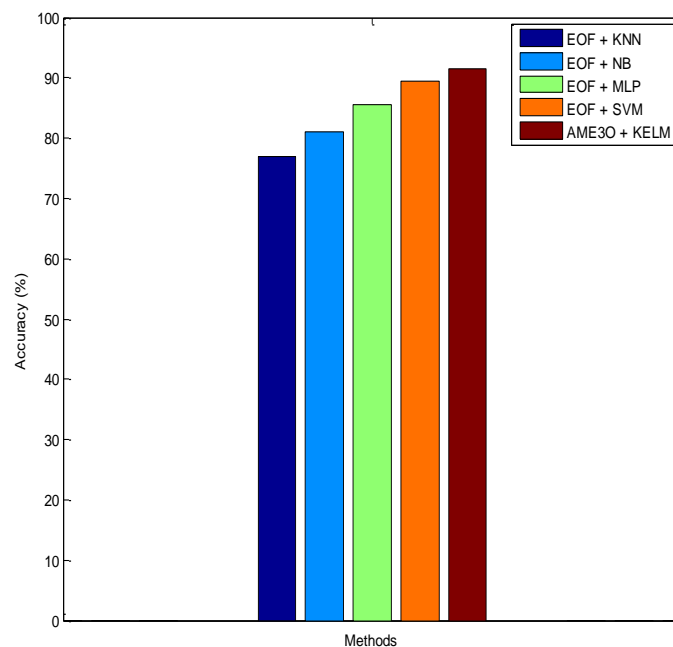**(a) F-measure comparison of WDBC dataset vs. Classifiers**

(b) F-measure comparison of WOBC dataset vs. Classifiers

FIGURE 8. F-MEASURE ANALYSIS COMPARISON OF CLASSIFEIRS VS. BC DATASETS

Figure 8 shows the f-measure comparison results of five different classifiers such as EOF+KNN, EOF+NB, EOF+MLP, EOF+SVM and proposed AMEHO+KELM classifier. Figure 8(a) and Figure 8(b) shows the f-measure results of different classifiers under WDBC dataset and WOBC dataset. Figure 8(a) and figure 8(b), it concludes that the proposed AMEHO+KELM classifier gives higher f-measure results of 91.4141% and 91.4807% for WDBC and WOBC dataset respectively. The classifiers such as EOF+KNN, EOF+NB, EOF+MLP, and EOF+SVM produces f-measure value of 77.4845%, 79.5973%, 84.6528%, and 89.2251% for WDBC dataset(See Table 3).
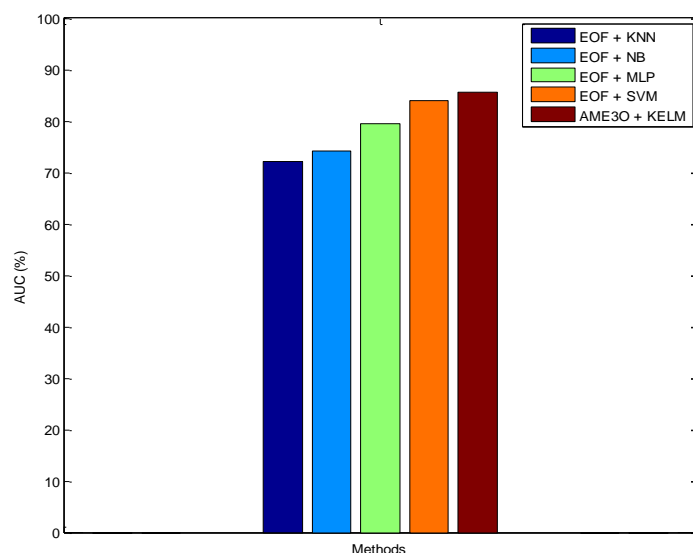
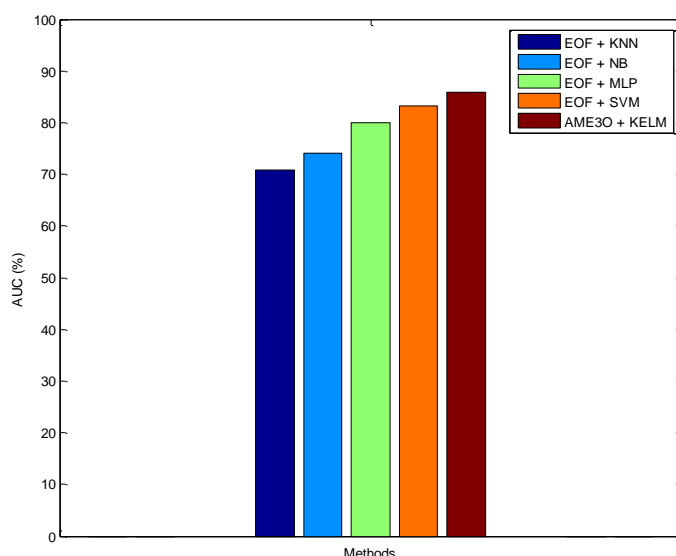**(a) Accuracy results comparison of WDBC dataset vs. Classifiers**

(b) Accuracy results comparison of WOBC dataset vs. Classifiers

FIGURE 9. ACCURACY ANALYSIS COMPARISON OF CLASSIFEIRS VS. BC DATASETS

Figure 9 shows the classifiers results of the accuracy metric. Figure 9(a) shows the accuracy results comparison of first dataset and figure 9(b) shows the accuracy results comparison of second dataset under classifiers. AMEHO+KELM classifier gives the accuracy results of 91.9156% and 91.5000% for WDBC and WOBC dataset respectively. The methods like EOF+KNN, EOF+NB, EOF+MLP, and EOF+SVM produces lesser accuracy value of 77.8559%, 80.3163%, 85.0615%, and 89.4552% for WDBC dataset (See Table 3).

(a) AUCresults comparison of WDBCDataset VS. Classifiers

(b) AUCresults comparison of WOBCDataset VS. Classifiers

FIGURE 10. AUC RESULTS COMPARISON OF CLASSIFEIRS VS. BC DATASETS

Figure 10(a)&(b) shows AUC metric results comparison of EOF+KNN, EOF+NB, EOF+MLP, EOF+SVM and proposed AMEHO+KELM classifier for both datasets. The proposed AMEHO+KELM classifier gives the AUC results of 85.6416% and 85.8514% for WDBC and WOBC dataset respectively. The methods like EOF+KNN, EOF+NB, EOF+MLP, and EOF+SVM produces lesser AUC value of 72.2338%,74.1946%, 79.4310%, and 84.0640% for WDBC dataset(See Table 3) .

6. CONCLUSION AND FUTURE WORK

Breast cancer is the greatest widespread cancer amongst women worldwide. Though, improved survival is due to the affected advances in the screening approaches, initial diagnosis, and discoveries in treatments. Accuracy of breast cancer analysis is essential to deliver suitable action or treatment.A novel feature selection and classifier is proposed to be executed in a Computer-Aided Diagnosis (CAD) system for breast cancer analysis.The proposed system has two phases. In the first phase, in order to remove irrelevant features, wrapper method by means of Adaptive Mutation Enhanced Elephant Herding Optimization (AMEHO)based on FS intended for selection of useful and important features. The examination and manipulation in AMEHOalgorithm are accomplished by the clan updating operator and the separating operator with mutation operator. Adaptive mutation parameter in AMEHOalgorithm is calculated via the fitness function.

In the second phase, Kernel Extreme Learning Machine (KELM) classifier is active to choose for dual different groups of focusses with or without breast cancer.In ELM, the original parameters of hidden layer essential not be modified and virtually all nonlinear piecewise constant functions can be used as the hidden nerve cell. This can be resolved by means of kernel function such as Gaussian kernel,hyperbolic tangent kernel, and wavelet kernel.KELM is classification of distinguished abnormality as benign or malignant. The proposed technique was estimated with BC dataset. Feature vectors are at first characterized according to their breast muscle kinds, and are then classified as benign, and malignant.These consequences obviously specify that by means of KELMclassifier is identical real for a CAD system and supportive for radiologists to make additional exact breast cancer diagnoses. In the future work, ensemble classification and ensemble feature selection are applied instead of single feature selection algorithm which will progress the accurateness of the BC system.

REFERENCES

1. Trimble, E.L., 2017. Breast cancer in sub-Saharan Africa. Journal of global oncology, 3(3), pp.187-188.

2. Wang, Q.X., Bai, Y., Lu, G.F. and Zhang, C.Y., 2017. Perceived health-related stigma among patients with breast cancer. Chinese Nursing Research, 4(4), pp.158-161.

3. Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A. and Jemal, A., 2018. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer journal for clinicians, 68(6), pp.394-424.

4. Hadgu, E., Seifu, D., Tigneh, W., Bokretsion, Y., Bekele, A., Abebe, M., Sollie, T., Merajver, S.D., Karlsson, C. and Karlsson, M.G., 2018. Breast cancer in Ethiopia: evidence for geographic difference in the distribution of molecular subtypes in Africa. BMC women's health, 18(1), pp.1-8.

5. Vanderpuye, V., Hammad, N., Martei, Y., Hopman, W.M., Fundytus, A., Sullivan, R., Seruga, B., Lopes, G., Sengar, M., Brundage, M.D. and Booth, C.M., 2019. Cancer care workforce in Africa: perspectives from a global survey. Infectious agents and cancer, 14(1), pp.1-8.

6. Chaurasia, V., Pal, S. and Tiwari, B.B., 2018. Prediction of benign and malignant breast cancer using data mining techniques. Journal of Algorithms & Computational Technology, 12(2), pp.119-126.

7. Diz, J., Marreiros, G. and Freitas, A., 2016. Applying data mining techniques to improve breast cancer diagnosis. Journal of medical systems, 40(9), pp.1-7.

8. Fogliatto, F.S., Anzanello, M.J., Soares, F. and Brust-Renck, P.G., 2019. Decision support for breast cancer detection: classification improvement through feature selection. Cancer Control, 26(1), p.1073274819876598.

9. Miao, J. and Niu, L., 2016. A survey on feature selection. Procedia Computer Science, 91, pp.919-926.

10. Nouira, K., Maalej, Z., Rejab, F.B., Ouerfelly, L. and Ferchichi, A., 2020, Analysis of breast cancer data: a comparative study on different feature selection techniques. In 2020 International Multi-Conference on:"Organization of Knowledge and Advanced Technologies"(OCTA) ,pp. 1-11.

11. Pirgazi, J., Alimoradi, M., Abharian, T.E. and Olyaee, M.H., 2019. An Efficient hybrid filter-wrapper metaheuristic-based gene selection method for high dimensional datasets. Scientific reports, 9(1), pp.1-15.

12. Ghosh, M., Adhikary, S., Ghosh, K.K., Sardar, A., Begum, S. and Sarkar, R., 2019. Genetic algorithm based cancerous gene identification from microarray data using ensemble of filter methods. Medical & biological engineering & computing, 57(1), pp.159-176.

13. Hassanien, A.E.; Emary, E. Swarm Intelligence: Principles, Advances, and Applications; CRC Press: Boca Raton, FL, USA, 2016.

14. Basir, M.A.; Ahmad, F. Comparison on Swarm Algorithms for Feature Selections Reductions. Int. J. Sci. Eng. Res. 2014, 5, 479–486.

15. Xue, B.; Zhang, M.; Browne, W.N.; Yao, X. A Survey on Evolutionary Computation Approaches to Feature Selection. IEEE Trans. Evol. Comput. 2016, 20, 606–626.

16. Wang, G.G., Deb, S. and Coelho, L.D.S. (2015b), "Elephant herding optimization", 3rd International Symposium on Computational and Business Intelligence (ISCBI), IEEE, Bali, pp. 1-5.

17. Wang, G.G., Deb, S., Gao, X.Z. and Coelho, L.D.S. (2016a), "A new metaheuristic optimisation algorithm motivated by elephant herding behaviour", International Journal of Bio-Inspired Computation, Vol. 8 No. 6, pp. 394-409.

18. Ismaeel, A.A., Elshaarawy, I.A., Houssein, E.H., Ismail, F.H. and Hassanien, A.E. (2019), "Enhanced elephant herding optimization for global optimization", IEEE Access, Vol. 7, pp. 34738-34752.

19. Ghosh M., S. Adhikary, K. K. Ghosh, A. Sardar, S. Begum, and R. Sarkar, ''Genetic algorithm based cancerous gene identification from microarray data using ensemble of filter methods,'' Med. Biol. Eng. Comput., vol. 57, no. 1, pp. 159–176, Jan. 2019

20. Jain, I., Jain, V.K. and Jain, R., 2018. Correlation feature selection based improved-binary particle swarm optimization for gene selection and cancer classification. Applied Soft Computing, 62, pp.203-215.

21. Jeyasingh, S. and Veluchamy, M., 2017. Modified bat algorithm for feature selection with the wisconsin diagnosis breast cancer (WDBC) dataset. Asian Pacific journal of cancer prevention: APJCP, 18(5), pp. 1257–1264.

22. Mafarja M. M. and S. Mirjalili, ''Hybrid whale optimization algorithm with simulated annealing for feature selection,'' Neurocomputing, vol. 260, pp. 302–312, Oct. 2017

23. Tubishat M., N. Idris, L. Shuib, M. A. M. Abushariah, and S. Mirjalili, ''Improved salp swarm algorithm based on opposition based learning and novel local search algorithm for feature selection,'' Expert Syst. Appl., vol. 145, May 2020, Art. no. 113122,

24. Al-Tashi Q., S. J. Abdul Kadir, H. M. Rais, S. Mirjalili, and H. Alhussian, ''Binary optimization using hybrid grey wolf optimization for feature selection,'' IEEE Access, vol. 7, pp. 39496–39508, 2019, doi: 10.1109/access.2019.2906757.

25. Emary E., H. M. Zawbaa, and A. E. Hassanien, ''Binary grey wolf optimization approaches for feature selection,'' Neurocomputing, vol. 172, pp. 371–381, Jan. 2016,

26. Ghosh, K.K., Ahmed, S., Singh, P.K., Geem, Z.W. and Sarkar, R., 2020. Improved binary sailfish optimizer based on adaptive β-hill climbing for feature selection. IEEE Access, 8, pp.83548-83560.

27. Hegazy, A.E., Makhlouf, M.A. and El-Tawel, G.S., 2020. Improved salp swarm algorithm for feature selection. Journal of King Saud University-Computer and Information Sciences, 32(3), pp.335-344.

28. Hegazy, A.E., Makhlouf, M.A. and El-Tawel, G.S., 2019. Feature selection using chaotic salp swarm algorithm for data classification. Arabian Journal for Science and Engineering, 44(4), pp.3801-3816.

29. ElShaarawy, I.A., Houssein, E.H., Ismail, F.H. and Hassanien, A.E., 2019. An exploration-enhanced elephant herding optimization. Engineering Computations, pp.1-18.

30. Li B., Y. Li, and X. Rong, "The extreme learning machine learning algorithm with tunable activation function," Neural Computing and Applications, vol. 22, pp. 531–539, 2013.

31. Huang G.-B., H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," IEEE Transactions on Systems, Man, and Cybernetics B, vol. 42, no. 2, pp. 513–529, 2012.

32. Huang W., N. Li, Z. Lin et al., "Liver tumor detection and segmentation using kernel-based extreme learning machine," in Proceedings of the 35th Annual International Conference of the IEEE EMBS, pp. 3662–3665, Osaka, Japan, 2013.

33. Ding S.F., Y. A. Zhang, X. Z. Xu, andL. N. Bao, "Anovel extreme learning machine based on hybrid kernel function," Journal of Computers, vol. 8, no. 8, pp. 2110–2117, 2013.

34. Li, B., Rong, X. and Li, Y., 2014. An improved kernel based extreme learning machine for robot execution failures. The Scientific World Journal, vol.2014, Volume 2014, no. 906546, pp.1-7.