How to Cite:
Priya, R. S. P., & Vadivu, P. S. (2022). Bio-inspired ensemble feature selection (biefs) and kernel extreme learning machine classifier for breast cancer diagnosis. *International Journal of Health Sciences*, *6*(S5), 1404–1429. https://doi.org/10.53730/ijhs.v6nS5.8976

Bio-inspired ensemble feature selection (biefs) and kernel extreme learning machine classifier for breast cancer diagnosis

R. S. Padma Priya
Assistant Professor, Department of Computer Technology, Dr.N.G.P.Arts and Science College, Tamilnadu, India
Corresponding author email: priypadma@gmail.com

P. Senthil Vadivu
Professor & Head, Department of Computer Applications (UG), Hindusthan College of Arts and Science, Tamilnadu, India
Email: sowjupavan@gmail.com

***Abstract*---**Breast cancer is a major disease identified in women, affecting 2.1 million women every year, and is the reason for most cancer-related mortality in women, as per the World Health Organization (WHO). For cancer researchers, accurately forecasting the life expectancy of breast cancer patients is a serious challenge. Machine Learning (ML) has acknowledged much interest in the hope of providing correct results, but due to irrelevant features, its modelling methodologies and prediction performance are still a difficulty. To solve this issue, Feature Selection (FS) was also done to verify whether comparable accuracy can be achieved even with lesser number of features or not. Bio-Inspired Ensemble Feature Selection (BIEFS) algorithm is introduced aimed at selecting a subset of features that increase the prediction performance of subsequent classification models while also simplifying their interpretability. BIEFS algorithm uses three feature selection methods such as Adaptive Mutation Enhanced Elephant Herding Optimization (AMEHO), Adaptive Mutation Butterfly Optimization Algorithm (AMBOA), and Adaptive Salp Swarm Algorithm (ASSA) and integrates their normalized outputs for getting quantitative ensemble importance. BIEFS algorithm depends upon the aggregation of multiple FS techniques by Pearson Correlation Coefficient (PCC).This BIEFS algorithm can improve the accuracy of analysis (benign and malignant). The Kernel Extreme Learning Machine (KELM) classifier is then used to select between two groups of cases: those with breast cancer and those without. Wisconsin Diagnosis Breast Cancer (WDBC) taken from University of California, Irvine (UCI) repository is evaluated on existing classifiers. Evaluation metrics like Precision, Recall, F-measure, Accuracy, and

Area Under Curve (AUC) is used to connect the proposed system to the existing system. All experiments are executed within a simulation environment and conducted in MATLAB tool.

***Keywords*---**Breast cancer, Data mining, Bio-Inspired Ensemble, Feature Selection, Kernel Extreme Learning Machine (KELM) classifier.

Introduction

As per the reports of the World Health Organization, breast cancer is a common form of cancer in women. It affects 2.1 million women per year, and is a leading cause of cancer-related mortality in women [1]. Breast cancer is a commonly reported cancer in India. Breast cancer amounts to 25% to 32% of cancer cases among female in big cities such as Mumbai, Delhi, and Bangalore. As it is increasingly more visible in younger adults, this disease has become even more severe. About 50% of such cases fall in the 25 to 50 age group [2]. The figures are alarming and steadily increasing [3]. As per reports of the Indian Council for Medical Research in 2016, the total count of fresh cancer cases was ab out $14.5* 10^5$ and it is expected to rise up to $17.3 * 10^5$ in 2020. As the rate of cases of breast cancer rises in India, so does the fear of cancer. Though it becomes unable to prevent breast cancer, it can at least minimize mortality rates by being made aware of and selecting the appropriate treatment at the appropriate time. For improving breast cancer prognosis and lifespan, early diagnosis is crucial which may be efficiently performed using Machine Learning (ML) and data mining techniques [4]. Use of Machine Learning algorithms such as classification techniques can be used to create a model to detect breast cancer as malignant or benign [5].

To manage the dataset and enhance the classification accuracy, several data mining techniques like class balancing, re-sampling, etc. can be utilized. A dataset includes several attributes/features. Not necessarily, all the features are required to extract relevant data from those datasets. Some features can be repetitive or useless, affecting the effectiveness of the model. Hence, to minimize the size of the original datasets while retaining the performance accuracy is the fundamental goal of feature reduction problem. Applying feature selection techniques are used to extract the most significant features that play a major part in the accuracy of the classification model and to decrease the computation time. As a result, to solve this issue, different techniques were presented. Exhaustive search, greedy search, random search etc. are some techniques that are employed in feature selection problems for getting the right subset. Majority of the approaches are plagued by premature convergence, excessive complexity, and high computing cost. As a result, metaheuristic algorithms are receiving a lot of attention in order to cope with these kinds of situations [7, 8]. They are considered as efficient and reliable approaches for selecting the best subset of characteristics and also preserving the accuracy for any model. During the past three decades, many metaheuristic algorithms are developed to solve different forms of optimization problems. Depending on the behavior of these metaheuristic algorithms, they are classified into four different groups; evolution-based, swarm

intelligence-based, physics-based, and human-related algorithms [9]. From that swarm intelligence-based algorithms are inspired by the social behaviors of insects, animals, fishes or birds etc. This research work also motivated from swarm-based methods for Feature Selection (FS). However single swarm-based algorithms will not give improved Feature Selection (FS).

Ensemble Feature Selection is a solution for improving classifier outputs, by integrating the result of different feature selectors, the performance can be enhanced and releases the user from selecting an individual method [10]. Ensembles for the feature selection are characterized as homogeneous (that use same base feature selector) or heterogeneous (that use different feature selectors). The homogeneous technique employs same feature selection strategy with different subsets for training data. These data subsets can be shared across multiple nodes or partitions, resulting in a reduction in temporal requirements. The partition size is actually a design component in this method. Data variation ensembles are another term for such methods. Few examples for homogeneous methods, with the purpose to manage large scale scenarios, are found in recent work [11,12], the latter with the extra aim to cope with imbalanced data sets. Several distinct feature selection approaches are used in the heterogeneous approach, but they are all applied to the same training data. As previously mentioned, the count of several feature selectors to be applied in this scheme can be considered as the design parameter. These techniques are sometimes known as function variation techniques. Heterogeneous feature selection ensembles are much prevalent than homogeneous, as shown by various instances are seen in recent work [13, 14]. In both techniques, based on feature selection method applied, a feature subset or else a feature ranking can be obtained, with the latter requiring an extra threshold process. Like in every ensemble, the base selector outputs are to be integrated via several aggregation methods to get the end result. In the work, Bio-Inspired Ensemble Feature Selection (BIEFS) algorithm is introduced for selecting most important feature subset to improve classification accuracy methods. BIEFS algorithm is performed based on combining three feature selection methods such as Adaptive Mutation Enhanced Elephant Herding Optimization (AMEHO), Adaptive Mutation Butterfly Optimization Algorithm (AMBOA), and Adaptive Salp Swarm Algorithm (ASSA) and combines their normalized outputs to quantitative ensemble significance. The BIEFS algorithm deals with multiple FS aggregation methods with Pearson Correlation Coefficient (PCC) in breast cancer datasets. This FS algorithm is applied to increase analytical accuracy (benign and malignant). The Kernel Extreme Learning Machine (KELM) classifier is then used for selecting between two groups of cases: those with and those without breast cancer.

Literature Review

Hengpraprohm et al.[15] proposed an Entropy and Signal to Noise Ratio(EnSNR) based ensemble filter feature selection approach for breast cancer data classification. The Microarray dataset applied in the experimental research has 50,739 features (genes) that is included for all 32 patients. The fundamental idea of the EnSNR approach is to merge informative features derived from two independent sets of feature evaluation criteria. Features that appear in both sets of evaluation findings are included in the EnSNR subset. The Entropy and SNR

feature subsets are generated using EnSNR evaluation functions.SNR is an useful function for quantifying feature discriminative power, while entropy refers to the amount of uncertainty in the results of a random experiment. For the EnSNR method, EnSNR functions give various benefits. The number of features for the SNR subset, for instance, is not user-defined; however, the EnSNR subset is automatically generated, and the EnSNR function operates independently of the type of classification algorithm.

Elgin Christo et al [16] proposed a methodology for medical diagnosis that applies bioinspired algorithms for feature selection and gradient descendant backpropagation neural network for classification. Missing values are handled by hot deck imputation, and data is transformed using min-max normalisation. Wrapper approach employed bioinspired algorithms, like, Differential Evolution (DE), Lion Optimization (LO), and Glowworm Swarm Optimization (GSO) having accuracy of AdaBoost Support Vector Machine (AdaBoostSVM) classifier because fitness function is applied for feature selection. All bioinspired algorithms choose each feature subset producing three feature subsets. To choose the best features from the three feature subsets, correlation-based ensemble feature selection is applied. A gradient descendent backpropagation neural network is trained using the optimum features determined using correlation-based ensemble feature selection. The classifier performance was trained and tested with ten-fold cross-validation technique. The classification accuracy is tested using the Hepatitis dataset and the Wisconsin Diagnostic Breast Cancer (WDBC) dataset from the University of California Irvine (UCI) Machine Learning repository.

Prince et al[17] purposed an ensemble method that selects major significant features to diagnose cancer as well to classify cancer pattern in accordance with the best necessary features. Set union operation was used to assess the benefits of Principal Component Analysis (PCA), Pearson Correlation Coefficient (PCC), and Chi Square (Chi2) for feature extraction. Using hybrid model of feature extraction, and classification, Naive Bayes (NBs), K Nearest Neighbors (KNNs), Decision Tree (J48, ID3) classifiers are applied for model classification. To get best expected outputs, individual classifier results are selected by majority voting.

Ji et al[18] developed an Improved Binary Particle Swarm Optimization (IBPSO) algorithm for solving formulated problem in classification. To boost the output of conventional PSO and make it appropriate for binary feature selection problems, the IBPSO algorithm incorporates a local search factor using Lévy flight, a global search factor on the basis of weighting inertia coefficient, a population diversity improvement factor on the basis of mutation mechanism, as well as a binary mechanism. Evaluations on the basis of 16 classic datasets were chosen for assessing the efficacy of new IBPSO algorithm, and its findings show that IBPSO outperforms previous comparative algorithms.

Kewat et al[19] developed a wrapper-based strategy with Particle Swarm Optimization (PSO), Genetic Algorithm(GA), as well as Greedy Hill Climbing(GHC) as an attribute oriented approach for medical dataset. The effectiveness of PSO search method is assessed using GA and GHC algorithm. This employed proposed technique on the three medical datasets. Results indicated that PSO and GA

techniques do increase classification performance and outclassed the GHC feature selection technique.

Abdullah Farid et al[20] suggested Composite Hybrid Feature Selection Learning-Based Optimization of Genetic Algorithm (CHFS-BOGA) for breast cancer dataset. The merits of three filter feature selection strategies are combined in this hybrid feature selection strategy, which uses Optimize Genetic Algorithm (OGA) to choose the important features and increase classification process efficiency and scalability. OGA algorithm, C4.5 decision tree classifier with classification accuracy is taken as fitness function rather than probability and random selection. It can be applied to Wisconsin dataset of UCI Machine Learning (ML) repository having a total of 569 rows and 32 columns. Outputs suggested the proposed CHFSBOGA-SVM system to be capable of precisely classifying breast cancer type as either malignant or benign. The outputs suggest the proposed CHFS-BOGA method to work superior than other single filter methods as well as PCA to get optimum feature selection.

Sayed et al. [21] suggested the Chaotic Salp Swarm Algorithm (CSSA), which was tested on 14 unimodal as well as multimodal benchmark optimization problems as well as 20 benchmark datasets. To improve the convergence rate and resultant precision, ten distinct chaotic maps are used. Simulation outputs reveal that t he proposed CSSA is an effective algorithm. The findings also show the ability of CSSA to determine an ideal feature subset that increases classifier accuracy and also reducing the total features used. Furthermore, outputs revealed the logistic chaotic map to be the best of ten maps tested, allowing the original SSA to perform remarkably better.

Allam et al [22] introduced a wrapper-based feature selection approach termed binary teaching learning based optimization (FS-BTLBO) that requires a typical control parameters such as population size and a count of generations to extract optimal features subset from any dataset. Different classifiers are applied as objective function for computing individuals' fitness to evaluate the effectiveness of the suggested model. Outputs show the FS-BTLBO to provide better accuracy using minimum set of features for the Wisconsin diagnosis breast cancer (WDBC) data set for classifying malignant as well as benign tumours.

Kiziloz et al [23] proposed a set of novel Teaching Learning Based Optimization (TLBO) algorithms combined with supervised classifiers methods to get the results of Feature Subset Selection (FSS) in Binary Classification Problems (FSS-BCP). TLBO is proposed as an FSS mechanism that makes use of an algorithm-specific parameter less idea that eliminates the need to adjust any parameters during optimization. Many classical metaheuristics like Genetic and Particle Swarm Optimization algorithms require extra efforts to tune the parameters such as crossover ratio, mutation ratio, velocity of particle, inertia weight, etc., that can result in negative impact performance-wise. Extensive studies are conducted on familiar machine learning datasets from UCI Repository.

Punitha et al[24] proposed a Intelligent Artificial Bee Colony and Enhanced Monarchy Butterfly Optimization Technique (IABC-EMBOT) to effectively diagnose breast cancer. This depends on two major improvements that, i) concentrates on

the alteration of Monarchy Butterfly Optimization (MBO) resulting in enhancement of the degree of assessment on the basis of utilisation rate in the search space, and ii) focuses to remove all constraints of the ABC scheme through increasing search diversification possibilities by remarkable updates made possible by the dynamic and adaptive butterfly operator which enhances the search on the whole. The proposed IABC-EMBOT approach has been shown to enhance average classification accuracy when used with the Wisconsin data set.

Emary et al[25] suggested a feature selection method using Binary Grey Wolf Optimization(bGWO). The two methodologies of bGWO are engaged in feature selection domain to find feature subset to maximize classifier accuracy and also to minimize the set of features selected. The suggested binary versions are matched with two widely applied optimizers for this domain, i.e. the Particle Swarm Optimizer (PSO) and Genetic Algorithms (GAs). A set of evaluation metrics is used to examine and analyzed the results of various methods over 18 distinct datasets of UCI repository. The suggested bGWO is proven to be capable of searching the feature space for optimal feature combinations irrespective of initialization or use of stochastic operators.

Zawabaa et al[26] suggested a hybrid bioinspired heuristic approach for feature selection that combined the Ant-Lion Optimization (ALO) and Grey Wolf Optimization (GWO) algorithms. By eliminating local optima and quickening search operation, given hybrid algorithm was able to achieve global optimization. The hybrid algorithm alone outperforms the ALO and GWO, which was tested with 18 datasets of UCI machine learning repository out of which Cleveland Heart Disease (CHD) and Wisconsin Diagnostic Breast Cancer (WDBC) dataset are of clinical domain. The proposed ALO-GWO algorithm presented the examination of search space and utilization of optimal solution in a more balanced manner. Researchers proposed a parallel distribution mode for improving the classifier's convergence time. Anter and Ali [27] developed a hybrid chaos theory Crow Search Optimization Algorithm integrated with and Fuzzy C-Means Algorithm (CFCSA) to select features in medical diagnosis. The Crow Search Algorithm (CSA) in the given CFCSA framework used global optimization to prevent the sensitivity of local optimization. The Fuzzy C-Means (FCM) algorithm's objective function is employed as a cost function for the Chaotic Crow Search Optimization Algorithm (CCSOA). The experiment results reveal that the proposed CFCSA technique outperforms competing systems in feature selection on medical diagnosis data sets.

Materials and Methods

In this paper, Bio-Inspired Ensemble Feature Selection (BIEFS) algorithm is introduced for selecting the feature subset which enhances classifier performance for classifiers and to minimize their interpretability. BIEFS algorithm is performed based on integrating three algorithms such as Adaptive Mutation Enhanced Elephant Herding Optimization (AMEHO), Adaptive Mutation Butterfly Optimization Algorithm (AMBOA), and Adaptive Salp Swarm Algorithm (ASSA). The results of these methods are combined and normalized outputs via Pearson Correlation Coefficient (PCC). Combination of three types of FS algorithms offers

the best performance for different types of Breast Cancer datasets. After that, the Kernel Extreme Learning Machine (KELM) classifier is used in selecting between two distinct groups of cases with or without breast cancer. The overall flow process of proposed BIEFS and KELM classifier is shown in the figure 1.

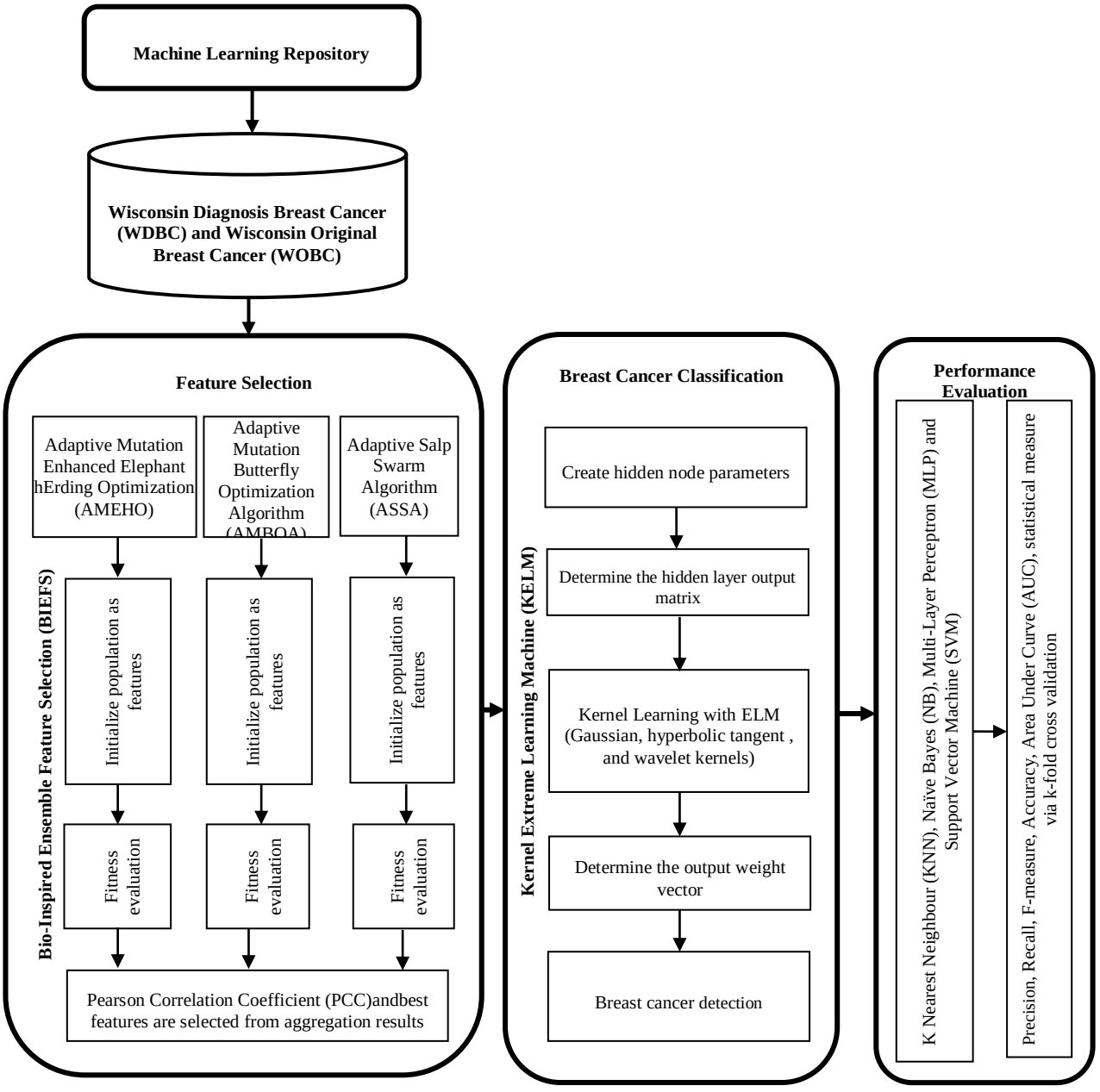


Figure 1. Overall flow of bio-inspired ensemble feature selection (BIEFS) and kernel extreme learning machine (KELM) classifier for breast cancer diagnosis model

Dataset Collection

Data collection necessitates collecting information. The data for this study made out of the Machine Learning (ML) repository in the University of California, Irvine (UCI). It's an open-source project. The source compiles commonly available BC datasets like WDBC and WOBC.

Bio-Inspired Ensemble Feature Selection (BIEFS)

Feature Selection (FS) refers to the technique for reducing the quantity of features of 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 which do not relate to the classification process. The aim of FS is to raise classification accuracy. Ensemble learning can be an efficient process to do FS. The main goal is to improve learning outcomes by integrating distinct FS models. The main aim of combined feature selection is to achieve many optimal features. BIEFS algorithm, firstly different feature selection methods such as Adaptive Mutation Enhanced Elephant Herding Optimization (AMEHO), Adaptive Mutation Butterfly Optimization Algorithm (AMBOA), and Adaptive Salp Swarm Algorithm (ASSA) are introduced for achieving candidate sets of several optimal feature subsets. After that, in relation to different rules, the learning outcomes of numerous optimal feature subset candidate sets are then merged to generate the optimal feature subsets via Pearson Correlation Coefficient (PCC).

Adaptive Mutation Enhanced Elephant Herding Optimization (AMEHO)

The Adaptive Mutation Enhanced Elephant Herding Optimization (AMEHO) algorithm, which is based on the FS algorithm, is proposed for choosing breast cancer features [28,29]. The clan informing operator and the separating operator with mutation operator are used in AMEHO to accomplish discovery and exploitation. In every assembly, the suitable individual elephant (Accuracy of the classifier for BC recognition (with best fitness)) in a clan $c_i$is selected as the matriarch (m_i) at time t as shown in equation (1),

$$m_i^t = F(x) \tag{1}$$

Where $x_i$is the group of single elephants in clan i(number of examples in the BC dataset).

Clan updating operator: In clan i, the distinct elephant j(features of the BC dataset) has previous position $x_{i,j}^t$. Its novel position $x_{i,j}^{t+1}$ is collapsed by the clan matriarch m_i^t as per 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 locating the significance of clan matriarch on the elephant (features of BC) at new position,$\beta \in [0,1]$are considered as the parameters the identifies the elephant's similarity to travel focussing the

clan center, $\gamma \in [0,1]$are considered as the parameters the identifies the elephant's similarityto roam around (walk in a random direction),

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

$x_{amut}$is one of the most sensitive parameters is the mutation rate parameter.

Separating operator: Equation (4) of an individual elephant can be used to measure the separating operator (feature position),

$$x_{i,worst}^{t+1} = x_{min} + (x_{max} - x_{min}) * rand * x_{amut} \quad (4)$$

Where x_{min} and x_{max} are considered to be the lower and upper limits of single elephant position (feature position) and $x_{i,worst}^{t+1}$ is the single 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)). 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 can be zero.

Adaptive Mutation Butterfly Optimization Algorithm (AMBOA)

Adaptive Mutation Butterfly Optimization Algorithm (AMBOA) is introduced to select the best features from the given breast cancer dataset. AMBOA is new nature-inspired algorithm that mimics food search (higher accuracy with selected features) and behavior of matting in butterflies, to overcome classification issues in BC disease diagnosis. The proposed AMBOA method is focused on foraging technique of butterflies, which uses the smelling sense for optimal selection of features in order to find out the nectar partner's location [30, 31]. Scientific observations made butterflies with accurate sense of identifying the fragrance (classification accuracy). A butterfly will emit fragrance with varying intensities that are related to its fitness (classification accuracy), i.e., as a butterfly keeps moving from one place to another, its fitness can change. In proposed AMBOA method, the processing modality and sensing depends on 3 significant terms viz. sensory modality (c), stimulus intensity (I) and power exponent (a) for optimal selection of features [31]. In proposed AMBOA method, I is correlated with the fitness (accuracy) for the selection of features. With the help of these, in proposed AMBOA method, the formulation of fragrance as a function of the stimulus's physical intensity is shown in equation (5),

$$bf = cI^a \quad (5)$$

Where bf is the perceived magnitude (strength) of the fragrance,c is the sensory modality which is generated via classification accuracy, I is the intensity of stimulus, and a is the power exponent that varies with modality. As a result, a and c are in the range$[0,1]$. If a = 0, it implies that the fragrance released by a butterfly could not be detected by all other butterflies. The parameter $a$maintains the algorithm's behaviour. Yet other parameter which is significant is c which is considered to be an important parameter to find the speed of convergence and the technique of AMBOA. To illustrate the aforementioned

discussions in the context of a search algorithm, the following characteristics for butterflies are framed:

1. All butterflies emit some fragrance which attracts other butterflies (features from BC dataset).
2. Every butterfly will travel in a random direction, focussing the butterfly that emits high fragrance.
3. The landscape of the objective function attracts or identifies the stimulus intensity for a butterfly.

There are 3 phases in AMBOA method such as (1) Initialization phase, (2) Iteration phase, and (3) Final phase. In every execution of AMBOA method, first the execution of initialization phase is done, then locating the optimal features from breast cancer dataset are carried out iteratively and in the final phase, it is terminated. At last, the optimal selection solution from breast cancer dataset is found. In the initialization phase, classification accuracy is computed in AMBOA method and its solution space of optimal features from breast cancer dataset. The parameters are used along with its value in AMBOA method are also assigned. The locations of butterflies (features) are generated on a random basis, in the search space of feature selection process, with their fragrance and fitness values [33,34]. After finishing initialization phase then algorithm starts the iteration phase. Every iteration consists of all butterflies in feature selection space shift to new location and then their classification accuracy data are measured. In the technique, the first fitness values are computed of all other butterflies on various locations in the solution space. The fragrance emitted by these butterflies will generate the fragrance at their feature positions utilizing equation (5). In search phase, the butterfly travels a step targeting the fittest solution (g∗)(optimal features) that is shown in equation (6),

$$x_i^{t+1} = x_i^t + (r^2 \times g^* - x_i^t) \times f_i * x_{amut} \tag{6}$$

where x_i^t is the vector's solution x_i for ith butterfly in the total number of iterations t. Here, g^*shows the current best selected feature solution which is identified when all solutions are considered in the iteration which is current. Fragrance of ith butterfly is shown as f_i and $r \in [0,1]$ is a random number Local search phase easily shown as equation (7),

$$x_i^{t+1} = x_i^t + \left(r^2 \times x_j^t - x_k^t\right) \times f_i * x_{amut} \tag{7}$$

x_{amut} is of the most sensitive parameters is the mutation rate parameter. x_j^t and x_k^t are jth and kth butterflies from the feature selection solution space. If x_j^t and x_k^t belongs to the same swarm and $r \in [0,1]$ is a random number then equation (6) becomes a local random walk. For the best selection of attributes from the dataset, butterflies can search for food and a mating partner at local as well as global scale. Switch probability p is used in AMBOA to exchange between the general searches high search locally. Until the criteria are paused, the continuation of iteration part takes place. Once the iteration phase is finished, the algorithm returns the optimal result with the highest fitness. In the equation (9, 10), x_{amut} is also added to AMBOA to choose the best number of features in the

dataset. The overall process in the presented AMBOA is shown in the algorithm 2. In the algorithm 2, initial population are generated via total features that exist in the dataset (Step 1), and then stimulus intensity In_i at x_i (Step 2) is computed based on the sensor modality$c$, power exponent a (Step 3). These factors are generated via the classification accuracy. Then it starts with stopping criteria (Step 4), for each butterfly in the dataset the fragrance value is computed (Step 6). After that find the best feature in the population(Step 8) and generated random number r (Step 10). If $r < p$ then locate good butterfly by equation (9), or move randomly by equation (7). Then update a value (Step 17), and evaluate individuals according to their new position (Step 18). Finally end the process via end while (Step 19).

Algorithm 1: adaptive mutation butterfly optimization algorithm (AMBOA)
Input: Breast Cancer dataset
Objective Function: Classifier accuracy, $f(x), x = (x_1, x_2, \dots., x_{dim}) dim = no. of dimesnions$
Output: Selection of optimal features

1. Develop the population initially consisting$n$ butterflies $x_i = (i = 1,2, \dots, n)$ via number of features in the dataset
2. Stimulus Intensity I_i at x_i is found by classification accuracy $f(x_i)$
3. Define sensor modality c, power exponent a and switch probability p
4. While stopping criteria not met do
5. For each butterfly bfin population do
6. Measure fragrance for bfusing equation (5)
7. End for
8. Identify the optimal butterfly
9. For each butterfly bf in population do
10. Produce the random number r
11. If $r < p$ then
12. Move towards the best butterfly (optimal features) by equation (6)
13. Else
14. Random movement is done using the equation (7)
15. End if
16. End for
17. The value of a to be updated
18. Evaluate individuals(features) according to their new position
19. End while
20. Output the identified best solution found
21.

Adaptive Salp Swarm Algorithm (ASSA)

The ASSA is nature-inspired optimizer and it is used to construct a population-based optimizer by imitating the behaviour of salps in nature for optimal selection of features from the dataset [32]. The group of salps X contains agents with$N$number with d-dimensions. This can be shown by$N \times d$-dimensional matrix as explained in equation (8),

$$X_i = \lfloor x_1^1 x_2^1 \cdots x_d^1 x_1^2 x_2^2 \cdots x_d^2 \ \vdots \vdots \cdots \vdots \ x_1^N x_2^N \dots x_d^N \rfloor \quad (8)$$

In SSA, the leader's position is measured by equation (9),

$$X_j^1 = \{F_j + C_1\big((ub_j - lb_j)\big)C_2 + lb_j) \quad C_3 \geq 0.5 F_j - C_1\big((ub_j - lb_j)\big)C_2 + lb_j) \quad C_3 < 0.5 \quad (9)$$

whereX_j^1denotes the leaders location and $F_j$reveals the feature position vector of food source in the jth dimension,ub_jshows the superior limit of jth feature dimension, and lb_jrepresents the lesser limit of jth dimension, $C_2$and $C_3$are random values within [0,1] is the significant parameter of the algorithm shown in equation (10),

$$C_1 = 2e^{-(T_{max})^2} \tag{10}$$

where $t$is the iteration, while T_{max}shows the total number of iterations. By maximizing the count of iterations, this parameter reduces. The location of followers is adjusted by equation (11),

$$X_j^i = \frac{X_j^i + X_j^{i-1}}{2} \tag{11}$$

Where $i \geq 2$ and x_j^i is the location of the i^{th} follower salp at the jth dimension. ASSA algorithm, the parameters are randomly initialized within a particular range as per the feature selection process. It is represented by equation (12),

$$x_{ij} = x_{min.j} + U(0,1) \times \left(x_{min.j} - x_{max.j}\right) \tag{12}$$

where $i$lies in the range [1, 2, 3.... n], $j$as [1, 2, 3.... D], $x_{ij}$is i^{th}feature selection solution for thej^{th}dimension, $x_{min.j}$and $x_{max.j}$ are lower and upper limits of the features in the dataset and $U(0,1)$is a even number which is random that lies within the limit [0,1].

Position updating: The algorithm modifies the operations like exploration and exploitation using the SSA algorithm. Exploration and exploitation are the types activities that determine how effective the technique on functionality basis. The technique having the content which is in balanced form that consists of both the operations has the ability to identify the optimal outputs in an effective way. It's accomplished in this current work through the usage of generation's division. The idea of division of generations is an innovative one and does not work more effectively in this ideology. This aids in the proper exploitation and exploration of the search space. In SSA, the commonly used equations are quite useful in exploitation but not so much in exploration. Hence with the help of iteration division method, dividing into two equal halves assists in algorithm, and the equations which are new are used in first half and the algorithm is divided into second half at the time of iterations. The solution space which is being used in the first half utilizes all the equations which are taken from the algorithms like (Glow Warm Optimization (GSO) [33] and Cuckoo Search (CS) [34]).These algorithms are more effective in performing the exploration operation, thereby the equation is modified depending on them will return optimal outputs. The new modified equations are generalized by equation (13, 14,15) ,

$$x_1 = x_i - A_1(C_1 x_{new} - x_i^t); x_2 = x_i - A_2(C_2 x_{new} - x_i^t)\ x_3 = x_i - A_3(C_3 x_{new} - x_i^t) \tag{13}$$

$$x_{new}^{t+1} = \frac{x_1 + x_2 + x_3}{3} \tag{14}$$

$$x_{new}^{t+1} = x_{new}^t + \alpha \times L(\lambda)(x_{best} - x_{new}^t) \tag{15}$$

where $x_{new}$is the new feature solution which is produced from the randomized population among the search space, for any problem with d dimensional and $A_1, A_2, A_3$and $C_1, C_2, C_3$are generated from $A = 2ar_1 - a$and$C_2 . r_2$and a here is minimized linearly within the limit [0, 2]. The α and L (λ) are distributed evenly and Lévy distributed random numbers, $r_1$and r_2 are random numbers are evenly distributed within the limit [0, 1]. Using equation (16), the Lévy flight-based step size is produced,

$$L\ (\lambda) \sim \frac{\lambda\Gamma(\lambda)sin(\frac{\pi\lambda}{2})}{\pi}\frac{1}{s^{1+\lambda}}(s \gg s_0 \gg 0) \tag{16}$$

where$s = \frac{U}{|V|1/\lambda}U \sim N(0, \sigma^2)$,$V\ \sim N(0\ ,1)$ and$\sigma^2 = \left\{\frac{\Gamma(1+\lambda)}{\frac{\lambda\Gamma(1+\lambda)}{2}} . \frac{sinsin\left(\frac{\pi\lambda}{2}\right)}{2^{\frac{\lambda-1}{2}}}\right\}$,also $\Gamma(\lambda)$is

gamma function and the value of λ equals to 1.5. N is derived from a typical Gaussian distribution with a mean of 0 and a variance of 2. Both GWO and CS processes are used in conjunction together in equations (14) and (15). These equations can only be utilized for the very first part of the cycles. The generic equations of SSA are utilised during the second half of the iterations to identify the best features from the dataset. This is due to the fact that SSA's exploitation abilities far outweigh its exploration capabilities.

Selection operation: The phase for selection for the presented technique is given here. In this case, greedy selection is executed to see if the newly produced outputs can outperform the best previous solution. For a quite common reduction methods having $F(x_i^t)$fitness for thex_i^tsolution, the selection procedure is given by equation (17),

$$x_{new}^{t+1} = \{x_{new} x_i^t if_{otherwise}^{F(x_{new}) < F(x_i^t)} \tag{17}$$

Controlling parameters: This step is described by the controlling parameterc_1. This parameter $c_1$is more promising and the value which is random in this parameter can move the algorithm from the operation namely, exploration to exploitation operation and vice versa. As a result, specifying appropriate values for this is very essential. In this, the authors have used variety of set of values for $c_1$also it is observed that any value within the limit lying between [0.05 - 0.95] can render better solutions if the algorithm does exploration in the early phases and exploitation in the last phase So here the lower $c_{min}$and the upper c_{max}range for the parameter used to control the limit [0.05 - 0.95] of values is c_1 and the general adaptive is given by equation (18),

$$c_1 = c_{max} + (c_{min} - c_{max}) \times \left(a + \frac{10t}{T_{max}}\right) \tag{18}$$

The equation (18) is about the logarithmic adaptive inertia weight and implements logarithmic minimization of number in a random basis instead of a value which remains constant. The inertia weight parameter is *c*and, the random number is a within the limit [0, 1],$t$and $T_{max}$are the existing and the iterations that occur in highest times. The main benefit of using a transition like this is to allow the algorithm to conduct exploration for a set number of iterations, and then progressively move to exploitation as the number of iterations increases, and eventually to use the spaces that are referred at the iterations limit.

Population adaptation: The level of complexities in computations that takes place during the run time and it identifies the feasible solution for the process of feature selection. The other techniques to minimize it, however instead of a constant population, the usage of linear minimization of population are adapted. Due to the complexity of computations, the size of population n, the size of dimension $d$and large number of evaluations in a function T_{max} and it is given by$O(n.d.T_{max})$. The population adaption is given by equation (19),

$$n(g+1) = round\left[\left(\frac{n_{min-}n_{max}}{T_{max}}\right).T_{max} + n_{max}\right] \quad (19)$$

where $n_{min}$and $n_{max}$are the greatest and the least sizes of population and T_{max}correspond to many times of iterations. The pseudo-code for ASSA is given by Algorithm 2.

Algorithm 2: Adaptive Salp Swarm Algorithm (ASSA)
Input: Breast Cancer dataset
Objective Function: Classifier accuracy, $f(x_i)$
Output: Selection of optimal features

1. **Begin**
   1.1. Generate initial population of n butterflies $x_i = (i = 1,2,\dots,n)$, $c_1$via number of features in the dataset
   1.2. The initial best is to be found
   1.2. Evaluate and select the relevant salp
2. **Until** termination criteria is met
3. **if iterations <MI/2**
   3.1. Find new solution using equation (17)
   3.2. Evaluate fitness $f(x_i)$
   3.3. Update c_1 using equation (18) and $\boldsymbol{n}$ using equation (19)
4. **Else**
   4.1. find x_{new}new solution using equation (15)
   4.2. Evaluate fitness $f(x_i)$
   4.3. Update c_1 using equation (18) and $\boldsymbol{n}$ using equation (19)
5. **end**
   find the current best feature in the dataset
6. **End until**
   Update final best features from the dataset
7. **End**

Pearson Correlation Coefficient (PCC) based aggregation function

Pearson Correlation Coefficient (PCC) based ensemble feature selector measures the values that are correlated from every features that are chosen from these 3 bioinspired methods such as Adaptive Mutation Enhanced Elephant Herding Optimization (AMEHO), Adaptive Mutation Butterfly Optimization Algorithm (AMBOA), and Adaptive Salp Swarm Algorithm (ASSA), and greater features which are similar are eliminated from every feature sets. After that, an ensemble feature selector is used to combine the features from all 3 approaches. The ensemble feature selector obtains its final optimal feature set through majority voting on the output of each individual feature set. The PCC feature selector which has the

following steps is given below. The features are chosen from the arithmetic mode utilizing AMEHO, AMBOA, and ASSA.
The correlation coefficient matrix is measured for the features that are chosen as the output of the ensemble feature selection, shown in equation (23),

$$PearsonCorrelationCoefficient(PCC) = \frac{N\sum xy - (\sum x)(\sum y)}{\sqrt{\left[N\sum x^2 - (\sum x)^2\right]\left[N\sum y^2 - (\sum y)^2\right]}} \quad (20)$$

Here x, y is values of attributes being considered, and N represents the instances (total number). Correlation values are related to pairs. Here x, y are considered to be the attributes which are related when the correlation value is higher than 0.95, then x and y are highly correlated, and one among them will be eliminated; Or else, the correlation-based ensemble feature selector will choose both. As an input, the classification subsystem receives the feature set chosen by the correlation depending on the ensemble feature selector.

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 [35, 36] is suggested to have high speed of learning and better results. In KELM, the hidden layer's initial parameters don't require to change, since nonlinear continuous function is used hidden neuron. Consequently, for N arbitrary separate breast cancer examples$\{(x_i, t_i)|x_i \in R^n, t_i \in R^m, i = 1,\ldots,N\}$, the result operation in ELM by Lscreened neurons are shown in equation (21),

$$f_L(x) = \sum_{i=1}^{L} \beta_i h_i(x) = h(x)\beta \quad (21)$$

where $\beta = [\beta_1, \beta_2, \ldots, \beta_L]$ is denoted as the resultant weight vector among the hidden layer of L and the output vector and $h(x) = [h_1(x), h_2(x), \ldots, h_L(x)]$ is denoted as the output vector from hidden layer for input x, it maps the samples from input sample to the KELM feature sample [37,38]. In order to reduce classification error and refining the generalization results of Neural Networks(NNs), the implementation error and the output weights must be reduced by equation (22),

$$Minimize: ||H\beta - T||, ||\beta|| \quad (22)$$

The least squares solution of equation (9) depending on KKT conditions is able to be written by equation (23),

$$\beta = H^T\left(\frac{1}{C} + HH^T\right)^{-1} T \quad (23)$$

Where $H$is the output matrix from hidden layer, $C$is the regulation coefficient, and $T$is the expected result matrix of examples. Subsequently, the result purpose of the KELM classifier by equation (24),

$$f(x) = h(x)H^T\left(\frac{1}{C} + HH^T\right)^{-1} T \tag{24}$$

If the feature mapping $h(x)$is unidentified and KELM, kernel matrix is formed from Mercer's circumstances be able to be described by equation (25),

$$M = HH^T : m_{ij} = h(x_i)h(x_j) = k(x_i, x_j) \tag{25}$$

Therefore, the output function (x) of the KELM be able to be written capably by equation (26),

$$f(x) = [k(x, x_1), \ldots . k(x, x_N)]\left(\frac{1}{c} + M\right)^{-1} T \tag{26}$$

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. The following three typical kernel functions are used in this work [39].

(1) Gaussian kernel is well-defined by equation (27),

$$k(x, y) = exp\ exp(-a||x - y||) \tag{27}$$

(2) Hyperbolic tangent (sigmoid) kernel by equation (28),

$$k(x, y) = tanh\ tanh(bx^T y + c) \tag{28}$$

(3) Wavelet kernel by equation (29),

$$k(x, y) = cos\ cos\left(d\frac{||x - y||}{e}\right) exp\ exp\left(-\frac{||x - y||^2}{f}\right) \tag{29}$$

All the above-mentioned kernels, the adjustable parameters a, b, c, e, and f play a foremost role in the usual of KELM and must be changed wisely based on the transmutation parameter.

Results and Discussions

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). KNN, NB, MLP and SVM are used for comparison**.** For individual dataset, 80% of the examples are utilized for exercise the perfect and rest 20% are for testing purpose. Functional the FS approaches on the data which is trained, and the features that are identifies for feature subsection. With the help of test data, the chosen features and the classification metrics are validated depending on the data which is tested with the classifiers.

Dataset Description

The WDBC and WOBC datasets from the (UCI) ML Repository are utilized 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 nign, 241 malignant

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) ($\boldsymbol{T_{pos}}$): Malignant people correctly recognized as malignant
True Negative (TN) ($\boldsymbol{T_{neg}}$): Benign people correctly recognized as benign
False Positive (FP)($\boldsymbol{F_{pos}}$): Benign people wrongly recognized as malignant
False Negative (FN)($\boldsymbol{F_{neg}}$): Malignant people wrongly recognized as benign

Precision: The fraction of correctly categorized instances or samples from those classified as positives is evaluated by precision. As a result, equation (30) provides the precision assessment technique,

Precision = True Positive/ (True Positive + False Positive) = $\boldsymbol{T_{pos}}$/ ($\boldsymbol{T_{pos}}$ + $\boldsymbol{F_{pos}}$) (30)

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 (31),

Recall = True Positive/ (False Negative True Positive) = $\boldsymbol{T_{pos}}$ / ($\boldsymbol{F_{neg}}$ + $\boldsymbol{T_{pos}}$) (31)

F-*Measure:* F-Measure is a method for combining precision and recall into a single measure that encompasses both. Equation (32) calculates the standard F-measure as follows,

F-Measure = 2* (Precision * Recall) / (Precision + Recall) (32)

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 are shown in equation (33).

$$\textbf{Accuracy} = \frac{T_{pos}+T_{neg}}{T_{pos}+T_{neg}+F_{pos}+F_{neg}} \tag{33}$$

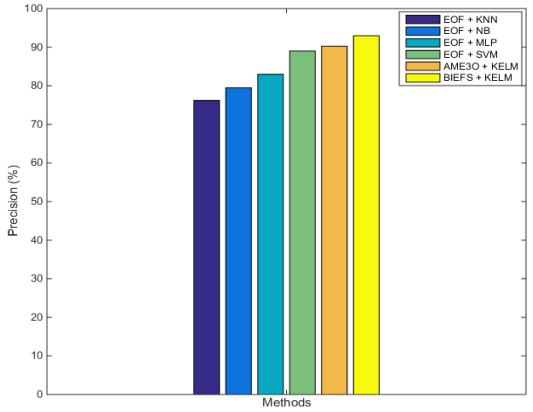
Area Under Curve (AUC) : The receiver - operating characteristic (ROC-curve) curve is represented by the area under the curve (AUC). The ROC-curve is a pictorial depiction of the True Positive Rate (TPR) vs. the False Positive Rate (FPR) for a binary classification with a wide-ranging discrimination criteria's.

Results Comparison Between Classifiers And Datasets

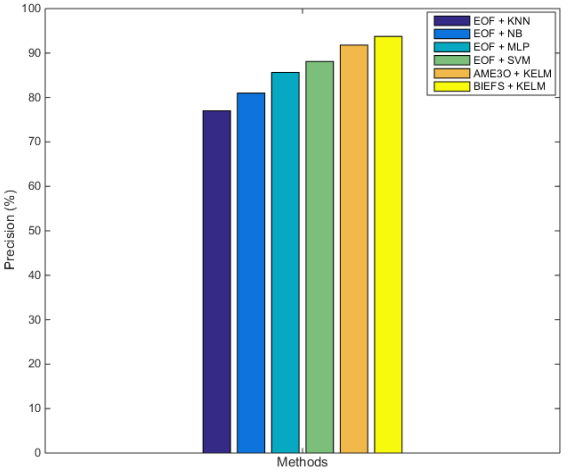
This section shows the performance outputs of comparison with various classifiers with respect to metrics mentioned above. The classifiers results are shown between two different datasets such as WDBC and WOBC. After completing the dataset and all missing data was obtained, the following indicators are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

Table 3. Evaluation metrics results comparison of classifiers for breast cancer datasets

Classifiers	First dataset- 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
BIEFS+ KELM	92.9477	93.1894	93.0684	93.4974	92.1894
Classifiers	Second dataset- 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
BIEFS+ KELM	93.7373	94.0066	93.8717	94.0000	93.0066



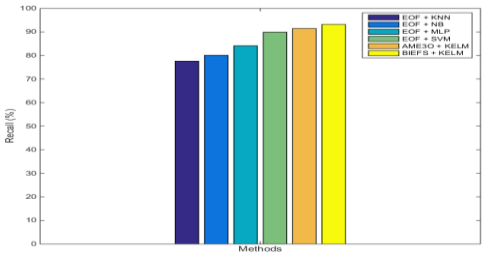
(a) Precision results of WDBC dataset vs. Classifiers



(b) Precision results of WOBC dataset vs. Classifiers

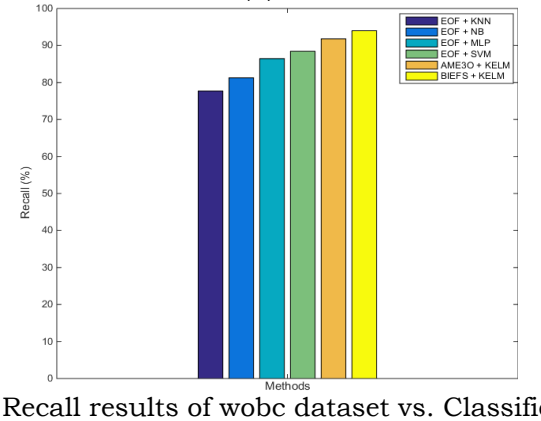
Figure 2. Precision metric comparison of classifiers vs. BC datasets

Two breast cancer datasets with respect to classifiers such as EOF+KNN, EOF+NB, EOF+MLP, EOF+SVM, AMEHO+KELM and proposed classifier are illustrated in figure 2. Figure 2(a) illustrates the precision results comparison of the different classifiers for first and second datasets with respect to classifiers are illustrated in figure 2(b). From these figures 2(a&b), however the proposed classifier shows improved precision results of 92.9477% and 93.7373% for WDBC and WOBC dataset. The classifiers such as EOF+KNN, EOF+NB, EOF+MLP, EOF+SVM, and AMEHO+KELM shows the reduced precision value of 76.7494%, 79.0087%, 83.9060%, 88.4017%, and 91.1877% for first dataset (Refer Table 3).



(a) Recall results of wdbc dataset vs. Classifiers

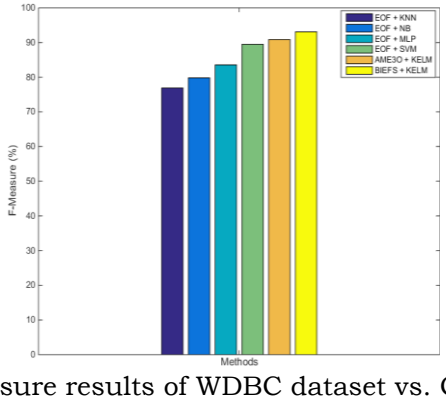
(b)



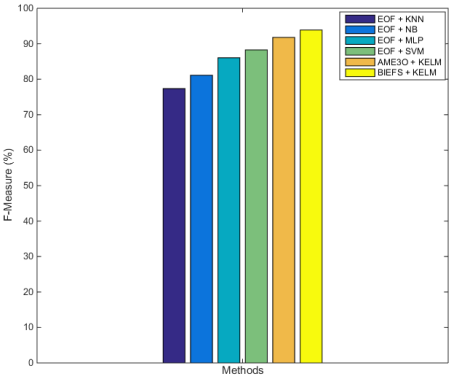
(c) Recall results of wobc dataset vs. Classifiers

Figure 3. Recall metric comparison of classifeirs vs. Bc datasets

Classification methods like EOF+KNN, EOF+NB, EOF+MLP, EOF+SVM, AMEHO+KELM, and proposed classifier for recall results are illustrated in the figure 3. Figure 3(a&b) shows the results of different classifiers with respect to WDBC dataset and WOBC dataset of recall. Figure 3(a) and figure 3(b), it demonstrates that the proposed classifier shows enhanced recall results of 93.1894%, and 94.0066% for first and second datasets respectively. At the same time the other classifiers like EOF+KNN, EOF+NB, EOF+MLP, EOF+SVM, and EOF+SVM shows the reduced recall value of 78.2338%, 80.1946%, 85.4130%, 90.0640%, and 91.6416% for first dataset(Refer Table 3).



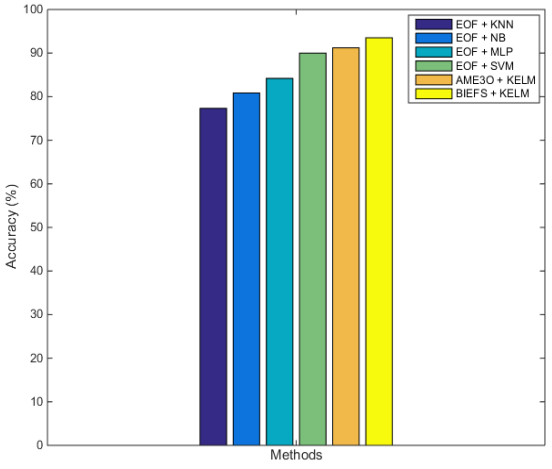
(a) F-measure results of WDBC dataset vs. Classifiers



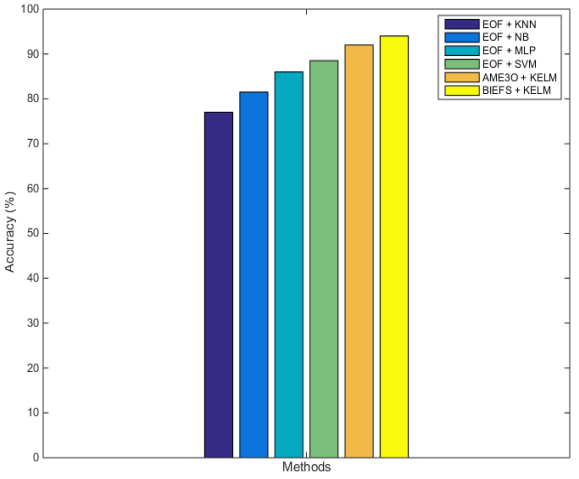
(b) F-measure results of WOBC dataset vs. Classifiers

Figure 4. F-measure metric comparison of classifiers vs. BC datasets

Figure 4 shows the classifiers results of methods like EOF+KNN, EOF+NB, EOF+MLP, EOF+SVM, AMEHO+KELM, and proposed classifier with respect to f-measure. Figure 4(a) and Figure 4(b) illustrated the f-measure results of different classifiers for WDBC and WOBC. Figure 4(a&b), it demonstrates that the proposed classifier gives increased f-measure results of 93.0684%, and 93.8717% for first and second datasets respectively. The classifiers such as EOF+KNN, EOF+NB, EOF+MLP, EOF+SVM, and AMEHO+KELM gives the reduced f-measure value of 77.4845%, 79.5973%, 84.6528%, 89.2251%, and 91.4141% for first dataset (Refer Table 3).



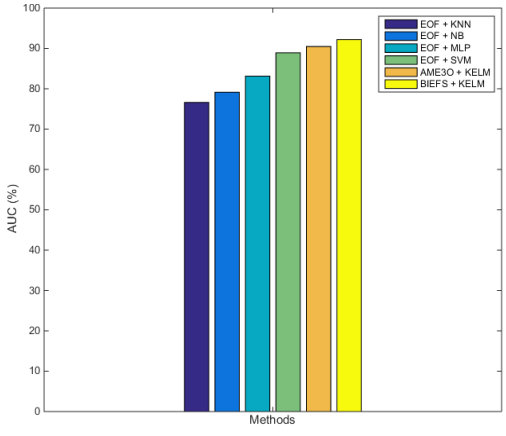
(b) Accuracy results of WDBC dataset vs. Classifiers



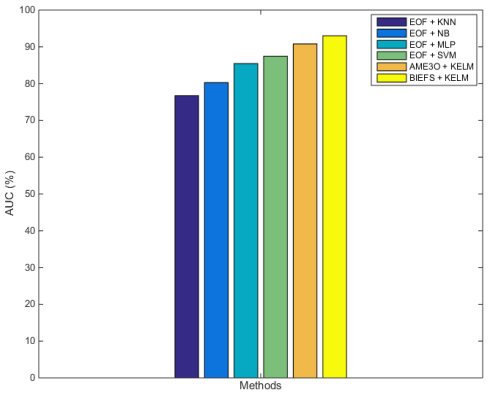
(a) Accuracy results of WOBC dataset vs. Classifiers

Figure 5. Accuracy metric comparison of classifiers vs. BC datasets

Figure 5 illustrates the accuracy results comparison of classifiers. Figure 5(a&b) shows the accuracy results comparison of first and second datasets under classifiers respectively. The proposed classifier produces improved accuracy results of 93.4974%, and 94.0000% for first and second datasets respectively. The methods like EOF+KNN, EOF+NB, EOF+MLP, EOF+SVM, and AMEHO+KELM produces reduced accuracy value of 77.8559%, 80.3163%, 85.0615%, 89.4552%, and 91.9156% for first dataset (Refer Table 3).



(a) AUC results of WDBC dataset vs. Classifiers



(b) AUC results of WOBC dataset vs. Classifiers

Figure 6. AUC metric comparison of classifiers vs. BC datasets

AUC metric results comparison of EOF+KNN, EOF+NB, EOF+MLP, EOF+SVM, AMEHO+KELM, and proposed AMEHO+KELM classifier under both datasets are illustrated in figure 6(a&b). The proposed AMEHO+KELM classifier shows the AUC results of 92.1894% and 93.0066% for first and second datasets respectively. The classifiers like EOF+KNN, EOF+NB, EOF+MLP, EOF+SVM, and AMEHO+KELM produces lesser AUC value of 72.2338%,74.1946%, 79.4310%, 84.0640%, and 85.6416% for WDBC dataset(Refer Table 3) .

Conclusion

Breast cancer is a quite common type of cancer that exists all over the world and is considered to be a threat to humans. In Computer-Aided Diagnosis (CAD) system, many techniques are introduced to diagnose and classify cancer from benchmark dataset. In CAD system, Feature Selection (FS), and classification plays a major important role for breast cancer analysis with improved accuracy. Thus the proposed system two major stages has been plays major part for disease diagnosis. In the first phase, Bio-Inspired Ensemble Feature Selection (BIEFS) is introduced for choosing the features that are relevant from dataset and ignore the irrelevant features, which is the primary objective for enhancing the accuracy of classification. BIEFS is performed based on combining three methods such as Adaptive Mutation Enhanced Elephant Herding Optimization (AMEHO), Adaptive Mutation Butterfly Optimization Algorithm (AMBOA), and Adaptive Salp Swarm Algorithm (ASSA). The proposed AMBOA method is focussed on butterflies' foraging strategy, which utilize their sense of smell for optimal selection of features in order to identify the nectar partners' location. In both AMBOA, and AMEHO mutation operator is proposed to adjust the random parameter so as to maximize searching ability. The ASSA is a optimizer which is based on population by imitating the behaviour of swarm of salps with two modifications such as position updating, and selection operation. The next phase consists of a classifier namely Kernel Extreme Learning Machine (KELM) classifier which is dynamic to select for dual different groups of focuses with or without breast cancer. In huge data, the necessity for improving and limitations in computational duration and the call for memory are the approaches that identify the ensemble sizes which is left as scope of future work.

References

1. Siegel, R.L.; Miller, K.D.; Jemal, A. Cancer statistics, 2019. CA. Cancer J. Clin. 2019, 69, 7–34.
2. Gupta, G.; Dang, R.; Gupta, S. Clinical presentations of carcinoma breast in rural population of North India: A prospective observational study. Int. Surg. J. 2019, 6, 1622–1635.
3. Kalarivayil, R.; Desai, P.N. Emerging technologies and innovation policies in India: How disparities in cancer research might be furthering health inequities? J. Asian Public Policy 2020, 13, 192–207.
4. Roy, S.; Kumar, R.; Mittal, V.; Gupta, D. Classification models for Invasive Ductal Carcinoma Progression, based on gene expression data-trained supervised machine learning. Sci. Rep. 2020, 10, 4113–4126.
5. Saba, T. Recent advancement in cancer detection using machine learning: Systematic survey of decades, comparisons and challenges. J. Infect. Public Health 2020, 13, 1274–1289.
6. Eedi, H.; Kolla, M. Machine Learning aproaches for healthcare data analysis. J. Crit. Rev. 2020, 7, 312–326.
7. Le, T.M., Vo, T.M., Pham, T.N. and Dao, S.V.T., 2020. A Novel Wrapper–Based Feature Selection for Early Diabetes Prediction Enhanced With a Metaheuristic. IEEE Access, 9, pp.7869-7884.
8. Mahendru, S. and Agarwal, S., 2019. Feature selection using Metaheuristic algorithms on medical datasets. In Harmony Search and Nature Inspired Optimization Algorithms (pp. 923-937). Springer, Singapore.
9. Mohamed A. W., A. A. Hadi, and A. K. Mohamed, ``Gaining-sharing knowledge based algorithm for solving optimization problems: A novel nature-inspired algorithm,'' Int. J. Mach. Learn. Cybern., vol. 11, pp. 1501-1529, 2020
10. Bolón-Canedo, V. and Alonso-Betanzos, A., 2019. Ensembles for feature selection: A review and future trends. Information Fusion, 52, pp.1-12.
11. Seijo-Pardo B., V. Bolón-Canedo , A. Alonso-Betanzos , Testing different ensemble configurations for feature selection, Neural Process. Lett. 46 (2017) 857–880 .
12. Pes B., N. Dess, M. Angioni, Exploiting the ensemble paradigm for stable feature selection: a case study on high-dimensional genomic data, Inf. Fusion 35 (2017) 132–147,
13. Seijo-Pardo B., I. Porto-Díaz , V. Bolón-Canedo , A. Alonso-Betanzos , Ensemble fea- ture selection: homogeneous and heterogeneous approaches, Knowl. Based Syst. 114 (2017) 124–139 .
14. Seijo-Pardo B., V. Bolón-Canedo , A. Alonso-Betanzos , On developing an automatic threshold applied to feature selection ensembles, Inf. Fusion (2019) . 1227–1245.
15. Hengpraprohm, S. and Jungjit, S., 2020. Ensemble Feature Selection for Breast Cancer Classification using Microarray Data. Inteligencia Artificial, 23(65), pp.100-114.
16. Elgin Christo, V.R., Khanna Nehemiah, H., Minu, B. and Kannan, A., 2019. Correlation-based ensemble feature selection using bioinspired algorithms and classification using backpropagation neural network. Computational and mathematical methods in medicine, Vol. 2019 , no.7398307, pp.1-7.

17. Prince, M.S.M., Hasan, A. and Shah, F.M., 2019, An efficient ensemble method for cancer detection. In 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT) ,pp. 1-6.
18. Ji, B., Lu, X., Sun, G., Zhang, W., Li, J. and Xiao, Y., 2020. Bio-inspired feature selection: An improved binary particle swarm optimization approach. IEEE Access, 8, pp.85989-86002.
19. Kewat, A.; Srivastava, P.N.; Kumhar, D. Performance Evaluation of Wrapper-Based Feature Selection Techniques for Medical Datasets. Algorithms Intell. Syst. 2020, 32, pp.619–633.
20. Abdullah Farid, A.; Selim, G.; Khater, H. A Composite Hybrid Feature Selection Learning-Based Optimization of Genetic Algorithm for Breast Cancer Detection. Preprints 2020, pp.1-21, 2020030298 (doi: 10.20944/preprints202003.0298.v1).
21. G. I. Sayed, G. Khoriba, and M. H. Haggag, "A novel chaotic salp swarm algorithm for global optimization and feature selection," Int. J. Speech Technol., vol. 48, no. 10, pp. 3462–3481, 2018.
22. Allam, M. and Nandhini, M., 2018. Optimal feature selection using binary teaching learning based optimization algorithm. Journal of King Saud University-Computer and Information Sciences, pp.1-13.
23. Kiziloz, H.E., Deniz, A., Dokeroglu, T. and Cosar, A., 2018, Novel multiobjective TLBO algorithms for the feature subset selection problem. Neurocomputing, 306, pp.94-107.
24. Punitha, S., Amuthan, A. and Joseph, K.S., 2019. Enhanced Monarchy Butterfly Optimization Technique for effective breast cancer diagnosis. Journal of medical systems, 43(7), pp.1-14.
25. Emary E., H. M. Zawbaa, and A. E. Hassanien, "Binary grey wolf optimization approaches for feature selection," Neurocomputing, vol. 172, pp. 371–381, 2016.
26. Zawbaa H. M., E. Emary, C. Grosan, and V. Snasel, "Large-dimensionality small-instance set feature selection: a hybrid bio-inspired heuristic approach," Swarm and Evolutionary Computation, vol. 42, pp. 29–42, 2018.
27. Anter A. M. and M. Ali, "Feature selection strategy based on hybrid crow search optimization algorithm integrated with chaos theory and fuzzy c-means algorithm for medical diagnosis problems," Soft Computing, pp. 1–20, 2019.
28. 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.
29. Li, J., Lei, H., Alavi, A.H. and Wang, G.G., 2020. Elephant herding optimization: variants, hybrids, and applications. Mathematics, 8(9), pp.1-25.
30. Arora, S. and Singh, S., 2019. Butterfly optimization algorithm: a novel approach for global optimization. Soft Computing, 23(3), pp.715-734.
31. Tubishat, M., Alswaitti, M., Mirjalili, S., Al-Garadi, M.A. and Rana, T.A., 2020. Dynamic butterfly optimization algorithm for feature selection. IEEE Access, 8, pp.194303-194314.
32. Mirjalili, S., Gandomi, A.H., Mirjalili, S.Z., Saremi, S., Faris, H. and Mirjalili, S.M., 2017. Salp Swarm Algorithm: A bio-inspired optimizer for engineering design problems. Advances in Engineering Software, 114, pp.163-191.

33. Mirjalili S., S.M. Mirjalili , A. Lewis , Grey wolf optimizer, Adv. Eng. Softw. 69 (2014) 46–61 .
34. Yang X.-S., and S. Deb , Engineering optimisation by cuckoo search, Int. J. Mathematical Modelling and Numerical Optimisation 1 (4) (2010) 330–343 .
35. 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.
36. 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.
37. Huang W., N. Li, Z. Lin et al., "Liver tumour 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.
38. Ding S.F., Y. A. Zhang, X. Z. Xu, and L. N. Bao, "A novel extreme learning machine based on hybrid kernel function," Journal of Computers, vol. 8, no. 8, pp. 2110–2117, 2013.
39. 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.