A Hybrid Artificial Bee Colony with Whale Optimization algorithm for improved breast cancer diagnosis

S.Punitha^a, Thompson Stephan^b*, Ramani Kannan^c, Ajith Abraham ^d

^aDepartment of Computer Science, Pondicherry University, Pondicherry, India ^b Department of Computer Science Engineering, Amity University, Noida, India ^c Department of Electrical and Electronics Engineering, Universiti Teknologi PETRONAS, Malaysia ^d Machine Intelligence Research Labs (MIR Labs), Auburn, Washington 98071, USA

Abstract

Breast cancer is the deadliest disease among women that leads to death if not diagnosed at early stages. Early diagnosis plays a vital role in decreasing the mortality rate globally. Manual methods for diagnosing breast cancers suffer from human errors, inaccuracy, and consumes time. A Computer-Aided Diagnosis (CAD) can overcome the disadvantages of manual methods and helps radiologists for accurate decision-making. A CAD system based on Artificial Neural Network (ANN) optimized using a swarm-based approach can improve the accuracy of breast cancer diagnosis due to its strong decision-making capacities. Artificial Bee Colony (ABC) and Whale Optimization are metaheuristic search algorithms used to solve combinatorial optimization problems. This paper proposes a Hybrid Artificial Bee Colony with whale Optimization algorithm (HAW) by integrating the exploitative employee bee phase of ABC with the bubble net attacking method of whale optimization to propose an employee bee attacking phase. In the employee bee attacking phase, employee bees use exploitation of humpback whales for finding better food source positions. The weak exploration of standard ABC is improved using the proposed mutative initialization phase that forms the explorative phase of the HAW algorithm. HAW algorithm is used in simultaneous feature selection and parameter optimization of an ANN model. HAW is implemented using back-propagation learning that includes resilient back-propagation (HAW-RP), Levenberg Marquart (HAW-LM), and momentum-based gradient descent (HAW-GD). These hybrid variants are evaluated using various breast cancer datasets in terms of accuracy, complexity, and computational time. HAW-RP variant achieved high accuracy with low complexity ANN model when compared to HAW-LM and HAW-GD.

Keywords: Artificial neural networks, Artificial Bee Colony, Multilayer Perceptron, Levenberg Marquardt, Resilient backpropagation, momentum-based Gradient Descent backpropagation, Whale Optimization Algorithm.

1. Introduction

Early-stage diagnosis plays a major role in increases the chance of recovery from breast cancer. World Health Organization (WHO) has reported that breast cancer occurs in millions among women and they die due to later stage detection. Each year about 400,000 die among the female population due to breast cancer according to WHO [1]. About 12.4% of women died in the United States due to breast cancer in 2019. In 2020, invasive breast cancer expected is 276,480, and non-invasive breast cancer expected is 48,530 [2]. Metropolitan Cities in India such as Mumbai, Chennai, Delhi, Bangalore, Ahmadabad, and Bhopal are affected by 28% to 35% among the women population [3]. Thus, breast cancer has become a serious health issue around the globe and early detection plays a vital role in speedy recovery [4]. Early detection can be done using various scanning methods such as magnetic resource imaging, ultrasound imaging, self-check-up, mammography, and biopsies [5]. Traditionally followed breast cancer methods require high diagnosis and they fail because of inaccurate diagnosis caused by human errors. Automated computer-based diagnosis schemes are much helpful in overcoming the problems that are caused by manual diagnosis methods and they avoid unnecessary surgeries and biopsies [6]. Expert systems built using ANN have strong decision-making capabilities because of which it can be used for building medical diagnosis systems [7, 8]. ANN-based decision-making systems have outperformed the traditional technique used for classifying patterns [9].

Evolutionary algorithms (EA) are based on that evolution process occurs naturally. Commonly used evolutionary methods for optimization problems that are population-based are the Differential Evolution (DE) [10], Evolution Strategy (ES) [11], Genetic Algorithms (GA) [12]. These algorithms are stochastic which is based on survival to fittest theory. A lot of researches has concentrated on improving the standard evolutionary methods by modifying its phases or else by hybridizing. A biography-based EA is introduced which is applicable for recombination and crossover operators inspired by GA [13]. An algorithm by combining DE with GA is proposed that can be used for path synthesis [14]. Another improved version of GA is introduced for fuzzy rules and membership function optimization for an Adaptive Networkbased Fuzzy Inference Systems [15]. An integrated algorithm combining DE and eagle strategy is introduced for constrained problems [16]. Metaheuristic based swarm intelligence approach is used for real-time optimization problem solving [17, 18, 19]. Commonly used swarm intelligence approaches are the Ant colony Optimization (ACO) [20], and the Particle Swarm Optimization (PSO) [21] inspired by the foraging behavior of ants and social behavior of birds respectively. The echolocation capability of the microbats available in nature forms the basis of the Bat Algorithm (BA) [22]. A population-based swarm technique introduced based on the foraging behavior of honey bees [23]. The dynamic and static behavior of dragonflies forms the basis of a new metaheuristic algorithm called Dragonfly Algorithm (DA) [24]. Based on the herding behavior of krill, another swarm technique is proposed called Krill Herd (KH) algorithm [25].

A technique based on migration behavior is called the Monarch Butterfly Optimization (MBO) is introduced [26]. The foraging behavior for the survival of E. coli bacteria forms the basis of the Bacterial Foraging Optimization (BFO) [27]. Another swarm technique called the Artificial Immune System (AIS) is inspired by the biological immune system of the human body [28]. An algorithm for global optimization based on interior design and decoration [29]. A Salp Swarm Algorithm (SSA) based on the swarming behavior of the Salp in the ocean is introduced that can be used to solve multidimensional optimization problems [30]. Based on the Brownian movements and levy movements of the predators during their foraging process, another swarm technique called Marine Predictor Algorithm (MPA) was proposed [31]. This paper focussed on hybridizing Artificial Bee Colony Optimization with the Whale Optimization algorithm to introduce the HAW algorithm. The proposed HAW algorithm integrates the employee bee phase of the ABC with the encircling prey/ bubble net attacking method to have an enhanced exploitative phase called the employee attacking phase. In the employee attacking phase, the bees follow the bubble net attacking method of the whales to find out better food sources. The explorative phase of the HAW is driven by a mutative initialization phase of the standard ABC algorithm.

Appropriate selection of ANN topology design parameters such as the number of hidden layers, numbers of hidden nodes, initial weight values between the connections, learning rate, and algorithm plays a vital role in building a successful ANN model [32]. The convergence of the back-propagation learning process can be affected by the improper selection of weights making the learning process to be trapped in the local optimal locations [33, 34]. Improper selection of the hidden nodes may make the ANN classifier to deal with the problems of underfitting and overfitting. If the usage of hidden nodes in an ANN model is not appropriate to the amount of learning required for accurate diagnosis, then the ANN classifier may either be overtrained where the ANN model can give accurate results in case of training

and fails with inaccurate results in the case of testing or undertrained where the prediction rate decreases. Based on the above discussion, this paper focuses on optimal selection of the value of initial weights and the optimal selection of the hidden node sizes of an ANN model using the proposed HAW algorithm with the help of a wrapper architecture such that the proposed work aims at improving the learning performance of an ANN avoiding the problems of overfitting and underfitting with increased predictive capabilities.

Feature Selection (FS) deals with the deletion of irrelevant, redundant, and noisy features present in the input dataset of a classifier. FS improves the generalization of an ANN classifier system with reduced computational time, as demonstrated in [35,36]. Hence, simultaneously optimizing the input features and design parameters of ANN such as the initial weights and hidden node size can increase the predictability of the ANN classifier. Swarm based intelligent systems are used for coupled optimization of input features and ANN design parameters [37]. Due to the importance of simultaneous optimization of ANN design parameters and feature selection process that improves the convergence of the training process and prevent the ANN from being under-trained and over-trained and because of the complex design issue of ANN, the ANN topology optimization process can be coupled with a swarm-based metaheuristic optimization, such as ABC due to its powerful local and global search capabilities in finding out global optimal solutions.

1.1 Artificial Bee Colony (ABC) Optimization

A meta-heuristic swarm-based search mechanism called ABC is introduced by Karaboga in 2005. It is a population-based approach, inspired by the foraging nature of honey bees that solves multidimensional and multimodal real-time optimization problems for different applications, as demonstrated in [38]. ABC is based on a stochastic process, that is robust and highly flexible with a lesser number of control parameters that make it simple. The algorithmic steps of the ABC optimization process are described in Algorithm (1):

Algorithm 1: Artificial Bee Colony algorithm

Step 1: Initialization:

Food sources are randomly produced using Equation (1).

$$A_{k}^{l} = A_{k}^{l} + random(0,1) * (A_{max}^{l} - A_{min}^{l})$$
(1)

 A_j^i represents k^{th} food source with l^{th} parameter and $j=1, 2, \dots, N$, in which N represents maximum food sources. $l = 1, 2, \dots, dim$, in which 'dim' represents the dimension representing the number of parameters in the optimization problem. A_{max}^l

and A_{min}^{l} are the minimum and maximum bound of the l^{th} parameter of the optimization problem respectively.

Step 2: Quality Evaluation of food source:

The fitness values are identified for each food source A_k .

Step 3: Employed bee Phase:

Food sources are assigned to employee bees or worker bees. The employee bees use Equation (2) to search neighborhood food sources surrounding the current food sources A_k^l .

$$E_k^l = A_k^l + random[-1,1] * (A_k^l - A_d^l)$$
(2)

 A_d is a random food source where $d \in \{1, 2..., N\}$. '*l*' is a random integer, and $i = \{1, 2..., dim\}$ and '*d*' should not be equal to '*l*' for proper exploitation. If the quality of E_k^l is greater than A_k^l , then bee discards A_k^l saving E_k^l , or vice versa.

Step 4: Onlooker Bee Phase:

Information regarding the selected food sources is shared with the onlooker bees. The probability value Z_i of each food source received from the employee bee is calculated using Equation (3).

$$Z_k = \frac{fitness(A_k)}{\sum_{k=1}^{N} fitness(A_k)}$$
(3)

The quality of the food source A_k is represented as $fitness(A_k)$. The value Z_k of food source is compared with a random(0,1). Food sources with a Z_i value greater than random(0,1) are selected by the onlooker bees.

Step 5: Food source memorization:

The food source with the highest $fitness(A_k)$ is selected and memorized.

Step 6: Scout Bee Phase:

In the scout bee phase, unimproved food sources are identified based on a counter value and they are replaced by a randomly generated food source according to Equation (1).

1.2 Whale Optimization Algorithm (WOA)

WOA is a population-based swarm intelligence metaheuristic algorithm introduced by Mirjalili and Lewis [39] which is inspired by the foraging behaviour of humpback whales. The humpback whales' hunts group of krill or fishes using shrinking circle and producing bubbles in a circle '9'-shaped path. The exploitation phase is carried out using encircling prey and bubble-net attacking based on the spiral. A random search of prey is used for exploration. The exploitation phase of WOA is explained in Algorithm (2).

Algorithm 2: Encircling Prey / Bubble Net Attacking of WOA

To hunt the prey for survival, humpback whales encircle around the prey which can be mathematically represented using Equation (4) and (5).

$$Y = |\vec{L} \cdot A^{*}(t) - A(t)|$$
(4)

$$\vec{A}(t+1) = \vec{A^*}(t) - \vec{M}.Y$$
(5)

Where *t* represents the current iteration, A^* represents the best solution found so far, A gives the position vector, | | represents the absolute value, . gives multiplication of two elements. L and M represents the coefficient vectors that can be obtained using Equation (6 & 7).

$$\vec{M} = 2\vec{m}.\vec{r} - \vec{m} \tag{6}$$

$$\vec{L} = 2.\vec{r} \tag{7}$$

Where *m* is reduced linearly starting from 2 till 0 as the iteration proceeds. *r* represents a random vector from a uniform distribution between [0,1]. Each whale that represents a solution updates its position using Equation (5) where the updated new position of the whale depends on the best position (prey) found so far. The position of the whales can be controlled by the adjustment of vectors L and M. The value of *m* is decreased to achieve the shrinking encircling behaviour using the Equation (8).

$$m = 2 - t \frac{2}{MaxIterat} \tag{8}$$

Where *t* represents the current iteration and *MaxIterat* represents maximum iterations. The new position of the whale on the spiral path can be calculated using Equation (9);

$$\vec{A}(t+1) = Y' \cdot e^{ws} \cdot \cos(2\pi l) + \vec{A} * (t)$$
(9)

Where $Y' = |\vec{A}^*(t) - \vec{A}(t)|$ which indicates the distance of a whale and the best solution (prey). *w* is a constant that represents the shape of the logarithmic spiral. *s* is the random number generated between [-1,1]. Hence, the updated new position of the whale is calculated using 50% probability using a random number P_i generated between [0,1] as represented by Equation (10).

$$\vec{A}(t+1) = \begin{cases} \vec{A}(t+1) = \vec{A^*}(t) - \vec{M}. Y & \text{if } (P_i < 0.5) \\ \vec{A}(t+1) = Y'. e^{ws}. \cos(2\pi l) + \vec{A} * (t) & \text{if } (P_i \ge 0.5) \end{cases}$$
(10)

1.3 Comparative Investigation of ABC and WOA in terms of exploration and exploitation

. In the context of exploration, WOA uses the search of prey phase for exploration that completely depends on a random search agent which is a stochastic strategy. In the same way, ABC incorporates scout bees for exploration with the help of a random search. This makes both the algorithms to produce solutions concentrated in a local area at the initialization phase, losing its diversification. Hence, the search process prematurely converges returning sub-optimal solutions in both of the algorithms. Hence, both ABC and WOA are weak at exploration.

In the context of exploitation, the local search process is incorporated using the encircling prey and the bubble net attacking method. The WOA exploitative phase guarantees convergence since positions of the whale are updated using the best solution (prey) obtained so far. Hence, proper exploitation is guaranteed by the encircling prey and the bubble net attacking method in the direction towards the prey since the search process is always guided by the best solution found so far. Comparatively, ABC exploitation is carried out using the employee bee phase and the onlooker bee phase where the positions of the food sources are updated by changing the single parameter of the old solution (food source) that causes the existence of similar food sources that converge at the same optimum locations. Also, the local search of ABC cyclically revisits the same solutions that create the problem of looping making the search converge prematurely. Hence, WOA is better at exploitation as compared to ABC.

1.4 Problems that are addressed by the proposed HAW

Many researches have used ABC and WOA to develop optimal classifiers that can be used for medical diagnosis purposes, but still, the standard ABC suffers from the following issues which the proposed HAW addresses.

- (i) The local search by the employee and onlooker bee cyclically revisits similar solutions inducing the problem of looping making the search process converge prematurely.
- (ii) ABC optimization makes the solutions to be concentrated in local regions due to a lack of diversified solutions at initialization.
- (iii) The food source positions are updated by changing the single parameter of the old solution (food source) which causes the existence of similar food sources that converge at the same optimum locations.
- (iv) Exploitation is performed by two phases namely the employee bee phase and the scout bee phase, whereas the exploration process is done only by scout bees, which leads to an imbalance in exploration and exploitation.

The proposed HAW that is capable of resolving the above issues can be used for generating an optimized ANN classifier that can accurately and efficiently be used for breast cancer diagnosis. This paper focus on the following objectives:

- (i) A Hybrid ABC-WOA Optimization (HAW) that integrates the encircling prey and the bubble net attacking method of WOA with the employee bee phase of standard ABC to form an employee attacking phase. The exploration of the HAW is enhanced using the proposed mutative exploration phase of ABC at initialization.
- (ii) Implementation of proposed HAW in optimal feature subset selection and ANN parameter optimization using Wisconsin breast cancer dataset. The HAWoptimized ANN model is evaluated in terms of accuracy, complexity, and computational time.

2. Related Works

A classification method for breast cancer diagnosis using principal component analysis and ANN is proposed [40]. The feature selection is done using a screen test, Kaiser Guttman rule, and cumulative variance. Then testing subset from WBCD is used to check ANN performance which attained classification accuracy of 95.68% with a set of 5 features selected. Even though the discriminating accuracy is high, the computational time required is more. Another feature selection scheme for breast cancer diagnosis is introduced [41]. The system used a support vector machine and ANN for classification. The method utilized WBCD that used 10-fold cross-validation of data portioning. The prediction accuracy was 97.14% for SVM and 96.71% for ANN. No parameter optimization of the classifier and hence can't be used for high dimensional datasets. A diagnosis system using entropy measures for feature selection is proposed [42]. It used artificial neural networks in which weights optimized by PSO and Levenberg–Marquardt. The ANN evaluated for WBCD where accuracy reached up to 98.83%. The proposed system requires high computational time for large datasets.

The integrated algorithm of ABC and ACO for finding optimal feature subsets of medical datasets is proposed [43]. The global search followed by ABC is improved by using feature subsets generated by ACO to the ABC optimization process. The approach yielded an accuracy of 99.07% using WBCD. The proposed exploitation uses traditional greedy selection making algorithm to prematurely converge at the local optimal locations. The proposed algorithm is only used for feature selection and no parameter optimization. The hybrid algorithm of DE and ABC for optimal binary subsets selection is introduced [44]. The proposed algorithm combines the high exploration property of DE with an improved onlooker bee phase of the ABC. The proposed approach achieved F-measure of 92.2, 96.4, 97.6 for decision tree

classifier, Naive Bayes classifier, and RBF networks classifier respectively using WBCD. The proposed algorithm is only used for feature selection and no parameter optimization.

A modified ABC is proposed for feature selection [45]. The exploitation of the employee bees is improved where the neighborhood search process is improved using the global best solution. The modified ABC is used for feature selection using benchmark datasets called MIAS and DDSM for breast cancer diagnosis. Classification is carried out using Self Adaptive Resource Allocation Network. The accuracy evaluated as 96.89% and 97.17% for MIAS and DDSM respectively. The algorithm has not focused on the explorative phase and it has used only randomized initial solutions with loss of diversification. A feature selection algorithm using ABC and decision trees based on the gradient boosting model is proposed [46]. The features are selected from the Wisconsin breast cancer dataset and Haberman's survival dataset. A regression tree is used as the classifier where Gradient descent finds the direction of the gradient of residuals. The classification accuracy is 74.3% for Haberman's cancer dataset and 92.8% for WBCD. It has not been evaluated in terms of complexity.

An efficient ABC is proposed for optimal learning of Deep Neural Networks (DNN) [47]. The proposed used ABC and Broyden–Fletcher–Goldfarb–Shannon (BFGS) with limited memory. This proposed ABC tuned the parameters of DNN with cascaded autoencoder layers. The classification accuracy using WBCD is 73.03%. The step size of the neighborhood search is kept static throughout the entire search process affecting the convergence. The optimized classification of DNA microarrays using ABC is proposed [48]. The optimal feature subsets from breast cancer datasets are selected using ABC. Then the selected optimal feature subsets are given to MLP, Radial Basis Function Neural Network, and SVM. The accuracy attained is 94.7% for MLP, for SVM accuracy is 89.5% and for RBF is 73.7%. The algorithm used standard ABC without any improvement. ABC based feature selection is proposed for UCI datasets [49]. The proposed system chooses 2 features from 9 attributes from WBCD and yielded an accuracy of 96.69%. The proposed system is simple but used only the standard ABC.

Optimal feature selection using ABC for UCI repository datasets is proposed [50]. The employee bee phase is modified using a modification rate where the feature is selected if the random number greater than the modification rate. The classification accuracy is 75.87%. The algorithm has not focused on the explorative phase and it has used only randomized initial solutions with loss of diversification. Two-hybrid algorithms are proposed based on ABC and PSO [51]. In the first algorithm, the employee bee phase is hybridized with PSO to find new velocity position updates. In the second algorithm, the onlooker and scout bee phase are improved using mutations of the genetic algorithm. Both of the algorithms have the highest

accuracy of 99.14% with an optimal selection of 13 features using WBCD. The proposed algorithm is only used for feature selection and no parameter optimization. A hybrid algorithm for feature selection using branch and bound approach and ABC is proposed [52]. The algorithm first applies the branch and bound and finds the first set of features. Then, it applies ABC to identify the second set of features. A union operation is done to form a new set of optimal features. The proposed has not focussed on classification.

Optimal feature selection using Whales Bubble Net Hunting strategy for UCI repository datasets is proposed [53]. The algorithm handles exploitation using the Bubble-Net Attacking Method phase. Further, a global search is carried out by the Search for the Prey phase. During the evaluation, the SVM classifier attained an accuracy of 98.77%, precision of 99.15%, recall of 98.64%, and f-score of 98.9%. The proposed algorithm is only used for feature selection and no parameter optimization. Optimal feature selection using the integration of ACO and Cuckoo Search is proposed [54]. Local search behavior of ACO is improved using the exploitation of Cuckoo Search. The proposed algorithm selected feature set that is optimal from the set of 78 texture features derived using GLCM. The input is taken from the MIAS dataset. In the proposed approach, 5 features were selected with 94% accuracy. The proposed algorithm uses the SVM classifier for prediction. The proposed showed increased performance of 4% and 2% when compared with PSO and ACO respectively.

3. Materials and Methodologies

A wrapper-based method that eliminates the use of statistical methods such as information gain or F-score is used for implementing the proposed HAW. The proposed architecture is depicted in Figure (1). The input dataset is the breast cancer dataset where the total set is divided into three subsets. The first set that contains 50% of samples is used for training. The next 25% of samples are used for the validation and the rest of 25% is used for testing. With the help of the optimal set of input features generated by the proposed HAW, the optimal features are selected from the three subsets where the other features are rejected. The optimal selected features of the training set are used for training the underlying ANN classifier.

The optimal initial weights and hidden node size generated by the proposed HAW are used as the initial parameter settings of ANN. The ANN error is calculated with the help of the validation set. If the validation error increases for six iterations continuously, the training of ANN is stopped. The fitness of trained ANN is calculated using Equation (12 &13). The ANN with high fitness (best) is selected and tested using the testing set with optimal feature subsets.

The complexity (number of connections) of final ANN network achieved is calculated using Equation (11).

$$Connect = U * V + V * W + V + W$$
(11)

'U gives input features (size), 'V' indicates hidden node size, and 'W' indicates output nodes (size). The resulting ANN with the least connections guarantees less complexity. Fitness of ANN is calculated by Equation (6). A higher value of *ANN Err* indicates low fitness ANN.

$$ANN \ Err = \left(\frac{P_{max} - P_{min}}{L^{*w}}\right) \sum_{i=1}^{l} \sum_{j=1}^{w} \left(A_j^i - B_j^i\right)^2 \tag{12}$$

$$Fitness = \frac{1}{ANN \ Error} \tag{13}$$

'*l*' and '*w*' is the size of the output nodes and validation examples, respectively. P_{max} and P_{min} are maximum and minimum actual output, respectively. B_j^i and A_j^i is the target output and actual output, respectively.



Figure 1: Proposed Wrapper Architecture

The initial solution representation is given using Figure (2). I bits give the random initial weights, in which 2^I different initial weights can be explored. J bits give the hidden node size so that 2^J hidden node size can be explored. K bits give the feature bits that represent the total features. If a feature is selected, then 'K' bit is one; otherwise, it's zero.



Figure 2: Initial solution representation

3.1 Breast cancer datasets used by the proposed Wrapper Architecture

The proposed HAW is evaluated using breast cancer datasets such as the Wisconsin breast cancer dataset (WBCD), Wisconsin Diagnostic Breast Cancer dataset (WDBC), Wisconsin Prognostic Breast Cancer dataset (WDBC), Digital Database for screening mammography (DDSM), Mammographic Image Analysis Society (MIAS) and INbreast database. The description of the datasets used is given in Table (1):

DATASET	INSTANCES	FEATURES	CLASSES
Wisconsin Breast Cancer Dataset(WBCD)	699- Total	9	2(Malignant, Benign)
Wisconsin Diagnostic Breast Cancer Dataset (WDBC)	569-Total 357 - Malignant 212-Benign	30	2(Malignant, Benign)
Wisconsin Prognostic Breast Cancer Dataset (WPBC)	198-Total 151-Non-recurrent 47-recurrent	32	2(Non-recurrent, recurrent)
Digital Database for Screening Mammography (DDSM)	480-Total 170-Malignant 310 –Benign	21	2(Malignant, Benign)
Mammographic Image Analysis Society (MIAS)	338-Total 138-Malignant 200-Benign	21	2(Malignant, Benign)
INBREAST Database	275-Total 95 –Malignant 180-Benign	21-	2(Malignant, Benign)

Table 1: Breast Cancer datasets

3.2 Detailed Description of the proposed HAW algorithm

The proposed HAW optimization algorithm is framed by the integration of a mutative initialization phase of ABC optimization with the exploitation phase of the whale optimization technique. The standard ABC is weak in exploration because of localized initial food sources due to the poor random search process. Hence, the HAW algorithm has used a mutative exploration phase at its initialization such that the algorithm can explore the entire problem space and finds out new promising regions. This employee bee phase of the ABC optimization process is integrated with the exploitative of WOA such that the employee bees follow the encircling prey/ bubble attacking method of whales to update the positions of the food sources. The best food source found at each iteration is considered as the target prey of WOA. The HAW involves two stages: In the first stage, HAW uses a mutative initialization phase is proposed using different mutations and it derives a possible set of diversified solutions. In the second stage, an employee bee attacking phase is proposed such that the optimum set of solutions derived by the mutative initialization phase forms the initial food source positions of the employee attacking phase that follows the attacking method of whales for the prey. The simulated annealing technique is used in the employee bee attacking phase to make the algorithm to escape from the local optimum locations and avoid looping problems. A flowchart representation of the HAW algorithm is shown in Figure (3). The proposed HAW optimization is summarized as follows:

- (i) A mutative initialization phase is proposed to derive a set of diversified solutions to expedite the search speed at the exploration phase.
- (ii) An employee attacking phase is proposed so that the employee bees adapt the encircling prey/ bubble net attacking method of whales for updating the current food source positions during their foraging process. The exploitation of the employee bee attacking phase is guided by the best food source (prey of the whales) found so far.
- (iii) To escape from suboptimal location and to avoid looping problems, Simulated Annealing (SA) based employee attacking phase is proposed.
- (iv) The onlooker bee phase and scout bee phase are followed in the same way as that of the standard ABC optimization.





Figure 3: Flowchart representation of the proposed HAW algorithm

3.2.1 Initialization and fitness calculation

A food source indicates a possible solution of the underlying optimization problem. Each food source is generated using the '*dim*' number of variables that represent the dimension of problem space considered. The generation of the initial population is done through the random distribution of food sources using Equation (14). A_k^l represents the kth variable of food source k and $k=1, 2, \dots, N$ where N represents the maximum size of the food sources.

$$A_{k}^{l} = A_{min}^{l} + random(0,1) * (A_{max}^{l} - A_{min}^{l})$$
(14)

Where N=1, 2,...,dim, and 'dim' represents the dimension based on the number of parameters of the underlying optimization problem. random(0,1) is the random number generated between 0 to 1. A_{max}^{l} represents the maximum bound of the l^{th} variable of the optimization problem and A_{min}^{l} gives the minimum bound of the l^{th} variable of an optimization problem. The algorithm for the initial generation of food sources is given in Algorithm (3).

Algorithm 3: Initialization of the proposed HAW Generation of Initial population

For m = 1 to i do For n = 1 to d do

$$Y_m^n = Y_{min}^n + random(0,1) * (Y_{max}^n - Y_{min}^n)$$

End for *n*

End for *m*

Affinity Evaluation and sorting

For
$$m = 1$$
 to i do

Calculate fitness (Y_m)

Sort (Y_m) based on fitness

End for

3.2.2 Proposed Mutative exploration phase

The mutative exploration phase detects multiple food sources based on its quality. Better food sources are selected from the total population. Further, they are divided into three subpopulations based on the fitness difference between each food source and the best food source in the population with the help of three different threshold values such as limit1, limit2, and limit3. The three different subpopulations of food sources are subjected to different mutations where higher fitness food sources are mutated less and low fitness food sources are mutated high. Thus, the amount of mutation is inversely proportional to the fitness value of food sources. Better food sources with high fitness values are grouped as A_{k1} food sources whose fitness is close to the fitness of the best food source of the total population. The A_{k1} food sources are best; hence, a local search process is facilitated around the best food sources, which are mutated using Gaussian mutations where Equation (15) is utilized to make small random changes to A_{k1} food sources.

$$A'_{k1} = A_{k1} + \mu_g.\,G(0,1) \tag{15}$$

 A'_{k1} , the mutated food source, which is generated after the Gaussian mutation; μ_g is the mutation rate indicating the strength of Gaussian noise added; A_{k1} is the original food source; and G(0,1) is the random number of Gaussian distribution whose mean is zero and variance is one.

Intermediate food sources with intermediate fitness values are grouped as A_{k1} food sources whose fitness shares an intermediate fitness difference from the best food source of the total population. The A_{k2} food sources are intermediate; hence, a uniform search process is facilitated around the intermediate food sources, and are mutated using uniform mutations where Equation (16) is utilized to generated uniformly mutated food sources of the A_{k2} represented as A'_{k2} in which a random value is chosen from a solution and replaced with a uniform random value between the user-defined upper *(Ub)* and lower *(Lb)* limits.

$$A'_{k2} = A_{k2} \tag{16}$$

The worst food sources with low fitness values are grouped as A_{k3} food sources whose fitness values are at a larger difference from the fitness of the best food source in the population. The A_{k3} food sources are worse than A_{k1} and A_{k2} . Hence, a global search is facilitated around the worst food sources, in which A_{k3} food sources are mutated using Levy Mutations (LM) using Equation (17) that are more probable to escape from the local optimum.

$$A'_{h3} = A_{h3} + \mu_c. C(0,1) \tag{17}$$

 X'_{h3} is the mutated food source generated after Cauchy mutation and μ_c is the Cauchy mutation rate that indicates the strength of the mutation. The mutated food sources along with the replaced worst food sources A_x forms the new set of food source A_{new} eligible for exploitation. The modified antibody detection phase is described by the algorithm (4).

Algorithm 4: Modified antibody detection phase

Replace the worst 'x' food source by randomly generated food source A_x

For each better food source (A_k)

Select A_{k1} , A_{k2} , and A_{k3} food sources from better food sources.

 A_{k1} if {|Fitness (Ak) – Fitness (A_{best}) |} \leq limit1

 A_{k2} if {|Fitness (A_k) – Fitness (A_{best}) |} \leq limit2 and if {| Fitness (A_k) – Fitness (A_{best}) |} \geq limit1

 A_{k3} if {|Fitness (A_k) – Fitness(A_{best}) |} $\geq limit2$

```
If A_{k1} food sources
```

```
Mutate using equation (15)
```

```
Else If A_{k2} food sources
```

```
Mutate using equation (16)
```

Else

Maturate using equation (17)

Set of mutated food sources generated are $\overline{A_k}$

Set $A_{new} = \overline{A_k}$

End for

Generate $A_{new} = A_{new} \cup A_x$

3.2.3 Proposed Employee bee attacking phase

The food sources generated from the mutative exploration phase along with the replaced randomly generated food sources form the initial food source positions of the exploitation phase. Each food source is assigned with an employee bee where a bee searches a better food source around the current food source following the same way a whale search for the prey. An employee bee follows a circular path during searching for a neighborhood food source where the best food source found so far is considered as the center of the circle. This exploitative mechanism makes the search process to be always guided by the best optimal locations. The step size β of the search process is kept high at the initial stages of the iteration on the circular path to facilitate the process of exploration whereas in the later iterations the step size is gradually decreased to facilitate the process of exploitation. This dynamic step size

 β of the search process makes the employee bees to search the entire problem space such that the bees can reach remote locations that were not reached in the exploration phase. The problem of oscillations and local optima can be eliminated with the dynamic step size in the circular path. A random number *ran* is generated using uniform distribution and checked against the control variables C_1 and C_2 . Then, if a randomly produced number less than C_1 then the food source position is found using Equation (18).

$$N_{k}^{l} = A_{k}^{l} + \beta * (A_{k}^{l} - A_{d}^{l})$$
(18)

Where β = random (-1,1) is the step size that is dynamically varied across the iterations. N_k^l is the new food source. A_k^l is the current food source, A_d^l is the randomly selected food source. Then if the randomly produced number is less than C_2 and a probability check is done using P_i where if $P_i < 0.5$ then new neighborhood position is found searching around a circular path keeping the best food source found so far A_k^* as the center of the circle using Equation (19).

$$N_{k}^{l} = \vec{A}_{k}^{*} - \vec{M}.Y$$
(19)

M represents the coefficient vectors obtained using Equation (3 & 4). Y is the distance between the current food source and the center of the circle (best food source) represented in the Equation (20).

$$\mathbf{Y} = \begin{bmatrix} A_k^* - A_k^l \end{bmatrix} \tag{20}$$

Another probability check, if $P_i \ge 0.5$ then new neighborhood position is found searching around a circular path keeping the best food source A_k^* found so far as the center of the circle using the Equation (21)

$$N_k^l = Y. e^{aw}. \cos(2\pi l) + A_k^*$$
(21)

a is a constant that represents the shape of the logarithmic spiral. w is the random number generated between [-1,1].

If fitness of N_k^l (neighborhood food source) is lesser than fitness A_k^l (current food source), the proposed employee bee attacking phase accepts N_k^l by accepting downhill movements to make the search process to escape from its local optimum in the fitness landscape. The acceptance of the worst food sources is based on the probability value $e^{\frac{-\Delta E}{t}}$. This is done using a simulated annealing-based selection mechanism where better and worst solutions are accepted in the proposed employee bee attacking phase. Acceptance or rejection of worst solutions is based on the controlling parameter called the simulated annealing temperature where the probability of accepting the worst solutions decreases as the iteration proceeds where the temperature gets decreased. At the initial stages the simulated annealing

temperature 'T' is set to have a high value where the value $\frac{-\Delta E}{T}$ tends towards 0 making the probability value $e^{\frac{-\Delta E}{t}}$ towards 1 allowing the acceptance of the worst solutions. As the iteration grows, the value $\frac{-\Delta E}{T}$ tends towards 1 making the probability value $e^{\frac{-\Delta E}{t}}$ towards 0 allowing the acceptance of better solutions. This makes the proposed HAW algorithm to allow both the uphill and downhill movements in the fitness landscape making the algorithm to eliminate the problem of striking at local optimum locations. When the iterations grow, the simulated annealing temperature 'T' is cooling down using the Equation (22).

$$T(t+1) = \emptyset * T(t) \tag{22}$$

Where, T(t + 1) is the new temperature and T(t) is the temperature of the previous iteration. ' \emptyset ' is the simulated annealing constant which is set close to 1. The proposed employee bee attacking phase is described by the algorithm (5).

Algorithm 5: Proposed Employee bee attacking phase

For k=1 to n do For l=1 to dim do If $ran \ge C_1$ Search for the neighbouring food source using Equation (18) Else If $ran < C_2$ If $(P_i < 0.5)$ Search for the neighbouring food source using Equation (19) Else *if* $(P_i \ge 0.5)$ Search for the neighbouring food source using Equation (21) End if

End if

End for *l*

If $fitness(N_k^l) > fitness(A_k^l)$

```
Accept N_k^l
```

Else

Calculate $\Delta E = fitness (A_k^l) - fitness(N_k^l)$ Accept N_k^l with probability $e^{\frac{-\Delta E}{t}}$

End if

End for k

3.2.4 Onlooker Bee Phase

The information about the newly generated food sources N_k^l is shared with onlooker bees. Further, onlooker bees produce probability value VS_i for a food source, it receives from the employee bee attacking phase using Equation (23).

$$VS_k = \frac{fitness(A_k)}{\sum_{k=1}^{N} fitness(Y_i)}$$
(23)

Where *fitness* (A_k) is the quality of food source A_k . A random (0,1) is generated and compared with VS_k for each food source. Those food sources whose VS_k is greater than random (0,1) is selected by onlooker bees. The detailed description for the onlooker bee phase given using Algorithm (6)

Algorithm 6: Onlooker Bee Phase

For k=1 to N do

```
Calculate probability using Equation (3)
```

```
If random(0,1) < VS_k
```

Accept N_k^l

Else

Reject N_k^l

End for

Memorize the best food source A_{best}

3.2.4 Scout Bee Phase

The abandoned food sources are replaced and new food sources are introduced by the scout bees. In each iteration, if the food source isn't improved then the limit value associated with the food source is incremented and if the food source doesn't improve for certain iterations and crosses the threshold limit value then those food sources are replaced by the scout bees by random generation process using Equation (1). The detailed description of the scout bee phase is given in the Algorithm (7).

Algorithm 7: Scout Bee Phase

```
While (A_k does not improve)
```

 $\operatorname{count} = \operatorname{count} + 1$

If count > $count_{max}$

Replace A_k with the newly generated food source

Else

count = 0

Iteration=Iteration+1

4 Performance Evaluation of the proposed HAW Optimization

The HAW algorithm is used to generate the optimal feature subsets, initial weights, and hidden node size of an ANN which is trained using three different learning mechanisms based on backpropagation such as Resilient Backpropagation (RP), Levenberg Marquardet backpropagation (LM), and momentum-based Gradient Decent backpropagation (GD). The HAW performance is analysed to find out the backpropagation variant that achieves the least ANN error with low complexity using WBCD. Further, the best backpropagation variant that is selected is used for training the ANN when tested for the rest of the datasets. The HAW optimized ANN is analysed for ten independent runs using various generation sizes and the mean accuracy, mean connections, and mean computational time are calculated.

4.1 Parameter Settings and Experimental Setup using WBCD

The proposed wrapper architecture was implemented utilizing MATLAB 8.5 software. Backpropagation training is done using a neural network toolbox. The implementation of backpropagation is done using default training parameters. The winner takes all approach in output nodes is used for classification. The parameter settings of the ANN classifier are shown in Table (2). HAW utilized the cancer1 dataset [55] which contains pre-processed WBCD samples. The values in between 0 to 1 are rescaled and attribute values that are missing are filled using mean values of non-missing attributes. The HAW parameter settings are shown in Table (3).

Parameter	Value
Training	Backpropagation
Input nodes size	9
Output nodes size	2
Initial weights (Number of bits)	15
Hidden node size (Number of bits)	2
Input features (Number of bits)	9
Activation function (hidden node)	Hyperbolic Tangent
Activation function (output node)	Pure Linear
Training set (no of samples)	349 (50%)
Validation set (no of samples)	175 (25%)
Testing set (no of samples)	175 (25%)

Table 2: Parameter settings of ANN

Fable 3:	Parameter	settings	of HAV	N
-----------------	-----------	----------	--------	---

Parameter	Value
Employee Bees (size)	30
Onlooker Bees (size)	30
Scout bees (size)	1
Total colony (size)	60
Count limit (<i>Count_{max}</i>)	10

Better food source selection (A_k)	80%
Worst food source selection (A_x)	20%
Control Threshold (limit1)	0.0001
Control Threshold (limit2)	0.5
Mutation rate (μ_{g})	0.4
Mutation rate (μ_c)	0.6

4.2 Evaluation of Proposed HAW using WBCD with respect to accuracy, complexity and computational time

The proposed HAW is evaluated for the complexity in terms of the number of connections and accuracy using ten runs for generation sizes 10, 20, 30 as given in Table (4). HAW-RP achieved the highest mean accuracy of 99.25% at the 20th generation size. The mean connections were 10.40 for HAW-RP, which is low when compared to HAW-LM and HAW-GD. HAW gained the less complexity ANN trained with RP. The validation error convergence of the optimized ANN network for RP, LM, and GD for various generations is depicted in Figure (3).

	Proposed			Proposed			Proposed					
Max	Max HAW-RP		HAW-LM			HAW-GD						
Generation	Classi	fication	Nu	mber of	Classi	fication	Nun	nber of	Classi	fication	Nun	nber of
Size	accura	acy (%)	con	nections	accur	acy (%)	con	nections	accur	acy (%)	coni	nections
	Best	Mean	Best	Mean	Best	Mean	Best	Mean	Best	Mean	Best	Mean
10	98.42	97.55	11	12.56	97.75	96.73	11	16.42	97.83	96.91	11	14.65
20	99.55	99.25	10	10.40	97.95	96.71	12	15.34	98.12	97.38	11	14.82
30	99.15	98.48	11	12.23	98.45	98.22	13	14.46	98.91	98.57	12	14.21

Table 4: Performance of the proposed HAW algorithm for various backpropagations

Followed by HAW-RP, HAW-GD is 98.57% accurate, with average connections of 14.21 in the 30th generation. Next to HAW-GD, HAW-LM achieved high accuracy of 98.22% in the 30th generation with average connections of 15.46. The classification accuracy of the optimized ANN network for RP, LM, and GD for various generations is depicted in Figure (4). The accuracy of HAW-RP was 1.05% more than HAW-LM and 0.69% more than HAW-GD.

The confusion matrix based on the True Positive, True Negative, False Positive, and False Negative of HAW-RP, HAW-LM, HAW-GD for the best network achieved is shown in Table (5). Figure (5) gives the complexity of the ANN network achieved in terms of the number of connections for various generation sizes. HAW-RP achieved lower complexity with lesser connections, followed by HAW-GD and HAW-LM. The mean hidden node count of HAW-

RP was 22.5% less than HAW-LM and 20.78% less than HAW-GD. HAW-RP produced ANN complexity of having 28.08% fewer connections than HAW-LM and 26.81% fewer connections than HAW-GD.



Figure 3: Convergence of validation error for HAW



Figure 4: Performance of HAW with respect to classification accuracy Table 5: Confusion Matrix of the HAW Optimized ANN network across ten runs

Methods of	Methods of A stual		Test outcome-Predicted			
comparison	Actual	Number of cases	Malignant	Benign		
HAWADD	Malignant	650	647(TP)	10 (FN)		
HAW-RP	Benign	1100	3 (FP)	1090(TN)		
HAW-LM	Malignant	650	640(TP)	21(FN)		
	Benign	1100	10 (FP)	1079(TN)		
HAW-GD	Malignant	650	642(TP)	17 (FN)		



Figure 5: Performance of HAW with respect to the number of connections

The performance metrics of the optimized ANN network generated using HAW-RP, HAW-LM, HAW-GD are shown in Table (6). HAW-RP showed higher sensitivity and specificity, followed by HAW-GD and HAW-LM. Figure (6) shows the hidden node count for HAW-RP, HAW-LM, and HAW-GD over generation sizes.

Metrics	Proposed HAW-RP	Proposed HAW-LM	Proposed HAW-GD
Sensitivity (%)	98.47	96.82	97.42
Specificity (%)	99.72	99.08	99.26
Accuracy (%)	99.25	98.22	98.57
Precision (%)	99.53	98.46	98.76
Negative predictive Value (NPV) (%)	99.09	98.09	98.45
F-measure	0.99	0.97	0.98

Table 6: Performance of HAW Optimized ANN network based on different metrics

Table (7) gives the importance of feature selection on the performance of the proposed HAW optimized ANN network in terms of classification. From Table (7), it can be concluded that HAW optimized ANN with feature selection improved prediction accuracy with less complexity when compared with the performance without the feature selection.

Table 7: Performance of HAW-RP based on feature selection

Max Generation Size	Feature Selection (FS)	Average Hidden Node Count	Average Number of Selected Features	Average Number of connections	Average Accuracy (%)
10	With FS	1.4	5.8	12.56	97.55
10	Without FS	2.5	9	27.5	88.34

20	With FS	1.2	5	10.40	99.25
20	Without FS	2.2	9	26.6	90.42
20	With FS	1.3	5.1	12.23	98.48
30	Without FS	2.2	9	24.8	89.05

Table (8) shows the confusion matrix for best network achieved using HAW-RP, with selected features such as Uniformity of cell size, Single epithelial cell size, Bare nuclei.

Table 8: Confusion Matrix of HAW-RP optimized best ANN with features selected

Feature Selection			Predicted Cases			
	Actual Cases		Benign	Malignant	Selected Feature Set	
With Feature	Benign	110	110	1	Uniformity of cell size, Single	
Selection	Malignant	65	0	64	epithelial cell size, Bare nuclei	
	Benign	110	102	7	Clump Thickness, Uniformity of cell shape	
Without Feature Selection	Malignant	65	8	58	Marginal adhesion, Single epithelial cell size, Bare nuclei, Bland chromatin, Normal nucleoli, Mitoses	



Figure 6: Evolution of hidden node count with feature selection for HAW across different generations

The evaluation of HAW optimized ANN with respect to computational time is given in Table (9). The mean computational time for 10 runs was estimated for various sizes of 10, 20, and 30. The computational time of ANN optimized using HAW-LM was low in comparison to the computational time of HAW-RP and HAW-GD.



Maximum	Average CPU Time(s)					
Generation Size	Proposed HAW-RP	Proposed HAW-LM	Proposed HAW-GD			
10	427.5	301.5	502.8			
20	820.2	789.6	845.9			
30	1105.9	1036.5	1254.1			

Figure (7) compares HAW with PSO, DE, ABC, BA, ACO, BFO, DA, GA, AIS, MBO, WOA, and SSA using the WBCD. HAW attained highest accuracy, which was 9.67%, 7.53%, 10.89%, 12.27%, 12.78%, 8%, 10.03%, 0.97%, 8.83%, 9.91%, 5.03%, and 3.39% respectively higher than when compared to above mentioned algorithms.



Figure 7: Comparison between evolutionary methods using WBCD dataset

Figure (8) compares HAW and existing hybrid algorithms that are ABC-based using WBCD, which includes ABC-ACO [43], ABC-DE [44], ABC-DA [56], and ABC-Gradient Decision Tree [46]. The accuracy of HAW was 0.07%, 1.58%, 2.36%, and 2.02% respectively higher when compared to the above-mentioned algorithms.



Figure 8: Comparison of ABC-based hybrid algorithms using WBCD datasets

Table (10) gives a comparison of the proposed HAW optimized classifier with existing breast cancer diagnosis schemes using WBCD datasets. Based on Table (10), it is concluded that the proposed Optimized ANN classifier using HAW outperformed existing breast cancer diagnosis schemes.

WBCD Dataset					
First author, Year	Method	Accuracy (%)			
Quinlan,1996 [53]	C4.5 classifier	94.74			
Hamilton,1996 [54]	RAIC classifier	95.00			
Nauck,1999 [55]	NEFCLASS	95.06			
Pena-Reyes,1999 [56]	FUZZY-GA classifier	97.36			
Setino, 2000 [57]	Neuro-rule classifier	98.10			
Albrecht, 2002 [58]	LSA classifier Machine	98.80			
Fogel, 1995 [59]	ENN classifier	98.05			
Abonyi, 2003 [60]	SFC classifier	95.57			
Polat, 2007 [61]	LS-SVM classifier	98.54			
Gujaro-Berdinas, 2007 [62]	LIS classifier	96.00			
Karabatak, 2009 [8]	AR+ANN classifier	97.40			
Monirul, 2010 [63]	CAFS+ANN classifier	98.76			
Stoean, 2013 [64]	SVM+EA classifier	97.07			
Fadzil Ahmad, 2014 [65]	GANN-ANN classifier	98.29			
Shunmugapriya, 2017 [43]	ABC-ACO-J48 classifier	99.04			
H. Rao, 2018 [46]	ABC+ GD classifier	92.8			
Ghanem, 2018 [56]	ABC-DA+ANN classifier	98.29			
Karthik, 2018 [70]	DNN classifier	98.62			
Proposed Work	HAW-ANN classifier	99.25			

Table 10: Comparison with existing breast cancer diagnosis schemes

4.3 Performance evaluation of HAW using different breast cancer datasets

The proposed HAW-RP yielded the best ANN networks with high accuracy and low complexity. Hence, HAW-RP was tested for the rest of the breast cancer datasets taken for investigation. Table (11) shows the performance of the HAW-RP Optimized ANN network for different breast cancer datasets. The average classification accuracy, F-measure, number of connections, and computational time has been calculated using ten different independent runs for different generation sizes and the best of the average values has been recorded. The proposed HAW algorithm optimized ANN model namely HAW-RP has shown better performance with a smaller number of connections in comparison with standard ABC and WOA for the datasets taken for investigation.

Datasets	Average classification Datasets accuracy (%)		Average F-measures		Average number of connections		Average computational time (Secs)					
	ABC	WOA	HAW-RP	ABC	WOA	HAW-RP	ABC	WOA	HAW-RP	ABC	WOA	HAW-RP
WBCD	89.5	94.5	99.2	0.8812	0.9327	0.990	25.21	28.11	10.40	640.6	700.5	820.2
WDBC	87.5	91.5	98.5	0.8648	0.9005	0.975	24.67	22.17	19.82	590.3	600.2	750.2
WPBC	86.5	93.5	96.3	0.8532	0.9206	0.953	24.89	23.19	18.52	300.1	350.7	425.5
DDSM	87.2	92.5	98.8	0.8641	0.9187	0.981	36.23	37.87	16.8	535.6	600.2	700.6
MIAS	89.4	91.2	98.7	0.8894	0.9072	0.978	33.78	35.78	25.34	450.2	500.9	656.8
INBREAST	86.3	89.3	99.1	0.8570	0.8831	0.986	30.82	31.12	25.12	400.1	455.7	600.4

Table 11: Performance of HAW-RP Optimized ANN network for different datasets

The Figure (9 & 10) shows the graphical representation of the performance of the standard ABC and WOA optimized ANN model and the proposed HAW optimized ANN model for various breast cancer datasets taken for investigation with respect to classification accuracy and an average number of connections. The HAW optimized ANN model has shown higher accuracy and low complexity in comparison to the standard ABC and WOA optimized ANN model.



Figure 9: Comparison based on classification accuracy for breast cancer datasets



Figure 10: Comparison based on number of connections for breast cancer datasets Table (12) shows the confusion matrix with a frequently selected feature set for the various datasets used for investigation of the proposed HAW-RP based ANN after the execution of ten independent runs.

	W	DBC datas	et		Frequently selected features	
Methods of	Actual	Number	Test outcome- Predicted		Texture, Radius, Compactness Concavity, Fractal dimension	
comparison		of cases	Malignant	Benign		
HAW-RP	Malignant	420	415(TP)	16 (FN)		
	Benign	1000	5 (FP)	984(TN)		
	W	PBC datas	et		Radius, Texture, Area, Concavity,	
			Test outcome-		symmetry, fractal dimension	
Methods of	Actual	Number	Pred	icted		
comparison	Actual	of cases	Recurrent	Non- Recurrent		
HAW-RP	Recurrent	190	183(TP)	11 (FN)		
	Non-	300	7 (FP)	289(TN)		
	Recurrent					
	D	DSM datas	et		Skewness, Kurtosis, Correlation, Contrast,	
Mothods of	Number		Test ou	itcome-	Area, Major Axis length, Skeleton	
comparison	Actual	of cases	Predicted	Predicted		
comparison		of cuses	Malignant	Benign		
HAW-RP	Malignant	400	392(TP)	7(FN)		
	Benign	900	8(FP)	893(TN)		
	N	IIAS datas	et		Solidity, Perimeter, Extent, Skewness,	
Methods of	Actual	Number	Test outcome- Predicted		Entropy, Mean, Variance, Eccentricity	
comparison		UI Cases	Malignant	Benign		
HAW-RP	Malignant	250	247(TP)	8(FN)]	
	Benign	600	3(FP)	592(TN)		
	INB	REAST da	taset			

Table 12:	Confusion	Matrix	after ten	runs for	various	datasets
1 4010 120	Comasion	TARGET THE	witcer cerr	I WIND I'VI		a a construction of the second

Methods of	Actual	Number	Test outcome- Predicted		Skewness, Kurtosis, Correlation, Area, Major Axis length, Convex Area,
comparison		of cases	Malignant	Benign	Eccentricity
HAW-RP	Malignant	290	287(TP)	3(FN)	
	Benign	400	3(FP)	397 (TN)	

Table (13) shows the performance comparison of the proposed HAW optimized ANN model in terms of average classification accuracy for ten independent runs with various classifiers for different datasets. The proposed HAW-ANN model has shown higher accuracy compared with existing classifiers for all the datasets taken for investigation.

Average classification accuracy (%) Datasets SVM **Naive Bayes Random Forest** MLP HAW-ANN WBCD 92.3 83.4 85.9 88.6 99.2 WDBC 87.5 98.5 89.7 88.4 82.4 WPBC 84.5 89.1 84.9 88.1 96.3 DDSM 92.2 80.9 82.5 92.2 98.8 MIAS 80.3 77.9 98.7 88.4 86.2 **INBREAST** 88.3 82.6 88.2 91.1 99.1

Table 13: Average classification accuracy for various classifiers

4.4 Comparison with existing approaches using WDBC

Figure (11) compares HAW with PSO, DE, ABC, BA, ACO, BFO, DA, GA, AIS, MBO, WOA, and SSA using the WDBC. Each algorithm is executed for 10 runs various generation sizes 10, 20, and 30 and the best is taken for comparison. HAW based ANN attained highest accuracy, which was 7.65%, 6.83%, 12.96%, 17.12%, 8.84%, 5.69%, 9.2%, 7.42%, 12.96%, 11.3%, 7.65%, and 5.35% respectively higher than when compared to above mentioned algorithms.



Figure 11: Comparison with other evolutionary methods using WDBC dataset

Table (14) gives a comparison of the proposed HAW optimized classifier with existing breast cancer diagnosis schemes using WDBC datasets. Based on Table (6.10), it is concluded that the proposed Optimized ANN classifier using HAW outperformed existing breast cancer diagnosis schemes.

WDBC dataset					
First author, Year	Method	Accuracy (%)			
Bamakan, 2014 [71]	CFS -SVM classifier	87.84			
Bamakan, 2014 [71]	SVM-Filtered classifier	87.84			
Bamakan, 2014 [71]	Logistic Regression-CFS classifier	95.95			
Bamakan, 2014 [71]	Logistic Regression-Filtered classifier	96.62			
Xue, 2012 [72]	BPSO classifier	92.98			
Xue, 2014 [73]	PSO classifier	93.98			
Maldonado, 2011 [74]	KP-SVM classifier	97 ±0.9			
Maldonado, 2011 [74]	REF-SVM classifier	95.25±1.0			
Miao, 2011 [75]	Training-self classifier	85.12			
Miao, 2011 [75]	Training- Random classifier	83.54			
Luukka, 2006 [76]	LDA classifier	97.19			
Luukka, 2006 [76]	C4.5 classifier	94.06			
Luukka, 2006 [76]	DIMLP classifier	96.92			
Luukka, 2006 [76]	SIM classifier	98.2			
Razieh, 2015 [77]	PSO-KDE classifier	98			
Djellali, 2018 [51]	ABC-GA classifier	91.67			
Proposed Work	HAW-ANN classifier	98.5			

Table 14: Comparison with existing breast cancer schemes for WDBC dataset

4.5 Comparison with existing approaches using Wisconsin Prognosis Breast Cancer Dataset (WPBC)

Figure (12) compares HAW with PSO, DE, ABC, BA, ACO, BFO, DA, GA, AIS, MBO, WOA, and SSA using the WPBC. Each algorithm is executed for 10 runs various generation sizes 10, 20, and 30 and the best is taken for comparison. HAW based ANN attained highest accuracy, which was 7.6%, 6.06%, 11.33%, 13.03%, 6.41%, 9.31%, 10.44%, 5.02%, 10.44%, 7.6%, 2.99%, and 2.67% respectively higher when compared to above mentioned.



Figure 12: Comparison with other evolutionary methods using WPBC dataset

Table (15) gives a comparison of the proposed HAW optimized classifier with existing breast cancer diagnosis schemes using WPBC datasets. Based on Table (6.10), it is concluded that the proposed Optimized ANN classifier using HAW outperformed existing breast cancer diagnosis schemes.

WPBC dataset				
First author, Year	Method	Accuracy (%)		
Belciug, 2012 [78]	MLP-GA Hybrid classifier	81.11		
Belciug, 2012 [78]	MLP-BP classifier	60.21		
Belciug, 2012 [78]	PNN classifier	74.43		
Belciug, 2012 [78]	PCNN classifier	72.10		
Salama, 2012 [79]	SMO classifier	76.28		
Salama, 2012 [79]	J48 classifier	76.28		
Sridevi, 2014[80]	Correlation +Rough Set classifier	85		
Proposed Work	HAW-ANN classifier	96.3		

Table 15: Comparison with existing breast cancer diagnosis schemes for WPBC

4.6 Comparison with existing approaches using DDSM, MIAS and INBREAST databases

The Figure (13) shows the comparison of HAW with evolutionary algorithms such as PSO, DE, ABC, BA, ACO, BFO, DA, GA, AIS, MBO, WOA, and SSA. Each algorithm is executed for 10 runs various generation sizes 10, 20, and 30 and the best is taken for comparison. Using DDSM database, HAW based ANN achieved accuracy which is 7.98%, 7.16%, 13.3%, 17.48.1%, 9.17%, 6.01%, 9.53%, 7.74%, 13.3%, 11.64%, 6.81%, and 6.35% more than above mentioned algorithms respectively. Using MIAS database, HAW based ANN achieved highest accuracy which is 10.53%, 10.28%, 10.4%, 8.2%, 8.22%, 12.67%,

10.28%,7.17%, 9.42%, 10.28%, 8.22%, and 7.87% more than more than above mentioned algorithms respectively. Using INBREAST database, HAW based ANN achieved highest accuracy which is 7.6%, 9.14%, 14.83%, 9.26%, 5.99%, 11.47%, 8.66%, 5.99%, 8.07%, 12.36%, 10.97%, and 10.11% more than more than above mentioned algorithms respectively.

Table (16) gives a comparison of the proposed HAW optimized classifier with existing breast cancer diagnosis schemes using DDSM, MIAS, and INBREAST datasets. Based on Table (16), it is concluded that the proposed Optimized ANN classifier using HAW outperformed existing breast cancer diagnosis schemes.



Figure 13: Comparison with other evolutionary methods using DDSM, MIAS, and INBREAST

 Table 16: Comparison with existing breast cancer diagnosis schemes for DDSM, MIAS,

First author, Year	Method	Database	Accuracy (%)
Wang, 2014 [81]	SVM classifier	DDSM	92.74
Liu, 2013 [82]	SVM classifier	DDSM	93.00
Saki, 2013 [83]	OWBPE classifier	MIAS	89.28
Buciu, 2011 [84]	PSVM classifier	MIAS	96.43
Tahmasbi, 2011 [85]	ANN-MLP classifier	MIAS	96.43
Tahmasbi, 2010 [86]	ANN-MLP classier	MIAS	92.80
Zhang, 2012 [87]	SVM classifier	DDSM	72.00
Verma, 2010 [88]	SCNN classifier	DDSM	94.28
Verma, 2009 [89]	SCBDL classifier	DDSM	97.50
Rojas, 2009 [90]	Bayesian classifier	DDSM, MIAS	81
Dheeba, 2012 [91]	PSO-ANN classifier	MIAS	97.61
Dheeba, 2012 [92]	DE-WNN classifier	MIAS	97.84
Rahimeh, 2015 [93]	GA-CNN classifier	MIAS, DDSM	88.15, 94.99

and INBREAST datasets

Andrik, 2017 [94]	CED classifier	INBREAST	98.9 ± 0.4
Ribl, 2018 [95]	DNN classifier	INBREAST	95
		DDSM	98.8
Proposed Work	HAW-ANN classifier	MIAS	98.7
		INBREAST	99.1

4.6 Comparison with existing approaches that doesn't involve complex tuning process for breast cancer diagnosis

Table (17) gives a comparison of the proposed HAW-ANN scheme with existing breast cancer diagnosis schemes that involves either feature selection or parameter tuning process. Based on Table (17), it is concluded that having both feature selection and parameter optimization of ANN in parallel is vital and the proposed HAW-ANN outperformed existing breast cancer diagnosis schemes that has used either feature selection or the parameter tuning of the classifier involved.

First author, Year	Method	Database	Accuracy
			(%)
Ghanem, 2018 [56]	ABC-DA based feature selection +ANN	WBCD	98.29
H. Rao, 2018[56]	Feature selection + Decision Tree	WBCD	97.18
B. Xue, 2014 [73]	PSO based feature selection + KNN	WBCD	94.74
Liu, 2013 [82]	Feature selection + SVM	DDSM	93.00
Buciu, 2011 [84]	Feature reduction + Proximal SVM	MIAS	96.43
Dheeba, 2012 [91]	PSO based parameter tuning of ANN	MIAS	97.61
Dheeba, 2012 [92]	PSO based parameter tuning of WNN	MIAS	97.84
Rahimeh, 2015 [93]	GA based feature selection +CNN	MIAS,	88.15
		DDSM	94.99
Kermani, 1995 [96]	GA based feature selection +ANN	WBCD	94.7
Verma, 2009 [97]	GA based feature selection + ANN	DDSM	85
Abbass, 2002 [98]	PSO based parameter tuning of WNN	WBCD	98.1
M. Supriya, 2019 [100]	DA based feature selection + Gray wolf-	WBCD	99
	based parameter tuning of ANN		
Dhanya, 2020 [101]	Feature selection + KNN	WBCD	93.57
		WBCD	99.2
		WDBC	98.5
Proposed Work	HAW based simultaneous feature selection	WPBC	96.3
reposed ord	and parameter tuning + ANN	DDSM	98.8
		MIAS	98.7
		INBREAST	99.1

Table 16: Comparison with existing breast cancer diagnosis schemes

5 Conclusions

This paper proposed a hybrid algorithm by integrating the standard ABC with WOA for simultaneous feature subset selection and parameter optimization of ANN for breast cancer diagnosis. It is noted that in the exploration phase of HAW, the food source positions are found randomly with different mutation strategies whereas in the exploitation phase the bees search in the circular path moving guided by a goal point in the center which is a deterministic approach. These two combinations of approaches make HAW avoid local optimal problems and have a high convergence speed. The advantage of HAW lies in deriving an optimal ANN network by the optimal searching of hidden node size, the initial value of weights, and input features using a simple wrapper approach. Performance evaluation of HAW using backpropagation variants such as RP, LM, and GD is done for different breast cancer datasets. Based on the results achieved, RP achieved the best prediction accuracy with a low complexity ANN network. From the results of this study, it is concluded that having both feature selection and parameter optimization of ANN is vital. In comparison with existing published researches, the average accuracy attained by HAW is promising. In comparison with existing evolutionary algorithms, the proposed HAW algorithm was more accurate and less complex. Further, the limitations of the proposed HAW-ANN classifier are based on the "no free lunch theorem" which states that classifier trained on a particular feature set may not be used for other feature sets. Based on this, the proposed HAW optimized classifier that is trained on the breast cancer datasets has not been yet tested for other medical datasets and other high dimensional datasets. The proposed wrapper approach of the HAW-ANN scheme is complex and requires more computational time than the existing breast cancer scheme since it involves feature selection and parameter tuning process parallel.

REFERENCES

- [1] N. R. Pal, B. Bhowmick, S. K. Patel, S. Pal, and J. Das, "A multi-stage neural network aided system for detection of microcalcifications in digitized mammograms," Neurocomputing, vol. 71, no. 13-15, pp. 2625–2634, 2008.
- [2] https://www.breastcancer.org/symptoms/understand_bc/statistics
- [3] Population-Based Cancer report. http://ncrpindia.org/annual_reports.aspx [Online]
- [4] M. M. Eltoukhy, I. Faye, and B. B. Samir, "A comparison of wavelet and curvelet for breast cancer diagnosis in digital mammogram," Computers in Biology and Medicine, vol. 40, no. 4, pp. 384–391, 2010.
- [5] M. Zhao, C. Fu, L. Ji, K. Tang, and M. Zhou, "Feature selection and parameter optimization for support vector machines: A new approach based on genetic algorithm

with feature chromosomes," Expert Systems with Applications, vol. 38, no. 5, pp. 5197–5204, 2011.

- [6] M. F. Akay, "Support vector machines combined with feature selection for breast cancer diagnosis," Expert Systems with Applications, vol. 36, no. 2, pp. 3240–3247, 2009.
- [7] L. C. Jain, "Advances in design and application of neural networks," Neural Computing and Applications, vol. 19, no. 2, pp. 167–168, 2010.
- [8] M. Karabatak and M. C. Ince, "An expert system for detection of breast cancer based on association rules and neural network," Expert Systems with Applications, vol. 36, no. 2, pp. 3465–3469, 2009.
- [9] E. D. Übeyli, "Implementing automated diagnostic systems for breast cancer detection," Expert Systems with Applications, vol. 33, no. 4, pp. 1054–1062, 2007.
- [10] Storn R, Price K. Differential evolution-a simple and efficient heuristic for global optimization over continuous spaces. J Global Optim 1997; 11:341-59.
- [11] Rechenberg I. Evolution strategy: optimization of technical systems by means of biological evolution, 104. Stuttgart: Fromman-Holzboog; 1973.
- [12] Goldberg DE, Holland JH, "Genetic algorithms and Machine Learning," Machine Learning, vol. 3, no. 2-3, pp. 95–99, 1988.
- [13] Simon D. Biogeography-based optimization. Evol Comput IEEE Trans 2008; 12:702–13.
- [14] W.-Y. Lin, "A GA–DE hybrid evolutionary algorithm for path synthesis of four-bar linkage," Mechanism and Machine Theory, vol. 45, no. 8, pp. 1096–1107, 2010.
- [15] A. Sarkheyli, A. M. Zain, and S. Sharif, "Robust optimization of ANFIS based on a new modified GA," Neurocomputing, vol. 166, pp. 357–366, 2015.
- [16] A. H. Gandomi, X.-S. Yang, S. Talatahari, and S. Deb, "Coupled eagle strategy and differential evolution for unconstrained and constrained global optimization," Computers & Mathematics with Applications, vol. 63, no. 1, pp. 191– 200, 2012.
- [17] S. Punitha, A. Amuthan, and K. Suresh Joseph, "Enhanced Monarchy Butterfly Optimization Technique for effective breast cancer diagnosis," *Journal of Medical Systems*, vol. 43, no. 7, pp. 1–14, 2019.
- [18] Thompson and K. Suresh Joseph, "Particle Swarm Optimization-Based Energy Efficient Channel Assignment Technique for Clustered Cognitive Radio Sensor Networks," *The Computer Journal, Oxford University Press*, vol. 61, no. 6, pages 926-936, Jun. 2018.
- [19] S. Thompson and K. Suresh Joseph, "PSO assisted OLSR Routing for Cognitive Radio Vehicular Sensor Networks," in ACM International Conference on Informatics and Analytics, pages 1–8, Aug. 2016.
- [20] Colorni A, Dorigo M, Maniezzo V. Distributed optimization by ant colonies. In: Proceedings of the first European conference on artificial life; 1991.
 p. 134-42.
- [21] Eberhart RC, Kennedy J. A new optimizer using particle swarm theory. In: Proceedings of the sixth international symposium on micro machine and human science; 1995. p. 39-43.
- [22] Yang X-S. A new metaheuristic bat-inspired algorithm. In: Nature inspired cooperative strategies for optimization (NICSO 2010). Springer; 2010. p. 65-74
- [23] Karaboga D, Basturk B. A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm. J Global Optim 2007; 39:459-71

- [24] Mirjalili S, Mirjalili S, Hatamlou A. Multi-verse optimizer: a nature-inspired algorithm for global optimization. Neural Comput Appl 2015:1-19 /03/17 2015.
- [25] A. H. Gandomi and A. H. Alavi, "Krill herd: A new bio-inspired optimization algorithm," Communications in Nonlinear Science and Numerical Simulation, vol. 17, no. 12, pp. 4831–4845, 2012.
- [26] G.-G. Wang, S. Deb, and Z. Cui, "Monarch butterfly optimization," *Neural Computing and Applications*, vol. 28, no. 3, pp. 1–20, 2015.
- [27] P. K. M, "Biomimicry of bacterial foraging for distributed optimization and control," *IEEE Control Systems*, vol. 22, no. 3, pp. 52–67, 2002.
- [28] Castro, L. & Timmis, "Artificial immune systems as a novel soft computing paradigm", *Journal of Soft Computing*7: 526,2003.
- [29] A. H. Gandomi, "Interior search algorithm (ISA): A novel approach for global optimization," ISA Transactions, vol. 53, no. 4, pp. 1168–1183, 2014.
- [30] S. Mirjalili, A. H. Gandomi, S. Z. Mirjalili, S. Saremi, H. Faris, and S. M. Mirjalili, "Salp Swarm Algorithm: A bio-inspired optimizer for engineering design problems," *Advances in Engineering Software*, vol. 114, pp. 163–191, 2017.
- [31] A. Faramarzi, M. Heidarinejad, S. Mirjalili, and A. H. Gandomi, "Marine Predators Algorithm: A Nature-inspired Metaheuristic," *Expert Systems with Applications*, p. 113377, 2020.
- [32] D. Furundzic, M. Djordjevic, and A. J. Bekic, "Neural networks approach to early breast cancer detection," Journal of Systems Architecture, vol. 44, no. 8, pp. 617–633, 1998.
- [33] J.F. Kolen, J.B. Pollack, "Back propagation is sensitive to initial conditions," Adv Neural Information Process System, vol 3, pp. 860–867, 1991.
- [34] K. P. Ferentinos, "Biological engineering applications of feedforward neural networks designed and parameterized by genetic algorithms," Neural Networks, vol. 18, no. 7, pp. 934–950, 2005.
- [35] R. Setiono and H. Liu, "Neural-network feature selector," IEEE Transactions on Neural Networks, vol. 8, no. 3, pp. 654–662, 1997.
- [36] A. Verikas and M. Bacauskiene, "Feature selection with neural networks," Pattern Recognition Letters, vol. 23, no. 11, pp. 1323–1335, 2002.
- [37] M. M. Kabir, M. M. Islam, and K. Murase, "A new wrapper feature selection approach using neural network," Neurocomputing, vol. 73, no. 16-18, pp. 3273–3283, 2010.
- [38] A. Telikani, A. H. Gandomi, A. Shahbahrami, and M. N. Dehkordi, "Privacypreserving in association rule mining using an improved discrete binary artificial bee colony," Expert Systems with Applications, vol. 144, p. 113097, 2020.
- [39] S. Mirjalili and A. Lewis, "The Whale Optimization Algorithm," *Advances in Engineering Software*, vol. 95, pp. 51–67, 2016.
- [40] H. Hasan and N. M. Tahir, "Feature selection of breast cancer based on Principal Component Analysis," 2010 6th International Colloquium on Signal Processing & its Applications, 2010.
- [41] R. Alyami, J. Alhajjaj, B. Alnajrani, I. Elaalami, A. Alqahtani, N. Aldhafferi, T. O. Owolabi, and S. O. Olatunji, "Investigating the effect of correlation based feature selection on breast cancer diagnosis using artificial neural network and support vector machines," 2017 International Conference on Informatics, Health & Technology (ICIHT), 2017.
- [42] M.-L. Huang, Y.-H. Hung, and W.-Y. Chen, "Neural Network Classifier with Entropy Based Feature Selection on Breast Cancer Diagnosis," *Journal of Medical Systems*, vol. 34, no. 5, pp. 865–873, 2009.

- [43] P. Shunmugapriya and S. Kanmani, "A hybrid algorithm using ant and bee colony optimization for feature selection and classification (AC-ABC Hybrid)," *Swarm and Evolutionary Computation*, vol. 36, pp. 27–36, 2017.
- [44] E. Zorarpacı and S. A. Özel, "A hybrid approach of differential evolution and artificial bee colony for feature selection," *Expert Systems with Applications*, vol. 62, pp. 91–103, 2016.
- [45] S. Shanthi and V. M. Bhaskaran. "Modified artificial bee colony-based feature selection: A new method in the application of mammogram image classification," *International Journal of Science, Engineering and Technology Research (IJSETR)*, Vol 3(6), pp.1664-1667, 2014.
- [46] H. Rao, X. Shi, A. K. Rodrigue, J. Feng, Y. Xia, M. Elhoseny, X. Yuan, and L. Gu, "Feature selection based on artificial bee colony and gradient boosting decision tree," *Applied Soft Computing*, vol. 74, pp. 634–642, 2019.
- [47] H. Badem, A. Basturk, A. Caliskan, and M. E. Yuksel, "A new efficient training strategy for deep neural networks by hybridization of artificial bee colony and limited– memory BFGS optimization algorithms," *Neurocomputing*, vol. 266, pp. 506–526, 2017.
- [48] B. A. Garro, K. Rodríguez, and R. A. Vázquez, "Classification of DNA microarrays using artificial neural networks and ABC algorithm," *Applied Soft Computing*, vol. 38, pp. 548–560, 2016.
- [49] S. Palanisamy, S. Kanmani, "Artificial Bee Colony Approach for Optimizing Feature Selection,", *International Journal of Computer Science Issues*, vol.9, no.3, pp.432-438, 2012.
- [50] M. Schiezaro and H. Pedrini, "Data feature selection based on Artificial Bee Colony algorithm," *EURASIP Journal on Image and Video Processing*, vol. 2013, no. 1, 2013.
- [51] H. Djellali, A. Djebbar, N. G. Zine, and N. Azizi, "Hybrid Artificial Bees Colony and Particle Swarm on Feature Selection," *Computational Intelligence and Its Applications IFIP Advances in Information and Communication Technology*, pp. 93–105, 2018.
- [52] G. Nagarajan, R. Minu, B. Muthukumar, V. Vedanarayanan, and S. Sundarsingh, "Hybrid Genetic Algorithm for Medical Image Feature Extraction and Selection," *Procedia Computer Science*, vol. 85, pp. 455–462, 2016.
- [53] G. I. Sayed, A. Darwish, A. E. Hassanien, and J.-S. Pan, "Breast Cancer Diagnosis Approach Based on Meta-Heuristic Optimization Algorithm Inspired by the Bubble-Net Hunting Strategy of Whales," *Advances in Intelligent Systems and Computing Genetic and Evolutionary Computing*, pp. 306–313, 2016.
- [54] J. Jona and N. Nagaveni, "Ant-cuckoo Colony Optimization for Feature Selection in Digital Mammogram," *Pakistan Journal of Biological Sciences*, vol. 17, no. 2, pp. 266– 271, 2014.
- [55] L. Prechelt Proben1: a set of neural network benchmark problems and benchmarking rules. Technical Report, University of Karlsruhe, Karlsruhe, Germany, 1994
- [56] H. Rao, X. Shi, A. K. Rodrigue, J. Feng, Y. Xia, M. Elhoseny, X. Yuan, and L. Gu, "Feature selection based on artificial bee colony and gradient boosting decision tree," Applied Soft Computing, vol. 74, pp. 634–642, 2019.
- [57] J. R. Quinlan, "Improved Use of Continuous Attributes in C4.5," Journal of Artificial Intelligence Research, vol. 4, pp. 77–90, 1996.
- [58] Hamilton HJ, Shan N, Cercone N, RIAC: a rule induction algorithm based on approximate classification. "International conference on engineering applications of neural networks", University of Regina,1996.

- [59] D. Nauck and R. Kruse, "Obtaining interpretable fuzzy classification rules from medical data," Artificial Intelligence in Medicine, vol. 16, no. 2, pp. 149–169, 1999.
- [60] C. A. Peña-Reyes and M. Sipper, "A fuzzy-genetic approach to breast cancer diagnosis," Artificial Intelligence in Medicine, vol. 17, no. 2, pp. 131–155, 1999.
- [61] R. Setiono, "Generating concise and accurate classification rules for breast cancer diagnosis," Artificial Intelligence in Medicine, vol. 18, no. 3, pp. 205–219, 2000.
- [62] A.Albrecht, G. Lappas, S. Vinterbo, C. Wong, and L. Ohno-Machado, "Two applications of the LSA machine," Proceedings of the 9th International Conference on Neural Information Processing, 2002. ICONIP 02., pp. 184–189.
- [63] D. B. Fogel, E. C. Wasson, and E. M. Boughton, "Evolving neural networks for detecting breast cancer," Cancer Letters, vol. 96, no. 1, pp. 49–53, 1995.
- [64] J. Abonyi and F. Szeifert, "Supervised fuzzy clustering for the identification of fuzzy classifiers," Pattern Recognition Letters, vol. 24, no. 14, pp. 2195–2207, 2003.
- [65] K. Polat and S. Güneş, "Breast cancer diagnosis using least square support vector machine," Digital Signal Processing, vol. 17, no. 4, pp. 694–701, 2007.
- [66] B. Guijarro-Berdiñas, O. Fontenla-Romero, B. Pérez-Sánchez, and P. Fraguela, "A Linear Learning Method for Multilayer Perceptrons Using Least-Squares," Intelligent Data Engineering and Automated Learning - IDEAL 2007 Lecture Notes in Computer Science, vol. 4881, pp. 365–374.
- [67] M. M. Kabir, M. M. Islam, and K. Murase, "A new wrapper feature selection approach using neural network," *Neurocomputing*, vol. 73, no. 16-18, pp. 3273–3283, 2010.
- [68] R. Stoean and C. Stoean, "Modeling medical decision making by support vector machines, explaining by rules of evolutionary algorithms with feature selection," Expert Systems with Applications, vol. 40, no. 7, pp. 2677–2686, 2013.
- [69] F. Ahmad, N. A. M. Isa, Z. Hussain, M. K. Osman, and S. N. Sulaiman, "A GA-based feature selection and parameter optimization of an ANN in diagnosing breast cancer," Pattern Analysis and Applications, vol. 18, no. 4, pp. 861–870, 2014.
- [70] S. Karthik, R. S. Perumal, and P. V. S. S. R. Chandra Mouli, "Breast Cancer Classification Using Deep Neural Networks," *Knowledge Computing and Its Applications*, pp. 227–241, 2018.
- [71] S. M. H. Bamakan and P. Gholami, "A Novel Feature Selection Method based on an Integrated Data Envelopment Analysis and Entropy Model," *Procedia Computer Science*, vol. 31, pp. 632–638, 2014.
- [72] B. Xue, M. Zhang, and W. N. Browne, "New fitness functions in binary particle swarm optimisation for feature selection," 2012 IEEE Congress on Evolutionary Computation, pp. 1–8, 2012.
- [73] B. Xue, M. Zhang, and W. N. Browne, "Particle swarm optimisation for feature selection in classification: Novel initialisation and updating mechanisms," *Applied Soft Computing*, vol. 18, pp. 261–276, 2014.
- [74] S. Maldonado, R. Weber, and J. Basak, "Simultaneous feature selection and classification using kernel-penalized support vector machines," *Information Sciences*, vol. 181, no. 1, pp. 115–128, 2011.
- [75] D. Miao, C. Gao, N. Zhang, and Z. Zhang, "Diverse reduct subspaces-based co-training for partially labeled data," *International Journal of Approximate Reasoning*, vol. 52, no. 8, pp. 1103–1117, 2011.

- [76] P. Luukka and T. Leppälampi, "Similarity classifier with generalized mean applied to medical data," *Computers in Biology and Medicine*, vol. 36, no. 9, pp. 1026–1040, 2006.
- [77] R. Sheikhpour, M. A. Sarram, and R. Sheikhpour, "Particle swarm optimization for bandwidth determination and feature selection of kernel density estimation-based classifiers in diagnosis of breast cancer," *Applied Soft Computing*, vol. 40, pp. 113– 131, 2016.
- [78] S. Belciug and F. Gorunescu, "A hybrid neural network/genetic algorithm applied to breast cancer detection and recurrence," *Expert Systems*, vol. 30, no. 3, pp. 243–254, 2012.
- [79] G. I. Salama, M. B. Abdelhalim, and M. A.-E. Zeid, "Experimental comparison of classifiers for breast cancer diagnosis," 2012 Seventh International Conference on Computer Engineering & Systems (ICCES), 2012.
- [80] T. Sridevi and A. Murugan, "A Novel Feature Selection Method for Effective Breast Cancer Diagnosis and Prognosis," *International Journal of Computer Applications*, vol. 88, no. 11, pp. 28–33, 2014.
- [81] W. Wang, L.-J. Yang, Y.-T. Xie, and Y.-W. An, "Edge detection of infrared image with CNN_DGA algorithm," *Optik*, vol. 125, no. 1, pp. 464–467, 2014.
- [82] X. Liu and J. Tang, "Mass Classification in Mammograms Using Selected Geometry and Texture Features, and a New SVM-Based Feature Selection Method," *IEEE Systems Journal*, vol. 8, no. 3, pp. 910–920, 2014.
- [83] F. Saki, A. Tahmasbi, H. Soltanian-Zadeh, and S. B. Shokouhi, "Fast opposite weight learning rules with application in breast cancer diagnosis," *Computers in Biology and Medicine*, vol. 43, no. 1, pp. 32–41, 2013.
- [84] I. Buciu and A. Gacsadi, "Directional features for automatic tumor classification of mammogram images," *Biomedical Signal Processing and Control*, vol. 6, no. 4, pp. 370–378, 2011.
- [85] A. Tahmasbi, F. Saki, and S. B. Shokouhi, "Classification of benign and malignant masses based on Zernike moments," *Computers in Biology and Medicine*, vol. 41, no. 8, pp. 726–735, 2011.
- [86] A. Tahmasbi, F. Saki, and S. B. Shokouhi, "Mass diagnosis in mammographyimages using novel FTRD features," 2010 17th Iranian Conference of Biomedical Engineering (ICBME), 2010.
- [87] Y. Zhang, N. Tomuro, J. Furst, and D. S. Raicu, "Building an ensemble system for diagnosing masses in mammograms," *International Journal of Computer Assisted Radiology and Surgery*, vol. 7, no. 2, pp. 323–329, 2011.
- [88] B. Verma, P. Mcleod, and A. Klevansky, "Classification of benign and malignant patterns in digital mammograms for the diagnosis of breast cancer," *Expert Systems with Applications*, vol. 37, no. 4, pp. 3344–3351, 2010.
- [89] B. Verma, P. Mcleod, and A. Klevansky, "A novel soft cluster neural network for the classification of suspicious areas in digital mammograms," *Pattern Recognition*, vol. 42, no. 9, pp. 1845–1852, 2009.
- [90] A.Rojas-Dominguez and A. K. Nandi, "Development of tolerant features for characterization of masses in mammograms," *Computers in Biology and Medicine*, vol. 39, no. 8, pp. 678–688, 2009.

- [91] J. Dheeba and S. T. Selvi, "A Swarm Optimized Neural Network System for Classification of Microcalcification in Mammograms," *Journal of Medical Systems*, vol. 36, no. 5, pp. 3051–3061, 2011.
- [92] J. Dheeba, N. A. Singh, and S. T. Selvi, "Computer-aided detection of breast cancer on mammograms: A swarm intelligence optimized wavelet neural network approach," *Journal of Biomedical Informatics*, vol. 49, pp. 45–52, 2014.
- [93] R. Rouhi, M. Jafari, S. Kasaei, and P. Keshavarzian, "Benign and malignant breast tumors classification based on region growing and CNN segmentation," *Expert Systems with Applications*, vol. 42, no. 3, pp. 990–1002, 2015.
- [94] A. Rampun, P. J. Morrow, B. W. Scotney, and J. Winder, "Fully automated breast boundary and pectoral muscle segmentation in mammograms," *Artificial Intelligence in Medicine*, vol. 79, pp. 28–41, 2017.
- [95] D. Ribli, A. Horváth, Z. Unger, P. Pollner, and I. Csabai, "Detecting and classifying lesions in mammograms with Deep Learning," *Scientific Reports*, vol. 8, no. 1, 2018.
- [96] Kermani BG, White MW, Nagle HT (1995) Feature extraction by genetic algorithms for neural networks in breast cancer classification. In: Proceedings of the 17th annual conference on IEEE engineering in medicine and biology society, vol 831, pp 831– 832
- [97] Verma B, Zhang P (2007) A novel neural-genetic algorithm to find the most significant combination of features in digital mammograms. Applied Soft Computing 7(2):612–625.
- [98] Abbass HA (2002) An evolutionary artificial neural networks approach for breast cancer diagnosis. Artif Intell Med 25(3):265–281.
- [99] R. Dhanya, I. R. Paul, S. S. Akula, M. Sivakumar, and J. J. Nair, "F-test feature selection in Stacking ensemble model for breast cancer prediction," *Procedia Computer Science*, vol. 171, pp. 1561–1570, 2020.
- [100] M. Supriya and A. J. Deepa, "A novel approach for breast cancer prediction using optimized ANN classifier based on big data environment," *Health Care Management Science*, 2019.
- [101] R. Dhanya, I. R. Paul, S. S. Akula, M. Sivakumar, and J. J. Nair, "F-test feature selection in Stacking ensemble model for breast cancer prediction," *Procedia Computer Science*, vol. 171, pp. 1561–1570, 2020.