Tech Science Press

Adaptive Dynamic Dipper Throated Optimization for Feature Selection in Medical Data

Ghada Atteia[1], El-Sayed M. El-kenawy[2,3], Nagwan Abdel Samee[1,*], Mona M. Jamjoom[4], Abdelhameed Ibrahim[5], Abdelaziz A. Abdelhamid[6,7], Ahmad Taher Azar[8,9], Nima Khodadadi[10,11], Reham A. Ghanem[12] and Mahmoud Y. Shams[13]

[1]Department of Information Technology, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, P. O. Box 84428, Riyadh, 11671, Saudi Arabia
[2]Department of Communications and Electronics, Delta Higher Institute of Engineering and Technology, Mansoura, 35111, Egypt
[3]Faculty of Artificial Intelligence, Delta University for Science and Technology, Mansoura, 35712, Egypt
[4]Department of Computer Sciences, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, P. O. Box 84428, Riyadh, 11671, Saudi Arabia
[5]Computer Engineering and Control Systems Department, Faculty of Engineering, Mansoura University, Mansoura, 35516, Egypt
[6]Department of Computer Science, Faculty of Computer and Information Sciences, Ain Shams University, Cairo, 11566, Egypt
[7]Department of Computer Science, College of Computing and Information Technology, Shaqra University, 11961, Saudi Arabia
[8]Faculty of Computers and Artificial Intelligence, Benha University, Benha, 13518, Egypt
[9]College of Computer and Information Sciences, Prince Sultan University, Riyadh, 11586, Saudi Arabia
[10]Department of Civil Engineering, University of Science and Technology, Miami, 33101, USA
[11]Department of Civil and Environmental Engineering, Florida International University, Miami, USA
[12]Oral Biology Department, Faculty of Oral and Dental Medicine, Delta University for Science and Technology, Gamasa, Egypt
[13]Faculty of Artificial Intelligence, Kafrelsheikh University, Kafrelsheikh, 33511, Egypt
*Corresponding Author: Nagwan Abdel Samee. Email: nmabdelsamee@pnu.edu.sa
Received: 25 April 2022; Accepted: 08 June 2022

Abstract: The rapid population growth results in a crucial problem in the early detection of diseases in medical research. Among all the cancers unveiled, breast cancer is considered the second most severe cancer. Consequently, an exponential rising in death cases incurred by breast cancer is expected due to the rapid population growth and the lack of resources required for performing medical diagnoses. Utilizing recent advances in machine learning could help medical staff in diagnosing diseases as they offer effective, reliable, and rapid responses, which could help in decreasing the death risk. In this paper, we propose a new algorithm for feature selection based on a hybrid between powerful and recently emerged optimizers, namely, guided whale and dipper throated optimizers. The proposed algorithm is evaluated using four publicly available breast cancer datasets. The evaluation results show the effectiveness of the proposed approach from the accuracy and speed perspectives. To prove the superiority of the proposed algorithm, a set of competing feature

selection algorithms were incorporated into the conducted experiments. In addition, a group of statistical analysis experiments was conducted to emphasize the superiority and stability of the proposed algorithm. The best-achieved breast cancer prediction average accuracy based on the proposed algorithm is 99.453%. This result is achieved in an average time of 3.6725 s, the best result among all the competing approaches utilized in the experiments.

1 Introduction

The proper diagnosis of some crucial information is a critical challenge for bioinformatics or medical research [1]. In medicine, disease diagnosis is a demanding and challenging task. Many diagnostic hospitals and research institutes, as well as various websites, have a vast amount of medical diagnostic data. It is barely essential to categorize them to automate and speed up disease diagnosis. The medical planning officer's expertise and skill in the medical profession are frequently used to diagnose the ailment. As a result, there are instances of inaccuracies, unintended biases, and the requirement for a lengthy time for a precise disease diagnosis [2].

Breast cancer affects more women than all other malignancies, according to the American Cancer Society. In 2017, around 252,710 women in the United States will be diagnosed with invasive breast cancer, while approximately 63,410 women will be diagnosed with in situ breast cancer. Breast cancer is more common in men than in women. In 2017, it was estimated that roughly 2470 males in the United States would be diagnosed with prostate cancer. According to another estimate, around 41,070 people would die in 2017 from this malignancy. Breast cancer affects 41,000 women in the United Kingdom each year, but just 300 males. The most frequent malignancy among women globally is breast cancer. Breast cancer is caused by the abnormal multiplication of specific breast cells. Several techniques have been developed for the reliable diagnosis of breast cancer. Breast cancer is detected through mammography, often known as breast screening. The status of a woman's nipple is assessed using X-rays. Breast cancer is difficult to detect in its early stages because of the small size of the cancer cell when viewed from the outside. Mammography is a quick and painless procedure that can identify cancer in its early stages [3–5].

Ultrasound is a well-known tool for identifying breast cancer that uses sound waves to inspect the inside status of the body. The bounce of sound waves from a transducer that generates sound waves and is placed on the skin captures the echoes of the body's tissues. The echoes are transformed to grayscale, which is a binary value that can be understood by a computer. F-fluorodeoxyglucose positron emission tomography (PET) imaging helps doctors see where a tumor is in the body. It works by looking for radiolabeled cancer-specific tracers. Dynamic magnetic resonance imaging (MRI) has been used to develop a detection tool for breast abnormalities. The modality forecasts the pace at which contrast enhancement occurs by boosting tumor angiogenesis. Magnetic reasoning imaging has been associated with contrast enhancement metastases in those with breast cancer. Elastography is an imaging-based technique that is fresh new. This method can be employed when the breast cancer tissue is greater than the surrounding parenchyma. To identify benign from malignant types, this approach employs a color map of probe compression [6–8].

The mortality of breast cancer can be reduced effectively if the diagnosis and detection are performed early. However, the lack of sufficient medical resources and inaccurate treatment make

the timely diagnosis of breast cancer hard to achieve. Therefore, computer-aided diagnosis systems present an effective solution to radiologists and pathologists to help them achieve an efficient and accurate medical diagnosis. Currently, existing machine learning techniques offer accurate and effective assistance in diagnosing breast cancer. They can achieve low positive and negative classification rates. These machine learning techniques include ensemble learning, deep learning, computer vision, feature selection, and feature fusion. To generate highly accurate annotations for medical images, a substantial economic cost is usually required. In addition, the availability of a large set of samples is limited by individual privacy and ethical issues [9].

Moreover, the acquisition of high-quality samples is affected by the large gap between the fields of computer science and medicine. On the other hand, scientists tried to find the actual reasons causing breast cancer by studying the factors that can raise the probability of growing breast cancer in a woman. These risk factors include family background, age, genetic properties, smoking, gene variation, obesity, taking alcohol, etc. Based on these studies, four datasets have emerged as a collection of data describing the cases of normal and breast cancer based on these risk factors. These datasets are the breast cancer Wisconsin (diagnostic) data set, breast cancer Wisconsin (original) data set, breast cancer Coimbra data set, and breast cancer data set [10–13].

Prediction of breast cancer based on the data of these four datasets requires building a machine learning model that processes the risk factors to find the best decision. However, not all the risk factors have the same relevance in finding the best decision. Therefore, the process of feature selection of essential to keep only the most relevant risk factors in achieving better results. There are many feature selection methods in the literature. From these methods, the binary grey wolf optimization (bGWO), binary guided whale algorithm (bGWA), binary particle swarm optimization (bPSO), binary dipper throated optimization (bDTO), and many others. The efficiency of these feature selectors is usually evaluated in terms of the classifier that determines whether the selected feature can achieve the best prediction rates or not. On the other hand, these feature selection methods are based on parameter optimization algorithms that generate continuous values, converted to binary to match the feature selection problem [14].

Parameter optimization is defined as the problem of finding the best parameter set that can achieve the highest classification/prediction values. The optimization problem can be utilized in the problem of breast cancer prediction by employing the exploration and exploitation processes of the optimizer to find the best set of risk factors that can be used in breast cancer prediction [15]. In addition, many hybrid optimization algorithms recently emerged to perform better than the single optimizer—these hybrid optimizers such as PSO-GWA and PSO-GWO. Consequently, binary hybrid optimizers have emerged for better feature selection and better classification accuracy. A summary of the research papers that addressed the task of breast cancer diagnosis is listed in Table 1. As presented in this table, the best accuracy achieved was 97.89%, and the speed of the corresponding approach is not available.

Table 1: Summary of recent research papers on breast cancer diagnosis

Paper	Method	Dataset	Findings
[16]	The deep feature fusion method	Histopathology images dataset.	The maximum achieved accuracy is 97% for differentiating two combined groups of breast cancer classes.
[17]	Probabilistic graphical models and deep belief network	Ljubljana breast cancer dataset, breast cancer dataset, and Netherlands cancer institute dataset.	The authors employed a deep belief network and probabilistic graphical models.

(Continued)

Table 1: Continued

Paper	Method	Dataset	Findings
[18]	Deep learning	Breast cancer Wisconsin dataset from the University of California, Irvine (UCI) repository.	The maximum achieved accuracy is 96.99% accuracy using deep learning.
[19]	Machine learning techniques	Breast cancer Wisconsin dataset from UCI repository.	Authors proved that employing recent enhancements in machine learning achieved better results.
[20]	Deep learning	Digital database for screening mammography dataset.	The maximum achieved accuracy is 96% using deep learning.
[21]	Machine learning techniques	Mammogram images.	The maximum achieved accuracy is 96% using a deep neural network.
[22]	Machine learning techniques	Breast cancer Wisconsin dataset from UCI repository.	The authors applied machine learning techniques to the breast cancer dataset.
[23]	Support vector machines (SVM), k-nearest neighbors (KNN), Naïve bayes, and Decision tree.	Breast cancer from From UCI repository datasets.	The maximum achieved accuracy is 97.89% using support vector machines.
[24]	Deep learning	Mammographic image analysis database	The maximum achieved accuracy is 98% using convolutional neural network (CNN).
[25]	Deep learning CNN	BreakHis dataset is used	The maximum achieved accuracy is 95.4% using DenseCNN.
[26]	Data mining techniques	Breast cancer Wisconsin dataset from UCI repository	Authors could detect hidden cancer-associated for classification.
[27]	Random forest, KNN and Naïve bayes.	Breast cancer Wisconsin dataset from UCI repository.	Authors achieved high accuracy using the KNN classifier.

We propose in this paper a new feature selection algorithm for selecting the significant features from four breast cancer datasets publicly available in UCI Repository. The proposed algorithm is compared with a set of recently emerged feature selection algorithms in the literature. The achieved results proved the efficiency of the proposed algorithm in terms of speed and accuracy.

This paper is organized as follows. The materials and methods employed in this work are presented in Section 2. Section 3 illustrates the proposed feature selection algorithm, followed by a detailed explanation of the achieved experimental results in Section 4. The conclusions and future perspectives are finally presented in Section 5.

2 Materials and Methods

This section starts by presenting the datasets employed in the conducted experiments, then the basic algorithms used in developing and testing the proposed algorithm are explained.

2.1 Datasets

In this work, four datasets from the UCI repository are employed to evaluate the proposed approach. These datasets are the breast cancer data set, which is referred to as D1 [10], breast cancer data set, which is referred to as D2 [11], breast cancer Wisconsin (original) data set which is referred to as D3 [12], and breast cancer Wisconsin (diagnostic) data set which is referred as D4 [13]. The features of each dataset are listed in Table 2. As shown in the table, the first nine rows represent the dataset features (denoted by F1, F2, . . ., F9), whereas the last two rows represent the classes to which the feature values belong and the size of the dataset.

Table 2: Attributes of the four datasets incorporated in this research

	D1 [10]	D2 [11]	D3 [12]	D4 [13]
F1	Age	Age	Clump thickness	Radius
F2	Menopause	BMI	Uniformity of Cell size	Texture
F3	Tumor-size	Glucose	Uniformity of Cell shape	Perimeter
F4	Inv-nodes	Insulin	Marginal adhesion	Area
F5	Node-caps	HOMA	Single epithelial cell size	Smoothness
F6	Deg-malig	Leptin	Bare nuclei	Compactness
F7	Breast	Adiponectin	Bland chromatin	Concavity
F8	Breast-quad	Resistin	Normal nucleoli	Concave points
F9	Irradiat	MCP-1	Mitoses	Fractal dimension
Class	Recurrence?	Healthy, Patients	Malignant, benign	Malignant, Benign
Size	286 samples	116 samples	699 samples	569 samples

2.2 Preprocessing

Before using the data of each dataset to learn the machine learning model and feature selector, two main preprocessing steps are applied to ensure the integrity of the recorded features. These preprocessing steps replace the null values in a column with the average of the surrounding values in the same feature column. In addition, a min-max scalar is employed to normalize the importance of the features to lay in the range between zero and one [28,29].

2.3 Dipper Throated Optimization Algorithm

The inspiration for the dipper throated optimization (DTO) algorithm is based on the dipper throated bird, which is renowned for its bobbing or dipping motions while perched and belongs to the Cinclidsae family of birds. The ability of a bird to dive, swim, and hunt beneath the surface sets it apart from other passerines. It can fly straight and rapidly with no stops or glides because of its tiny flexible wings. The dipper throated bird has a distinct hunting style, quick bowing motions, and a white breast. It rushes headlong into the water to get its prey, regardless of how turbulent or fast-flowing it is. As it descends and picks up pebbles and stones, aquatic invertebrates, aquatic insects, and tiny fish perish. The great white shark uses its hands to move on the ocean floor. By bending your body at an angle and traveling down the bottom of the water with your head lowered, you might be able to locate prey. It can also dive into the water and submerge itself, using its wings to propel itself through the water and stay submerged for an extended time. The DTO algorithm assumes that a flock of birds is swimming and flying about looking for food. More details about the DTO algorithm are explained in [30].

2.4 Guided Whale Optimization Algorithm

The guided whale (GW) optimization approach is effective in various optimization situations. According to the literature, GW is one of the most successful optimization approaches. On the other hand, its exploration capabilities may be restricted. For mathematical computations, let n as the number of variables or the dimension of the search space to be searched by the whales. If the agents' positions (solutions) in the space search are updated over time, the best food solution will be determined. The update of the agents' positions can be performed using the following equation.

$$\vec{W}(t+1) = \vec{W}^*(t) - \vec{A} \cdot \vec{D} \tag{1}$$

$$\vec{D} = \left| \vec{C} \cdot \vec{W}^*(t) - \vec{W}(t) \right|$$

where the found solution at iteration t is denoted by $\vec{W}(t)$. The position of the optimal solution is referred as $\vec{W}^*(t)$. The agent's new position is denoted by $\vec{W}(t+1)$. $\vec{A}$ and $\vec{C}$ are two vectors calculated as follows: $\vec{A} = 2\vec{a} \cdot r_1 - \vec{a}$, and $\vec{C} = 2 r_2$, where $\vec{a}$ is a vector whose values are linearly changed from 2 to 0, and r_1, r_2 are randomly changing between 0 and 1. More details about this algorithm can be found in [31,32].

2.5 Feature Selection

The selection of essential features from the recorded features set is the first step in adequately training classification models. As a result, training time may be cut in half, and the classification model can be simplified significantly. Furthermore, by employing feature selection, overfitting may be prevented. To execute feature selection, the output solution from the continuous optimizer should be changed to a binary solution using the integers 0 and 1. To convert a continuous optimizer solution to a binary solution for feature selection, use the following function [33,34].

$$W^{(t+1)} = \begin{cases} 0 & \text{if } Sigmoid(W_{Best}) < 0.5 \\ 1 & \text{otherwise} \end{cases}$$

$$Sigmoid(W_{Best}) = \frac{1}{1 + e^{-10(W_{Best} - 0.5)}} \tag{2}$$

2.6 Fitness Function

The quality of an optimizer's output may be determined using fitness functions. The fitness function's essential elements include classification and regression errors, as well as the features chosen for the input pictures. Rather than a solution based on a broad list of specific traits, select one with the lowest potential classification error rate [35–40]. The following equation is used to evaluate the quality of the specified attributes.

$$F_n = \alpha \, Err(O) + \beta \frac{|s|}{|f|} \tag{3}$$

where α is a number in the range [0, 1], $\beta = 1 - \alpha$, $|s|$ is the number of selected features, and $|f|$ is the number of features in the dataset. $Err(O)$ represents the error of the optimizer (O). In our case, this optimizer is the proposed bGW-DTO algorithm.

Algorithm 1 : The proposed binary-guided-whale-dipper-throated-optimizer (bGW-DTO) for feature selection

1: **Initialize** population $\overrightarrow{W}_i(i = 1, 2, ..., n)$ with size n, maximum iterations Max_{iter}, fitness function F_n.
2: **Initialize** GW parameters $\overrightarrow{A}$, $\overrightarrow{D}$, $\overrightarrow{r_1}$, $\overrightarrow{w_1}$, $\overrightarrow{w_2}$, $\overrightarrow{w_3}$
3: **Initialize** DTO parameters, bird velocities $\overrightarrow{V}$, $\overrightarrow{C_1}$, $\overrightarrow{C_2}$, $\overrightarrow{r_2}$
4: **Set** t = 1
5: **Calculate** fitness F_n for each $\overrightarrow{W}_i$
6: **Find** best individual $\overrightarrow{W^*}$
7: **while** $t \leq Max_{iter}$ (Termination condition) **do**
8:    **if** $(t\%2 == 0)$ **then**
9:      **for** $(i = 1 : i < n + 1)$ **do**
10:        **if** $(\overrightarrow{r_1} \leq 0.5)$ **then**
11:          **if** $(|\overrightarrow{A}| < 1)$ **then**
12:            **Update** position of current search agent as $\overrightarrow{W}(t + 1) = \overrightarrow{W}^*(t) - \overrightarrow{A}.\overrightarrow{D}$
13:          **else**
14:            **Select** three random search agents $\overrightarrow{W}_{rand1}$, $\overrightarrow{W}_{rand2}$, and $\overrightarrow{W}_{rand3}$
15:            **Update** $(\overrightarrow{z})$ by the exponential form of
                $\overrightarrow{z} = 1 - \left(\frac{t}{Max_{iter}}\right)^2$
16:            **Update** position of current search agent as
                $\overrightarrow{W}(t + 1) = \overrightarrow{w_1} * \overrightarrow{W}_{rand1} + \overrightarrow{z} * \overrightarrow{w_2} * (\overrightarrow{W}_{rand2} - \overrightarrow{W}_{rand3}) + (1 - \overrightarrow{z}) * \overrightarrow{w_3} * (\overrightarrow{W} - \overrightarrow{W}_{rand1})$
17:          **end if**
18:        **else**
19:          **Update** velocity of DTO agent as
              $\overrightarrow{V}(t + 1) = \overrightarrow{C_1}\overrightarrow{V}(t) + \overrightarrow{C_2}r_2(\overrightarrow{W}_{best}(t) - \overrightarrow{W}_{second}(t))$
20:          **Update** position of DTO agent as
              $\overrightarrow{W}(t + 1) = \overrightarrow{W}(t) + \overrightarrow{S}(t + 1)$
21:        **end if**
22:      **end for**
23:      **Calculate** fitness F_n for each $\overrightarrow{W}_i$ from GW
24:    **else**
25:      **Calculate** fitness F_n for each $\overrightarrow{W}_i$ from DTO
26:    **end if**
27:    **Update** $\overrightarrow{r_1}$, $\overrightarrow{r_2}$, $\overrightarrow{A}$, $\overrightarrow{D}$, $\overrightarrow{C_1}$, $\overrightarrow{C_2}$
28:    **Find** best individual $\overrightarrow{W^*}$
29:    **Set** t = t + 1
30: **end while**
31: **Convert** best solution $\overrightarrow{W^*}$ to binary using:
     $X = \begin{cases} 1 & \text{if } Sigmoid(\overrightarrow{W^*}) \geq 0.5 \\ 0 & otherwise \end{cases}$
     $Sigmoid(\overrightarrow{W^*}) = \dfrac{1}{1 + e^{-10(\overrightarrow{W^*} - 0.5)}}$
32: return X

3 Proposed Methodology

The details of the proposed binary guided whale dipper throated optimizer (bGW-DTO) for feature selection are depicted in Algorithm 1. To perform feature selection using this algorithm, the resulting solution is converted to a binary solution based on the sigmoid function, as shown in line 31 of the algorithm. In addition, the K-nearest neighbor is employed to assess the quality of the selected features that achieve the best fitness for the breast cancer diagnosis.

In addition, a high-level overview of the proposed feature selection algorithm is illustrated in Algorithm 2. In this algorithm, the process goes through iterations at which the bGW-DTO is performed to retrieve binary values representing the selected features. These selected features are

assessed using k-NN with the help of a fitness function. At the end of these iterations, the algorithms return the set of features that achieve the best fitness among all other collections of features.

Algorithm 2 : The proposed feature selection algorithm

1: **Initialize** bGW-DTO population, parameters, and configurations
2: **while** $t \leq Max_{iter}$ **do**
3:　**Apply** bGW-DTO algorithm
4:　**Calculate** objective function and select the best solution
5:　**Train** k-NN and calculate error
6:　**Calculate** Fitness
7:　**Update** Fitness
8: **end while**
9: **Return** features that achieve the best fitness

---

4 Experimental Results

To evaluate the efficiency of the proposed algorithm, several experiments were conducted to assess its performance. The conducted experiments are based on several evaluation criteria presented in Table 3. These criteria include average error, best fitness, worst fitness, average fitness size, mean, and standard deviation.

Table 3: Evaluation metrics

Metrics	Equation
Average error	$\dfrac{1}{M} \displaystyle\sum_{j=1}^{M} \dfrac{1}{N} \sum_{i=1}^{N} mse\,(C_i,\ L_i)$
Average select size	$\dfrac{1}{M} \displaystyle\sum_{i=1}^{M} size\,(g_*^i)$
Average fitness	$\dfrac{1}{M} \displaystyle\sum_{M}^{1} g_*^i$
Best fitness	$min_{i=1}^{M} g_*^i$
Worst fitness	$Max_{i=1}^{M} g_*^i$
Standard deviation (STD)	$\sqrt{\dfrac{1}{M-1} \displaystyle\sum \left(g_*^i - Mean\right)^2}$

The configuration parameters of the proposed algorithm are presented in Table 4. These values are used as initial values to start the optimization algorithm. It can be noted that these parameters are set once, and their values are not changed during the running of the algorithm.

Table 4: Configuration parameters of the proposed bGW-DTO algorithm

Parameter	Value
Number of search agents	10
Number of iterations	80

(Continued)

Table 4: Continued

Parameter	Value
Number repetitions of runs	20
Problem dimension	Number of features
Search domain	[0, 1]
K-neighbors	5
K-fold cross-validation	10
Maximum diffusion level	1

To prove the superiority of the proposed algorithm, a set of eleven feature selectors were applied to the same four datasets incorporated in the conducted experiments, and the results were recorded. Table 5 presented the average error resulting from selecting the best features using the proposed algorithm (bGW-DTO) and the other eleven approaches. As shown in the table, the proposed algorithm could achieve the minimum average error of (**0.57186**) for the dataset D1, (**0.50924**) for the dataset D2, (**0.58111**) for the dataset D3, and (**0.52502**) for the dataset D4. In addition, the overall average error is (**0.5468**), which is the minimum among all the employed feature extraction methods. In addition, a visual representation of the general average error is depicted, the minimum value is recorded by the proposed algorithm. Moreover, the average select size is presented in Table 6, the average fitness is presented in Table 7, the best fitness is presented in Table 8, the worst fitness is presented in Table 9, and the standard deviation in Table 10. As shown in these tables, the proposed algorithm has been able to find the superiority fitness for all datasets.

Table 5: Average error

Dataset	D1	D2	D3	D4	Average
bGW-DTO	0.57186	0.50924	0.58111	0.52502	0.5468
bGWO	0.58558	0.52224	0.59422	0.54002	0.5605
bGWO_PSO	0.57872	0.52326	0.59227	0.52784	0.5555
bPSO	0.57591	0.55001	0.59956	0.53176	0.5643
bSFS	0.58456	0.52935	0.59615	0.52536	0.5589
bWAO	0.57284	0.53506	0.59131	0.53415	0.5583
bMGWO	0.58225	0.51083	0.58290	0.55653	0.5581
bMVO	0.57284	0.54403	0.60732	0.53545	0.5649
bSBO	0.59048	0.53035	0.61363	0.53914	0.5684
bGWO_GA	0.59440	0.56967	0.60004	0.53219	0.5741
bFA	0.57971	0.54061	0.60587	0.54241	0.5671
bGA	0.57284	0.52993	0.59519	0.53002	0.5570

Table 6: Average select size

Dataset	D1	D2	D3	D4	Average
bGW-DTO	0.7158	0.5111	0.7813	0.64694	0.663818
bGWO	0.7798	0.6414	0.8748	0.80337	0.774861
bGWO_PSO	0.7748	0.6657	0.8748	0.83194	0.786811
bPSO	0.9098	0.8657	0.8831	0.90694	0.891395
bSFS	0.7347	0.6888	0.8779	0.86051	0.790479
bWAO	1.0048	0.7066	0.8831	1.05337	0.91198
bMGWO	0.8261	0.6289	0.8635	0.74622	0.766157
bMVO	0.87982	0.8460	0.8914	0.88909	0.876593
bSBO	0.89481	0.8051	0.8078	0.94622	0.863482
bGWO_GA	0.85482	0.7991	1.0748	0.90337	0.907993
bFA	0.91481	0.8626	0.8831	0.89981	0.890103
bGA	0.87481	0.7657	0.8831	0.88194	0.851395

Table 7: Average fitness

Dataset	D1	D2	D3	D4	Average
bGW-DTO	0.59959	0.52125	0.5918	0.6878	0.6001
bGWO	0.63548	0.55622	0.76536	0.70265	0.6649
bGWO_PSO	0.6178	0.5228	0.71738	0.527843	0.5965
bPSO	0.62189	0.58372	0.77064	0.69447	0.6677
bSFS	0.6358	0.5291	0.6191	0.515365	0.5748
bWAO	0.62286	0.56891	0.76247	0.69684	0.6628
bMGWO	0.6013	0.5529	0.5978	0.556539	0.5771
bMVO	0.62286	0.57779	0.77833	0.69813	0.6693
bSBO	0.6298	0.5298	0.7128	0.539148	0.6029
bGWO_GA	0.5938	0.5688	0.6998	0.532191	0.5986
bFA	0.62965	0.57441	0.77689	0.70502	0.6715
bGA	0.62286	0.56383	0.76632	0.69275	0.6614

Table 8: Best fitness

Dataset	D1	D2	D3	D4	Average
bGW-DTO	0.50445	0.46949	0.68271	0.65121	0.5769
bGWO	0.52386	0.48641	0.70192	0.66391	0.5940
bGWO_PSO	0.60150	0.51179	0.74997	0.67252	0.6339
bPSO	0.52386	0.52025	0.69231	0.65961	0.5990
bSFS	0.58509	0.47621	0.74036	0.67582	0.6194

(Continued)

Table 8: Continued

Dataset	D1	D2	D3	D4	Average
bWAO	0.56268	0.49487	0.69231	0.65531	0.6013
bMGWO	0.55209	0.51956	0.69114	0.66821	0.6078
bMVO	0.58209	0.47795	0.71153	0.67252	0.6110
bSBO	0.58209	0.50333	0.74997	0.68112	0.6291
bGWO_GA	0.58209	0.55410	0.72114	0.68112	0.6346
bFA	0.56268	0.46949	0.71153	0.66822	0.6030
bGA	0.52386	0.50333	0.69231	0.66391	0.5959

Table 9: Worst fitness

Dataset	D1	D2	D3	D4	Average
bGW-DTO	0.69898	0.62072	0.82448	0.72167	0.71646
bGWO	0.71798	0.63025	0.80765	0.77152	0.73185
bGWO_PSO	0.67915	0.57948	0.76920	0.71556	0.68585
bPSO	0.73739	0.63872	0.83648	0.75431	0.74172
bSFS	0.65974	0.66242	0.79842	0.69404	0.70365
bWAO	0.73739	0.65564	0.83648	0.73709	0.74165
bMGWO	0.71797	0.63871	0.84609	0.81026	0.75326
bMVO	0.69856	0.62179	0.87493	0.71987	0.72878
bSBO	0.69856	0.60486	0.87492	0.71126	0.72240
bGWO_GA	0.69856	0.66410	0.81726	0.71126	0.72279
bFA	0.75681	0.67256	0.98066	0.87052	0.82013
bGA	0.69856	0.68102	0.87493	0.72848	0.74574

Table 10: Standard deviation fitness

Dataset	D1	D2	D3	D4	Average
bGW-DTO	0.40767	0.40777	0.40078	0.39023	0.4016
bGWO	0.42669	0.41732	0.40362	0.40683	0.4136
bGWO_PSO	0.41028	0.41311	0.40339	0.39366	0.4051
bPSO	0.43437	0.40814	0.41477	0.39674	0.4135
bSFS	0.42669	0.44944	0.42328	0.40161	0.4253
bWAO	0.42104	0.41714	0.41249	0.39843	0.4123
bMGWO	0.41724	0.41251	0.41237	0.40885	0.4127
bMVO	0.41458	0.41445	0.41516	0.39019	0.4086
bSBO	0.42234	0.41235	0.42593	0.39154	0.4130
bGWO_GA	0.41906	0.42129	0.41037	0.39412	0.4112
bFA	0.42537	0.42278	0.43883	0.42032	0.4268
bGA	0.42514	0.42234	0.41976	0.39269	0.4150

On the other hand, the speed of finding the single iteration performed for finding the best set of features is measured are recorded in Table 11. As shown in the table, the processing time of the proposed algorithm for D1 is (3.7327), for D2 is (4.1507), for D3 is (3.4027), and for D4 is (3.4037). In addition, the overall average processing time of the proposed algorithm is (3.6724). Based on these results, it can be noted that the proposed algorithm can perform the feature selection process in the shortest time among the other competing feature selection algorithms. It is clearly shown in this table that the proposed algorithm is considered the fastest when compared with the other eleven feature extraction algorithms.

Table 11: Processing time

Dataset	D1	D2	D3	D4	Average
bGW-DTO	3.7327	4.1507	3.4027	3.4037	3.6724
bGWO	4.3767	6.0817	4.5987	4.0697	4.7817
bGWO_PSO	3.9687	5.2187	4.8387	4.1987	4.5562
bPSO	4.0587	5.6537	4.6777	4.9147	4.8262
bSFS	5.4587	5.4087	4.7687	4.4287	5.0162
bWAO	3.8657	4.9867	3.4137	3.5057	3.9429
bMGWO	4.1487	4.6887	4.0997	3.0212	4.3123
bMVO	4.3197	5.5937	4.3707	4.6827	4.7417
bSBO	4.7887	5.6187	4.5587	4.1487	4.7787
bGWO_GA	4.5087	5.8887	4.8787	4.0487	4.8312
bFA	5.0867	5.6707	4.9267	4.0067	4.9227
bGA	4.3327	5.6067	4.0777	4.1647	4.5454

On the other hand, the statistical difference between the proposed bGW-DTO and the other competing algorithm is tested. To realize this test, a one-way analysis of variance (ANOVA) test is employed. Two main hypotheses are set in this test, namely, the null hypothesis and the alternate hypothesis. For the null hypothesis denoted by H0, the mean values of the algorithm is set equal, μbGW-DTO = μbGW = μbGW-PSO = μbSFS = μbPSO = μbWAO = μbMGWO = μbMVO = μbSBO = μbGW-GA = μbFA = μbGA). Whereas in the alternate hypothesis denoted by H1, the means of the algorithms are not equal. The results of the ANOVA test are presented in Table 12. As shown in the table, the expected effectiveness of the proposed algorithm is confirmed when compared with other feature selection methods. In addition, a statistical analysis of the achieved results is presented in Tables 13 and 14.

Table 12: ANOVA table

ANOVA table	SS	DF	MS	F (DFn, DFd)	P value
Treatment (between columns)	0.006185	11	0.0005623	F (11, 120) = 18.77	$P < 0.0001$

(Continued)

Table 12: Continued

ANOVA table	SS	DF	MS	F (DFn, DFd)	*P* value
Residual (within columns)	0.003595	120	0.00002996		
Total	0.009781	131			

Table 13: Statistical analysis (part 1)

	bGW-DTO	bGWO	bGWO_PSO	bPSO	bSFS	bWAO
Number of values	11	11	11	11	11	11
Minimum	0.5709	0.5756	0.5787	0.5759	0.5846	0.5728
25% percentile	0.5719	0.5856	0.5787	0.5759	0.5846	0.5728
Median	0.5719	0.5856	0.5787	0.5759	0.5846	0.5728
75% percentile	0.5719	0.5856	0.5787	0.5759	0.5846	0.5728
Maximum	0.5719	0.5996	0.5987	0.5959	0.5996	0.5928
Range	0.001	0.024	0.02	0.02	0.015	0.02
10% percentile	0.5711	0.5776	0.5787	0.5759	0.5846	0.5728
90% percentile	0.5719	0.5988	0.5967	0.5939	0.5986	0.5908
95% CI of median						
Actual confidence level	98.83%	98.83%	98.83%	98.83%	98.83%	98.83%
Lower confidence limit	0.5719	0.5856	0.5787	0.5759	0.5846	0.5728
Upper confidence limit	0.5719	0.5956	0.5887	0.5859	0.5946	0.5828
Mean	0.5718	0.5869	0.5814	0.5786	0.5868	0.5756
Std. deviation	0.0003	0.00615	0.0064	0.0064	0.0051	0.0064
Std. error of mean	0.0001	0.001854	0.0019	0.0019	0.0015	0.0019
Skewness	−3.317	0.6713	2.42	2.42	2.127	2.42
Kurtosis	11	2.013	5.51	5.51	3.492	5.51
Sum	6.289	6.455	6.396	6.365	6.455	6.331

Table 14: Statistical analysis (part 2)

	bMGWO	bMVO	bSBO	bGWO_GA	bFA	bGA
Number of values	11	11	11	11	11	11
Minimum	0.5823	0.5728	0.5905	0.5944	0.5797	0.5728
25% percentile	0.5823	0.5728	0.5905	0.5944	0.5797	0.5728
Median	0.5823	0.5728	0.5905	0.5944	0.5797	0.5728
75% percentile	0.5823	0.5728	0.5905	0.5944	0.5797	0.5728
Maximum	0.5993	0.5928	0.5995	0.5994	0.5997	0.5928

(Continued)

Table 14: Continued

	bMGWO	bMVO	bSBO	bGWO_GA	bFA	bGA
Range	0.0171	0.0221	0.0091	0.0051	0.0211	0.0221
Mean	0.5847	0.5756	0.5918	0.5951	0.5824	0.5756
Std. deviation	0.0056	0.00646	0.0029	0.0016	0.0064	0.0064
Std. error of mean	0.0017	0.00195	0.0008	0.0005	0.0019	0.0019
Skewness	2.2491	2.4211	2.3092	2.2292	2.4221	2.4223
Kurtosis	4.3411	5.5122	4.7521	4.2012	5.5121	5.5123
Sum	6.4322	6.3311	6.5091	6.5462	6.4071	6.3311

Moreover, the statistical difference between every two algorithms is performed to get the *p*-values between the proposed bGW-DTO algorithm and the other competing approaches to prove that the proposed method has a significant difference. To realize this test, Wilcoxon's rank-sum test is employed. Two main hypotheses are set in this test, namely, the null hypothesis and the alternate hypothesis. For the null hypothesis denoted by H0, the mean values of the algorithm is set equal, μbGW-DTO = μbGW, μbGW-DTO = μbGW-PSO, μbGW-DTO = μbSFS, μbGW-DTO = μbPSO, μbGW-DTO = μbWAO, μbGW-DTO = μbMGWO, μbGW-DTO = μbMVO, μbGW-DTO = μbSBO, μbGW-DTO = μbGW-GA, μbGW-DTO = μbFA, and μbGW-DTO = μbGA). Whereas in the alternate hypothesis denoted by H1, the means of the algorithms are not equal. The results of Wilcoxon's rank-sum test are presented in Tables 15 and 16. As shown in the table, the *p*-values are less than 0.05 between the proposed algorithm and other algorithms, confirming the superiority of the proposed bGW-DTO algorithm and its statistical significance.

Table 15: Wilcoxon signed-rank test (part 1)

	bGW-DTO	bGWO	bGWO_PSO	bPSO	bSFS	bWAO
P value (two tailed)	0.001	0.001	0.001	0.001	0.001	0.001
Exact or estimate?	Exact	Exact	Exact	Exact	Exact	Exact
P-value summary	***	***	***	***	***	***
Significant (alpha = 0.05)?	Yes	Yes	Yes	Yes	Yes	Yes

Table 16: Wilcoxon signed-rank test (part 2)

	bMGWO	bMVO	bSBO	bGWO_GA	bFA	bGA
P value (two tailed)	0.001	0.001	0.001	0.001	0.001	0.001
Exact or estimate?	Exact	Exact	Exact	Exact	Exact	Exact
P-value summary	***	***	***	***	***	***

(Continued)

Table 16: Continued

	bMGWO	bMVO	bSBO	bGWO_GA	bFA	bGA
Significant (alpha = 0.05)?	Yes	Yes	Yes	Yes	Yes	Yes

A more investigation of the results achieved by the proposed algorithm is represented by the plots in Fig. 1. In this figure, four plots are depicted to emphasize the effectiveness of the proposed algorithm. The first plot is the residual plot, in which the predicted values are plotted with the residual error. In this plot, it can be shown that the residual error is minor and lies between -0.01 and 0.02, which indicates the high accuracy of the resulting values. The second plot is the QQ plot in which the actual and predicted values approximately fit a line, and this proves the effectiveness of the proposed algorithm. in which the proposed approach is shown to outperform the other methods.
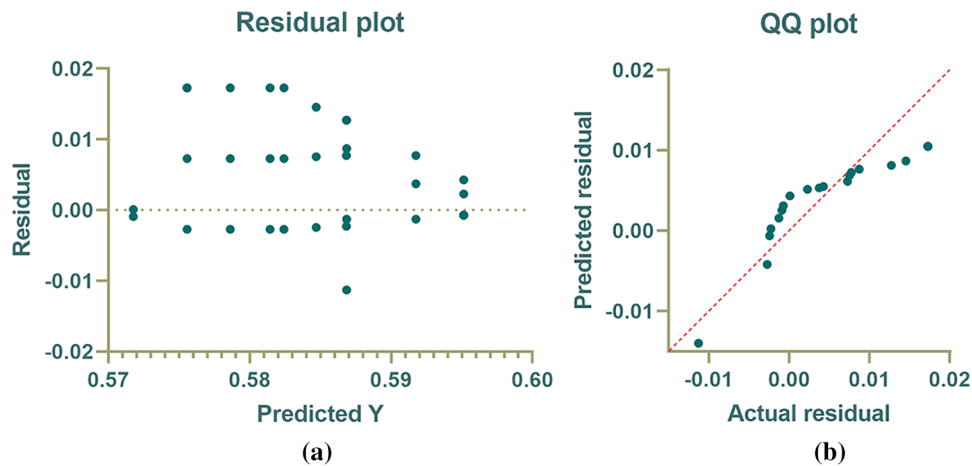
Figure 1: Analytical plots of the achieved results

Fig. 2 depicts the histogram of the average error of the achieved results using the proposed and other approaches. In this figure, the error values achieved by the proposed bGW-DTO are smaller than those achieved by the different methods. In addition, the number of occurrences of these error values is more enormous than the corresponding values reached in the other approaches. These results confirm the effectiveness of the proposed method in selecting the significant features that could achieve the minimum error values.
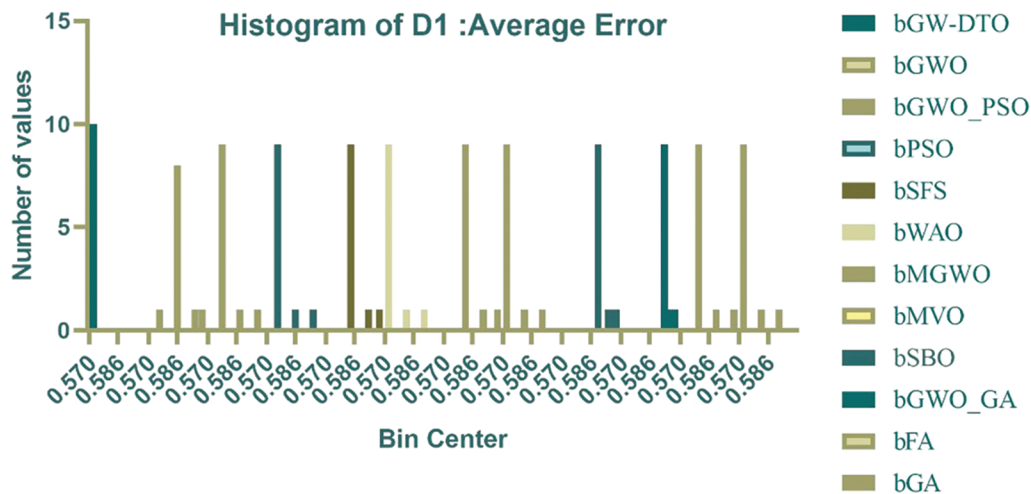
Figure 2: Histogram of the average error

5 Conclusions

In this paper, we proposed a new algorithm for feature selection based on a hybrid guided whale optimization algorithm and dipper throated optimizer. The proposed algorithm exploits the advantages of both optimizers in terms of exploration and exploitation properties. The proposed algorithm is applied to the problem of a breast cancer diagnosis. Four datasets from the UCI repository were incorporated in the conducted experiments. To prove the effectiveness of the proposed algorithm, a set of experiments were conducted and evaluated using several evaluation criteria. The achieved results showed the superiority of the proposed algorithm and its stability in classifying breast cancer cases. In addition, statistical analysis is performed in terms of the Wilcoxon and analysis of variance tests to emphasize the effectiveness of the proposed method. The overall results achieved by the proposed method outperform the corresponding results using the other eleven feature selection methods. The future perspective of this research is to evaluate the proposed algorithm using thermal images of breast cancer cases.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1]   A. Ibrahim, S. Mohammed, H. Ali and S. Hussein, "Breast cancer segmentation from thermal images based on chaotic salp swarm algorithm," *IEEE Access*, vol. 8, pp. 122121–122134, 2020.

[2]   M. Etehadtavakol and E. Ng, "Breast thermography as a potential non-contact method in the early detection of cancer: A review," *Journal of Mechanics in Medicine and Biology*, vol. 13, no. 2, pp. 1–20, 2013.

[3]   J. Shan, "A fully automatic segmentation method for breast ultrasound images," *Ph.D. dissertation*, Department of Computer Science, Utah State Univ., Logan, UT, USA, 2011.

[4]   C. Sehgal, S. Weinstein, P. Arger and E. Conant, "A review of breast ultrasound," *Journal of Mammary Gland Biol Neoplasia*, vol. 11, no. 2, pp. 113–123, 2006.

[5]   T. Nelson, L. Cervino, J. Boone and K. Lindfors, "Classification of breast computed tomography data," *Medical Physics*, vol. 35, no. 3, pp. 1078–1086, 2008.

[6]   A. Jalalian, S. Mashohor, R. Mahmud, B. Karasfi, M. Iqbal Saripan *et al.,* "Computer-assisted diagnosis system for breast cancer in computed tomography laser mammography (CTLM)," *Journal of Digital Imaging*, vol. 30, no. 6, pp. 796–811, 2017.

[7]   P. Pavithra, R. Ravichandran, S. Sekar and M. Manikandan, "The effect of thermography on breast cancer detection: A survey," *Systematic Reviews in Pharmacy*, vol. 9, no. 1, pp. 10–16, 2018.

[8]   S. Prabha, C. Sujatha and S. Ramakrishnan, "Asymmetry analysis of breast thermograms using BM3D technique and statistical texture features," in *Proc. of Int. Conf. on Informatics, Electronics and Visualization*, Dhaka, Bangladesh, pp. 1–4, 2014.

[9]   L. Silva, D. Saade, G. Sequeiros, A. Silva, A. Paiva *et al.,* "A new database for breast research with infrared image," *Journal of Medical Image and Health Informatics*, vol. 4, no. 1, pp. 92–100, 2014.

[10] M. Tan and J. Schlimmer, M. Zwitter and M. Soklic, "Breast cancer data set," 1988, Accessed: April. 1, 2022. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/Breast+Cancer.

[11] M. Patrício, J. Pereira, J. Crisóstomo, P. Matafome, M. Gomes *et al.,* "Breast cancer coimbra data set," 2018, Accessed: April. 1, 2022. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Coimbra#.

[12] O. Mangasarian and W. Wolberg, "Breast cancer Wisconsin (Original) data set," 1990, Accessed: April. 1, 2022. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Original%29.

[13] K. Bennett and O. Mangasaria, "Breast cancer Wisconsin (Diagnostic) data set," 1992, Accessed: April. 1, 2022. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29.

[14] S. Mambou, P. Maresova, O. Krejcar, A. Selamat and K. Kuca, "Breast cancer detection using infrared thermal imaging and a deep learning model," *Sensors*, vol. 18, no. 9, pp. 2799, 2018.

[15] L. Ehsan and F. Mohammad, "Early breast cancer detection in thermogram images using AdaBoost classifier and fuzzy C-means clustering algorithm," *Middle East Journal of Cancer*, vol. 7, no. 3, pp. 113–124, 2016.

[16] A. Lashkari and M. Firouzmand, "Developing a toolbox for clinical preliminary breast cancer detection in different views of thermogram images using a set of optimal supervised classifiers," *Scientific Iranica*, vol. 25, no. 3, pp. 1545–1560, 2018.

[17] F. Li, C. Ou, Y. Gui and L. Xiang, "Instant edit propagation on images based on bilateral grid," *Computers Materials & Continua*, vol. 61, no. 2, pp. 643–656, 2019.

[18] N. Mohamed, "Breast cancer risk detection using digital infrared thermal images," *International Journal Bioinformatics and Biomedical Engineering*, vol. 1, no. 2, pp. 185–194, 2015.

[19] S. Kamath, K. Prasad and K. V. Rajagopal, "Segmentation of breast thermogram images for the detection of breast cancer: A projection profile approach," *Journal of Image and Graphics*, vol. 3, no. 1, pp. 47, 2015.

[20] H. Minh, M. Van and T. Lang, "Deep feature fusion for breast cancer diagnosis on histopathology images," in *11th Int. Conf. on Knowledge and Systems Engineering (KSE)*, Da Nang, pp. 1–6, 2019.

[21] M. Khadermi and S. Nedialkov, "Probabilistic graphical models and deep belief networks for prognosis of breast cancer," in *IEEE 14th Int. Conf. on Machine Learning and Applications*, Miami, FL, USA, pp. 727–732, 2015.

[22] P. Mekha and N. Teeyasuksaet, "Deep learning algorithms for predicting breast cancer based on tumor cells," in *Int. Conf. on Digital Arts, Media and Technology and 2nd ECTI Northern Section Conf. on Electrical, Electronics, Computer and Telecommunication Engineering*, Nan, Thailand, pp. 343–346, 2019.

[23] N. Rane, J. Sunny, R. Kanade and S. Devi, "Breast cancer classification and prediction using machine learning," *International Journal of Engineering Research & Technology*, vol. 9, no. 02, pp. 576–580, 2020.

[24] W. Fathy and A. Ghoneim, "A deep learning approach for breast cancer mass detection," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 1, pp. 175–182, 2019.

[25] R. Chtihrakkannan, P. Kavitha, T. Mangayarkarasi and R. Karthikeyan, "Breast cancer detection using machine learning," *International Journal of Innovative Technology and Exploring Engineering*, vol. 8, no. 11, pp. 3123–3126, 2019.

[26] S. Nallamala, P. Mishra and S. Koneru, "Breast cancer detection using machine learning way," *International Journal of Recent Technology and Engineering*, vol. 8, no. 2S3, pp. 1402–1405, 2019.

[27] A. Kumar, R. Sushil and A. Tiwari, "Comparative study of classification techniques for breast cancer diagnosis," *International Journal of Computer Science and Engineering*, vol. 7, no. 1, pp. 234–240, 2019.

[28] W. Wang, Y. Li, T. Zou, X. Wang, J. You *et al.,* "A novel image classification approach via dense-MobileNet models," *Mobile Information Systems*, vol. 2020, no. 1, pp. 1–8, 2020.

[29] A. A. Abdelhamid and S. R. Alotaibi, "Robust prediction of the bandwidth of metamaterial antenna using deep learning," *Computers, Materials & Continua*, vol. 72, no. 2, pp. 2305–2321, 2022.

[30] A. Takieldeen, E., El-kenawy, E. Hadwan and M. Zaki, "Dipper throated optimization algorithm for unconstrained function and feature selection," *Computers, Materials & Continua*, vol. 72, no. 1, pp. 1465–1481, 2022.

[31] M. M. Eid, E.-S. M. El-kenawy and A. Ibrahim, "A binary sine cosine-modified whale optimization algorithm for feature selection," in *National Computing Colleges Conf.*, Taif, Saudi Arabia, pp. 1–6, 2021.

[32] M. Hassib, I. El-Desouky, M. Labib and E.-S. M. El-kenawy, "WOA + BRNN: An imbalanced big data classification framework using whale optimization and deep neural network," *Soft Computing*, vol. 24, no. 1, pp. 5573–5592, 2020.

[33] A. Ibrahim, H. A. Ali, M. M. Eid and E.-S. M. El-Kenawy, "Chaotic harris hawks optimization for unconstrained function optimization," in *2020 16th Int. Computer Engineering Conf. (ICENCO)*, Cairo, Egypt, IEEE, pp. 153–158, 2020.

[34] A. A. Salamai, E.-S. M. El-kenawy and A. Ibrahim, "Dynamic voting classifier for risk identification in supply chain 4.0," *Computers, Materials & Continua*, vol. 69, no. 3, pp. 3749–3766, 2021.

[35] N. Prerita, A. Rana and A. Chaudhary, "Breast cancer detection using machine learning algorithms," in *Int. Conf. on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, Noida, India, pp. 114–118, 2021.

[36] S. S. M. Ghoneim, T. A. Farrag, A. A. Rashed, E.-S. M. El-Kenawy and A. Ibrahim, "Adaptive dynamic meta-heuristics for feature selection and classification in diagnostic accuracy of transformer faults," *IEEE Access*, vol. 9, no. 1, pp. 78324–78340, 2021.

[37] E.-S. M. El-Kenawy, S. Mirjalili, A. Ibrahim, M. Alrahmawy, M. El-Said *et al.,* "Advanced meta-heuristics, convolutional neural networks, and feature selectors for efficient COVID-19 X-ray chest image classification," *IEEE Access*, vol. 9, no. 1, pp. 36019–36037, 2021.

[38] E.-S. M. El-kenawy and M. Eid, "Hybrid gray wolf and particle swarm optimization for feature selection," *International Journal of Innovative Computing, Information & Control*, vol. 16, no. 1, pp. 831–844, 2020.

[39] W. Wang, X. Huang, J. Li, P. Zhang and X. Wang, "Detecting COVID-19 patients in X-ray images based on MAI-nets," *International Journal of Computational Intelligence Systems*, vol. 14, no. 1, pp. 1607–1616, 2021.

[40] Y. Gui and G. Zeng, "Joint learning of visual and spatial features for edit propagation from a single image," *The Visual Computer*, vol. 36, no. 3, pp. 469–482, 2020.