A Selected Deep Learning Cancer Prediction Framework

**NADIA G. ELSEDDEQ[1], SALLY M. ELGHAMRAWY[ID][2], (Senior Member, IEEE),
MOFREH M. SALEM[1], AND ALI I. ELDESOUKY[1]**
[1]Department of Computer Engineering and Control Systems, Faculty of Engineering, Mansoura University, Mansoura 35516, Egypt
[2]Department of Computer Engineering, MISR Higher Institute for Engineering and Technology, Mansoura 35516, Egypt

Corresponding author: Sally M. Elghamrawy (sally_elghamrawy@ieee.org)

ABSTRACT Deep learning (DL) algorithms are crucial for predicting various diseases because they can analyze a large amount of healthcare data within a short prediction time. One of these diseases is cancer, which causes one out of six deaths worldwide. Many researchers have adopted predictive frameworks such as machine learning and DL to predict cancer prognosis, in addition to the probability of its recurrence, progression, and the patients' survival estimation. Currently, all stakeholders are interested in the accuracy of cancer prognosis prediction. This study selected a framework within high accuracy and short prediction time from three DL frameworks for improving the performance of cancer prognosis prediction. This prediction requires a quick and high-accuracy optimizer, so we propose a binary version of the continuous AC-parametric whale optimization algorithm. This version is built on S-shaped transfer functions to identify the minimal optimal subset of features and maximize the classification accuracy. These frameworks proposed have the following forms: the first is a Feed-Forward Neural Network (FFNN) in which the input is the optimal set of feature selection. The second is an optimized parameter FFNN. The third is composed of a feature selection layer in which the best subset of selected features is for use as inputs in the optimized FFNN. We compared these frameworks using a comparative study. Our results show that, under all conditions, the third framework is superior to the others with an average accuracy of 100%, whereas the first and second frameworks achieved 94.97% and 93.12% accuracy, respectively.

INDEX TERMS BACP-WOA-S, cancer diagnosis, deep learning, exploitation, exploration, feature selection, optimization.

I. INTRODUCTION

Cancer is the second disease that causes one out of six deaths worldwide [1]. In 2020, the International Agency for Research on Cancer predicted 19,300,000 new cases and 10,000,000 deaths [2]. The reason for growing concern about cancer is that it weakens the immune system and causes an imbalance in other biological changes. The most common kinds of cancer are breast, colon, cervical, lung, prostate, and ovarian cancer.

Several previous studies have introduced frameworks for predicting the prognosis of cancer, the probability of recurrence and progression, and estimating patients' survival [1]. All stakeholders, including patients, their caregivers, and providers, are interested in the accuracy of cancer prognosis prediction. One of the factors that contribute to the

The associate editor coordinating the review of this manuscript and approving it for publication was Muhammad Sharif[ID].

effective treatment of patients is prediction accuracy [3]. Disease detection involves the classification of tumor types and identification of cancer symptoms to train a machine that can identify new metastatic tumor types or diagnose a disease at an early stage because treatment in later stages is more difficult. However, due to the enormous number of gene expression levels in a person, diagnosing cancer might be challenging. The basic difficulties linked to the treatment and prevention of illnesses are recognized to be inscribed by gene expression levels [4].

As precision medicine and early detection procedures have developed recently [4]–[6] with many detector screens reaching 70%–80% [7], the demand for new machine learning (ML) approaches for discovering new biomarkers has become one of the primary drivers of most biomedical research.

Deep sequencing is a DNA fractionation technique that has changed genomic science significantly. The progression

of deep sequencing over the past decade has continuously generated huge volumes of data making genomics among the top fields of data generation [8]. Although the sequence itself cannot explain ready-to-use information, it can be converted using a complicated procedure that deduces protein drawn from the sequence. As the designed genome sequence agrees with historically identified cancer genome sequences, it evaluates the expression of the protein and checks whether it is cancerous [9]. The genomic data collection has caused several difficulties in providing a logical description of cancer's genetic origin. Moreover, cancer prognosis is complex due to the existence of genomic datasets containing several features but comparatively few samples. Early diagnosis increases the chances of healing; thus, its study is crucial.

The H_2O framework is a multi-layer neural network (NN) system suitable for DL tasks [10]. The Deep Learning Architecture (DLA), including multiple levels of non-linearity, is a hierarchical model for extracting features. DL models can learn to represent the usable original data. Moreover, they show the best output for complicated data such as text, images, and audio [11].

Single-solution and population-based algorithms are two classes of meta-heuristic algorithms. In first-class, an optimization algorithm performs the optimization process using only one candidate solution that evolves and gets updated during iterations, whereas the second class performs the optimization process with an initial random search agent representing the population. The solution to the optimization problem is a candidate for each search agent. Individuals exchange data on the search area and work together to prevent local optima stagnation and coverage toward a global purpose. Many studies have used different optimization algorithms for resolving decision-making problems [12]–[16]. Notably, a metaheuristic algorithm's quality is determined by its ability to achieve proper control and considerable balance between exploitation and exploration [17]. Exploitation is the ability to discover optimal solutions surrounding the best-known solutions. Exploration is the effort of using meta-heuristics to locate novel locations in a search space having better points. Most metaheuristics use exploration early in the optimization process to thoroughly examine the feasible region and prevent a recession in the local optima.

Based on the above-mentioned reasons, multiple meta-heuristic techniques have been adopted with wrapper methods to produce an acceptable solution in an acceptable time.

In addition to using a single optimization algorithm to resolve the Feature Selection (FS) problem, researchers have proposed different hybrid approaches to solve binary optimization problems. A hybrid approach between Whale Optimization Algorithm (WOA) and simulated annealing is studied in [18] and that of Genetic Algorithm (GA) and Particle Swarm Optimizer (PSO) is reported in [19]. Note that a hybrid approach between filter and wrapper methods of FS has been previously studied [20], [21].

In the FS problem, there is no guarantee that a better subset of features will be identified. Moreover, no optimizer is ideally suitable to solve any optimization issues using the No Free Lunch (NFL) theorem [22]. This explains why certain optimizers poorly work when solving certain optimization problems.

This paper is structured as follows. The motivations and contributions of the study are in Section 2. A literature review, a short overview of the FFNN, and AC-parametric WOA (ACP-WOA) are presented in Section 3. The proposed BACP-WOA-S and designed frameworks are described in Section 4. The experimental results are discussed in Section 5. Finally, the conclusion and future work are presented in Section 6.

II. MOTIVATIONS & CONTRIBUTIONS OF THE STUDY

The motivations of this study are as follows:

1) Select a framework from three proposed DL H_2O frameworks that deal with big data for improving the performance and prediction of cancer prognosis.

2) Propose a very accurate and fast optimizer, Binary AC-Parametric WOA (BACP-WOA-S), which is required for FS to reduce the size of the dataset used and also to tune FFNN (number of layers and number of neurons per layer).

The relevant contributions of this study are as follows:

- DL frameworks support a large amount of data in various formats. Because patient health data contains multi-source data, suitable for cancer prognosis prediction.
- FS is an approach to efficiently selecting optimal features for NN training; thus, potentially improving cancer prognosis prediction and reducing the size of input data to FFNN.
- The three proposed frameworks can be used in biomedical diagnostic applications to improve the prediction accuracy of the diagnosis of disease. However, the third framework has the highest accuracy at the same time, it has been the nearest time of the first framework.
- The frameworks have been evaluated on six benchmark datasets, including breast, cervical, colon, lung, prostate, and ovarian cancers, to demonstrate their reliability and efficiency. These datasets are publicly available and are still used in most current studies [23]–[29].
- The advantages and efficiency of BACP-WOA-S are compared with other common optimizers.
- 100% accuracy is achieved in predicting all types of cancers, which benefits patients, as the earlier the treatment begins, the better the chance of cure.

This study will help in selecting the most suitable framework and accurate one based on the case's severity (case of critical patients or cases of non-critical patients).

III. RELATED WORK

Several studies reported on cancer diagnosis have used various methods for cancer diagnosis prediction; some of these methods have demonstrated significant prediction accuracies. Table 1 lists the results of previously reported methods.

TABLE 1. Overview of recent studies using different methods and their corresponding results.

Ref	Year	Methods	Accuracy					
			Breast	Colon	Cervical	Lung	Prostate	Ovarian
[30]	2008	NN	-	88.7%	-	97%	-	97%
[31]	2009	Cuckoo Search (CS)	-	-	-	87.7%	85.3%	-
[32]	2014	The multi objective cuckoo search (MOCS)	-	-	-	87.7%	91.2%	-
[33]	2018	• ReliefF • PSO and Support Vector Machine (SVM)	-	98.1%	-	86.5%	-	-
[34]	2018	A Novel Complex Network and Hybrid GP+PSO	-	95.52%	-	100%	-	-
[35]	2020	SVM Classifier	-	85.71%	-	70.13%	88.75%	100%
[35]	2020	Levenberg-Marquardt Backpropagation (LMBP) Classifier	-	81.43%	-	96.25%	97.65%	100%
[36]	2020	Their multi-objective search algorithm has been designed with evolutionary operators	-	-	-	93.7%	92.2%	97.5%
[37]	2020	• GA • Random Forest (RF)	86.60%	96.77%	-	-	98.04%	100%
[38]	2020	• ANN • ANN-GA	-	86%	-	100%	82%	-
[23]	2021	• Decision tree (DT) • RF • XGBoost	-	-	93.33%	-	-	-
[24]	2021	RF + GA	96.82%	-	-	-	-	-
[25]	2021	Logistic Regression (LR) technique	99.27%	-	-	-	-	-
[26]	2021	DT	-	94%	-	-	-	-
[27]	2021	MultiLayer perceptron (MLP)	-	-	94.5%	-	-	-
[28]	2021	SVM Classifier based on selected genes by XGBoost-MOGA	82.33%	90.24%	-	100%	98%	100%
[29]	2021	1-Dimensional Convolution Neural Network (1D-CNN)	-	-	-	-	99.07%	-

Some researchers use different ML classifiers such as k-nearest neighbor (KNN), LR, DT, RF, and SVM to improve treatment and medication discovery for diagnosis. In [23] performed on cervical cancer dataset. Also, in [24], [25] performed on four, six different breast cancer datasets respectively. But, in [26] performed on two different colon cancer datasets.

In [29] performed on different microarray data using multi-layered DL algorithm to detect the type of disease.

In [30] proposed an ensemble of Artificial Neural Network (ANN) classifiers for the classification of microarray data. They used four different cancer datasets.

In addition to other forms of omics data, [34] proposes combining gene FS with cancer classification for gene expressions.

To dimension reduction method, in [35] performed on SVM and LMBP using Principal Component Analysis (PCA). Also [38] worked on ANN and GA using PCA.

The reason for using DL over other traditional methods is that we expect to handle the big data size problems. Also, the prediction time for cancer diagnoses is very critical as the patient's life depends on this time, especially in critical cases. DL is the most suitable methodology to use in this case.

As it needs to have the high-end infrastructure to train in a reasonable time.

The details of the description of the artificial neural network used in our proposal and the optimized Whale algorithm (AC-PARAMETRIC WOA) used in the proposed framework is are shown in appendix A.

IV. THE PROPOSED CANCER PREDICTION FRAMEWORK

The most crucial factors for a cancer patient are accuracy and prediction time since the higher the prediction accuracy of the disease, the higher the chance of the patient receiving treatment. Meanwhile, a long prediction time decreases the patient's chance of receiving treatment.

This paper proposes three different cancer prediction frameworks. These frameworks are used to select the most suitable one based on the case's severity (case of critical patients or cases of non-critical patients).

The main contributions of this work are to A) propose three different frameworks using the FFNN. B) propose a new binary optimizer, namely S-SHAPED BINARY AC-PARAMETRIC WOA (BACP-WOA-S) for the binary classification problem, Feature Selection and tuning the FFNN. First, the details of these frameworks are described below.

A. THE PROPOSED FRAMEWORKS BASED ON THE FFNN

The main objective is to propose three different deep learning frameworks then select the framework that has the best accuracy and less processing time for improving the performance of cancer prediction based on the severity of patients.

An FFNN was used for all frameworks. The whale algorithm was modified for this purpose. We built simple NN frameworks with three to four layers. The first framework is an NN, and its input is the optimal set of FS, whereas the second framework is an optimized parameter NN. The third framework comprises the above two frameworks.

These frameworks presented take the following form: the first is an FFNN in which the input is the optimal set of feature selection. The second is an optimized parameter FFNN. The details of the first and second frameworks are shown in appendix B.

The third framework, namely, the FFNN framework with FS and best configuration, is composed of a feature selection layer in which the best subset of features is selected for use as inputs in the optimized FFNN.

FIGURE 1. Block diagram of the 3rd framework.

In this framework, H$_2$O is used for processing big data. The framework consists of four main layers. The processing, feature selection, deep learning, and prediction layer, as shown in figure 1.

1) THE PROCESSING LAYER

Often, the medical datasets collected from PCs or sensors may be incomplete, inconsistent, and contain errors. This may cause a classification problem.

Normalization methods: features having values of varying degrees of magnitude, may hurdle the performance of some

ML algorithms. So, these methods may be used to scale their values between 0 and 1.

Data imbalance reduction: features imbalance refers to a classification problem where some features are highly underrepresented. This causes the classifier to bias towards the majority of features. The algorithm mentioned in the research [39] was used to treat the problem of the imbalance dataset by over-sampling the minority class instances.

These error values should be eliminated before dividing the dataset into training, testing, and testing subsets. To evaluate the performance of this layer, six benchmark datasets are used. The first dataset is for breast cancer. The second and third are for colon, cervical cancer, whereas the fourth is for lung cancer. The fifth is prostate cancer. The sixth is ovarian cancer. We perform normalization on certain columns in these datasets during the first phase of pre-processing. The values of these columns have been rescaled. The second stage of pre-processing involves removing unnecessary columns such as the id column. The labels for each entry in the dataset are changed from string values to numeric values in the third and final phase of pre-processing. Each dataset has two classes (Malignant and Benign), which are converted to 0 and 1, respectively.
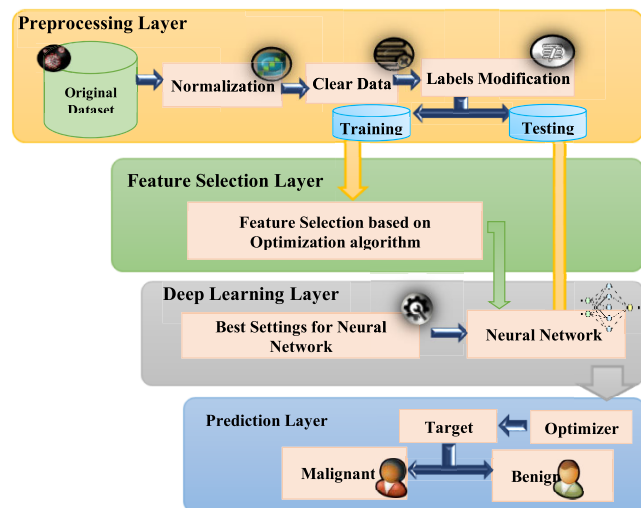
2) THE FS LAYER

In this layer, a recent binary variant of WOA, BACP-WOA-S, was proposed to solve the issue of FS. Therefore, the number *N* of features in the dataset will have to be *2N*, which represented a large area of features to be extensively searched for feature reduction. The proposed algorithm was employed to adapt the search space for achieving the optimal feature combination. The proposed algorithm was employed to adapt the search space for achieving the optimal feature combination. Moreover, the lesser the selected features, the better the solution. A special fitness function was used to measure each solution; the function was based on two primary objectives: the number of selected features in the solution and the error rate. For these aims to be achieved, we employed an NN. Moreover, these stated features were used to train the NN for achieving the most efficient model using our fitness function, which is defined as follows:

$$Fitness = \alpha E_R(D) + \beta \frac{|L|}{|N|} \tag{1}$$

$E_R(D)$ is the error rate, $|N|$ is the number of features, $|L|$ is the chosen feature's subset length, α determines the importance of the classification quality, and β corresponds to the feature reduction. We illustrate them as follows: $\alpha \in [0; 1]$, $\beta = 0.01$, and $\beta = 1 - \alpha$ are adopted from [40]. This fitness function indicates that classification quality and feature subset length are of distinct importance for the FS problem. In our experiment, we select parameters based on trial and error in modest and common simulations such as α where a high value ensures the optimal position or at least is reduced by a real rough set.

3) THE DL LAYER

This layer represents FFNN with the best settings using the proposed optimizer. The FFNN is trained to use a selected subset of features with structure parameters such as the number of layers, the number of neurons in the hidden layer, biases, and activation function are 3, 10, random, and TanH, respectively. As for the initial weights, H_2OFrame IDs initialize the weights such that the default initial_weight_distribution and initial_weight_scale parameters are uniform adaptive and one, respectively. Moreover, the training parameters learning rule and sum-squared error are Levenberg–Marquardt and 0.01, respectively. The FFNN is then trained to use the features and tested using the validation data. Subsequently, the error rate that is utilized to measure the fitness value is resolved. All the iterations and solutions in the population were achieved with previous tasks. Furthermore, the proposed BACP-WOA-S, binary WOA (BWOA), binary PSO (BPSO), binary GA (BGA), and binary gray wolf optimizer (BGWO) algorithms are examined in this layer. Each optimizer generates the best solution, and it is verified using the test data after the optimization process is performed. During the last testing process, various metrics were enlisted for comparison. The BACP-WOA-S uses training and validation data portions during the optimization process and for testing data after optimization. Therefore, we ensure that every optimizer examines the same data set portions in every iteration. In this manner, a fair comparison is obtained.

4) THE PREDICTION LAYER

The proposed optimizer, in the next subsection, is used in this layer to select a solution to an optimization problem that requires the resolution of the two functions: exploration and exploitation. to identify the infected cases.

B. THE PROPOSED ALGORITHM: S-SHAPED BINARY AC-PARAMETRIC WOA (BACP-WOA-S)

A continuous version of ACP-WOA uses (17) to move search agents inside the search space for adjusting their positions to any point. This process is called continuous space. Naturally, the FS problem is binary; if there is no alteration, the continuous version of ACP-WOA cannot be used to resolve the FS problem. Therefore, we propose a BACP-WOA-S version that is appropriate for resolving the FS problem. BACP-WOA-S binary indicates that only binary solutions [0, 1] are required for the candidate's options. The feature is not selected if it has a value of 0; however, it is selected if it has a value of 1.

To convert the solutions of ACP-WOA from continuous to binary, we first scale the values to be in the interval [0,1]. As per a previous study, the conversion is achieved using an S (Sigmoid)-shaped transfer function (TF). Its family has four TFs: S1, S2, S3, and S4 (Table 2). Thus, the elements of the location vectors must be transformed from 0 to 1 and vice versa, forcing agents to move in a binary space.

Table 2 lists the mathematical formulations for each TF, whereas figure 2 shows the mathematical curve of S-shaped. Algorithm 1 shows the steps involved in BACP-WOA-S. We use the following formula, proposed by Kennedy and Eberhart [41], to convert scaled continuous values to binary values.

$$X_d^{(i+1)} = \begin{cases} 1, & 0.5 \leq sigmoid(x) \\ 0, & otherwise \end{cases} \quad (2)$$

where $X_d^{(i+1)}$ is the updated binary position at specific iterations i and dimensions d, and $sigmoid(x)$ is given by (3)-(6) in table 2.

TABLE 2. S-shaped transfer functions.

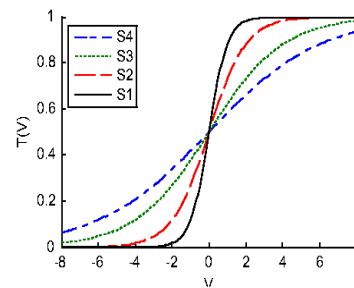| Name	Transfer Functions	
S1	$T(X) = \frac{1}{1+e^{-2x}}$	(3)
S2	$T(X) = \frac{1}{1+e^{-x}}$	(4)
S3	$T(X) = \frac{1}{1+e^{(-x/2)}}$	(5)
S4	$T(X) = \frac{1}{1+e^{(-x/3)}}$	(6)

FIGURE 2. S-shaped transfer functions.

V. EXPERIMENTAL RESULTS AND DISCUSSION

We conducted four experiments. The first experiment was to test the performance of the proposed optimizer, whereas the rest was to test the performance of the three frameworks. All experiments were conducted on Intel® Core™ i7-2.90 GHz processor with 32-GB RAM and an NVIDIA Quadro M2000M GPU. Based on datasets, four experiments were conducted for this purpose.

The packages which are used for prediction and visual representation of the findings are as follows: NumPy, pandas, time, random, math, h2o, sklearn.metrics, matplotlib.pyplot and h2o.estimators.deeplearning. (See appendix D, figure 12).

A. DATASET DESCRIPTION

The experiments were conducted on six benchmark datasets that have two labels: B = Benign and M = Malignant.

- Breast Cancer: This dataset is from the UCI ML Repository [42], [24]. These data were provided by the Hospital of Wisconsin University and created

Algorithm 1 Pseudo Code of BACP-WOA-S

1 **Input:**
 n number of whales in the population
 MaxIter number of iterations
2 **Output:**
     Position of the Optimal whale.
3 Initialize *a* and *n*.
4 Find X^*.
5 **while** *Criteria for halt are not met* **do**
6   **for** *Each Whale_i* **do**
7   Update *a, a2, A, C,* as in equations 25,26,27,28 respectively;
8   Update *l*, and *p*;
9   **if** *p < 0.5* **then**
10   **if** *A < 1* **then**
11   Update position by Eq.(16);
12   **else**
13   Select a random search agent (X_{rand});
14   Update position of current search agent by the Eq.(24);
15   **end**
16   **else** (*p ≥ 0.5*)
17   Update position by the Eq.(20);
18   **end**
19   Update X(t + 1) from the Eq.(3) or   the Eq.(4) or the Eq.(5) or the Eq(6);
20   **end**
21   Check if any agent;
22   Calculate the fitness of each agent;
23   Update X^* if there is a better solution;
24   *t = t + 1*;
25   **end**

by Dr. William H. Wolberg. Features are derived from a group of digitized images utilizing a biopsy of the breast. It contains information on 699 samples for 10 genes.
- Colon Cancer: This dataset was collected from Alba *et al.* [43]. It contains information on 62 samples for 2000 genes [26].
- Cervical Cancer: This dataset is from the UCI ML Repository [44]. The data was provided via a questionnaire distributed at a specialist facility in Jakarta (Indonesia). It contains information on 72 samples for 19 genes [27].
- Lung Cancer: This dataset was collected from Brigham and Women's Hospital, Harvard Medical School. It contains 181 samples. Each sample is described by 12533 genes [28].
- Prostate Cancer: The dataset was taken from the public Kent Ridge Bio-medical Data Repository. It contains 136 samples. Each sample is described by 12600 genes [29].

TABLE 3. List of datasets utilized in the research results.

Number	Dataset	Number of Attributes	Number of Instances	Number of classes Malignant	Benign
1	Breast	10	699	239	460
2	Colon	2000	62	22	40
3	Cervical	19	72	21	51
4	Lung	12533	181	31	150
5	Prostate	12600	136	77	59
6	Ovarian	15154	253	162	91

- Ovarian Cancer: This dataset contains 253 samples. Each sample is described by 15154 genes. It was collected from ovarian cancer patients [28].

As descriptive examples of issues that the proposed frameworks could solve, the datasets were selected to include several features, classes, and instances (Table 3). The instances were randomly categorized into three equivalent subsets in each dataset. These subsets are called train (80% of the data), test (5% of the data), and valid (15% of the data) in a cross-validation manner.

B. EVALUATION METRICS
Different evaluation matrices may be used to evaluate the performance of the trained model, such as the confusion matrix illustrated in figure 3, which is commonly used to offer various classification metrics and performance evaluation parameters.

FIGURE 3. Confusion matrix.

True Positive (TP): is the number of correct positive predictions. False Negative (FN): is the number of incorrect negative predictions. False Positive (FP): is the number of incorrect positive predictions. True Negative (TN): is the number of correct negative predictions. The following equations show some advanced evaluation metrics deduced from the confusion matrix.

$$Recall = \frac{TP}{(TP + FN)} \qquad (7)$$

$$Specificity = \frac{TN}{(TN + FP)} \tag{8}$$

$$Precision = \frac{TP}{(TP + FP)} \tag{9}$$

$$Accuracy = \frac{TP + TN}{(TP + FN + FP + TN)} \tag{10}$$

$$F1Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{11}$$

$$Mean\ Square\ Error\ (MSE)$$
$$= \frac{\sum_{i=1}^{N}(Predicted_i - Actual_i)^2}{N} \tag{12}$$

$$logloss$$
$$= -\frac{\sum_{i=1}^{N}\sum_{j=1}^{M} Actual_{ij} * \log\left(probability_{ij}\right)}{N} \tag{13}$$

C. PERFORMANCE ANALYSIS OF THE PROPOSED BACP-WOA-S

1) EXPERIMENT NUMBER 1

The first experiment depends on the second layer of the first framework and the third framework. The proposed BACP-WOA-S was tested against BWOA, BPSO, BGA, and BGWO algorithms to verify its performance. The algorithms were compared in terms of the average error, average fitness value, average selection size, and average standard deviation.

The configuration values are listed in table 4. The results of this experiment are summarized in tables 5–8.

TABLE 4. The configuration values.

Configuration	Value
Number of iterations	500
Search_agents_number	50
Dimension	Number of features in the data
Search domain	[0,1]
Number repetitions of runs	20
α Parameter in the fitness function	0.99
β Parameter in the fitness function	0.01

TABLE 5. Average error obtained from compared algorithms.

Number	WOA	BACP-WOA-S	GA	PSO	GWO
1	0.037	**0.013**	0.037	0.032	0.033
2	0.033	**0.008**	0.088	0.075	0.018
3	0.029	**0.000**	**0.000**	0.004	0.017
4	0.092	**0.000**	0.031	**0.000**	0.014
5	0.032	**0.007**	0.030	0.044	0.013
6	**0.000**	**0.000**	0.014	0.028	0.014

Tables 5-8 show cumulative results for all optimizers on six datasets. BACP-WOA-S showed superior results as it achieved the lowest average error, average selection, and standard deviation values, indicating its superior stability

TABLE 6. Average selection size obtained from compared algorithms.

Number	WOA	BACP-WOA-S	GA	PSO	GWO
1	0.825	0.501	0.505	0.625	0.6
2	0.204	0.127	0.503	0.48	0.521
3	0.476	0.361	0.476	0.418	0.532
4	0.403	0.400	0.548	0.481	0.594
5	0.223	0.221	0.451	0.475	0.585
6	0.141	0.490	0.518	0.497	0.594

TABLE 7. Average fitness value obtained from compared algorithms.

Number	WOA	BACP-WOA-S	GA	PSO	GWO
1	0.035	**0.019**	0.047	0.042	0.0443
2	0.038	**0.000**	0.092	0.080	0.0256
3	0.032	0.040	**0.000**	0.007	0.017
4	0.100	0.042	**0.005**	0.008	0.022
5	0.035	**0.011**	0.032	0.048	0.015
6	0.014	**0.000**	**0.000**	0.014	**0.000**

TABLE 8. Average standard deviation for fitness value obtained from compared algorithms.

Number	WOA	BACP-WOA-S	GA	PSO	GWO
1	0.016	**0.009**	**0.009**	0.011	0.018
2	0.062	**0.000**	0.068	0.081	0.076
3	0.042	0.040	**0.000**	0.021	0.213
4	0.082	**0.072**	0.080	0.090	0.098
5	0.031	**0.026**	0.036	0.040	0.042
6	0.010	**0.000**	**0.000**	0.010	**0.000**

to other algorithms. In other words, BACP-WOA-S outperformed other algorithms, such as the FS algorithm for the fitness value on all datasets, except for cervical and lung cancer datasets because they were relatively modest in terms of features, which indicated that BACP-WOA-S could select the optimal subset of features showing the lowest error.

2) EXPERIMENT NUMBER 2

The accuracy of the proposed optimizer is checked against the accuracy resulting from WOA, GA, PSO, and GWO optimizers. This is made to assert that the proposed method improves the accuracy. The results obtained are shown in the following figures for 50 iterations. These figures show the superiority of the proposed algorithms over the others.

The more results obtained are given in appendix E.

D. PERFORMANCE ANALYSIS OF THE PROPOSED FRAMEWORKS

In Experiments 3-5, the proposed frameworks were compared in terms of accuracy, precision, recall, F1-score, specificity, confusion matrix values, computational time, mean squared error (MSE), and logarithmic loss (log-loss) values.
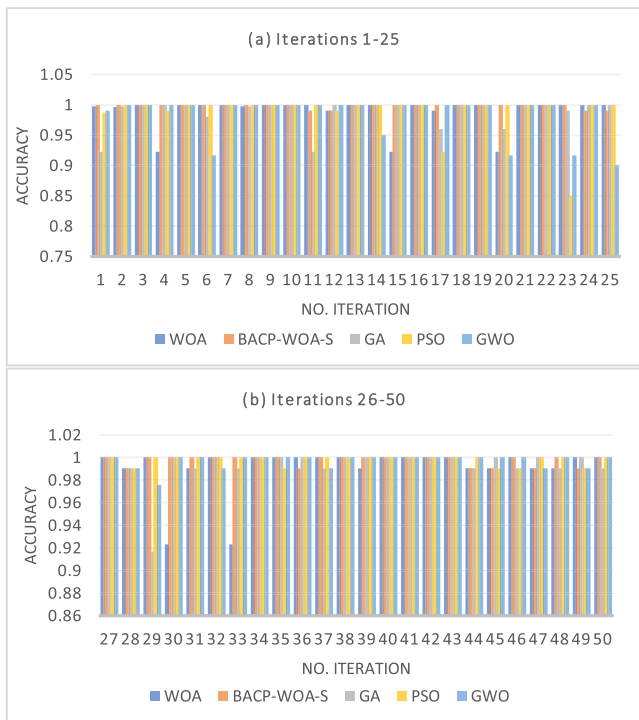
FIGURE 4. The various iterations of training the NN with a different set of features for "Breast" as selected by algorithms.
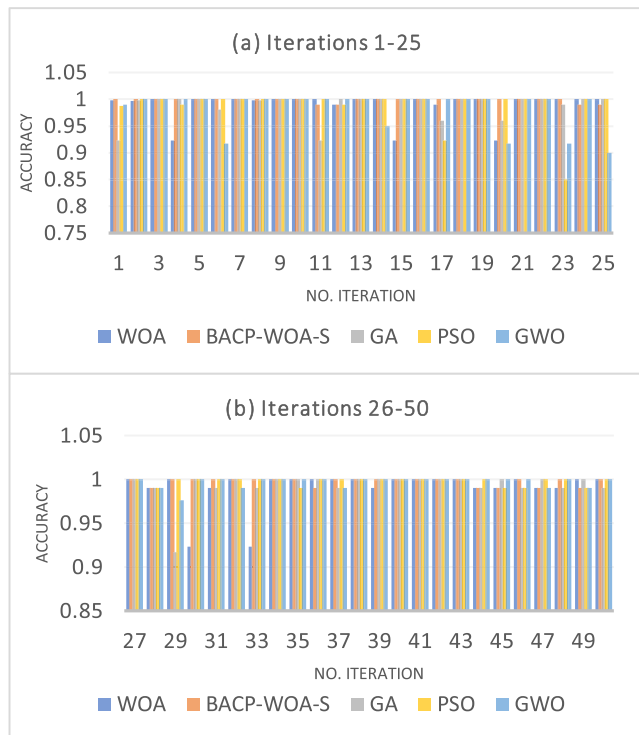

FIGURE 5. The various iterations of training the NN with a different set of features for "Colon" as selected by algorithms.

FIGURE 6. The various iterations of training the NN with a different set of features for "Cervical" as selected by algorithms.
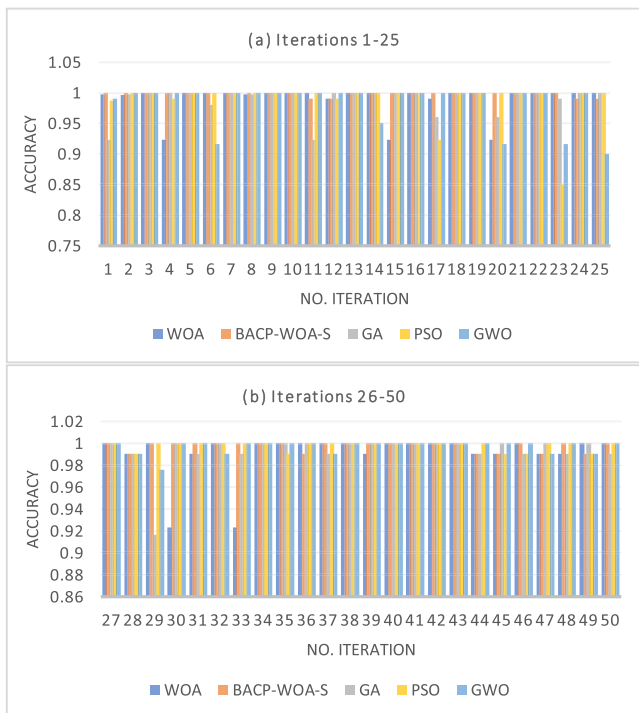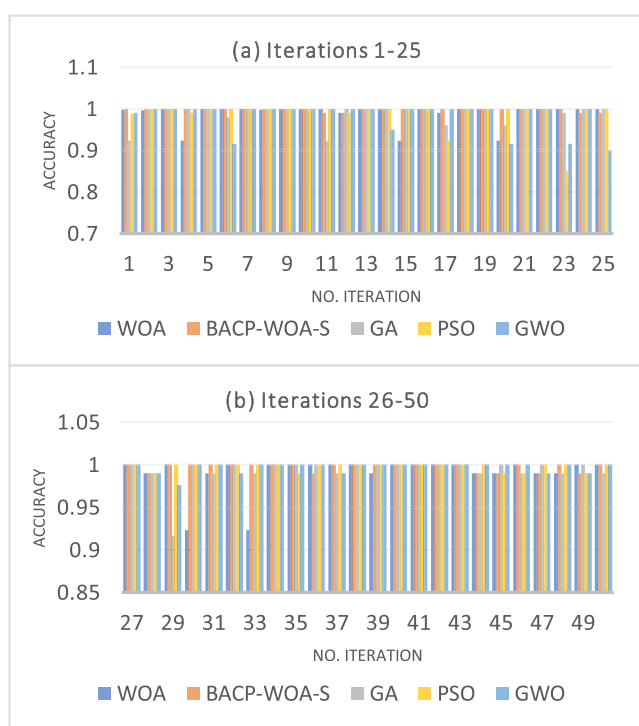

FIGURE 7. The various iterations of training the NN with a different set of features for "Lung" as selected by algorithms.

1) EXPERIMENT NUMBER 3 TESTS THE BEHAVIOR OF THE 1ST FRAMEWORK

This procedure demonstrates the effectiveness of the first framework in classifying cancer cases into four layers:

preprocessing, FS, DL without optimizing FFNN, and prediction.

Table 9 summarizes the performance of cancer datasets. As shown in this table, the first framework achieved an

TABLE 9. Performance obtained from the 1st framework cancer prediction.

Number	Accuracy	Precision	Recall	F1-score	Specificity	TP	FN	FP	TN	Computational Time (second)	MSE	Logloss
1	99.09%	100%	98.34%	99.16%	100%	61	1	0	48	61	0.033	0.132
2	90.90%	80%	100 %	88.88%	85.71%	4	0	1	6	1351	0.062	0.245
3	91.67%	100%	90.9 %	95.23%	100%	10	1	0	1	15	0.258	0.835
4	96.67%	100%	83.33%	90.90%	100%	5	1	0	24	24710	0.099	0.435
5	96%	90.9%	100%	95.23%	93.33%	8	0	1	16	18666	0.201	0.685
6	95.5%	100%	85.7%	100%	100%	12	2	0	31	18700	0.022	0.119

TABLE 10. Performance obtained from the 2nd framework cancer prediction.

Number	Accuracy	Precision	Recall	F1-score	Specificity	TP	FN	FP	TN	Computational Time (second)	MSE	Logloss
1	97.97%	96.72%	100 %	98.33%	95%	59	0	2	38	0.05	0.001	0.002
2	100%	100%	100 %	100%	100%	1	0	0	5	1	0.003	0.007
3	100%	100%	100 %	100%	100%	10	0	0	2	0.012	0.004	0.014
4	100%	100%	100%	100%	100%	8	0	0	20	19	0.001	0.004
5	68.42%	55.56%	71.43%	62.50%	80%	5	2	4	8	14	0.014	0.048
6	92.3%	82.3%	100%	100%	88%	14	0	3	22	16	0.053	0.273

TABLE 11. Performance obtained from the 3rd framework cancer prediction.

Number	Accuracy	Precision	Recall	F1-score	Specificity	TP	FN	FP	TN	Computational Time (second)	MSE	Logloss
1	100%	100%	100 %	100%	100%	65	0	0	33	500	0.009	0.038
2	100%	100%	100 %	100%	100%	6	0	0	8	13400	0.106	0.545
3	100%	100%	100 %	100%	100%	11	0	0	2	100	0.048	0.172
4	100%	100%	100%	100%	100%	7	0	0	16	246000	0.012	0.051
5	100%	100%	100%	100%	100%	10	0	0	15	180000	0.282	0.936
6	100%	100%	100%	100%	100%	13	0	0	32	180280	0.050	0.110

TABLE 12. p-values of 3rd framework using Wilcoxon's rank-sum.

Dataset	3rd framework	
	p-value	Accuracy
Breast Cancer	0.017239581	99.98%
Colon Cancer	0.013958892	99.99%
Cervical Cancer	0.047823524	99.97%
Lung Cancer	0.013958892	99.9%
Prostate Cancer	0.047823524	99.9%
Ovarian Cancer	0.000671471	99.99%

average of 94.97% accuracy, 95.15% precision, 93.05% recall, 94.9% F1-score, 96.51% specificity, 0.113 MSE, and 0.409 log-loss for all cancer datasets (See appendix C, figure 13).

average of 93.12%, 89.1%, 95.24%, 93.47%, 93.83%, 0.013, 0.058 for accuracy, precision, recall, F1-score, specificity, MSE, and log-loss, respectively, for all cancer datasets (See appendix C, figure 14).

2) EXPERIMENT NUMBER 4 TESTS THE BEHAVIOR OF THE 2nd FRAMEWORK

This procedure demonstrates the effectiveness of the second framework in classifying cancer cases into three layers: pre-processing, DL with optimizing FFNN, and prediction.

Table 10 summarizes the performance of cancer datasets. As shown in this table, the second framework achieved an

3) EXPERIMENT NUMBER 5 TESTS THE BEHAVIOR OF THE 3rd FRAMEWORK

This procedure demonstrates the effectiveness of the third framework for classifying cancer cases into four layers: pre-processing, FS, DL with optimizing FFNN, and prediction.

Table 11 summarizes the performance of cancer datasets. As shown in this table, the third framework achieved an

TABLE 13. A comparison between the three proposed frameworks on cancer datasets in terms of accuracy, precision, and computational time.

Number	1st Framework			2nd Framework			3rd Framework		
	Accuracy	Precision	Computational Time (second)	Accuracy	Precision	Computational Time (second)	Accuracy	Precision	Computational Time (second)
1	99.09%	100%	61	97.97%	96.72%	**0.05**	100%	100%	500
2	90.90%	80%	1351	100%	100%	**1**	100%	100%	13400
3	91.67%	100%	15	100%	100%	**0.012**	100%	100%	100
4	96.67%	100%	24710	100%	100%	**19**	100%	100%	246000
5	96%	90.9%	18666	68.42%	55.56%	**14**	100%	100%	180000
6	95.5%	100%	18700	92.3%	82.3%	**16**	100%	100%	180280

TABLE 14. Summary of the studies introduced in the literature review section.

Ref	Year	Methods	Accuracy					
			Breast	Colon	Cervical	Lung	Prostate	Ovarian
Y. Chen et al. [30]	2008	NN	-	88.7%	-	97%	-	97%
X. S. Yang et al. [31]	2009	Cuckoo Search (CS)	-	-	-	87.7%	85.3%	-
X. S. Yang et al. [32]	2014	The multiobjective cuckoo search (MOCS)	-	-	-	87.7%	91.2%	-
Liu, M. et al. [33]	2018	• ReliefF • PSO and SVM	-	98.1%	-	86.5%	-	-
Wu, P. et al. [34]	2018	A Novel Complex Network and Hybrid GP+PSO	-	95.52%	-	**100%**	-	**-**
Adiwijaya et al. [35]	2020	SVM Classifier	-	85.71%	-	70.13%	88.75%	**100%**
Adiwijaya et al. [35]	2020	LMBP classifier	-	81.43%	-	96.25%	97.65%	**100%**
Othman, M. S. et al. [36]	2020	Their multi-objective search algorithm has been designed with evolutionary operators	-	-	-	93.7%	92.2%	97.5%
Saqib, P., et al. [37]	2020	• GA • RF	86.60%	96.77%	-	-	98.04%	**100%**
Cahyaningrum, K., et al. [38]	2020	• ANN • ANN-GA	-	86%	-	**100%**	82%	-
Akter et al. [23]	2021	• DT • RF • XGBoost	-	-	93.33%	-	-	-
Sahar, A. [24]	2021	RF + GA	96.82%	-	-	-	-	-
S. Jijitha et al.[25]	2021	LR	99.27%	-	-	-	-	-
Murad, A., et. al. [26]	2021	DT	-	94%	-	-	-	-
Francesco Curia [27]	2021	MLP	-	-	94.5%	-	-	-
Xiongshi, D. el. al.[28]	2021	SVM Classifier based on selected genes by XGBoost-MOGA	82.33%	90.24%	-	**100%**	98%	**100%**
Samson, A., et al[29]	2021	1D-CNN	-	-	-	-	99.07%	-
Proposed 3rd framework	2021	FFNN	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**

average of 100% accuracy, 100% precision, 100% recall, 100% F1-score, 100% specificity, 0.085 MSE, and 0.309 log-loss for all cancer datasets (See appendix C, figure 15).

Notably, the best performance was obtained using the third framework (Tables 9–11) (See appendix D, figure 16).

4) EXPERIMENT NUMBER 6 WILCOXON'S RANK-SUM

The p-values of the proposed 3rd framework are calculated using Wilcoxon's rank-sum test. This test determines if the outputs of the proposed have a significant difference or not. If the p-value < 0.05, it means that the

3rd framework results have significantly different. Otherwise, a p-value > 0.05 means that the results have no significant difference. Table 12 shows the results of the p-value and mean of accuracy by using the t-Test. Where the p-value values are smaller than 0.05. This proves the superiority of the 3rd proposed framework and it is statistically significant.

E. DISCUSSION

In table 13, the accuracy of the first framework is less than the third, despite their proximity in time. Moreover, the third framework is better than the second framework, but with a comparatively longer prediction time. The accuracy of the second framework is slightly lower than that of the third framework, but with a much shorter prediction time. As a result, for noncritical patients, the third framework is suitable because of its high accuracy, whereas the second framework is suitable for critical patients because of its shorter prediction time. The bolded letters in the table represent these results.

In table 14, the proposed method uses the optimizer BACP-WOA-S to select the optimal set of features that are used as input to optimize FFNN (best number of layers and number of neurons). The table shows the results obtained from the proposed framework (100% breast, 100% colon, 100% cervical, 100 lung, 100% prostate, and 100% ovarian cancer datasets). The bolded letters in the table represent the best results. The 3rd framework outperforms the other frameworks for the breast, colon, cervical, prostate cancer datasets. Regarding the lung and ovarian cancer datasets, the 3rd framework, Deng *et al.* [28], Wu and Wang [34], Adiwijaya *et al.* [35], Saqib *et al.* [37], and Cahyaningrum and Astuti [38] outperform the other framework.

F. DATA ANALYSIS

First, as we mentioned earlier, cancer is the second disease that causes one out of six deaths worldwide. In 2021, ovarian cancer was one of the most common causes of cancer death, whereas 13,770 women died [45]. In addition, lung (1,800,000 deaths), colon (935,000 deaths), and breast (685,000 deaths) were some of the most common causes of cancer death in 2020 [45]. Prostate and cervical cancer were the most common causes of cancer deaths in 2019 and 2018, causing the deaths of 31,638 males and 34,000 women, respectively [45]. These large numbers of deaths can be reduced if cases are detected and treated early because the most crucial factors for a cancer patient are accuracy and prediction time, since the higher the prediction accuracy of the disease, the higher the chance of the patient receiving treatment. Meanwhile, the cost of treatment is lower.

Second, according to previously mentions reasons, this paper proposed three different cancer prediction frameworks each with a different prediction time. According to the experimental results, for noncritical patients, the third framework is suitable because of its high accuracy, whereas the second framework is suitable for critical patients because of its shorter prediction time.

Third, in proposed frameworks, the modified optimizer is required for FS to select optimal features and also to tune FFNN (number of layers and number of neurons per layer). When using DL, the accuracy increases as the number of inner layers increases, but the more the layers increase, the longer the processing time becomes. Therefore, we propose three frameworks, 1st with one layer, 2nd with two layers, and 3rd with three layers, to choose one of them with high accuracy and less time.

VI. CONCLUSION AND FUTURE DIRECTIONS

In this study, we proposed a selected DL cancer prediction framework based on the dynamic group to achieve a balance between exploration and exploitation. The proposed BACP-WOA-S divides the solutions into two methods: the first group is responsible for exploration and the second is responsible for exploration. The first group applies two techniques: search around individuals and mutation. Similarly, the second group applies two techniques: move toward the leader and search around the leader.

In this study, we selected a framework from three DL H_2O frameworks based on their accuracy and processing time. The performance of the FFNN framework was better in breast, colon, cervical, and prostate cancer but [28], [34], [35], [37], [38] in the lung, and ovarian cancer datasets, which are the same as our result. The proposed algorithm achieved an accuracy of 100%, 100%, 100%, 100%, 100%, and 100% for breast, colon, cervical, lung, prostate, and ovarian cancer datasets, respectively. Although the selected DL framework achieved higher accuracy in the six benchmark datasets, we ensured the stability of the control scheme because it was designed in the proposed framework.

In the future, new metaheuristic algorithms will be tested against our proposed algorithm. Further, the proposed algorithm will be applied to solve an additional binary classification problem. The impact of increasing the difficulty of datasets used in DL is still under study. From tables 5–8, we can obtain the average 0.0046 error, 0.350 selection size, 0.0186 fitness value, 0.0245 standard deviations for the proposed algorithm which outperforms other algorithms such as the FS algorithm for the fitness value on all datasets, except for cervical cancer and lung cancer datasets. Therefore, we plan to apply the broad learning system (BLS) to improve the result.

APPENDICES
APPENDIX A
A. FEED-FORWARD NEURAL NETWORK (FFNN)

The ANN was inspired by the human neural system. It is used in various applications, including pattern recognition, optimization, and control. The FFNN is similar to ANN in that nodes' connections do not build a loop [46]. The important reason for selecting this layer is that the input data are simply generalized and time scales where disruptions are easier to categorize.

In the FFNN, the neuron is the essential part; however, data flows in numeric form between m neurons in the past layer to one neuron i, and the data flow over as an aggregate as follows:

$$Z_i = a(\sum_{j=1}^{m} q_{ij}x_{ij} - c_{i0}) \tag{14}$$

where q_{ij} is the weight of contact between neuron j of the past layer to the present neuron i. x_{ij} = relating data and c_{i0} = ingrained threshold for neuron i is considered as a standard weight.

A popular approach is to use activation function "a" to generalize hidden neurons or output neurons (Table 15). The network comprises three layers, input, hidden, and output layers, with the neurons of the input and output layers as m and n, respectively (Figure 8).

TABLE 15. Activation functions.

Name	Activation Function
Linear	$A(x) = x$
Sigmoid	$A(x) = \frac{1}{1 + e^{-x}}$
TanH	$A(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$
Step	$A(x) = \begin{cases} -1 & \text{if } x < 0 \\ 1 & \text{if } x \geq 0 \end{cases}$

The aim is to identify an FFNN weight set that correctly reflects the relationship between the input vector and desired output vector. The network has been training on a collection of P input–output vector combined with an error back propagation algorithm [47] that can minimize the execution work for the pattern using the following equation:

$$E = \frac{1}{2Z} \sum_{z=1}^{Z} \sum_{g=1}^{G} (y_g^z - d_g^z)^2 \tag{15}$$

where E is the total mean sum squared error between the measured outputs, y^z is actual state, and d^z is desired state. z and g denote the values for the zth training set and g^{th} component of the output vector.

B. AC-PARAMETRIC WOA (ACP-WOA)

In 2016, Mirjalili and Lewis [48] proposed a wrapper-based FS algorithm (WOA) with better performance than some existing algorithms.

1) ENCIRCLING PREY

In this strategy, the humpback whale starts with the best applicant solution, which is the target prey. Consequently, the remaining search agents update their positions to those of the best search agents. This behavior can be mathematically formulated as follows:

$$\vec{D} = \left| \vec{C}.\vec{X}^*(i) - \vec{X}(i) \right| \tag{16}$$

$$\vec{X}(i+1) = \vec{X}^*(i) - \vec{A}.\vec{D} \tag{17}$$

where i is iteration, $\vec{A}$ and $\vec{C}$ are coefficient vectors, and $\vec{X}*$ is the location vector of the optimal solution obtained.
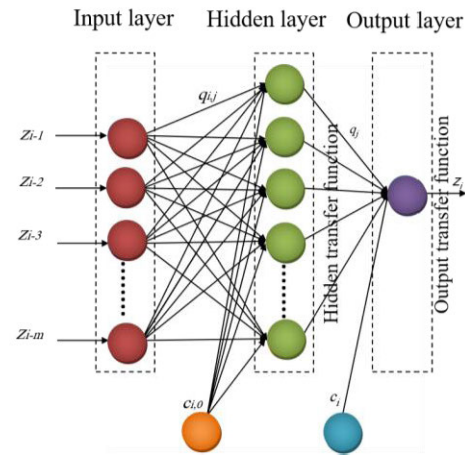

FIGURE 8. FFNN model.

We measure $\vec{A}$ and $\vec{C}$ by

$$\vec{A} = 2\vec{a}.\vec{r_1} - \vec{a} \tag{18}$$

$$\vec{C} = 2.\vec{r_1} \tag{19}$$

where $\vec{r_1}$ is defined as vectors in [0;1]. We require to linearly reduce $\vec{a}$ vector from two to zero values through the iterations.

2) BUBBLE-NET ATTACKING METHOD (EXPLOITATION PHASE)

The bubble-net attacking behavior is divided as follows.

1) Shrinking encircling mechanism: This behavior of humpback whale was achieved by reducing the value of $\vec{a}$ from two to zero in (18) over the number of processes, where $\vec{A}$ is a random value in the interval $[-\vec{a}, \vec{a}]$. Between the original positions and the current best position, a search agent's new positions can be described.

2) Spiral updating mechanism: The distance between the whale and the prey can be described by

$$\vec{X}(i+1) = \vec{X}^*(i) + \vec{D}' \times e^{bl} \times cos(2\pi l) \tag{20}$$

$$\vec{D}' = \left| \vec{C}.\vec{X}^*(i) - \vec{X}(i) \right| \tag{21}$$

where $\vec{D}'$ is the distance from the population optimal solution to the current individual whale, b is a fixed value used to describe the logarithmic spiral shape, and l is a value inside range $[-1, 1]$. To switch between these two mechanisms, 50% of it is assumed the shrinking mechanism, and the other 50% is the spiral mechanism. Consequently, the mathematical model is as follows:

$$\vec{X}(i+1) = \begin{cases} \vec{X}^*(i) - \vec{A}.\vec{D}, & p < 0.5 \\ \vec{D}'e^{bl}.cos(2\pi l)) + \vec{X}^*(i), & p \geq 0.5 \end{cases} \tag{22}$$

where p is a value defined in [0, 1].

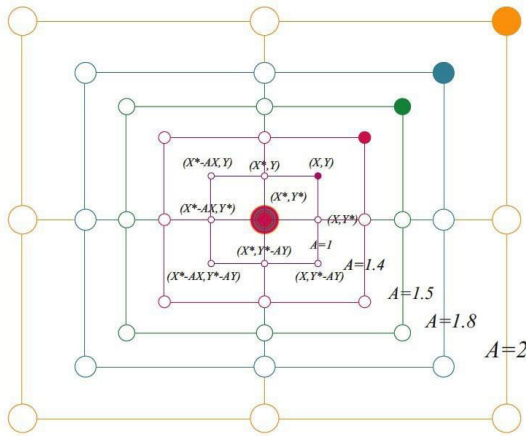**FIGURE 9.** The exploration mechanism implemented in WOA ($\bar{X}^*$ is a randomly chosen search agent).
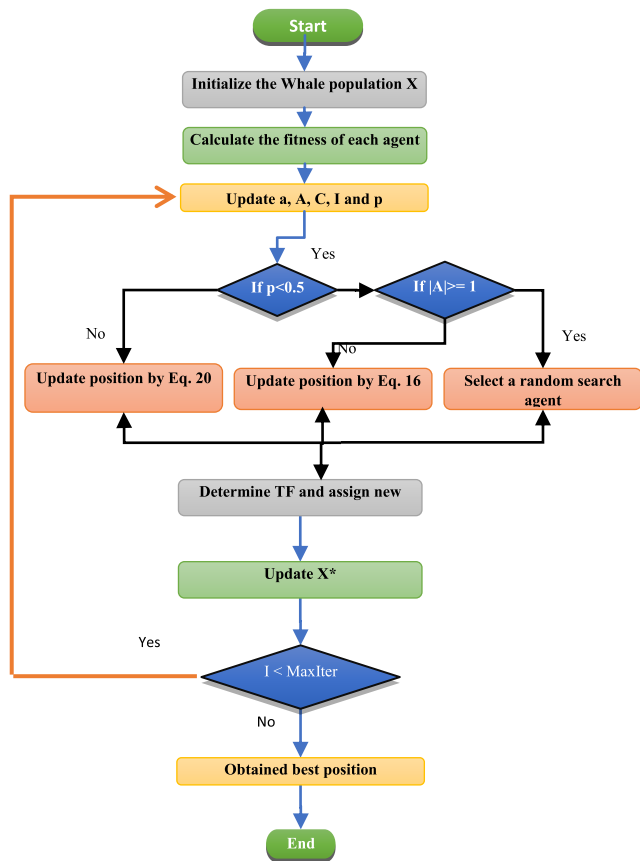

FIGURE 11. Block diagram of the 1st framework.

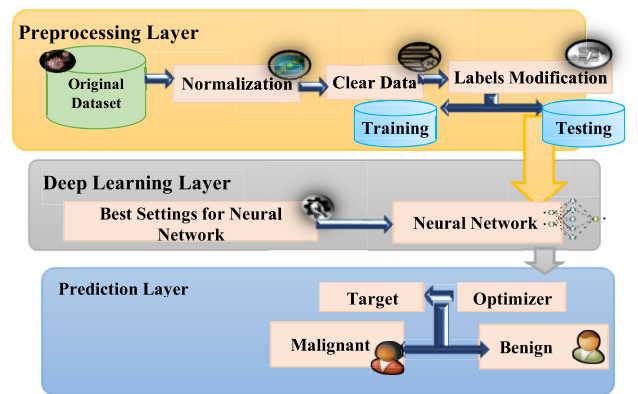**FIGURE 10.** ACP-WOA flowchart.

FIGURE 12. Block diagram of the 2nd framework.

where $\vec{X}_{rand}$ is a random position vector selected from the current population.

Multiple parameters are used to customize the algorithm during these two phases; some of them have a considerable impact on efficiency. Most whale variants are capable of achieving higher efficiency. Moreover, researchers had devoted little interest to the intelligibility of the typical whale.

In [49], the suggested revisions to the parameters *a, a2, A,* and *C* could affect the WOA exploration and exploitation. The following equations describe these modifications.

$$a = 2 - t \times \frac{2}{Max\_iter^2} \tag{25}$$

$$a2 = -1 - t \times \frac{1}{Max\_iter^2} \tag{26}$$

$$A = -2 \times a \times sinr1 - a \tag{27}$$

$$C = -2 \times sinr2 \tag{28}$$

As realized, *a* and *a2* are time versions that slowly reduce the change range when the denominator is changed to the square of maximum iterations. The fluctuation of a transition is more effective than standard randomization after the alteration of *A* and *C* is transformed into a sinusoidal change. Figure 10 shows the flowchart of the modified ACP-WOA.
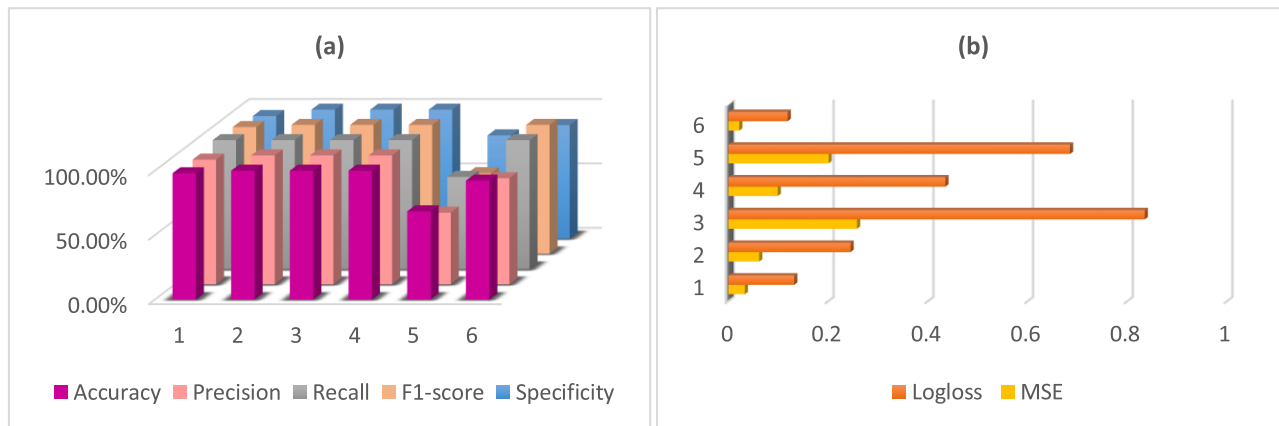
3) SEARCH FOR PREY (EXPLORATION PHASE)

Figure 9 shows the exploration phase, $\vec{A}$ can use the random values within $1 \prec A \prec -1$ to move the agent away from the reference whale. The new position of a search agent is discovered by selecting an agent randomly that allows the WOA to perform a global search.

$$\vec{D} = |\vec{C} \cdot \vec{X}_{rand} - \vec{X}| \tag{23}$$

$$\vec{X}(i+1) = \vec{X}_{rand} - \vec{A} . \vec{D} \tag{24}$$

FIGURE 13. (a) Accuracy, precision, recall, f1-score, and specificity and (b) log-loss, and MSE for 1st framework.
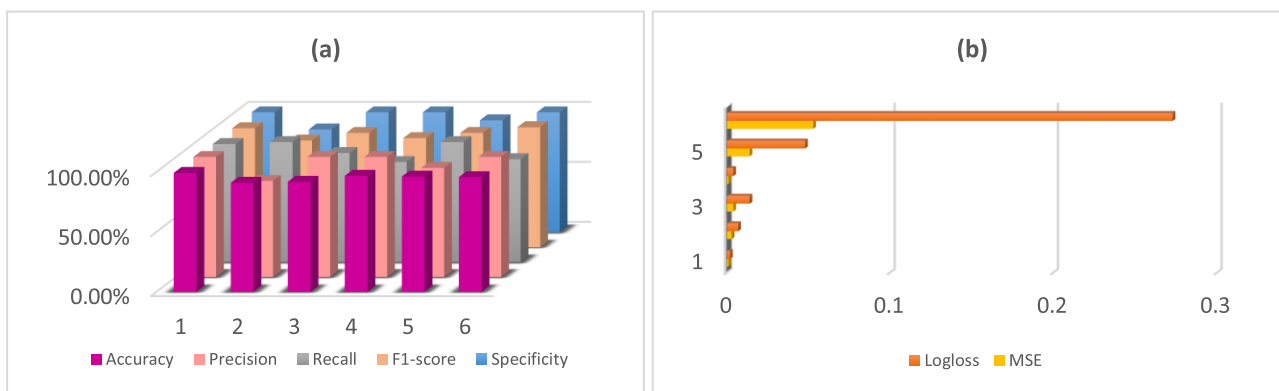

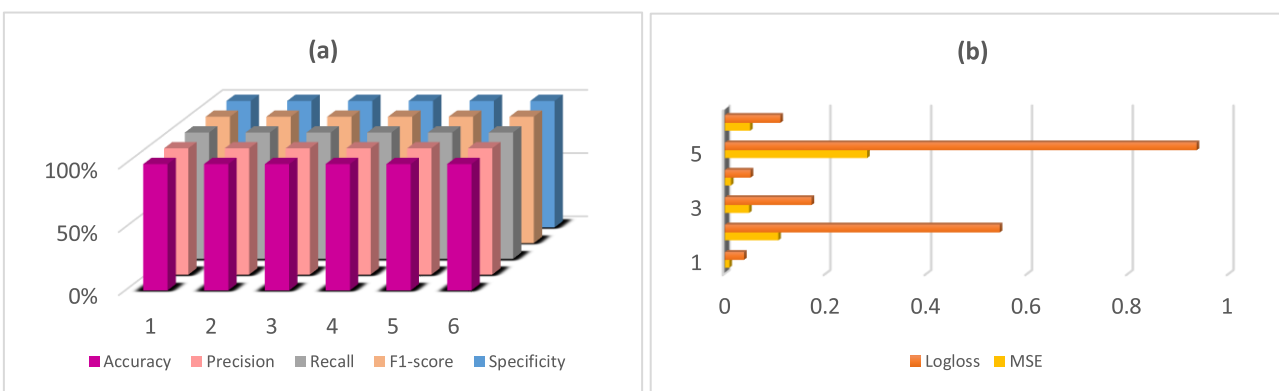**FIGURE 14.** (a) Accuracy, precision, recall, f1-score, and specificity and (b) log-loss, and MSE for 2nd framework.

FIGURE 15. (a) Accuracy, precision, recall, f1-score, and specificity and (b) log-loss, and MSE for 3rd framework.

APPENDIX B

C. FFNN FRAMEWORK WITH FEATURE SELECTION

In the first framework, H$_2$O is based on four layers. The first layer is the pre-processing layer. The second layer is where BACP-WOA-S is used with the FS layer to select the optimal features that will be used in the third layer, which is the DL layer. To train the NN, the third layer uses the optimal-selected features and NN's default settings. The presented optimizer is used in the final layer to identify the infected cases. (Appendix B -figure 11).

1) FFNN FRAMEWORK WITH BEST THE CONFIGURATION

In the second framework, H$_2$O comprises three layers. The second layer is where the DL uses the best settings of the
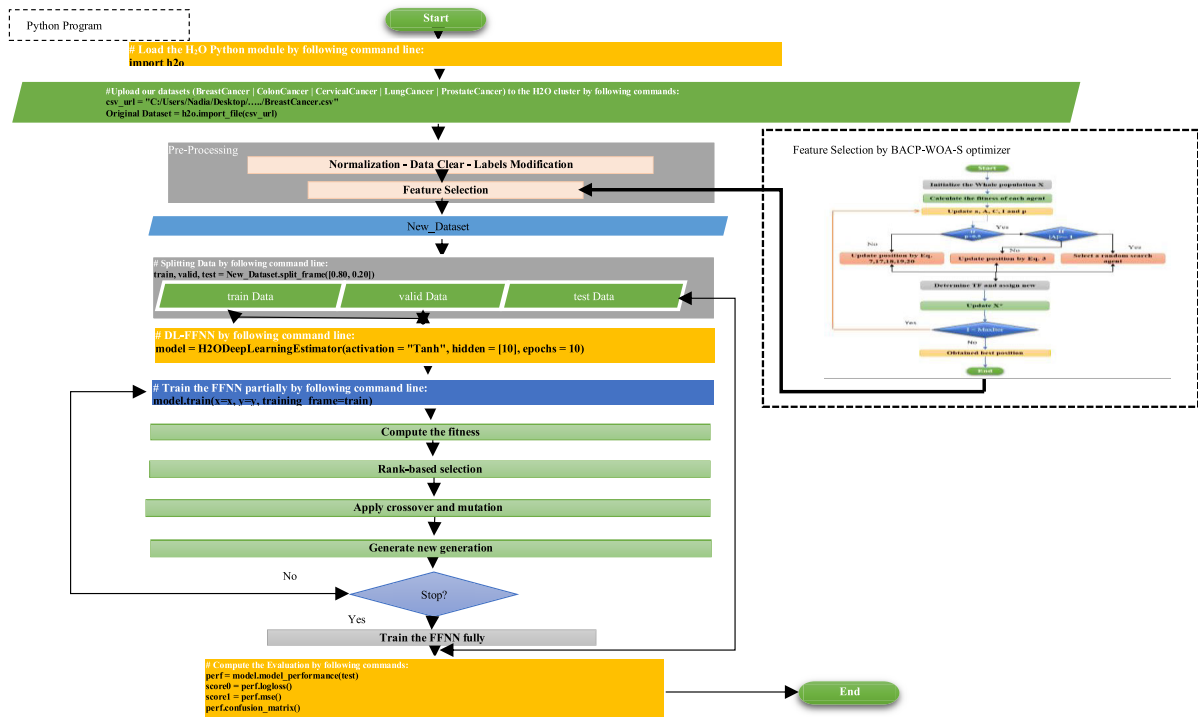
FIGURE 16. Flowchart of the 3rd proposed framework.

FIGURE 17. The various iterations of training the NN with a different set of features for "Prostate" as selected by algorithms.

FIGURE 18. The various iterations of training the NN with a different set of features for "Ovarian" as selected by algorithms.

NN to train the FFNN for obtaining the best classification accuracy. (Appendix B, figure 12)
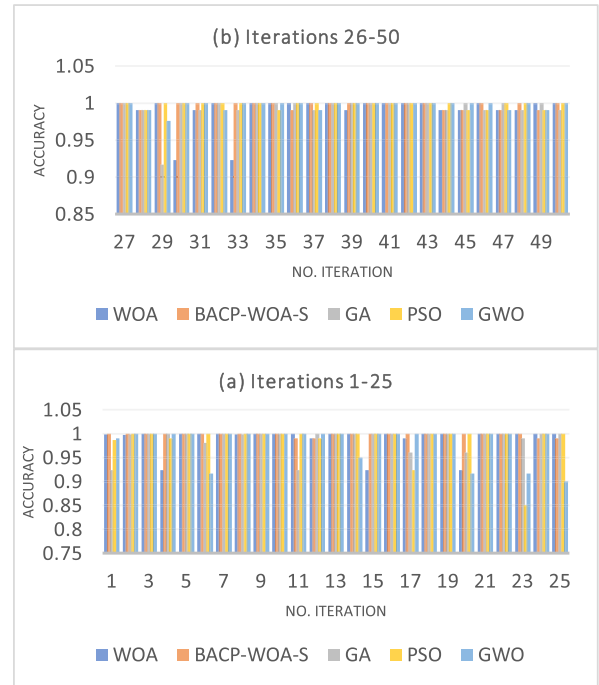
IPENDIX C
See Fig. 13–15.

APPENDIX D
The following shows the python program flowchart in detail for the third proposed framework, see Fig. 16.

APPENDIX E
See Fig. 17 and 18.

## REFERENCES

[1] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2019," *CA, Cancer J. Clin.*, vol. 69, no. 1, pp. 7–34, 2019.

[2] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray, "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA, Cancer J. Clin.*, vol. 71, no. 3, pp. 209–249, May 2021.

[3] I. Hameed, S. R. Masoodi, P. A. Malik, S. A. Mir, K. Ghazanfar, and B. A. Ganai, "Genetic variations in key inflammatory cytokines exacerbates the risk of diabetic nephropathy by influencing the gene expression," *Gene*, vol. 661, pp. 51–59, Jun. 2018.

[4] J. D. Cohen, L. Li, Y. Wang, C. Thoburn, and B. Afsari, "Detection and localization of surgically resectable cancers with a multi-analyte blood test," *Science*, vol. 359, no. 6378, pp. 926–930, Feb. 2018.

[5] M. A. Haendel, C. G. Chute, and P. N. Robinson, "Classification, ontology, and precision medicine," *New England J. Med.*, vol. 379, no. 15, pp. 1452–1462, Oct. 2018.

[6] J. Phallen, "Direct detection of early-stage cancers using circulating tumor DNA," *Sci. Transl. Med.*, vol. 9, p. 403, Aug. 2017.

[7] J. D. Schiffman, P. G. Fisher, and P. Gibbs, "Early detection of cancer: Past, present, and future," *Amer. Soc. Clin. Oncol. Educ. Book*, no. 35, pp. 57–65, May 2015.

[8] S. Behjati and P. S. Tarpey, "What is next generation sequencing?" *Arch. Disease Childhood-Educ. Pract.* vol. 98, no. 6, pp. 236–238, 2013.

[9] L. Ding, M. C. Wendl, D. C. Koboldt, and E. R. Mardis, "Analysis of next-generation genomic data in cancer: Accomplishments and challenges," *Hum. Mol. Genet.*, vol. 19, no. R2, pp. R188–R196, Oct. 2010.

[10] *Welcome to H2O 3-H2O 3.30.0.3 Documentation*. Accessed: Aug. 2021. [Online]. Available: http://docs.h2o.ai/h2o/latest-stable/h2o-docs/welcome.html

[11] Y. Bengio, *Learning Deep Architectures for AI*. New York, NY, USA: Now, 2009.

[12] S. M. El-Ghamrawy and I. A. El-Desouky, "Distributed multi-agent communication system based on dynamic ontology mapping," *Int. J. Commun. Netw. Distrib. Syst.* vol. 10, no. 1, pp. 1–24, 2013.

[13] S. M. El-Ghamrawy and A. I. Eldesouky, "An agent decision support module based on granular rough model," *Int. J. Inf. Technol. Decis. Making*, vol. 11, no. 4, pp. 793–820, Jul. 2012.

[14] S. M. El-Ghamrawy, "A knowledge management framework for imbalanced data using frequent pattern mining based on Bloom filter," in *Proc. 11th Int. Conf. Comput. Eng. Syst. (ICCES)*, Dec. 2016, pp. 226–231.

[15] E. A. El-Shafeiy, A. I. El-Desouky, and S. M. Elghamrawy, "Prediction of liver diseases based on machine learning technique for big data," in *Proc. Int. Conf. Adv. Mach. Learn. Technol. Appl.* Cham, Switzerland: Springer, 2018, pp. 362–374.

[16] S. Elghamrawy, "Healthcare informatics challenges: A medical diagnosis using multi agent coordination-based model for managing the conflicts in decisions," in *Proc. Int. Conf. Adv. Intell. Syst. Inform.* Cham, Switzerland: Springer, 2020, pp. 347–357.

[17] K. C. Tan, S. C. Chiam, A. A. Mamun, and C. K. Goh, "Balancing exploration and exploitation with adaptive variation for evolutionary multi-objective optimization," *Eur. J. Oper. Res.*, vol. 197, no. 2, pp. 701–713, Sep. 2009.

[18] M. M. Mafarja and S. Mirjalili, "Hybrid whale optimization algorithm with simulated annealing for feature selection," *Neurocomputing*, vol. 260, pp. 302–312, Oct. 2017.

[19] Y. Zhou, K. Zhou, and W. Bi, "Feature selection based on hybridization of genetic algorithm and competitive swarm optimizer," *Soft Comput.*, vol. 24, no. 15, pp. 11663–11672, Aug. 2020.

[20] H.-H. Hsu, C.-W. Hsieh, and M.-D. Lu, "Hybrid feature selection by combining filters and wrappers," *Expert Syst. Appl.*, vol. 38, no. 7, pp. 8144–8150, 2011.

[21] T. Butler-Yeoman, B. Xue, and M. Zhang, "Particle swarm optimisation for FS: A hybrid filter-wrapper approach," in *Proc. IEEE Congr. Evol. Comput. (CEC)*, Dec. 2015, pp. 2428–2435.

[22] Y. C. Ho and D. L. Pepyne, "Simple explanation of the no-free-lunch theorem and its implications," *J. Optim. Theory Appl.*, vol. 115, no. 3, pp. 549–570, Dec. 2002.

[23] L. Akter, Ferdib-Al-Islam, M. M. Islam, M. S. Al-Rakhami, and M. R. Haque, "Prediction of cervical cancer from behavior risk using machine learning techniques," *Social Netw. Comput. Sci.*, vol. 2, no. 3, pp. 1–10, May 2021.

[24] S. A. El-Rahman, "Predicting breast cancer survivability based on machine learning and features selection algorithms: A comparative study," *J. Ambient Intell. Hum. Comput.*, vol. 12, no. 8, pp. 8585–8623, Aug. 2021.

[25] S. Jijitha and T. Amudha, "Breast cancer prognosis using machine learning techniques and genetic algorithm: Experiment on six different datasets," in *Evolutionary Computing and Mobile Sustainable Networks*. Singapore: Springer, 2021, pp. 703–711.

[26] M. Al-Rajab, J. Lu, and Q. Xu, "A framework model using multifilter feature selection to enhance colon cancer classification," *PLoS ONE*, vol. 16, no. 4, Apr. 2021, Art. no. e0249094.

[27] F. Curia, "Cervical cancer risk prediction with robust ensemble and explainable black boxes method," *Health Technol.*, vol. 11, no. 4, pp. 875–885, Jul. 2021.

[28] X. Deng, M. Li, S. Deng, and L. Wang, "Hybrid gene selection approach using XGBoost and multi-objective genetic algorithm for cancer classification," 2021, *arXiv:2106.05841*.

[29] S. Parisapogu, C. S. R. Annavarapu, and M. Elloumi, "1-Dimensional convolution neural network classification technique for gene expression data," in *Deep Learning for Biomedical Data Analysis*. Cham, Switzerland: Springer, 2021, pp. 3–26.

[30] Y. Chen and Y. Zhao, "A novel ensemble of classifiers for microarray data classification," *Appl. Soft Comput.*, vol. 8, no. 4, pp. 1664–1669, Sep. 2008.

[31] X.-S. Yang and S. Deb, "Cuckoo search via Lévy flights," in *Proc. World Congr. Nature Biologically Inspired Comput. (NaBIC)*, Dec. 2009, pp. 210–215.

[32] X.-S. Yang and S. Deb, "Cuckoo search: Recent advances and applications," *Neural Comput. Appl.*, vol. 24, no. 1, pp. 169–174, Jan. 2014.

[33] M. Liu, L. Xu, J. Yi, and J. Huang, "A feature gene selection method based on ReliefF and PSO," in *Proc. 10th Int. Conf. Meas. Technol. Mechatronics Autom. (ICMTMA)*, Feb. 2018, pp. 298–301.

[34] P. Wu and D. Wang, "Classification of a DNA microarray for diagnosing cancer using a complex network based method," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 16, no. 3, pp. 801–808, May 2019.

[35] A. Adiwijaya, U. N. Wisesty, E. Lisnawati, A. Aditsania, and D. S. Kusumo, "Dimensionality reduction using principal component analysis for cancer detection based on microarray data classification," *J. Comput. Sci.*, vol. 14, no. 11, pp. 1521–1530, Nov. 2018.

[36] M. S. Othman, S. R. Kumaran, and L. M. Yusuf, "Gene selection using hybrid multi-objective cuckoo search algorithm with evolutionary operators for cancer microarray data," *IEEE Access*, vol. 8, pp. 186348–186361, 2020.

[37] P. Saqib, U. Qamar, R. A. Khan, and A. Aslam, "MF-GARF: Hybridizing multiple filters and GA wrapper for feature selection of microarray cancer datasets," in *Proc. 22nd Int. Conf. Adv. Commun. Technol. (ICACT)*, Feb. 2020, pp. 517–524.

[38] K. Cahyaningrum, Adiwijaya, and W. Astuti, "Microarray gene expression classification for cancer detection using artificial neural networks and genetic algorithm hybrid intelligence," in *Proc. Int. Conf. Data Sci. Appl. (ICoDSA)*, Aug. 2020, pp. 1–7.

[39] N. B. Abdel-Hamid, S. ElGhamrawy, A. E. Desouky, and H. Arafat, "A dynamic spark-based classification framework for imbalanced big data," *J. Grid Comput.*, vol. 16, no. 4, pp. 607–626, Dec. 2018.

[40] E. Emary, H. M. Zawbaa, and A. E. Hassanien, "Binary grey wolf optimization approaches for feature selection," *Neurocomputing*, vol. 172, pp. 371–381, Jan. 2016.

[41] K. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proc. Int. Conf. Neural Netw.*, vol. 4, Nov. 1995, pp. 1942–1948.

[42] W. H. Wolberg and O. L. Mangasarian. (1992). *Breast Cancer Wisconsin (Original)*. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/breast+cancer+wisconsin+(original)

[43] E. Alba, "Gene selection in cancer classification using PSO/SVM and GA/SVM hybrid algorithms," in *Proc. IEEE Congr. Evol. Comput.*, Sep. 2007, pp. 284–290.

[44] Sobar. (2021). *Cervical Cancer Behavior Risk Data Set (2016)*. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/Cervical+Cancer+Behavior+Risk

[45] *American Cancer Society*. Accessed: Oct. 2021. [Online]. Available: https://www.cancer.org/

[46] S. K. Pal and S. Anand, "Cryptography based on RGB color channels using ANNs," *Int. J. Comput. Netw. Inf. Secur.*, vol. 10, no. 5, pp. 60–69, May 2018.

[47] T. Sree Kala and A. Christy, "HFFPNN classifier: A hybrid approach for intrusion detection based OPSO and hybridization of feed forward neural network (FFNN) and probabilistic neural network (PNN)," *Multimedia Tools Appl.*, vol. 80, no. 4, pp. 6457–6478, Feb. 2021.

[48] S. Mirjalili and A. Lewis, "The whale optimization algorithm," *Adv. Eng. Softw.*, vol. 95, pp. 51–67, May 2016.

[49] A. S. Elsayad, A. I. E. Desouky, M. M. Salem, and M. Badawy, "A deep learning H_2O framework for emergency prediction in biomedical big data," *IEEE Access*, vol. 8, pp. 97231–97242, 2020.

NADIA G. ELSEDDEQ received the B.Sc. degree in computer engineering from the Department of Computer Engineering and Control Systems, MISR Higher Institute for Engineering and Technology (MET), Mansoura, Egypt. She is currently pursuing the M.Sc. degree in computer engineering with the Department of Computer Engineering and Control Systems, Faculty of Engineering, Mansoura University. Her major research interests include optimization techniques, big data, and computing technologies.
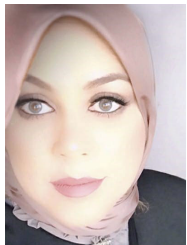
SALLY M. ELGHAMRAWY (Senior Member, IEEE) received the B.Sc. degree in computers engineering and systems, the M.Sc. degree in automatic control systems engineering, and the Ph.D. degree in distributed decision support systems based on multi-intelligent agents from the Department of Computer Engineering, Faculty of Engineering, Mansoura University, Egypt, in 2003, 2006, and 2012, respectively. She is currently the Head of the Department of Communications and Computer Engineering, MISR Higher Institute for Engineering and Technology, and a part-time Associate Professor with the Department of Electrical and Computer Engineering, Faculty of Engineering, The British University in Egypt (BUE), and the Department of Computer Engineering, Faculty of Engineering, Mansoura University. She is also a member of the Scientific Research Group, Egypt. She was delivering lectures as well as supervising graduation projects, master's thesis, and doctoral dissertations. She was delivering lectures and gave practical training in the grants from the Ministry of Communications and Information Technology with the collaboration of IBM. She received a certificate A+ International Inc., CompTIA. She has authored number peer-reviewed publications and received the best paper awards. Her research interests include big data analysis, No-SQL databases, Hadoop, MapReduce techniques, and software engineering. She is a reviewer of a number of international journals, and a judge on the IEEE Young Professionals Competitions.

MOFREH M. SALEM received the Ph.D. degree from the University of Strathclyde, U.K., in 1985. He was a member of High Scientific Consul for the Colleges of Technology, Saudi Arabia, from 1998 to 2001. He was the Director of the Software Engineering Unit, Faculty of Engineering, from 2001 to 2006. He was the Head of the Department of Computers Engineering and Control, Faculty of Engineering, Mansoura University, Egypt, from 2004 to 2008, where he is currently a member of the Computer Center Council. He was the Dean of the High Institute for Computers, Mansoura, from 2008 to 2011. He is also a Consultant with the Studies and Researches and Consultant Engineering Center, Egypt. He is an Arbitration and Examine Research Faculty Member of promotion to the Rank of a Professor and an Associate Professor with the Universities of Egypt, Saudi Arabia, and Jordan. He is an arbitration scientific researcher of publication in international journals, periodicals, and conferences. He has supervised 42 M.Sc. and Ph.D. degree students in the fields of computer engineering. He has approved and discussed 157 of M.Sc. and Ph.D. thesis. He has published 92 scientific articles in international journals or periodicals and conferences of computers engineering. His current research interests include software engineering, computer systems design, parallel processing, computer networks, cloud computing, and big data. He received the Encouragement Award in computer engineering at Egypt, in 1992.

ALI I. ELDESOUKY received the M.A. and Ph.D. degrees from the University of Glasgow, USA. He is currently a Full Professor with the Department of Computers Engineering and Systems, Faculty of Engineering, Mansoura University, Egypt. He is also a part-time Visiting Professor with MET Academy. He teaches in American and Mansoura Universities, and has taken over many positions of leadership and supervision of many scientific papers. He has published hundreds of articles in well-known international journals.

• • •