

Received 22 October 2022, accepted 7 November 2022, date of publication 15 November 2022, date of current version 28 November 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3222365

## RESEARCH ARTICLE

# A Fast Accurate Deep Learning Framework for Prediction of All Cancer Types

MAGDY M. FADEL<sup>1</sup>, NADIA G. ELSEDEEQ<sup>1</sup>, REHAM ARNOUS<sup>2</sup>, ZAINAB H. ALI<sup>3</sup>, AND ALI I. ELDESOUKY<sup>1</sup>

<sup>1</sup>Department of Computer Engineering and Control Systems, Faculty of Engineering, Mansoura University, Mansoura 35516, Egypt

<sup>2</sup>Department of Communications and Electronics, Delta Higher Institute for Engineering and Technology (DHJET), Mansoura 35111, Egypt

<sup>3</sup>Department of Embedded Network Systems and Technology, Faculty of Artificial Intelligence, Kafrelsheikh University, Kafrelsheikh 33512, Egypt

Corresponding author: Magdy M. Fadel (mfares73@yahoo.com; mailto:mfares@mans.edu.eg)

**ABSTRACT** The mortality rate of cancer is among the highest in the world. One death occurs every six in the world. Both machine learning (ML) and deep learning (DL) have been used by scientists to predict cancer. In addition, DL can analyze a huge amount of healthcare data in a short period of time to study the chances of recurrence, progression and patient survival. An accurate and quick framework for improving cancer prognosis prediction is presented in this study. A fast and accurate optimizer is necessary to predict both critical and non-critical cases, so a modified binary version of the Whale Optimization Algorithm (WOA) is proposed. Based on sigmoid transfer functions, this version identifies the subset of features that is minimally optimal while maximizing classification accuracy. This framework is composed of an optimized parameter Long-Short Term Memory (LSTM) Neural Network, with the input being the optimal set of feature selection layer. The proposed framework performs better than previous frameworks having an average accuracy of 100% and an execution time of 4113 seconds.

**INDEX TERMS** Binary modified WOA S-shaped (BMWAOA-S), cancer diagnosis, exploration, exploitation, LSTM, feature selection, optimization.

## I. INTRODUCTION

By 2020, Cancer will be the second most deadly illness, killing one in six individuals, according to the International Agency for Research on Cancer, which predicts 19.3 million new cases and 10 million deaths [1], [2]. Breast, lung, prostate, colon, ovarian and cervical cancers are the most prevalent forms, and they weaken the immune system and alter other biological processes, which is why there is increased worry about this illness.

Disease detection entails classifying tumor kinds and identifying cancer symptoms in order to train a machine capable of detecting new metastatic tumor forms or diagnosing the disease early, when treatment is more difficult. Several earlier researches have proposed frameworks for predicting cancer prognosis, recurrence risk, progression, and patient survival as researchers have demonstrated that prediction accuracy is an aspect that contributes to the efficient treatment of patients [1], [3].

The associate editor coordinating the review of this manuscript and approving it for publication was Gustavo Callico<sup>1</sup>.

Recent advances in Deoxyribonucleic Acid (DNA) fragmentation technologies have created large amount of data, making genomics one of the first fields to generate data [4]. Sequences cannot convey ready-to-use information; however, they can be translated by a complex method that uses the sequence to generate protein. Because the constructed genome sequence matches previously recognized cancer genome sequences, it assesses the protein's expression and determines if it is malignant [5]. The gathering of genomic data has created various challenges in presenting a rational definition of cancer's genetic basis.

Moreover, the identification of the main difficulties associated with the treatment and prevention of diseases by the vast number of gene expression levels in a person containing many features but relatively few samples [6]. The chances of a successful recovery increase with early diagnosis. Its study is therefore crucial [7].

In a Deep Learning Architecture (DLA), features are extracted hierarchically with several degrees of nonlinearity. The H2O framework is a multi-layer neural network (NN) architecture designed for DL tasks [8]. It is possible

to train DL models to represent original data in a useful way. Furthermore, they produce the greatest results for complex data [9].

It is possible to categorize meta-heuristic algorithms into two types: single-solution algorithms or population-based algorithms. The single-solution method utilizes only one candidate solution the optimization process, which evolves and is updated over time. Whereas the population-based uses a random search agent to begin the optimization process. Every search agent has its own candidate for solving the optimization issue. Sharing information about the search area and collaborating with each individual will prevent local optima from stagnating and ensure that a global search goal is covered.

As a rule, metaheuristic algorithms are evaluated based on their ability to solve decision-making difficulties and their significant balance between exploring and exploiting [10], [11], [12], [13], [14], [15]. Exploitation refers to the capacity to find better alternatives to well-known answers. The exploration of search spaces means finding better-scoring places by employing metaheuristics.

The Feature Selection (FS) problem and binary optimization difficulties can both be solved using a single optimization technique. References [16], [17], [18], and [19] introduced several hybrid techniques between WOA and simulated annealing presented by researchers such as Genetic Algorithm (GA), Gray Wolf (GW) and Particle Swarm Optimizer (PSO), also a hybrid strategy combining the filter and wrapper approaches of FS.

There is no certainty that a better selection of characteristics will be discovered in the FS issue. Furthermore, no optimizer is appropriate for solving any optimization problems based on the No Free Lunch (NFL) theorem [20].

Based on the reasons mentioned above, there is still a need for a modern, accurate, and high-speed system to deal with cancer diseases, as in non-critical cases we need high accuracy in an appropriate time, while in critical cases we need high speed for less chance of life. And on this, the research provides a fast and accurate DL framework to overcome these previous problems.

The rest of this paper is organized as follows: Section 2 describes the motivations for the study and its contributions. A literature review and a brief overview of the LSTM and Modified WOA (MWOA) are presented in Section 3. The proposed BMWOA-S and its designed frameworks are described in Section 4. Section 5 discusses the experimental results. Section 6 concludes with a discussion of future work and conclusions.

## II. MOTIVATIONS & CONTRIBUTIONS OF THE STUDY

Following are the motivations for this study:

- 1) Propose a fast, accurate, and scalable framework for DL H2O that uses big data to improve cancer prognosis prediction.
- 2) Provide a binary modified WOA optimizer that is highly accurate and fast, which will allow FS to reduce

the dataset size and also tune LSTM (number of layers and number of neurons per layer).

This study provides the following relevant contributions:

- The DL H2O framework can handle a lot of data in many forms. Using patient health data to predict cancer prognosis is useful since it incorporates multi-source data. The frame has a high level of accuracy and quickness. To verify their dependability and efficacy, the frameworks were tested on six standard data sets, including breast, lung, prostate, colon, ovarian and cervical cancer. These data sets are freely available and continue to be utilized in the majority of contemporary investigations [21], [22], [23], [24], [25], [26], [27], [28].
- FS is a method for quickly picking the best features for NN training, possibly enhancing cancer prognostic prediction while also lowering the bulk of the input data to LSTM.
- BMWOA-S is compared to other popular optimizers for its benefits and efficiency.
- Cancers of every type can be predicted with 100% accuracy, which means the earlier treatment is started, the greater the chance of a cure.

Based on the severity of the patient's case (critical vs non-critical patients), this study will help propose the most appropriate framework.

## III. RELATED WORK

There has been considerable research into cancer diagnosis prediction using a variety of methodologies, with some showing high accuracy. Previous approaches have revealed the following findings:

To enhance therapy and medicine discovery for diagnosis, several researchers employ Machine Learning (ML) classifiers such as k-nearest neighbour (KNN), logistic regression (LR), decision trees (DT), random forest (RF), and support vector machine (SVM).

In [21] applied to cervical cancer dataset. In addition, in [22] and [23] performed on four, six different datasets related to breast cancer. But in [24], two types of datasets were used to study colon cancer. Further study with a larger data set will aid in the improvement of these models' performance.

In [27] S. Parisapogu et al. used a multi-layered DL algorithm on diverse microarray data to diagnose the kind of illness. To categorise illness samples, our model must use many deep learning classification approaches on biological data sets.

N. G. El-Seddeq et al. introduced in [28] one of these newly introduced three frameworks that used to improve the performance of cancer prognosis prediction. With the exception of the lung and cervical cancer datasets, the proposed optimizer outperforms the FS algorithm for the fitness value on all datasets.

Also, in [29] various feature selection methods, like correlation analysis and Fisher-ratio, are used to extract useful features and reduce dimensionality. Then, PSO (Particle

Swarm Optimization) algorithm is used to generate a pool of candidate base classifiers for learning the subsets re-sampled from the different feature subsets.

In [30] the classification of microarray data using artificial neural networks (ANNs) was suggested. More research with a larger data set will aid in improving the model's effectiveness.

In [31] global optimization, cuckoo search (CS) has been found to be an effective algorithm. Cuckoo search fundamentals and applications are reviewed along with the latest developments.

Liu et al. deployed in [32] ReliefF and PSO algorithms for selecting feature genes. The first step is to use ReliefF as a feature prefilter to eliminate genes that have a low correlation with the target class. Search is then carried out using PSO. For the final optimal subset of genes, the classification accuracy of SVM is used as the evaluation function.

According to [33], gene FS data should be integrated with cancer classification for gene expression, along with other types of genomic data. The use of other algorithms for parameter optimization can help with this model.

As in [34], Principal Component Analysis (PCA) was used for dimension reduction in SVM and Levenberg-Marquardt Back Propagation (LMBP). The deficiency that occurs as a result of the model's excessive time spent in the training process; the choice of architecture must still be done in a more organised manner.

Othman et al. showed in [35], the cuckoo search was combined with evolutionary operators for gene selection as a hybrid multi-objective search. Double mutations and single crossovers are the evolutionary operators used also. The aim is to improve the search capabilities and values of the dimensions.

Saqib et al. in [36], to ensure relevancy and remove irrelevant features, a Multiple Filters and GA Wrapper for Feature Selection (MF-GARF) hybrid approach consists of three phases relevance block; Information Gain, Gain Ratio and Gini Index. Second phase involves removing redundancy among features using Pearson Correlation statistics, followed by an Optimization Block. The Optimization Block consists of a Genetic Algorithm wrapper with Random Forest as a fitness evaluator, which provides a high predictive power feature subset based on the Genetic Algorithm wrapper.

Cahyaningrum et al. in [37] used PCA to work on ANN and GA. This model needs better tuning of the parameters of the genetic algorithm, such as mutation rate, crossover rate, elitism, and fitness normalization.

A hybrid model (CNN-LSTM) is proposed in [38] Using CNN layers with convolutional layers; it can extract features with a multi-time scale. By a wide margin, this hybrid model outperforms other existing methods.

Finally, in [39] a classification approach based on deep feature fusion and selection using whale optimization techniques was proposed. A PCA is used to select the best features, which are serially merged to produce a vector of  $N \times 2125$  features. Furthermore, WOA was used to select informative features

$N \times 1049$  among  $N \times 2125$  features and provide them to SVM, KNN, and Wide Neural Network (WNN) classifiers.

The reason for choosing DL over other standard approaches is that we anticipate dealing with large data size issues. Furthermore, the forecast time for cancer diagnosis is crucial since the patient's life depends on it, particularly in serious circumstances. In this scenario, DL is the best approach to utilise. Because it requires high-end infrastructure to train in a reasonable amount of time.

Next section describes the LSTM neural network used in our proposal and the Binary Modified Whale Optimization Algorithm (BMWOA-s) used.

#### IV. THE PROPOSED CANCER PREDICTION FRAMEWORK

Several Biomedical studies included deep learning frameworks that utilize different optimizers for selecting best features input to the FFNN and tuning it (no. of hidden layers and no. of neurons/layer) have been published [22], [23], [24], [25], [26], [27], [28], [29], [30], [31], [32], [33], [34], [35], [36], [37], [38], [39]. Accordingly, this thought has been succeeded in the application of deep learning biomedical techniques, especially in case of large datasets which includes two important cases, initial and severe states of the diseases, that happens in all diseases. It should be noted that, the initial state of the disease allows us some time to treat the patient with different attempts and different medicines, however in the severe case does not allow enough time for the prediction. On this, it imposed on these systems high accuracy in prediction as well as high speed in implementation. This always assured us of the need to search for a fast and accurate optimizer. The latest research published to us [28], on different types of cancer, we have obtained a 100% accuracy in most cases of the disease at reasonable processing time, encouraged us to search for the best so that we can reach 100% accuracy in all types of cancers at less processing time. Accordingly, this research proposes a Fast Accurate Deep Learning (FADL) Framework hoping it may results in predicting all types of cancers at a time that allows saving the patient's life, especially in severe cases. The framework is divided into four layers. As shown in Figure 1, these layers include processing, feature selection, deep learning, and prediction (classification).

##### A. PRE-PROCESSING LAYER

Medical data sets collected from several sources are often insufficient and riddled with errors that lead to misclassification. Certain machine learning algorithms are impaired by features of varying magnitudes. The first stage of the processing layer is the normalization method. Therefore, scaling (normalization) some feature's (column's) values between 0 and 1 can be achieved using these techniques.

Reducing data imbalance: As a second step of preprocessing, redundant columns, such as the ID column, are removed. A trait imbalance is a categorization problem in which some qualities are significantly underrepresented. As a result, the classifier prefers the majority of qualities. By over-sampling,

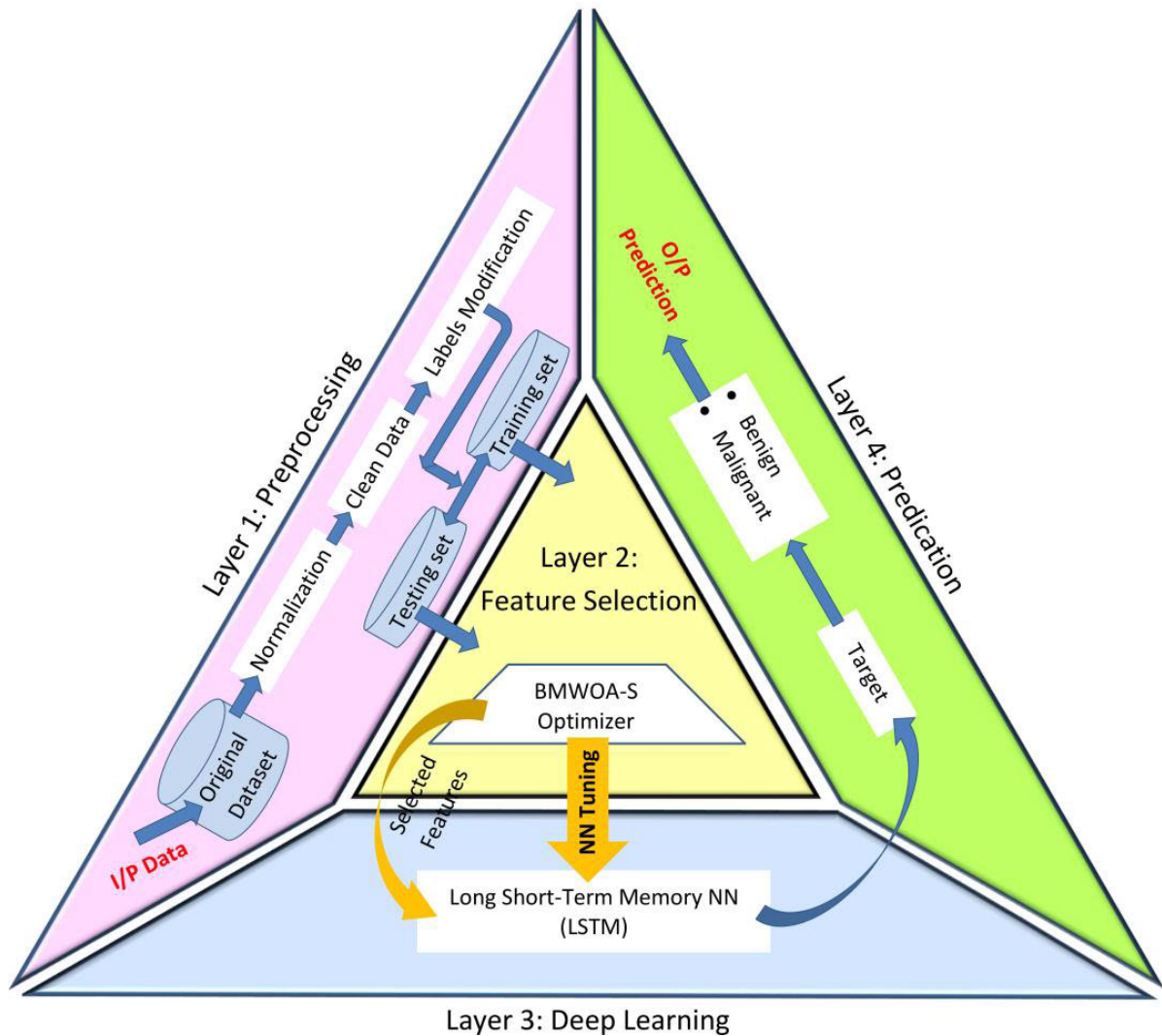


FIGURE 1. Block diagram of the fast accurate framework for all types of cancer prediction.

the occurrences of the minority class, the approach presented in the article [40] was employed to overcome the imbalance dataset problem.

It is essential to remove these incorrect values before separating the dataset into training and testing subsets. The performance of this layer is measured using six benchmark datasets. The first set of data is for breast cancer. The second and third are for lung and prostate cancers, respectively, while the fourth is for colon cancer. Ovarian cancer is the fifth. Cervical cancer is the sixth most common cancer. Finally, the labels for each element in the dataset are converted from textual values to numerical values in the third step of preprocessing. There are two classifications per dataset (Malignant and Benign), which are then transformed to the 0 and 1 values, respectively.

**B. THE FS LAYER**

To address the FS issue, in this layer, BMWOA-S is proposed as a binary variation of WOA. A dataset containing N features

must, therefore, contain 2N features, indicating that a large region of features should be explored for feature reduction thoroughly.

1) THE PROPOSED ALGORITHM: BINARY MODIFIED WOA S-SHAPED (BMWOA-S)

WOA [41] is a meta-heuristic algorithm effectively employs 2 attacking strategies, bubble-net and searching for Prey. To make the updating process, many suggestions have been tested, such as modifying the exploration equations, unfortunately all of these have been failed in proposing a fast, accurate optimizer. Therefore, in the following paragraphs, a new version of WOA named MWOA is explained in detail.

*a: ENCIRCLING PREY*

Based on mathematics, this behaviour can be described as follows:

$$\vec{D} = \left| \vec{C} \cdot \vec{X}^*(i) - \vec{X}(i) \right| \tag{1}$$

$$\vec{X}(i+1) = \vec{X}^*(i) - \vec{A} \cdot \vec{D} \tag{2}$$

$i$  = iteration,  $\vec{A}$  and  $\vec{C}$  = coefficients, and  $\vec{X}^*$  = the position vector of the best solution found.  $\vec{A}$  and  $\vec{C}$ :

$$\vec{A} = 2\vec{a} \cdot \vec{r}_1 - \vec{a} \tag{3}$$

$$\vec{C} = 2 \cdot \vec{r}_1 \tag{4}$$

$\vec{r}_1 = [0; 1]$ . Through iterations, we must linear decrease  $\vec{a}$  vector [2-0].

**b: BUBBLE-NET ATTACKING METHOD (EXPLOITATION PHASE)**

The bubble-net attacking behaviors are:

1) Shrinking encircling mechanism: humpback was reducing the  $\vec{a}$  from [2-0] in (3) over the series of processes.  $\vec{A}$  = random value  $[-\vec{a}, \vec{a}]$ .

2) Spiral updating mechanism: The space between both the whale and its prey:

$$\vec{X}(i+1) = \vec{X}^*(i) + \vec{D}' \times e^{wq} \times \cos(2\pi q) \tag{5}$$

$$\vec{D}' = \left| \vec{C} \cdot \vec{X}^*(i) - \vec{X}(i) \right| \tag{6}$$

$\vec{D}'$  = the gap between the population's best solution and the present individual whale,  $w$  = a fixed value, and  $q = [-1, 1]$ . As a result, the scientific formula looks like this:

$$\vec{X}(i+1) = \begin{cases} \vec{X}^*(i) - \vec{A} \cdot \vec{D}, & p < 0.5 \\ \vec{D}' e^{wq} \cdot \cos(2\pi q) + \vec{X}^*(i), & p \geq 0.5 \end{cases} \tag{7}$$

$p$  = a value [0, 1].

**c: SEARCH FOR PREY (EXPLORATION PHASE)**

Gaussian equation is used to increase likelihood of discovering the globally minimum while avoiding becoming caught in the minima. In the DLA growth process, a randomness mechanism is used to produce new particle that use the Gaussian methodology. Based on the improved solution, a sequence of diffusion operations may be calculated:

$$\vec{B}_i^* = \text{Gaussian}(\mu_{\vec{B}^*}, \omega) + (\alpha \times \vec{B}^* - \times \vec{b}_i) \tag{8}$$

$\vec{B}_i^*$  = the most improved diffusion procedure solution.  $\alpha$  and  $\omega$  = the random numbers [0; 1].  $\vec{B}^*$  and  $\vec{b}_i$  show the best position and the  $i^{\text{th}}$  point in the group.  $\mu_{\vec{B}^*}$  and  $\omega = \left| \vec{b}_i - \vec{B}^* \right|$  As the number of iterations reduces, a best alternative becomes available. It enhances the Search for Prey capability in the proposed WOA by employing the diffusion approach to discover an ideal solution [42].

The MWOA is an updated WOA version (See Figure 2). One of the ways to resolve this type of disadvantage, an advanced technique will substitute the search technique for the Exploration Group. A list of random walks about the optimal solution can be generated using the diffusion process. The enhancement of the WOA exploration potential to find the optimal solution through this diffusion procedure. In the modified version of WOA (MWOA) for enhancing exploration performance via applying diffusion process instead

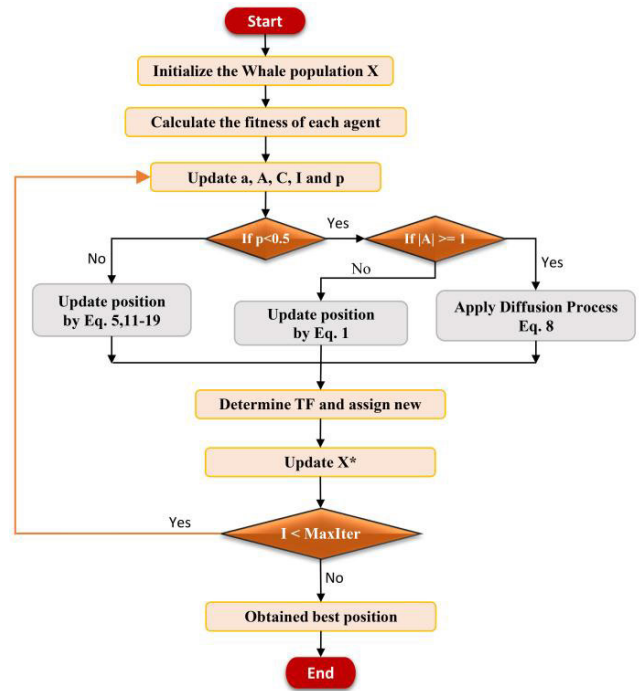


FIGURE 2. MWOA flowchart.

of the search space in WOA. This can need folks further exploring a promising location in the search area in order to avoid local stagnation by swapping:

$$\vec{X}(i+1) = \vec{X}_{rand} - \vec{A} \cdot \vec{D} \tag{9}$$

A continuous MWOA employs Eq. 2 to transfer search agents across the search space to adjust their locations to every location. This is referred to as space. The FS problem handles only binary data [43]; when there is no change, the continuous form of MWOA cannot be employed to solve the FS problem. As a result, we propose the BMWOA-S variant that really is suitable for solving the FS issue. BMWOA-S specifies that the candidate's choices require only binary solutions [0, 1]. If the feature has a value of 0, it is not picked; however, if it has a value of 1, it is selected.

We initially scale the data in the interval [0, 1] to change the MWOA solutions from continuous to binary. According to a prior study, the translation is accomplished by the use of an S-shaped (Sigmoid) transfer function (TF). S1, S2, S3, and S4 make up its family of TFs (Table 1). In order to travel in a binary space, the components of position vectors are converted from 0 to 1. Table 1 contains the theoretical formulas for each TF, and Figure 3 depicts the algebraic curve of an S-shaped curve. The stages involved in BMWOA are depicted in Algorithm 1. Converting scaled continuous data to binary values will be done by applying the formula provided by Kennedy and Eberhart [44].

$$X_d^{(i+1)} = \begin{cases} 1, & 0.5 \leq \text{sigmoid}(x) \\ 0, & \text{otherwise} \end{cases} \tag{10}$$

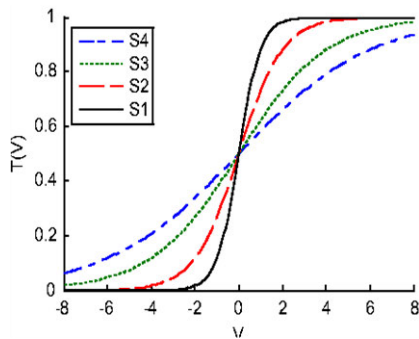


FIGURE 3. S-shaped transfer functions.

TABLE 1. S-shaped transfer functions.

Name	Transfer Functions
S1	$T(X) = \frac{1}{1 + e^{-2x}}$ (11)
S2	$T(X) = \frac{1}{1 + e^{-x}}$ (12)
S3	$T(X) = \frac{1}{1 + e^{-(x/2)}}$ (13)
S4	$T(X) = \frac{1}{1 + e^{-(x/3)}}$ (14)

$X_d^{(i+1)}$  = updated binary position  $i$  and dimensions  $d$ , and  $\text{sigmoid}(x)$  is given by Eqs. (11)-(14).

C. THE DL LAYER

1) LONG-SHORT TERM MEMORY (LSTM)

The LSTM neural network is a type of recurrent neural network (RNN). The RNN architecture is trained by using sequential information which travels from the input vector through the network to the output neurons. Errors are calculated and propagated backward to update the parameters of the network. This type of network incorporates loops of information in its hidden layer. Using loops, information can flow multi-directionally, so at a given time step, the hidden state represents past data held at a given time step. Consequently, all outputs depend on previously made predictions already known. There are, however, limitations to RNNs’ ability to bridge more than a specific number of steps. Gradient vanishing is the main reason for the prediction to capture short-term dependencies over time as the information from earlier steps decays. The gradient of the loss function approaches zero as the number of layers in the RNN containing activation functions increases. The LSTM neural networks (LSTM-NNs) enable learning long-term dependencies. By introducing a memory unit and gate mechanism, LSTM can capture sequences with long dependencies. Therefore, LSTM-NNs are capable of selectively remembering or forgetting information and of learning thousands of timesteps by using three gates and cell states.

Medical journals have published practical applications of LSTM. A variety of variants of RNN have been used for classification and prediction purposes in studies. According to one study, longitudinal medical records with irregular time

Algorithm 1: Pseudo Code of BMOA-S

```

1: Input:
2:   No. of whales (= 50)
3:   MaxIter no. of iterations (= 500)
4: Output:
5:   Location of the optimal whale
6:   Find  $X^*$ 
7: Initialize:  $a$  and  $n$ 
8: while ( $t < \text{MaxIter}$ )
9:   for each whale in population
10:    Update variables  $a, A, C, I$  and  $p$ 
11:    if ( $p < 0.5$ ) then
12:      if ( $|A| < 1$ ) then
13:        # Encircling Prey
14:        Update position by Eq. (1):
15:         $\vec{D} = |\vec{C} \cdot \vec{X}^*(i) - \vec{X}(i)|$ 
16:      else ( $|A| \geq 1$ )
17:        # Exploration Phase
18:        Select a random search agent ( $X_{rand}$ )
19:        Apply Diffusion Process from Eq. 8
20:         $\vec{B}_l^{*k} = \text{Gaussian}(\mu_{\vec{B}^*}, \omega) +$ 
21:          ( $\alpha \times \vec{B}^* - \times \vec{b}_l$ );
22:      end if
23:    else ( $p \geq 0.5$ )
24:      # Exploitation Phase
25:      Update location by the Eq.(5)
26:       $\vec{X}(i+1) = \vec{X}^*(i) + \vec{D}' \times e^{wq} \times \cos(2\pi q)$ 
27:    end if
28:    Update  $X(t+1)$  from the Eqs.(11-14)
29:  end for
30:  Get new agent
31:  Calculate the cost
32:  Update  $X^*$  if there is a better
33:   $i = i + 1$ 
34: end while
35: return  $X^*$ 

```

intervals were used in LSTM to predict Parkinson’s disease progression [45].

Based on Figure 4, we can see that an LSTM block extends the memory of an RNN by using structures like cell states and three gates that allow it to selectively remember and forget information. An LSTM block has four additional layers in addition to the hidden state in a RNN. Cell state ( $C_t$ ), input gate ( $i_t$ ), output gate ( $O_t$ ), and forget gate ( $f_t$ ) are the names of these layers. In order to generate knowledge from the training data, each layer interacts with the other in a unique way.

A sigmoid activation function (shown in Tab. 2) is implemented in the forget gate. For the input and output gates, however, a combination of sigmoid and hyperbolic tangent (TanH) are used to provide the necessary information to the cell state. In an LSTM neural network as shown in Eq. 15,

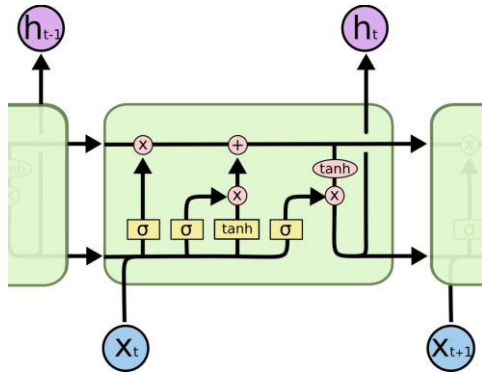


FIGURE 4. Block diagram of LSTM at any timestamp.

the information generated by each block flows through the cell state from block to block.

$$f_t = \sigma \left( W^{(f)} (p_t, h_{t-1}) + b^{(f)} \right) = \frac{1}{1 + e^{-(W^{(f)}(p_t, h_{t-1}) + b^{(f)})}} \tag{15}$$

where,  $\sigma$  is the sigmoid activation function,  $W^{(f)}$  and  $b^{(f)}$  are the weight matrix and bias vector, which will be learned from the input training data.

The function takes the old output ( $h_{t-1}$ ) at time  $t-1$  and the current input ( $p_t$ ) at time  $t$  for calculating the components that control the cell state and hidden state of the layer. The results are  $[0, 1]$ , where 1 represents “completely hold this” and 0 represents “completely throw this away”.

LSTM networks, like RNNs, produce outputs  $\hat{y}_t$  that are used to train the network via gradient descent at each time step. Iteratively, the parameters of the network are updated during the backward pass. A minor modification to the algorithm is the only fundamental difference between the RNN and LSTM back-propagation algorithms. At each time step, the calculated error term is  $E_t = -y_t \log \hat{y}_t$ . Similarly to RNN, the error is calculated as the sum of the errors from all time steps  $E = \sum_t -y_t \log \hat{y}_t$ .

For disease subtype analysis, music generation, text generation, handwriting recognition, language translation, time series analysis, and image captioning, LSTM is a very powerful ANN architecture. As information flows through the state of a cell, LSTM is effective to make predictions since it gives equal attention to all input sequences. The LSTM’s prediction accuracy is not affected by the small change in the input sequence due to the mechanism adopted.

An equation describing data flow as an aggregate between  $m$  neurons in the previous layer and one neuron  $i$  is as follows:

$$Z_i = a \left( \sum_{j=1}^m w_{ij} x_{ij} - c_{i0} \right) \tag{16}$$

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

In order for FFNN weights to represent the relationship between input vectors and desired output vectors correctly,

TABLE 2. 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	$A(x) = \begin{cases} -1 & \text{if } x < 0 \\ 1 & \text{if } x \geq 0 \end{cases}$

TABLE 3. Structure parameters.

Structure Parameters	Value
No. of layers	3
No. of neurons	10
Biases	Random
Activation function	TanH
Initial weights	Default

TABLE 4. Training parameters.

Training Parameters	Value
learning rule	Levenberg–Marquardt
sum-squared error	0.01

they should be identified accurately. To train the neural network and minimize the pattern’s execution work, use the following equation:

$$E = \frac{1}{2Z} \sum_{z=1}^Z \sum_{g=1}^G (y_g^z - d_g^z)^2 \tag{17}$$

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

This layer represents LSTM with the best settings using the proposed optimizer. The LSTM 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 (shown in Table 3). As for the initial weights, H2O Frame IDs initialize the weights such that the default initial\_weight\_distribution and initial\_weight\_scale parameters are uniform adaptive and one, respectively (shown in Table 4). Moreover, the training parameters learning rule and sum-squared error are Levenberg–Marquardt and 0.01, respectively. The LSTM 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 BMWOA-S, binary WOA (BWOA), binary (GWO), binary PSO (BPSO), and binary GA (BGA) 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 BMWOA-S uses training and validation data portions during the optimization process and for testing

**Algorithm 2:** Pseudo Code of Proposed DL H2O Framework

```

1: Input:
2:   Original_Dataset
3: Output:
   # Tuning parameters of best LSTM configuration
4:   No._of_Hidden_Layers (H)
5:   No._of_Neurons_per_Layer (N)
6: Initialization:
   # Load the H2O Deep Learning Module
7:   Load H2O_DeepLearning_Estimator
8:   No._of_Iterations (I) := 500
9:   No._of_Agents (G) := 50
10:  No._of_Hidden_Layers (H) := [1; 2]
11:  No._of_Neurons_per_Layer (N) := [5; 15]
   # Load the dataset
12:  csv_url := "./breast_cancer.csv"
13:  Original_Dataset := h2o.import_file(csv_url)
14:  y := 'label'
   # Dataset preprocessing
15:  Normalization
16:  Data_Clean
17:  Labels_Modifications
   # Feature selection using the proposed optimizer
18:  Selected_Features := Algorithm1: Pseudo code of BMOA-S
   # Splitting Dataset
19:  Train, Valid, Test := data.split_frame([0.8, 0.15])
   # Tuning the LSTM Classification NN
20: for each Iteration in (I)
21:   H := random[1; 2]
22:   N := random[5; 15]
23:   Model := H2O_DeepLearning_Estimator(activation
   := "Tanh", hidden := [H], Neurons := [N], epochs := 10)
24:   for each Agent in (G)
25:     Model.train(x := x, y := y, training_frame := train)
26:     Predication := Algorithm 1: Pseudo code of BMOA-S
27:     Compute the cost function
28:     Generate new agent
29:   end for
30: end for
   # Compute the evaluation by following commands
31:  perf := model.model_performance(test)
32:  score0 := perf.logloss()
33:  score1 := perf.mse()
34:  perf.confusion_matrix()

```

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.

**D. THE PREDICTION LAYER**

The optimizer suggested in earlier layers is employed to resolve a problem requiring two different functions: exploration and exploitation.

The algorithm presenting the fast accurate framework for all types of cancer prediction is shown above. Moreover, Figure 5 illustrates the 4 layers explained above of the framework visually.

**V. EXPERIMENTAL RESULTS AND DISCUSSION**

The four experiments we conducted were as follows. During the first experiment, we evaluated the performance of the proposed optimizer, while during the rest, we examined the performance of the three frameworks. All experiments were conducted on Intel® Core™ i7-2.90 GHz processor with

**TABLE 5.** The datasets used in this study are listed below.

Dataset no.	Dataset	No. of Records	No. of Features	No. of classes	
				Benign	Malignant
1	Breast	699	10	460	239
2	Lung	181	12533	150	31
3	Prostate	136	12600	59	77
4	Colon	62	2000	40	22
5	Ovarian	253	15154	91	162
6	Cervical	72	19	51	21

32-GB RAM and an NVIDIA Quadro M2000M GPU. The four experiments were conducted based on datasets.

Predictions and visual representations of the findings are performed using the following packages: pandas, NumPy, random, time, math, h2o, matplotlib.pyplot, sklearn.metrics, and h2o.estimators.deeplearning.

**A. DATASETS DESCRIPTION**

Six benchmark datasets were used to evaluate the framework's performance each with two classes, Benign and Malignant.

- Breast Cancer (No. 1): it is submitted by the UCI ML Repository [46], [22] having 699 records each with 10 features, these features are deduced from medical digital images of breast.
- Lung Cancer (No. 2): it is gathered from Harvard Medical School having 181 records each with 12533 features [26].
- Prostate Cancer (No. 3): it is collected by the public Kent Ridge Bio-medical Data Repository, having 136 records each with 12600 features [27].
- Colon Cancer (No. 4): it is gathered by Alon et al. [47]. Having 62 records each with 2000 features [24].
- Ovarian Cancer (No. 5): it is gathered from ovarian cancer patients, having 253 records each with 15154 features [26].
- Cervical Cancer (No. 6): it is presented by the UCI ML Repository [48] gathered from specialist's questionnaire in Jakarta. It has 72 records each with 19 features [25].

As shown above and summarized in Table 5, the selected datasets include diverse records, features and classes that could be solved by the proposed framework. The records in every dataset were randomly portioned into three categories, training, testing and validation with ratios of 80%, 5% and 15%, respectively.

**B. EVALUATION METRICS**

The confusion matrix shown in Figure 6 is considered as one of the most important tools for evaluating the performance of any DL frameworks. Additionally, some deduced advanced metrics such recall, accuracy and precision (shown in Eqs. 18-24) are used to help more accurately assessment the proposed framework.



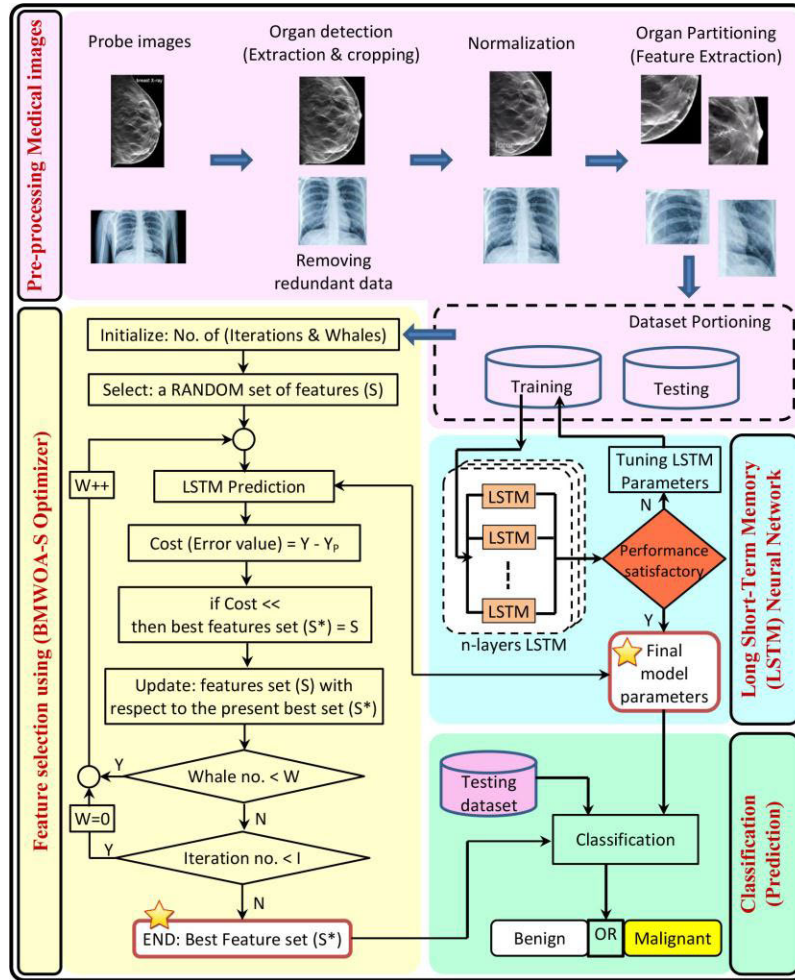


FIGURE 5. Details of the proposed framework.

		Predicted class		
		Positive (P)	Negative (N)	
Actual class	Positive (P)	True Positive (TP)	False Negative (FN)	Recall
	Negative (N)	False Positive (FP)	True Negative (TN)	Specificity
		Precision	Negative Predictive value	Accuracy

FIGURE 6. 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 (18-24) show some advanced evaluation metrics deduced from the confusion matrix.

$$Recall = \frac{TP}{(TP + FN)} \tag{18}$$

$$Specificity = \frac{TN}{(TN + FP)} \tag{19}$$

$$Precision = \frac{TP}{(TP + FP)} \tag{20}$$

$$Accuracy = \frac{TP + TN}{(TP + FN + FP + TN)} \tag{21}$$

$$F1Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{22}$$

$$Mean Square Error (MSE) = \frac{\sum_{i=1}^N (Predicted_i - Actual_i)^2}{N} \tag{23}$$

$$logloss = - \frac{\sum_{i=1}^N \sum_{j=1}^M Actual_{ij} * \log(\text{probability}_{ij})}{N} \tag{24}$$

### C. PERFORMANCE ANALYSIS OF THE PROPOSED BMWOA-S

#### 1) EXPERIMENT NUMBER 1

Based on the second layer of the DL H2O framework, we conduct the first experiment. To validate its performance, the proposed BMWOA-S was evaluated against the BWOA, BGWO, BPSO, and BGA algorithms. The mean error, mean

**TABLE 6.** The configuration values.

Configuration	Value
No. of iterations	500
Search_agents_no.	50
Dimension	No. of features in the dataset
Search domain	[0,1]
No. repetitions of runs	20
Fitness function ( $\alpha$ ) parameter	0.99
Fitness function ( $\beta$ ) parameter	0.01

**TABLE 7.** Compared algorithms’ average errors obtained.

Dataset no.	GA	PSO	GWO	WOA	BMWOA-S
1	0.037	0.032	0.033	0.037	<b>0.003</b>
2	0.031	<b>0.000</b>	0.014	0.092	<b>0.000</b>
3	0.030	0.044	0.013	0.032	<b>0.004</b>
4	0.088	0.075	0.018	0.033	<b>0.005</b>
5	0.014	0.028	0.014	<b>0.000</b>	<b>0.000</b>
6	<b>0.000</b>	0.004	0.017	0.029	<b>0.000</b>

**TABLE 8.** Compared algorithms’ average selection size obtained.

Dataset no.	GA	PSO	GWO	WOA	BMWOA-S
1	0.505	0.625	0.6	0.825	0.301
2	0.548	0.481	0.594	0.403	0.300
3	0.451	0.475	0.585	0.223	0.120
4	0.503	0.48	0.521	0.204	0.120
5	0.518	0.497	0.594	0.141	0.390
6	0.476	0.418	0.532	0.476	0.360

**TABLE 9.** Compared algorithms’ average fitness value obtained.

Dataset no.	GA	PSO	GWO	WOA	BMWOA-S
1	0.047	0.042	0.0443	0.035	<b>0.018</b>
2	0.005	0.008	0.022	0.100	<b>0.000</b>
3	0.032	0.048	0.015	0.035	<b>0.011</b>
4	0.092	0.080	0.0256	0.038	<b>0.000</b>
5	<b>0.000</b>	0.014	<b>0.000</b>	0.014	<b>0.000</b>
6	<b>0.000</b>	0.007	0.017	0.032	0.040

fitness value, mean size of choice, and mean standard deviation of the algorithms were compared. Table 6 shows the configuration values.

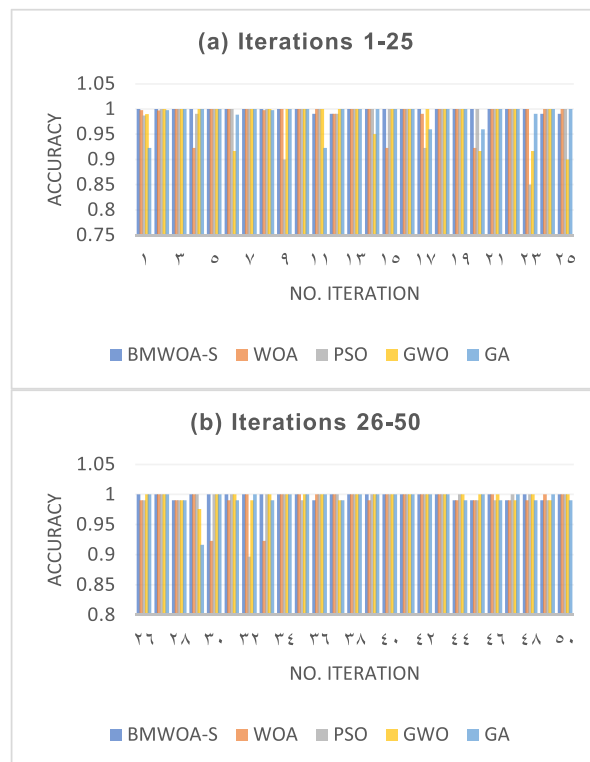
Tables 7–10 describe the outcomes of this investigation and provide the aggregate findings for all enhancers across the six data sets. BMWOA-S had the lowest values for mean error, mean selection, and standard deviation. In other words, BMWOA-S beat other algorithms, such as the FS method, for fitness value in all datasets except the cervical dataset, which had relatively few characteristics, demonstrating that BMWOA-S can select the best subset of features with the least error.

2) EXPERIMENT NUMBER 2

For the purpose of evaluating the proposed optimizer’s performance, its accuracy is compared to results obtained from

**TABLE 10.** Compared algorithms’ average standard deviation for fitness value obtained.

Dataset no.	GA	PSO	GWO	WOA	BMWOA-S
1	<b>0.009</b>	0.011	0.018	0.016	<b>0.009</b>
2	0.080	0.090	0.098	0.082	<b>0.000</b>
3	0.036	0.040	0.042	0.031	<b>0.026</b>
4	0.068	0.081	0.076	0.062	<b>0.000</b>
5	<b>0.000</b>	0.010	<b>0.000</b>	0.010	<b>0.000</b>
6	<b>0.000</b>	0.021	0.213	0.042	0.040



**FIGURE 7.** Breast cancer neural network iterations for different set of features.

GA, PSO, GWO and WOA optimizers. From the following figures showing the accuracy for 50 iterations for different optimizers, it is evident that the proposed framework is superior.

**D. FRAMEWORK’S PERFORMANCE ANALYSIS**

In Experiments 3-4, the framework was compared in terms of confusion matrix values, precision, accuracy, recall, specificity, F1-score, computational time, logarithmic loss (log-loss) and mean squared error (MSE) values.

1) EXPERIMENT NUMBER 3 TESTS THE BEHAVIOR OF THE DL H2O FRAMEWORK

The DL H2O framework for cancer case classification in this experiment comprises four layers: pre-processing, FS, DL with FFNN optimization, and prediction.

**TABLE 11.** Performance obtained from the DL H2O framework cancer prediction.

Dataset	TP	FN	FP	TN	Accuracy	Precision	Recall	F1-score	Specificity	Prediction Time (sec.)	MSE	Logloss
Breast	65	0	0	33	100%	100%	100 %	100%	100%	500	0.009	0.038
Lung	7	0	0	16	100%	100%	100%	100%	100%	6000	0.012	0.051
Prostate	10	0	0	15	100%	100%	100%	100%	100%	7300	0.282	0.936
Colon	6	0	0	8	100%	100%	100 %	100%	100%	3500	0.106	0.545
Ovarian	13	0	0	32	100%	100%	100%	100%	100%	7280	0.050	0.110
Cervical	11	0	0	2	100%	100%	100 %	100%	100%	100	0.048	0.172

**TABLE 12.** p-values and accuracy of the proposed framework using Wilcoxon’s rank-sum.

Cancer type dataset	p-value	Accuracy
Breast	0.017239579	99.98%
Lung	0.013958893	99.9%
Prostate	0.047823521	99.9%
Colon	0.013958889	99.99%
Ovarian	0.000671469	99.99%
Cervical	0.047823526	99.97%

**TABLE 13.** Comparison between the studies in the related work section and proposed framework.

Ref.	Year	Algorithm	Accuracy					
			Breast	Lung	Prostate	Colon	Ovarian	Cervical
Y. Chen et al. [29]	2008	NN	-	97%	-	88.7%	97%	-
X. S. Yang et al. [30]	2009	Cuckoo Search (CS)	-	87.7%	85.3%	-	-	-
X. S. Yang et al. [31]	2014	The multiobjective cuckoo search (MOCS)	-	87.7%	91.2%	-	-	-
Liu, M. et al. [32]	2018	• ReliefF • PSO and SVM	-	86.5%	-	98.1%	-	-
Wu, P. et al. [33]	2018	A Novel Complex Network and Hybrid GP+PSO	-	100%	-	95.52%	-	-
Adiwijaya et al. [34]	2020	SVM Classifier	-	70.13%	88.75%	85.71%	<b>100%</b>	-
Adiwijaya et al. [34]	2020	LMBP classifier	-	96.25%	97.65%	81.43%	<b>100%</b>	-
Othman, M. S. et al. [35]	2020	Their multi-objective search algorithm has been designed with evolutionary operators	-	93.7%	92.2%	-	97.5%	-
Saqib, P., et al. [36]	2020	• GA • RF	86.60%	-	98.04%	96.77%	<b>100%</b>	-
Cahyaningrum, K., et al. [37]	2020	• ANN • ANN-GA	-	<b>100%</b>	82%	86%	-	-
Akter et al. [21]	2021	• DT • RF • XGBoost	-	-	-	-	-	93.33%
Sahar, A. [22]	2021	RF + GA	96.82%	-	-	-	-	-
S. Jijitha et al.[23]	2021	LR	99.27%	-	-	-	-	-
Murad, A., et. al. [24]	2021	DT	-	-	-	94%	-	-
Francesco Curia [25]	2021	MLP	-	-	-	-	-	94.5%
Xiongshi, D. el. al.[26]	2021	SVM Classifier based on selected genes by XGBoost-MOGA	82.33%	<b>100%</b>	98%	90.24%	<b>100%</b>	-
Samson, A., et al[27]	2021	1D-CNN	-	-	99.07%	-	-	-
Nadia, G., et al. [28]	2021	FFNN	<b>100%</b> (500sec)	<b>100%</b> (24600sec)	<b>100%</b> (180000sec)	<b>100%</b> (13400sec)	<b>100%</b> (180280sec)	<b>100%</b> (100sec)
Proposed DL H2O framework	2022	LSTM	<b>100%</b> (500sec)	<b>100%</b> (6000sec)	<b>100%</b> (7300sec)	<b>100%</b> (3500sec)	<b>100%</b> (7280sec)	<b>100%</b> (100sec)

The performance of cancer-specific data sets is shown in Table 11. The DL H2O framework obtained mean accuracy of 100 %, accuracy of 100 %, recall of 100 %, F1 score of 100 %, specificity of 100 %, 0.085 MSE, and 0.309 log loss for all cancer data sets, as indicated in the table (see Figure 13).

F1 score of 100 %, specificity of 100 %, 0.085 MSE, and 0.309 log loss for all cancer data sets, as indicated in the table (see Figure 13).

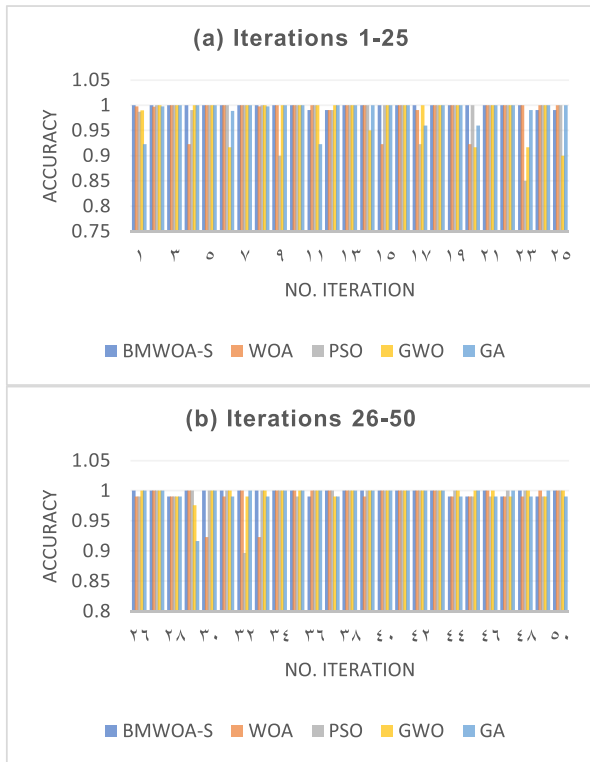


FIGURE 8. Lung cancer neural network iterations for different set of features.

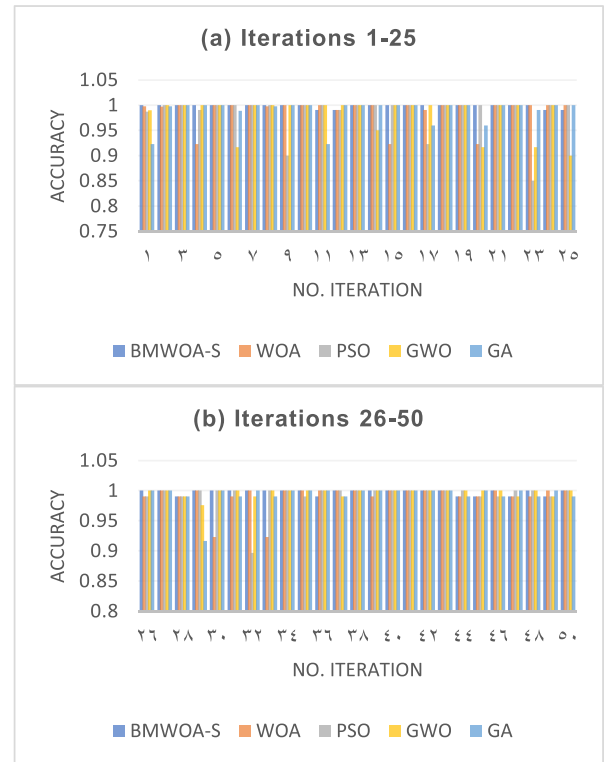


FIGURE 10. Colon cancer neural network iterations for different set of features.

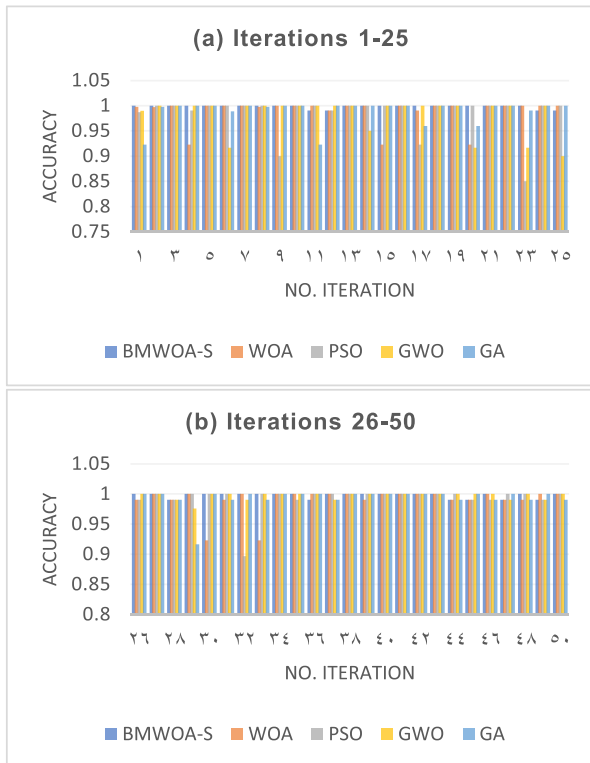


FIGURE 9. Prostate cancer neural network iterations for different set of features.

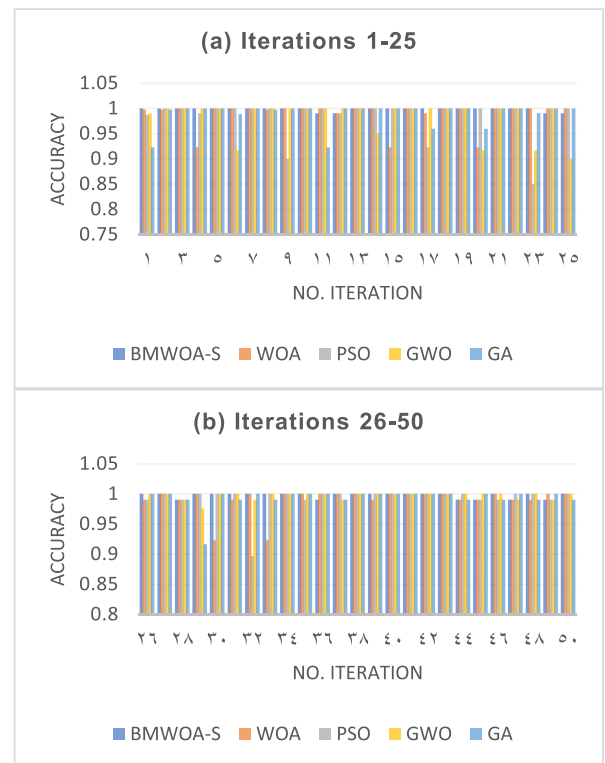


FIGURE 11. Ovarian cancer neural network iterations for different set of features.

2) EXPERIMENT NUMBER 4 WILCOXON'S RANK-SUM  
 Wilcoxon's rating test was used to calculate p-values for the proposed DL framework. Based on this outcome, the

proposal's outputs will determine whether or not it has made a significant difference. DL frameworks differ significantly from other frameworks with p-values below 0.05; however,

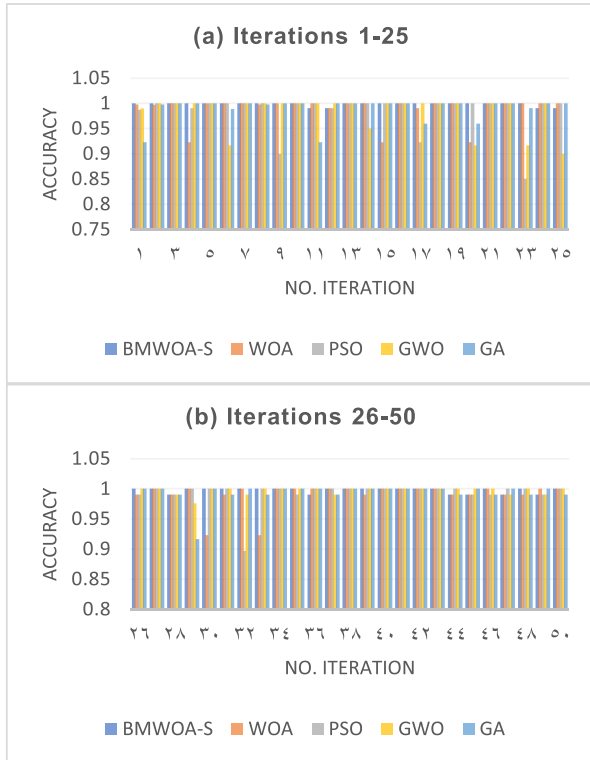


FIGURE 12. Cervical cancer neural network iterations for different set of features.

those with p-values above 0.05 are not significantly different. T-test results are presented in Table 12 along with their p-values and average accuracy. When the p-value is less than 0.05, This validates the suggested framework’s superiority and is statistically significant.

E. DISCUSSION

The BMWOA-S optimizer is used to determine the best collection of characteristics to utilize as FFNN inputs, as well as the ideal number of layers and neurons. Table 13 shows the outcomes of the suggested framework (100% for the six used cancer datasets). The table’s bold letters reflect the best outcomes. For Breast, Prostate, Colon and Cervical cancer data sets, the proposed system outperforms

TABLE 14. Summary of deaths during the last 4 years.

Year	Type	Deaths
2018	Cervical	34,000
2019	Prostate	31,638
2020	Breast	685,000
2020	Lung	1,800,000
2020	Colon	935,000
2021	Ovarian	13,770

others. In terms of lung and ovarian cancer data sets, the suggested framework by Xiongshi, Deng et al. [26], Wu and Wang et al. [33], Adiwijaya et al. [34], Saqib et al. [36], and Cahyaningrum et al. [37] surpasses the other framework. In [28], the proposed framework equals the result achieved in [28] but the proposed framework is faster than their framework.

F. DATA ANALYSIS

First, cancer is one of the diseases that causes the highest mortality rate. In the world, every six deaths are one death. In 2021, 2020, 2019, and 2018, ovarian, lung, prostate and cervical cancers were the most common causes of death worldwide, respectively. Early detection of cases increases the chances of the patient receiving treatment, as we discussed earlier. In the case of cancer patients, accuracy and prediction time are the most important factors (See Table 14).

Second, a framework for predicting cancer with high accuracy and a short prediction time has been proposed. According to the results, we advise that for non-critical patients, the frame is suitable with high accuracy, and for critical patients, the prediction time is shorter.

The modified optimizer performs the main step in the FS algorithm by selecting the most optimal features as well as by adjusting the layers and neurons per layer. The accuracy of DL increases as the inner layers are added and as the time is increased. So, we suggest a three-layer frame. The best choice is the one that takes the least amount of time and is high precision.

VI. CONCLUSION AND FUTURE DIRECTIONS

Our goal is to achieve a balance between exploration and exploitation in this paper by proposing a DL cancer

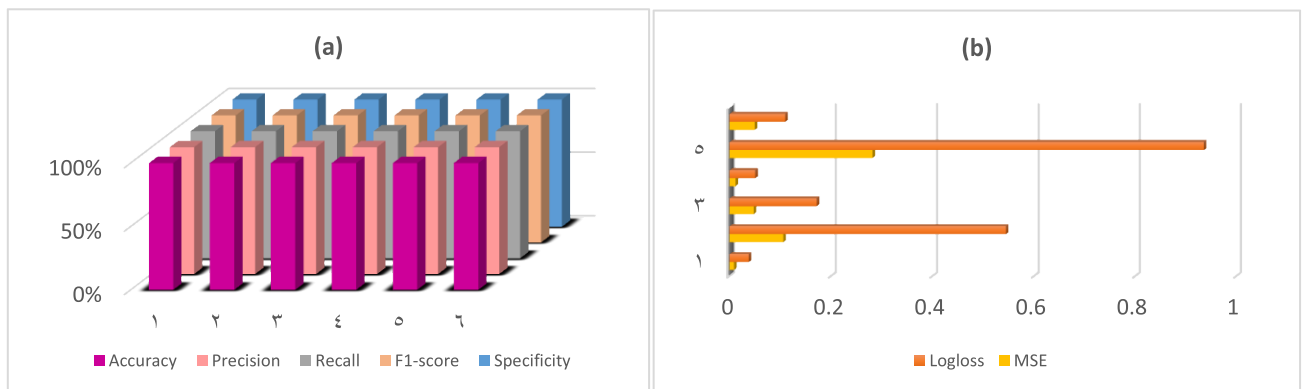


FIGURE 13. (a) and (b) various evaluation metrics for DL H2O framework.

prediction framework. The first applies: searching around for individuals and mutations. Moving towards and searching around the leader are the second applies. BMWOA's have been designed to be the most accurate and fastest prediction for all types of cancer diseases. The result of all trials on all cancer types was 100% accuracy. This has been done in less than 4113 seconds as this time is very important for critical cases. As the proposed framework consisted of the control scheme, we were able to assure the stability of it, despite DL H2O's superior accuracy in the six standard data sets.

The proposed approach will be evaluated in the future against other metaheuristic algorithms to address another binary classification problem. Data sets utilizing DL are being analyzed for the impact of increasing their difficulty level. Tables 6-9 show that the suggested method has a mean error of 0.002, a size of determination of 0.265, a fitness value of 0.0115, and a standard deviation of 0.0125, outperforming other algorithms in all datasets except the Cervical cancer dataset. As a result, we want to use the Broad Learning System (BLS) to enhance the outcome.

## REFERENCES

- [1] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2019," *CA, Cancer J. Clinicians*, vol. 69, no. 5, pp. 7–30, 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. Clinicians*, vol. 71, no. 3, pp. 209–249, May 2021, doi: 10.3322/caac.21660.
- [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, doi: 10.1016/j.gene.2018.03.095.
- [4] S. Behjati and P. S. Tarpey, "What is next generation sequencing?" *Arch. Disease Childhood Educ. Pract. Ed.*, vol. 98, no. 6, pp. 236–238, Dec. 2013, doi: 10.1136/archdischild-2013-304340.
- [5] 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. 2, pp. 188–196, 2010, doi: 10.1093/hmg/ddq391.
- [6] J. D. Cohen, L. Li, Y. Wang, C. Thoburn, B. Afsari, L. Danilova, C. Douville, A. A. Javed, F. Wong, A. Mattox, and R. H. Hruban, "Detection and localization of surgically resectable cancers with a multi-analyte blood test," *Science*, vol. 359, no. 6378, pp. 926–930, 2018, doi: 10.1126/science.aar3247.
- [7] F. Chen, Y. Tang, C. Wang, J. Huang, C. Huang, D. Xie, T. Wang, and C. Zhao, "Medical cyber-physical systems: A solution to smart health and the state of the art," *IEEE Trans. Computat. Social Syst.*, vol. 9, no. 5, pp. 1359–1386, Oct. 2022, doi: 10.1109/TCSS.2021.3122807.
- [8] *Welcome to H2O 3-H2O 3.30.0.3 Documentation*. Accessed: Nov. 1, 2022. [Online]. Available: <http://docs.h2o.ai/h2o/latest-stable/h2o-docs/welcome.html>
- [9] Y. Bengio, "Learning deep architectures for AI," *Found. Trends Mach. Learn.*, vol. 2, no. 1, pp. 1–127, 2009, doi: 10.1561/22000000006.
- [10] S. M. El-Ghamrawy and A. I. 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, doi: 10.1504/IJC-ND.2013.050503.
- [11] S. M. El-Ghamrawy and A. I. El-Desouky, "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, doi: 10.1142/S0219622012500216.
- [12] 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, doi: 10.1109/ICCES.2016.7822004.
- [13] 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. (AMLTA)*, vol. 723, Cham, Switzerland: Springer, 2018, pp. 362–374, doi: 10.1007/978-3-319-74690-6\_36.
- [14] S. M. 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.*, vol. 1261, Cham, Switzerland: Springer, 2020, pp. 347–357, doi: 10.1007/978-3-030-58669-0\_32.
- [15] 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, doi: 10.1016/j.ejor.2008.07.025.
- [16] M. M. Mafarja and S. Mirjalili, "Hybrid whale optimization algorithm with simulated annealing for feature selection," *Neuro Comput.*, vol. 260, pp. 302–312, Oct. 2017, doi: 10.1016/j.neucom.2017.04.053.
- [17] Y. Ding, K. Zhou, and W. Bi, "FS based on hybridization of genetic algorithm and competitive swarm optimizer," *Soft Comput.*, vol. 24, pp. 11663–11672, Feb. 2020.
- [18] 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, doi: 10.1016/j.eswa.2010.12.156.
- [19] B.-Y. Tony, B. Xue, and M. Zhang, "Particle swarm optimization for feature selection: A hybrid filter-wrapper approach," in *Proc. IEEE Congr. Evol. Comput. (CEC)*, Sendai, Japan, Jan. 2015, pp. 2428–2435, doi: 10.1109/CEC.2015.7257186.
- [20] 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, doi: 10.1023/A:1021251113462.
- [21] L. Akter, 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, doi: 10.1007/s42979-021-00551-6.
- [22] S. A. El-Rahman, "Predicting breast cancer survivability based on machine learning and features selection algorithms: A comparative study," *J. Ambient Intell. Humanized Comput.*, vol. 12, no. 8, pp. 8585–8623, Aug. 2021, doi: 10.1007/s12652-020-02590-y.
- [23] 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*, vol. 53, Singapore: Springer, 2021, pp. 703–711, doi: 10.1007/978-981-15-5258-8\_65.
- [24] M. Al-Rajab, J. Lu, and Q. Xu, "A framework model using multi-filter feature selection to enhance colon cancer classification," *PLoS ONE*, vol. 16, no. 4, 2021, Art. no. e0249094, doi: 10.1371/journal.pone.0249094.
- [25] 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, doi: 10.1007/s12553-021-00554-6.
- [26] 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*.
- [27] S. A. B. 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, doi: 10.1007/978-3-030-71676-9\_1.
- [28] N. G. Elseddeq, S. M. Elghamrawy, M. M. Salem, and A. I. El-Desouky, "A selected deep learning cancer prediction framework," *IEEE Access*, vol. 9, pp. 151476–151492, 2021, doi: 10.1109/ACCESS.2021.3124889.
- [29] 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, doi: 10.1016/j.asoc.2008.01.006.
- [30] X.-S. Yang and S. Deb, "Cuckoo search via Lévy flights," in *Proc. World Congr. Nature Biologically Inspired Comput. (NaBIC)*, Dec. 2009, pp. 210–214, doi: 10.1109/NaBIC.2009.5393690.
- [31] X.-S. Yang and S. Deb, "Cuckoo search: Recent advances and applications," *Neural Comput. Appl.*, vol. 24, no. 1, pp. 169–174, Jan. 2014, doi: 10.1007/s00521-013-1367-1.
- [32] M. Liu, L. Xu, J. Yi, and J. Huang, "A feature gene selection method based on ReliefF and PSO," in *Proc. 10th Int. Conf. Measuring Technol. Mechatronics Autom. (ICMTMA)*, Feb. 2018, pp. 298–301, doi: 10.1109/ICMTMA.2018.00079.

- [33] 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, doi: [10.1109/TCBB.2018.2868341](https://doi.org/10.1109/TCBB.2018.2868341).
- [34] 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, doi: [10.3844/jcssp.2018.1521.1530](https://doi.org/10.3844/jcssp.2018.1521.1530).
- [35] 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, doi: [10.1109/ACCESS.2020.3029890](https://doi.org/10.1109/ACCESS.2020.3029890).
- [36] 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, doi: [10.23919/ICACT48636.2020.9061234](https://doi.org/10.23919/ICACT48636.2020.9061234).
- [37] K. Cahyaningrum 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, doi: [10.1109/ICoDSA50139.2020.9213051](https://doi.org/10.1109/ICoDSA50139.2020.9213051).
- [38] S. D. Khan, L. Alarabi, and S. Basalamah, "Toward smart lockdown: A novel approach for COVID-19 hotspots prediction using a deep hybrid neural network," *Computers*, vol. 9, no. 4, p. 99, Dec. 2020, doi: [10.3390/computers9040099](https://doi.org/10.3390/computers9040099).
- [39] S. Malik, J. Amin, M. Sharif, M. Yasmin, S. Kadry, and S. Anjum, "Fractured elbow classification using hand-crafted and deep feature fusion and selection based on whale optimization approach," *Mathematics*, vol. 10, no. 18, p. 3291, Sep. 2022, doi: [10.3390/math10183291](https://doi.org/10.3390/math10183291).
- [40] 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, doi: [10.1007/s10723-018-9465-z](https://doi.org/10.1007/s10723-018-9465-z).
- [41] S. Mirjalili and A. Lewis, "The whale optimization algorithm," *Adv. Eng. Softw.*, vol. 95, pp. 51–67, May 2016, doi: [10.1016/j.advengsoft.2016.01.008](https://doi.org/10.1016/j.advengsoft.2016.01.008).
- [42] E.-S. M. El-Kenawy, M. M. Eid, M. Saber, and A. Ibrahim, "MbGWO-SFS: Modified binary grey wolf optimizer based on stochastic fractal search for feature selection," *IEEE Access*, vol. 8, pp. 107635–107649, 2020, doi: [10.1109/access.2020.3001151](https://doi.org/10.1109/access.2020.3001151).
- [43] 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, doi: [10.1016/j.neucom.2015.06.083](https://doi.org/10.1016/j.neucom.2015.06.083).
- [44] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proc. Int. Conf. Neural Netw. (ICNN)*, Nov. 1995, pp. 1942–1948, doi: [10.1109/ICNN.1995.488968](https://doi.org/10.1109/ICNN.1995.488968).
- [45] H. Okut, "Deep learning for subtyping and prediction of diseases: Long-short term memory," in *Deep Learning Applications*. London, U.K.: IntechOpen, 2021. [Online]. Available: <https://www.intechopen.com/chapters/75265>, doi: [10.5772/intechopen.96180](https://doi.org/10.5772/intechopen.96180).
- [46] W. H. Wolberg and O. L. Mangasarian. (1992). *Breast Cancer Wisconsin (Original) Dataset*. [Online]. Available: [https://archive.ics.uci.edu/ml/datasets/breast+cancer+wisconsin+\(original\)](https://archive.ics.uci.edu/ml/datasets/breast+cancer+wisconsin+(original))
- [47] E. Alba, J. Garcia-Nieto, L. Jourdan, and E.-G. Talbi, "Gene selection in cancer classification using PSO/SVM and GA/SVM hybrid algorithms," in *Proc. IEEE Congr. Evol. Comput.*, Sep. 2007, pp. 284–290, doi: [10.1109/CEC.2007.4424483](https://doi.org/10.1109/CEC.2007.4424483).
- [48] R. Machmud, and A. Wijaya. (2021). *Cervical Cancer Behavior Risk Data Set (2016)*. [Online]. Available: <https://archive.ics.uci.edu/ml/datasets/Cervical+Cancer+Behavior+Risk>



and computing technologies.

**NADIA G. ELSEDEQ** received the B.Sc. degree in computer engineering from the Computer Engineering and Control Systems Department, MISR Higher Institute for Engineering and Technology (MET), Mansoura, Egypt. She is currently pursuing the M.Sc. degree in computer engineering with the Computer Engineering and Control Systems Department, Faculty of Engineering, Mansoura University, Egypt. Her major research interests include optimization techniques, big data,



applied artificial intelligence. She serves as a Reviewer for *EURASIP Journal on Wireless Communications and Networking*.

**REHAM ARNOUS** received the M.S. and Ph.D. degrees from the Department of Computer Engineering and Control Systems, Faculty of Engineering, Mansoura University, Mansoura, Egypt, in 2008 and 2018, respectively. She is currently a Lecturer with the Delta Higher Institute for Engineering and Technology (DHIT), Mansoura. She has published in the field of resource management in cognitive radio networks. Her research interests include cognitive radio, wireless networking, and applied artificial intelligence.



of Graduate Studies for Statistical Research, Cairo University. She is a Lecturer with the Embedded Network Systems and Technology Department, Faculty of Artificial Intelligence, Kafrelsheikh University, Kafrelsheikh, Egypt. She received a lot of badges and certificates from Huawei, Cisco, and IBM in fields of the Internet of Things, artificial intelligence, cloud computing, and predictive analysis. She has served as a Reviewer for many high-quality journals, including *Journal of Artificial Intelligence* university, *Journal of Computer, Oxford Academic*, and *Journal of Supercomputing* (Springer).

**ZAINAB H. ALI** received the B.Sc. degree in computer science and engineering from the Faculty of Electronic Engineering, Menofia University, the master's degree in telecommunication and computer networks, in 2016, and the Ph.D. degree from the Department of Computer Engineering and Control Systems, Faculty of Engineering, Mansoura University, in 2020. Since 2019, she has been an Instructor with Huawei Company. In 2021, she works as a Lecturer with the Faculty



well-known international journals.

**ALI I. ELDESOUKY** received the M.A. and Ph.D. degrees from the University of Glasgow, U.K. He is currently a Full Professor with the Computers Engineering and Systems Department, Faculty of Engineering, Mansoura University, Egypt. He is also a Visiting Part-Time 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



**MAGDY M. FADEL** received the B.Sc., M.Sc., and Ph.D. degrees in computers and control engineering from the Faculty of Engineering, Mansoura University, in 1995, 2003, and 2018, respectively. He was a Lecturer at Prince Sattam University. He is a Chief Engineer with the Faculty of Engineering, Mansoura University. His research interests include network security, data security, artificial intelligence (AI), data mining, and deep learning (DP).