Influential Gene Selection from High-Dimensional Genomic Data using a Bio-inspired Algorithm wrapped Broad Learning System

Pouramasi Parhi¹, Ranjeeta Bisoi², and Pradipta Kishore Dash³, Member, IEEE
¹Department of Computer Science Engineering, Siksha ‘O’ Anusandhan Deemed to be University, Bhubaneswar, Odisha, India
²Multidisciplinary Research Cell, Siksha ‘O’ Anusandhan Deemed to be University, Bhubaneswar, Odisha, India
³Multidisciplinary Research Cell, Siksha ‘O’ Anusandhan Deemed to be University, Bhubaneswar, Odisha, India

ABSTRACT
The classification of high dimensional gene expression/ microarray data always plays an important role in various disease diagnoses and drug discovery. To avoid the curse of high dimensionality, the selection of the most influential genes remains a challenging task for the researchers in the machine learning field. As the extraction of influential features by a bio-inspired algorithm is a non-deterministic polynomial-time (NP-Hard) task, the possibility of applying new algorithm is always there. In this suggested work, a recently developed bio-inspired algorithm, Monarch Butterfly Optimization (MBO), is wrapped with the Broad Learning System (BLS), called MBO-BLS, to choose the most influential features and classify the microarray data simultaneously. In the first stage, a pre-selection method (Relief) is used to select a feature subset. Then, this selected feature subset undergoes further execution with the MBO-BLS model. To estimate the efficacy of the presented model, six cancerous microarray datasets are taken. Here, sensitivity, specificity, precision, F-score, Kappa, and MCC measures are used for an impartial comparison. Further, to prove the supremacy of the presented method, the basic BLS, Genetic Algorithm wrapped BLS (GA-BLS), Particle Swarm Optimization wrapped BLS (PSO-BLS), and the existing ten models are taken for comparison. Moreover, to examine the designed model statistically, Analysis of variance (ANOVA) test is also performed here. From the above qualitative and quantitative analysis, it is concluded that the proposed MBO-BLS model outclasses other considering models.

INDEX TERMS
High dimensional gene expression data, Influential genes, Classification, Relief, Monarch Butterfly Optimization, Broad Learning System

I. INTRODUCTION
Gene expression data allows monitoring the expression values of thousands of genes in a single experiment. This advanced technology is very much beneficial for analysing disease mechanism which set to improve the quality of health science. The classification of high dimensional gene expression data always plays an important role in various disease diagnoses and drug discovery [1]. But the drawback of gene expression data is its high dimensionality which comprises a small number of samples and a large number of features. Thus, the selection of the most relevant features is essential to reduce the classificational complexity and computational cost of this high dimensional gene expression data [2]. Researchers have been proposed mainly two techniques to select relevant genes, one is feature extraction and another is feature selection. In the feature extraction technique, the original high dimensional feature space is transformed to a lower-dimensional subset of reduced features by using nonlinear and linear techniques [3]. But in the feature selection technique, a small subgroup of essential features is chosen from the original massive set of attributes. Feature extraction techniques may lose some information due to the transformation of original data. So, in this proposed work feature selection technique is emphasized to deduct the irrelevant, redundant features and to increase the learning performance. By considering the evaluation aspect [4], the feature selection is classified as a filter, wrapper, and hybrid method. In the filter approach, the significant feature subset is selected by evaluating each feature through some independent test before the application of the classification
algorithm. But in the case of the wrapper approach, the vital feature subset is evaluated by using a classifier where classification accuracy is the key factor to evaluate the best feature space. The wrapper method is more effective than the filter method because in the filter method the evaluation process is not following any learning algorithm but the advantage of the filter approach is it is having less computational cost. In the wrapper approach, the metaheuristic algorithm is wrapped with any machine learning algorithm to search the optimal best feature subset. Some wrapper method is discussed here as genetic algorithm (GA) is hybridized with support vector machine (SVM) to select the best features of the gene expression data [5].

GA is embedded with an Extreme Learning Machine (ELM) for the classification of biomedical cancer data [6]. Particle swarm optimization (PSO) used K-Nearest Neighbor (KNN) as a learning algorithm for the selection of significant genes of cancer data [7]. Artificial Bee Colony (ABC) and Genetic Bee Colony (GBC) used SVM as a learning algorithm for gene selection of microarray cancer data [8,9]. GA is applied for gene selection and Naïve Bayes is considered as a classifier for Indian diabetes data [10]. Best First Search technique is wrapped with an Artificial Neural Network learning algorithm for feature selection and classification of colon cancer data [11]. Bat algorithm (BA) is wrapped with optimum path forest (OPF) learning algorithm for feature selection of biomedical data [12]. Cat swarm optimization (CSO) and kernel extreme learning machine (KELM) are used for biomedical data gene selection [13]. Mutual information combined with adaptive genetic algorithm for selection of informative genes [14]. Although the wrapper approach is more applicable for getting better classification accuracy than the filter approach as it has a high risk of data overfitting and it is more computationally complex than the filter approach. Considering the advantages of both approaches the hybrid feature selection method is formed by combining both filter and wrapper methods. In the hybrid method, the significant genes are selected initially applying the filter approach and then the wrapper method is applied in which the optimum subset of that significant genes is sorted out. For the past decade various traditional and metaheuristic machine learning techniques have been used to classify the complex high dimensional gene expression data. The high classification accuracy and low computational time is the key factor for motivating the researchers to explore various advanced techniques for better results. Various approved popular techniques have been introduced as a classifier to solve the issues such as Support Vector Machine (SVM) [15-17], Artificial Neural Network (ANN) [18], Fuzzy Set Theory [19], K-Nearest Neighbor (KNN) [20], Functional Link Artificial Neural Network (FLANN) [21], Backpropagation [22], Multi-Layer Perception (MLP) [23], Radial Basis Function Neural Network (RDFNN) [24], etc.

These discussed classifiers are very popular for their potential for various classification performances. Now a days, deep learning is also very much popular for its effectiveness in the classification task. It effectively promotes classification performance by deepening the layers in neural networks. But the deep network is very time-consuming because of its complicated deep network. To solve these issues, a single-layer feed-forward neural network (SLFNN) is more applicable to solve classification and regression problems [25-30]. The conventional gradient descent method is used as a learning algorithm to train SLFNN [31, 32]. But it suffers from various issues like slow converging and trapping in local minimum and overfitting [33]. So, a different non-iterative learning method called random vector functional link neural network (RVFLNN) is presented which promotes the generalization performance [29],[34]. It also eliminates the drawback of a long training process. So, the RVFLNN is very less time-consuming. But it has a drawback that it is not working well in large high-volume data [35]. Thus, a new method called broad learning system (BLS) is proposed [36] by taking the concept of single-layer feed-forward RVFLNN. According to BLS, the input features are mapped to expanded feature nodes which form a broad network to enhance the classification accuracy. Here in BL-RVFLNN, the input weights are chosen randomly which may affect the classification performance of the algorithm. To solve this issue researchers have been proposed an optimization algorithm that optimizes the input weight and enhances the performance of the algorithm. Many metaheuristic algorithms like PSO [37,38], Ant-colony [39,40], cuckoo search [41], Moth-Flame-Optimization (MFO) [42,43], Genetic Algorithm (GA) [44-46] have been proposed by the researchers to optimize weight, various parameters that can help to improve performance of the algorithm.

Here, Monarch butterfly optimization-based BLS (MBO-BLS) model [47] is used for the selection of notable genes and to get the enhanced classification accuracy with optimum weight and bias. The supremacy of the proposed model is estimated by considering the validation test as specificity, F-score, sensitivity, and Matthews Correlation Coefficient (MCC). Here, the performance of the suggested MBO-BLS model is compared with the conventional technique like GA-BLS, PSO-BLS, and the basic classifier BLS. The main goal of this paper is:

- To present a robust classification model i.e., MBO optimized BLS to classify the high dimensional cancerous data.
- To achieve high classification accuracy with a minimum number of features.
To show the supremacy of the presented method, other benchmark approaches such as GA-BLS, PSO-BLS, and basic BLS are considered for comparison.

Eventually, a statistical analysis i.e., ANOVA test has been performed to establish the superiority of the presented model with respect to other standard models.

The appearance of the rest of the work is aligned in the following way: Section 2 performs an analysis of the presented model. Section 3 explains all the supported methods and section 4 covers the proposed method. Section 5 and 6 explain the experimental setup with the experimentation and the result validation portion respectively. Eventually, Section 7 discusses the conclusion part.

II. MODEL ANALYSIS

The basic layout of the presented model (MBO-BLS) is described in Fig.1. Before the execution, the missing values of the dataset are replaced with the most occurring value of that attribute and then each value of the dataset is a mapping between 0 and 1 by min-max normalization. At first, the 10-Fold Cross-Validation technique is used to divide the whole sample into a training sample and a testing sample. Then, the Relief feature selection method is used to preselect the relevant genes. After getting the most significant features, the MBO-BLS wrapper method is employed to get the optimum subset of genes. In the presented model, the MBO algorithm is applied to optimize the weight and bias of the BLS classifier to select the most significant feature subset. Moreover, MBO optimized BLS (MBO-BLS) model undergoes the testing phase with the testing data having the most influential feature subset to achieve the classification accuracy.

III. SUPPORTED METHODS

In this section, all the supported methods are discussed.

A. BROAD LEARNING SYSTEM (BLS)

In BLS, the input data are mapped to generate feature nodes then the mapped feature nodes are expanded to form the enhancement nodes where the weights are randomly generated [36]. Fig. 2 describes the basic structure of BLS.

In BLS, the input data are mapped to generate feature nodes then the mapped feature nodes are expanded to form the enhancement nodes where the weights are randomly generated [36]. Fig. 2 describes the basic structure of BLS.

The original feature set \( X \in \mathbb{R}^{p \times q} \) is mapped randomly to new feature nodes of the network layer.

\[
Z_n = \varnothing_n(W_{tn} + \beta_{tn}), \quad n = 1 \ldots p
\]  

In (1), \( \varnothing_n \) represents the \( n \)th mapping function, \( \beta_{tn} \) and \( W_{tn} \) represents the bias and random weight respectively. Then the new feature nodes as \( Z_p = [Z_1 \ldots \ldots Z_p] \) is enhanced to form a new set of enhancement nodes which is represented as

\[
H_l = \xi_l(Z_p^l W_{hl} + \beta_{hl}), \quad l = 1 \ldots q
\]  

In (2), \( H_l \) is the \( l \)th enhancement node, where \( \beta_{hl} \) and \( W_{hl} \) denote the bias and random weight respectively. The newly generated enhancement nodes are considered as \( H^q = [H_1 \ldots \ldots H_q] \)

The final output is computed as per (3).

\[
Y = [z_1 \ldots \ldots z_p]\xi_1(z_p^p w_{h1} + \beta_{h1}) \ldots \ldots \xi_q(z_p^p w_{hm} + \beta_{hm}) w_p^q
\]

\[
=[z_1 \ldots \ldots z_p|H_1 \ldots \ldots H_q]w_p^q
\]

\[
=[z_p^p |H_q^q]w_p^q
\]

Here, the output layer weight \( w_p^q \) is represented in (4).

\[
w_p^q = [Z_p^p |H_q^q]^T Y
\]

Then, \( A_p^q = [Z_p^p |H_q^q] \) and the connecting weight \( W_p^q \) is calculated by getting the solution of L2 norm regularized least square problem.

\[
W_p^q = \text{argmin}_{W_p^q} \| A_p^q W_p^q - Y \|_F^2 + \lambda \| W_p^q \|_2^2
\]
In (5), \( \lambda \) is taken as the constraint constant. Then the solution is obtained by solving the above given optimal problem as per (6).

\[
W_p^q = \left( \lambda I + A_p^q A_p^{-1} \right)^{-1} A_p^q Y
\]  

(6)

Then, \( A_p^{q+} \) is taken as the inverse of \( A_p^q \) and is represented by (7).

\[
\lim_{\lambda} \left( \lambda I + A_p^q A_p^{-1} \right)^{-1} A_p^q
\]  

(7)

**B. MONARCH BUTTERFLY OPTIMIZATION ALGORITHM**

Monarch butterfly optimization (MBO) algorithm is a newly designed swarm intelligence optimization technique based on the migration strategies of monarch butterfly species found in North America [47,48]. Fig. 3 presents a pictorial representation of the MBO algorithm.

The implementation of this algorithm is simpler and easier which is based on two operators: migration operator and butterfly adjusting operator. According to this algorithm whole population is divided into two subpopulations of equal size. Half of the population has the best fitness value resides in population 1 and the rest half to population 2. After each iteration, the best value is reserved as global optimum and then the newly updated population is again divided into two subpopulations based on the new fitness function. This process is repeated until the stopping criteria are met.

1) MIGRATION OPERATOR

The main objective of the migration operator is to exchange the information between both populations and also in the subpopulation1. The updating of each butterfly in subpopulation 1 depends on the position of another butterfly in both populations and on migration ratio \( p \). The updating of \( l \)th butterfly in subpopulation 1 can be calculated:

\[
Y_{ln}^{t+1} = \begin{cases} 
Y_{1,ln}^t & \text{if } r < p \\
Y_{2,ln}^t & \text{else}
\end{cases}
\]  

(8)

Where \( Y_{ln}^{t+1} \) represents the position of \( Y_l \) on \( l \)th dimension in \( t+1 \) generation and \( l_1 \) and \( l_2 \) indices the integer index randomly selected from both subpopulation 1 and subpopulation 2. Here parameter \( r \) is taken as \( r=\text{Rand}*\text{Peri} \), where \( \text{Rand} \) is denoted as a random real number in [0,1] and \( \text{Peri} \) is taken as migration period.

2) ADJUSTING OPERATOR

The movement of each butterfly in subpopulation 2 is calculated based on adjusting ratio \( p \) and the adjusting rate of butterfly (BAR).

\[
y_{ln}^{t+1} = \begin{cases} 
Y_{\text{best},n}^{t} & \text{if } \text{Rand} \leq p \\
Y_{ln}^{t} + \xi \times (dY_n - 0.5) & \text{if } \text{Rand} > p \text{ and } \text{Rand} \leq \text{BAR} \\
Y_{ln}^{t} & \text{if } \text{Rand} > \text{BAR}
\end{cases}
\]  

(9)

In (9), \( Y_{\text{best},n}^{t} \) is the \( n \)th element of the global best \( Y_{\text{best},n}^{t} \) is the \( n \)th element of global best \( Y_{\text{best},n}^{t} \) at the current generation \( t \), \( Y_{ln}^{t} \) is the \( n \)th element of the randomly.

Selected butterfly from the subpopulation 2. \( \xi \) represents as a weighted factor which can be formulated as (10):

\[
\xi = \frac{Z_{\text{max}}}{t^2}
\]  

(10)

Here, \( Z_{\text{max}} \) = maximum walk step of individual butterfly in each step and \( t \) is the current generation and \( dY_n \) represents the walk step of the individual butterfly \( i \) which can be formulated by taking the concept of Levy flight method as (11).

\[
dx_n = \text{Levy}(x_n^t)
\]  

(11)

Here, \( \xi \) has a great impact on \( dY_n \) and \( Y_{ln}^{t} \). When \( \xi \) value will be more than it can accelerate the exploration in search space and if the \( \xi \) value will be small it can accelerate the exploitation in search space.

**Algorithm 1:** Monarch Butterfly Optimization (MBO)

**Input:** The population size of Monarch Butterfly NP and the various parameters as migration ratio \( p \), migration period \( \text{Peri} \), Butterfly Adjustment Rate (BAR), the maximum walk step of Levy flight \( Z_{\text{max}} \), and maximum iteration

**Output:** Global best solution

**Step:** 1 Begin
Initialize the whole population of Monarch Butterfly NP and the various parameters as migration ratio \( p \), migration period \( \text{Peri} \), Butterfly Adjusting Rate (BAR), the maximum walk step of Levy flight \( Z_{\text{max}} \), and maximum iteration.

Step 2: While the stopping condition is not reached do

Arrange the whole population in descending order based on the fitness value.

Design subpopulation 1 by taking the best half of the population as NP1 and rest as subpopulation 2 or NP2.

Modify subpopulation 1 using migration operator in (8).

Modify subpopulation 2 using the butterfly adjusting operator (9).

Remerge the newly updated subpopulation as a whole new population.

Get the best one.

End While.

Step 3: Get the global best one

End.

**C. RELIEF APPROACH**

The original Relief algorithm is formulated by Kira and Rendell [49] which is based on a feature approach. According to this instance-based learning algorithm, the features are ranked based on the assigned weight to each feature. It is an iterative method, where the feature weight of an individual feature is evaluated to estimate the "Relevance" of features. From the input dataset, a sample \( K \) is randomly selected at each iteration and then its nearest neighbor from the same class called nearest-Hit \( k_H \) and the nearest neighbor from the different class as nearest-Miss \( k_M \) is picked out. The updating of weight for each feature \( i \) is given below:

\[
W(i) = W(i) - \frac{\text{diff}(i, k_H)}{m} + \frac{\text{diff}(i, k_M)}{m} \tag{12}
\]

Where the heuristic estimation function \( W(i) \) is the relevance of feature \( i \) and it is initialized to zero. The difference of feature \( i \) between sample \( k \) and \( k' \) is denoted as \( \text{diff}(i, k, k') \). Equation (12) is repeated \( m \) times where \( m \) is the sample size of training data for getting the final update relevance of features.

**IV. SUGGESTED METHODS**

**A. SELECTION OF GENES BY RELIEF APPROACH**

In the proposed work, a filter approach is used for preselection of the most relevant genes, then it is followed by a wrapper method to find the optimal gene subset. Among several filter approaches, the Relief feature selection method [49] is used in this proposed model for gene preselection. According to the Relief approach, each feature is evaluated according to its rank. Here the topmost ranked genes (ranges of 500) are selected (as per [50]) for further execution. Then this top-ranked subset of genes is used in the MBO-BLS model to find the optimum gene subset and enhance the classification accuracy. At the same time, the MBO algorithm is applied to obtain the optimum subset of genes and optimize the weight and bias of the BLS classifier.

**B. PROPOSED MBO-BLS ALGORITHM**

In the proposed work, the MBO algorithm is used for the selection of genes and at the same time optimizes the weight and bias of BLS respectively. Fig.4 illustrates the schematic diagram of this proposed work and the proposed BLS-MBO model is described in algorithm 2. The individual solution of the selected \( K \) dimensional feature is, \( X^j = \{ w, b, X^j_1, X^j_2, X^j_3, \ldots, X^j_K \} \) where \( j = \{1,2,3, \ldots, K\} \). Here among the \( n \) bit solution the first reserved 2 bits are assigned with weight \( w \) and bias \( b \) and the remaining \( n-2 \) bits are fixed for the gene subset and assigned with the encoded value 0 and 1 (0 indicates for rejection and 1 for selection). Then \( X^j \) can be represented as \( X^j = \{ w, b, 1,0,0, \ldots, 1 \} \).

Here logistic conversion function is used to convert the value of each feature into binary form. This logistic function is expressed in (13) and (14).

\[
x^j_p = \begin{cases} 1, & \log \sigma(x^j_p) > 0.5 \\ 0, & \text{otherwise} \end{cases} \tag{13}
\]

In (13),

\[
\log \sigma(x^j_p) = \frac{1}{1 + e^{(-\gamma^m)}} \tag{14}
\]

In the proposed algorithm the fitness value is calculated in (15).

\[
fitt(\text{value}) = \text{Acc}(\text{avg}) = \frac{\sum_{i=1}^{10} \text{test Acc}_{i}}{10} \tag{15}
\]

Here, the 10-fold cross-validation method is used to compute the average testing accuracy of the proposed MBO-BLS classifier.
**FIGURE 4.** Flowchart of the presented MBO-BLS approach

**Algorithm 2:** Suggested MBO-BLS algorithm

| Input | Output |
|-------|--------|
| Population size (NP), migration ratio P, migration period Peri, Butterfly Adjusting Rate (BAR), the maximum walk step of Levy flight $Z_{\text{max}}$ and maximum iteration (N), Fitness (f), Ub and Lb (upper and lower bound) for w, b, number of features and the dimension size of the search space. | The final outcome as classification accuracy% with the length of the subset of features. |

1. for individual candidate solutions do
2. Apply the transformation function and convert the values of each candidate solution into (1,0) pattern (Set the initial two bits for w and b, and the remaining bits are considered for the gene subset. Here, 1 and 0 specify the selection and rejection of that gene respectively).
3. Determine the fitness employing w, b, and the selected gene subsets.
4. end for
5. Store the achieved values of the fitness in descending manner and choose the best and worst one.
6. As per the index of organized fitness value, arrange the population (NP).
7. Keep the best fitness with the location of the candidate solution.
8. Find out the mean value of the achieved fitness.
while I < N do
10:  if I == 1 then
11:     for m=1: NP do
12:         Change the solution location, w, and b by employing Eq. (8) and Eq. (9).
13:     end for
14:  else if (mean_value_fitness (current solution) – mean_value_fitness (previous solution))/ mean_value_fitness (current solution)> 0.001
15:     Then
16:         Repeat steps 11 to 14
17:     Else
18:         break.
19:  end if
20:  for every upgraded solution do
21:     Check the Lb and Ub of the solution location, w, and b.
22:     Repeat steps 1 to 4 for achieving the new values of fitness.
23:  if fitness_current_solution > fitness_previous_solution, then
24:      Change the fitness of the solution.
25:  Else
26:      keep the previous solution fitness.
27:  end if
28:  end for
29:  Repeat steps from 5 to 8
30: end while
31: Find the final outcome as classification accuracy% with the length of the subset of features.

V. SUGGESTED METHODS

A. SIMULATION ENVIRONMENT
The simulation environment of the entire work is given below:
Processing unit: Intel(R) Core (TM) i5-7200U with 2.5 GHz processing speed, Operating System: Windows 10, RAM capacity: 8 GB, and Programming Language platform: R2015b MATLAB.

B. SIMULATION ENVIRONMENT

The detailed description of the microarray data which are taken for the implementation is given in Table 1.

| Datasets          | Dimensions | Sample size | No. of Genes | Classes |
|-------------------|------------|-------------|--------------|---------|
| Leukemia [51]     | 72×7129    | 72          | 7129         | 2       |
| Colon tumor [52]  | 62×2000    | 62          | 2000         | 2       |
| Ovarian cancer [53]| 253×15154  | 253         | 15154        | 2       |
| Lymphoma-3[54]    | 62×4026    | 62          | 4026         | 3       |
| SRBCT [55]        | 88×2308    | 88          | 2308         | 4       |
| ALL-AML-3 [56]    | 72×7129    | 72          | 7129         | 3       |

C. PARAMETER INITIALIZATION
Here to prove the supremacy of the proposed work (MBO-BLS), some well-known models like PSO-BLS, GA-BLS are taken for comparison. To avoid ambiguity the equal value of both iteration numbers and also the population size are considered. The other parameters of all the algorithm which shows the best performance are considered as the initial value which is shown in Table 2.

| GA               | PSO          | MBO          |
|------------------|--------------|--------------|
| Population size  | Population size | Population size |
| ~100             | ~100         | ~100         |
| No. of Iterations (Max) | No. of iterations (Max) | Max. iteration |
| ~100             | ~100         | ~100         |
| Cross over.      | C1 and C2 ~2 | Migration Ratio ~ 5/12 |
| Probability ~ 0.8| Maximum velocity ~ 65% | Adjusting Rate ~ 5/12 |
| Mutation probability ~ IW (Inertia of weight) ~1 | Max Generation ~ 100 | Max walk step ~ 1.0 |

D. MODEL ESTIMATION MEASURES
Here Various estimation measuring values are taken to evaluate the performance of all the models taken in this paper like classification accuracy (In percentage), sensitivity, specificity, F-score, precision, Matthew's correlation coefficient (MCC), etc. These attributes are evaluated by taking the concept of confusion matrix which is explained in Table 3.

| Actual Output | Positive | Negative |
|---------------|----------|----------|
| Expected Output | TN | FN | Positive | TP | TN |

- **Sensitivity (Sn):** The validation test is used to identify correctly the patients having disease. It is computed by taking the ratio of true positive sample and the total sample. Equation (16) determines the Sn value.

  \[
  Sensitivity(Sn) = \frac{TP}{TP+FN} \tag{16}
  \]

- **Specificity (Sp):** This validation test is used to identify correctly the people having no disease.
is computed by taking the proportion between the true negative sample and the total samples. Equation (17) determine the $Sp$ value.

$$Sp = \frac{TN}{(TN + FP)}$$

- **Precision (Pr):** Equation (18) determines the $Pr$ value.

$$Pr = \frac{TP}{(TP + FP)}$$

- **F-Score (Fs):** Equation (19) determines the $Fs$ value.

$$Fs = \frac{2 \times Pr \times Sn}{Pr + Sn}$$

- **Matthew’s correlation coefficient (MCC):** The computation of the correlation coefficient between the observed and predicted class is MCC. The range of the MCC value is within $+1$ and $-1$. If the MCC value returns $+1$, then it implies ideal prediction. Equation (20) determines the MCC value.

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

- **Kappa (Kpa):** Kappa is a statistical measurement that is used to calculate inter-rater reliability. Its value can range between $-1$ and $+1$. The value 1 indicates perfect agreement and $-1$ indicates less than perfect agreement. Equations (21), (22), and (23) determine the Kappa value.

$$Kappa = \frac{\text{Observed}_{Acc} - \text{Expected}_{Acc}}{1 - \text{Expected}_{Acc}}$$

$$\text{Observed}_{Acc} = \frac{TP \times TN}{TP + FN + TN + FN}$$

$$\text{Expected}_{Acc} = \frac{(TP + FN) \times (TP + FN) - (FP + TN) \times (FN + TN) + (FP + TN) \times (FN + TN) - (TP + FN) \times (FN + TN)}{TP + FN + TN + FN}$$

### VI. RESULT DISCUSSION

#### A. OUTCOME OF GENE PRE-FILTRATION

Here, the most significant genes are initially selected using a filter approach (Relief algorithm). Moreover, the topmost N prominent genes having the range of 1,500 are selected from the whole dataset. Then this reduced dataset further undergoes classification with BLS classifier. TABLE 4 and TABLE 5 demonstrate the percentage of classification accuracy of six benchmark gene expression binary and multiclass data respectively. Moreover, from these Tables, it is concluded that the percentage of classification accuracy increases till a specific no. of selected genes (N), and then the accuracy percentage remains unchanged or decreased. When the best subset of a gene is found from each microarray data, it is further forwarded for classification with the BLS classification model.

#### B. OUTCOME OF GENE PRE-FILTRATION

In the preselection process, the most significant genes which are selected through the Relief approach like 100 top genes of Ovarian cancer, 200 top genes of Leukemia, colon tumor, Lymphoma, and SRBCT, 500 top genes of ALL-MLL3 are individually forwarded for further execution to MBO-BLS.
model. In the MBO-BLS model each dataset is executed 10-times by using 10-fold cross-validation.

In TABLE 6, the all-performance measures of six microarray datasets like accuracy, specificity, sensitivity, precision, F-Score, MCC, and Kappa with 10-fold cross-validation are shown. From Table 6, it is observed that the performance measure of binary class Ovarian cancer data outperforms the other dataset. Moreover, TABLE 6 shows 100% results of sensitivity and specificity in Leukemia and SRBCT and ALL-MLL3 datasets.

TABLE 6. RESULTS OBTAINED (PERFORMANCE MEASURING VARIABLES) FROM THE MBO-BLS MODEL

| Name of the Dataset | ACC % | Sen% | Spe % | MCC % | F-measure | Kapp a |
|---------------------|-------|------|-------|-------|-----------|--------|
| Leukemia            | 99.45 | 100  | 97.56 | 97.72 | 98.83     | 94.32  |
| Colon tumor         | 98.2  | 97.86| 96.83 | 93.44 | 97.18     | 94.84  |
| Ovarian             | 99.96 | 99.87| 98.72 | 99.56 | 99.93     | 99.8   |
| Lymphoma-3          | 99.64 | 99.25| 97.83 | 97.3  | 97.86     | 96.72  |
| SRBCT               | 99.87 | 100  | 97.83 | 97.48 | 99.6      | 97.18  |
| ALL-AML-3           | 99.78 | 98.57| 100   | 97.87 | 98.96     | 97.53  |

Here, the convergence graph of GA-BLS, PSO-BLS, MBO-BLS models with all these six-microarray data are shown in Fig. 5(a)-5(f). From these graphs, it is observed that the accuracy of all these six datasets is increased gradually up to a maximum 100 iterations. As in Leukemia data, the accuracy is being converged after $54^{th}, 69^{th}, 74^{th}$, and $81^{st}$ iterations in MBO-BLS, GA-BLS, PSO-BLS, and BLS models respectively. For the Colon dataset, the accuracy in MBO-BLS, GA-BLS, PSO-BLS, and BLS models are converging after the $43^{rd}, 63^{rd}, 72^{nd}$, and $84^{th}$ iteration respectively. In Ovarian, the accuracy of the data is converging after $44^{th}, 74^{th}, 75^{th}$, and $81^{st}$ iterations in MBO-BLS, GA-BLS, PSO-BLS respectively. In Lymphoma it is showing that the accuracy of MBO-BLS, GA-BLS, PSO-BLS, and BLS is converged after the $43^{rd}, 63^{rd}, 72^{nd}$, and $84^{th}$ iteration respectively. Like this, the accuracy is also converging in MBO-BLS, GA-BLS, PSO-BLS, and BLS models after $42^{th}, 69^{th}, 79^{th}$, and $88^{th}$ iteration respectively. For ALL-MLL3 data, the accuracy in MBO-BLS, GA-BLS, PSO-BLS, and BLS models are converging after $30^{th}, 53^{th}, 64^{th}$, and $78^{th}$ iteration respectively. From these discussed converging graphs, it is observed that the rate of convergence of the MBO-BLS model is earlier than another discussed model.
Here, the proposed (MBO-BLS) approach is compared with the other three benchmark methods such as GA hybridized BLS, PSO hybridized BLS, and BLS to achieve an impartial comparison. This comparison is illustrated in TABLE 7. According to this TABLE 7, in the MBO-BLS model, the discussed microarray dataset like Leukemia, Colon, Ovarian, SRBCT, ALL-MLL-3, and Lymphoma-3 are performed 99.45%, 98.2%, 99.96%, 99.87%, 99.78%, and 99.64% accuracy respectively. These accuracies are superior compared to the other three considered models for all the datasets.

### TABLE 7. ACC% COMPARISON BETWEEN ALL THE APPROACHES

| Dataset            | BLS | GA-BLS | PSO-BLS | MBO-BLS |
|--------------------|-----|--------|---------|---------|
| Leukemia           | 97.78 | 98.62  | 98.9    | 99.45   |
| Colon              | 92.82 | 93.96  | 95.85   | 98.2    |
| Ovarian            | 94.6  | 95.84  | 98.14   | 99.96   |
| Lymphoma-3         | 96.74 | 97.68  | 98.32   | 99.64   |
| SRBCT              | 94.98 | 95.8   | 97.92   | 99.87   |
| ALL-AML-3         | 96.87 | 97.7   | 98.89   | 99.78   |

### C. SELECTED SIGNIFICANT BIO-MARKERS BY MBO-BLS

The highly influenced genes are selected from the proposed MBO-BLS model having high classification accuracy are listed in TABLE 8 and TABLE 9. Here, the proposed model selects mostly 3 prominent genes (i.e., Y00787_s_at, X95735_at, M23197_at) for Leukemia, 5 genes (i.e., M63391, M26383, H08393, Z50753, and J02854) for colon tumor, 3 genes (i.e., MZ244.95245, MZ245.24466, and MZ245.8296) for Ovarian cancer. Similarly in multiclass data, 3 genes (i.e., GENE1622X, GENE2403X, GENE2152X) are selected from Lymphoma-3, 6 genes (i.e., gene 2, gene 554, gene 714, gene 1003, gene 129, gene 153) are selected from SRBCT, and 3 genes (i.e., M21624_at, X76223_s_at, X59871_at, M31523_at) are selected from ALL-MLL-3.

### TABLE 8. NO. OF HIGHLY INFLUENCED GENES SELECTED FROM THE PROPOSED MBO-BLS MODEL IN THREE BINARY CLASS MICROARRAY DATASETS

| Binary Class      | Index of Gene | Name of the Gene | # Genes selected |
|-------------------|---------------|------------------|------------------|
| Leukemia          | 6201          | Y00787_s_at      | 3                |
|                   | 4847          | X95735_at        |                  |
|                   | 1834          | M23197_at        |                  |
| Colon             | 249           | M63391           | 5                |
|                   | 1772          | H08393           |                  |
|                   | 377           | Z50753           |                  |
|                   | 1671          | M26383           |                  |
| Ovarian cancer    | 1679          | MZ244.95245      | 3                |
|                   | 1680          | MZ245.24466      |                  |
|                   | 1684          | MZ246.41524      |                  |
**TABLE 9. NO. OF HIGHLY INFLUENCED GENES SELECTED FROM THE PROPOSED MBO-BLS MODEL IN THREE MULTI-CLASS MICROARRAY DATASETS**

| Multi-Class Dataset | Index of Gene | Name of the Gene | # Genes selected |
|---------------------|---------------|------------------|------------------|
| Lymphoma-3          | 3763          | GENE1622X        | 3                |
|                     | 757           | GENE2403X        |                  |
|                     | 859           | GENE2152X        |                  |
| SRBCT               | 2             | gene 2           | 6                |
|                     | 554           | gene 554         |                  |
|                     | 714           | gene 714         |                  |
|                     | 1003          | gene 1003        |                  |
|                     | 129           | gene 129         |                  |
|                     | 153           | gene 153         |                  |
| ALL-AML-3           | 1809          | M21624_at        | 4                |
|                     | 6696          | X76223_s_at      |                  |
|                     | 4342          | X59871_at        |                  |
|                     | 6855          | M31523_at        |                  |

In this work, for a clear visualization of selected influential features w.r.t the samples discretely, heat maps in six microarray datasets are presented in Fig. 6 (a)-(f). In general, a heat map gives a clear view of the level of expression of genes above a number of related samples. Heat map allocates different colours to every selected gene, that helps a naive user to visualize the gene expression data clearly.
Fig. 6. Heat map of selected influential features over the no. of samples

D. EXECUTION TIME OF PRESENTED MODEL

Here, the presented approach has two parts, i.e., a feature extraction part (by Relief) and a wrapper part (by MBO-BLS). Henceforth, the complete execution time relies on the time consumed by the two parts. For six microarray datasets, TABLE 9 shows the execution time in both parts.

| Dataset       | Time consumed by Relief stage | Time by MBO-BLS wrapper stage | Complete execution time |
|---------------|-------------------------------|-------------------------------|-------------------------|
| Leukemia      | 0.185                         | 162.125                       | 162.31                  |
| Colon cancer  | 0.625                         | 148.32                        | 148.945                 |
| Ovarian       | 0.52                          | 52.615                        | 53.135                  |
| Lymphoma-3    | 0.465                         | 57.312                        | 57.777                  |
| SRBCT         | 0.125                         | 142.281                       | 142.406                 |
| ALL-AML-3     | 0.462                         | 132.925                       | 133.387                 |
E. COMPARATIVE STUDY
In the presented work, the accuracy percentage (%) is considered as one of the performance measures for this proposed model. Here, 10 benchmark models have taken for comparison to prove the supremacy of the proposed model. Though different performance measures are used by different algorithms, this comparison couldn’t conclude a perfect outcome. However, a comparison is needed to present a qualitative analysis among the existing models with the proposed one. A comparison of classification accuracy w.r.t no. of selected genes for six microarray datasets in different models is illustrated in TABLE 10.

TABLE 10. A QUALITATIVE ANALYSIS OF THE PRESENTED METHOD W.R.T ACC% AND NO. OF GENES SELECTED WITH OTHER STANDARD MODELS (THE ‘-‘ SIGN INDICATES DATA UNAVAILABILITY)

| Existing approaches | Microarray Datasets |
|---------|------------------|
|         | Leukemia | Colon | Ovarian | Lymphoma | a-3 | SRBCT | ALL-AML-3 |
| PSO- AKNN (7) | - | - | - | - | 94 | 90.66 |
| (8.5) | (3.3) |
| IWSS-MB- NB (55) | 97.1 | 86 | - | - | - | - |
| (6.4) | (5.2) |
| DRFO-CFS (56) | 91.18 | 90 | 100 | - | - | - |
| (13) | (10) | (16) |
| mRMR- ABC (9) | - | - | - | 96.96 (5) | 96.30 (10) | 96.12 (20) |
| CC- PSO(57) | - | - | - | 96.8(306) | - | 93.7(63) |
| 8-SPMSO (58) | 98.1 (20) | 94.2 (20) | - | - | - | - |
| GBC (8) | - | - | - | 98.48 (5) | 96.38 (6) | 95.83 (8) |
| (10) | (20) |
| MCSO (13) | - | - | - | - | 71.04 (100) | - |
| GEM (5) | 91.5 (3) | 91.2 (8) | - | - | - | - |
| BDE- XRankf(59) | 82.4 (6) | 75 (4) | 95 (3) | - | - | - |
| MBO-BLS | 99.45 | 98.2 | 99.46 | 99.64 | 99.87 | 99.78 |
| (3) | (5) | (3) | (3) | (6) | (4) |

From TABLE 10, it is concluded that this proposed model performs maximum classification accuracy with minimal genes and it is significantly better than others in all six-microarray datasets. One model like DRFO-CFS achieves 100% accuracy but it selects 16 genes from ovarian cancer. Proposed MBO-BLS algorithm selects 5 genes from colon tumor with 98.2% classification efficacy, 3 informative genes from Leukemia with 99.45% accuracy, 3 genes from ovarian cancer having 99.46% accuracy, 3 genes from Lymphoma-3 with 99.64% accuracy, 4 genes from ALL-MLL-3 with 99.78% accuracy and 6 genes from SRBCT with 99.87% efficacy. In colon tumor, BDE-XRankf selects 4 genes whereas the proposed MBO-BLS selects 5 genes but with higher classification accuracy, and for ovarian cancer both these algorithms select the same number of genes but here is also the proposed algorithm classifies the data with better accuracy. From the above analysis, it is derived that the proposed algorithm outperforms in all six microarray datasets.

F. STATISTICAL ANALYSIS BY ANOVA
To examine the mean value of the definite groups (whether they are similar or not), a most popular statistical approach i.e., Analysis of variance (ANOVA) has been applied in this work. This statistical approach helps to estimate the model statistically. Generally, a null or alternative hypothesis is taken in ANOVA. In this test, F-value is calculated first, then according to F-value, the p-value is determined. The obtained p-value of ANOVA finalizes either to keep or discard the alternative hypothesis. The null hypothesis is refused, if p-value ≤ 0.05 (taking 5% as significance level) and it can be decided that the accuracy percentages of all the models are undoubtedly dissimilar. Moreover, the statistical analysis of the ANOVA test is shown in TABLE 11 and TABLE12. In this work, the p-value is calculated as 0.047. It is quite smaller than the previously considered p-value (i.e., 0.05). Henceforth, the alternative hypothesis is refused. So, it is decided that the presented algorithm is statistically better than other models.

TABLE 11. 2-WAY ANOVA TEST W.R.T ACC%

| Dataset | BLS | GA-BLS | PSO-BLS | MBO-BLS | Total |
|---------|-----|--------|---------|---------|-------|
| Number of datasets | 6 | 6 | 6 | 6 | 24 |
| ΣX | 574.49 | 577.76 | 585.07 | 596.9 | 2314.84 |
| Mean | 95.7483333 | 96.2933 | 97.51166 | 99.48333 | 96.4516 |
| ΣX2 | 3 | 333 | 667 | 333 | 6 |
| | 51919.7706 | 55649.9 | 58987.71 | 59383.73 | 223434. |
| Standard deviation | 0.20761141 | 0.653809 | 0.89941 | 0.89941 | 0.89941 |
| | 1.52397 | 1.212265 | 0.653809 | | |

TABLE 12. P-VALUE CALCULATION BY 2-WAY ANOVA

| Source | SS | DF | MS |
|--------|----|----|----|
| Between- treatments | 49.35675 | 3 | 16.45225 | f value= 2.909656469 |
| Within- treatments | 113.087233 | 20 | 5.654361667 | p value=0.047994755 |
| Total | 172.435925 | 23 | | |

VII. CONCLUSION
In this proposed work, a nature-inspired algorithm MBO is wrapped with an efficient classifier, namely, BLS to select the most influential features with maximum classification.
efficacy. Here six microarray datasets are taken for evaluation, where 3 datasets are binary (i.e., Leukemia, Colon tumor, Ovarian cancer) and the rest three are multiclass (i.e., Lymphoma, SRBCT, and ALL-MLL-3). At the first stage, a pre-selection method (Relief-f) is used to select a feature subset and then this selected feature subset undergoes further execution with the MBO-BLS model. Here various performance measures (i.e., sensitivity, specificity, precision, F-score, MCC, Kappa) are applied for an impartial comparison. Further to present the supremacy of the suggested method, the benchmark models as GA-BLS, PSO-BLS, BLS, and existing ten standard models are taken for comparison. Moreover, to examine the mean values of specific groups (whether they are similar or not), a most popular statistical approach i.e., Analysis of variance (ANOVA) is applied in this work. From the above qualitative and quantitative analysis, it is concluded that the proposed MBO-BLS model can be a dependable framework for the diagnosis of various diseases.

REFERENCES

[1] Ayadenta, Husna, and Adiwijaya Adiwijaya. "A clustering approach for feature selection in microarray data classification using random forest." Journal of Information Processing Systems 14.5 (2018): 1167-1175.

[2] Wang, Hong, Xingjian Jing, and Ben Niu. "A discrete bacterial algorithm for feature selection in classification of microarray gene expression cancer data." Knowledge-Based Systems 126 (2017): 8-19.

[3] Biccioato, Silvio, A. Luchini, and Carlo Di Bello. "PCA disjoint models for multiclass cancer analysis using gene expression data." Bioinformatics 19.5 (2003): 571-578.

[4] Ang, Jun Chin, Mirzal, A., Haron, H., & Hamed, H. N. A. "Supervised, unsupervised, and semi-supervised feature selection: a review on gene selection." IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB) 13.5 (2016): 971-989.

[5] J.C.H. Hernandez, B. Duval, J.K. Hao, A genetic embedded approach for gene selection and classification of microarray data, in: European Conference on Evolutionary Computation, Machine Learning and Data Mining in Bioinformatics, Springer, 2007, pp.90-101.

[6] Shukla, Alok Kumar, Pradeep Singh, and Manu Vardhan. "A two-stage gene selection method for biomarker discovery from microarray data for cancer classification." Chemometrics and Intelligent Laboratory Systems 183 (2018): 47-58.

[7] Kar, Subhajit, Kaushik Das Sharma, and Madhubanti Maitra. "Gene selection from microarray gene expression data for classification of cancer subgroups employing PSO and adaptive K-nearest neighborhood technique." Expert Systems with Applications 42.1 (2015): 612-627.

[8] H.M. Alshamlan, G.H.Badr, Y. A. Alholali, Genetic Bee Colony (GBC) algorithm: A new gene selection method for microarray cancer classification, Comput. Biol.Chem.56(2015)49–60.

[9] H. Alshamlan, G. Badr, Y. Alholali, mRMR-ABC: A hybrid gene selection algorithm for cancer classification using microarray gene expression profiling, BioMedRes.Int.2015(2015)604910–604910.

[10] Choubey, Dilip Kumar, et al. "Classification of Pima indian diabetes dataset using naive bayes with genetic algorithm as an attribute selection." Communication and computing systems: proceedings of the international conference on communication and computing system (ICCCS 2016). 2017.

[11] Rahman, Md Akizur, and Ravie Chandren Muniyandi. "Feature selection from colon cancer dataset for cancer classification using artificial neural network." International Journal on Advanced Science, Engineering and Information Technology 8.4-2 (2018): 1387-1393.

[12] Rodrigues, Douglas, et al. "A wrapper approach for feature selection based on bat algorithm and optimum-path forest." Expert Systems with Applications 41.5 (2014): 2250-2258.

[13] Mohapatra, P., Sreejit Chakravarty, and P. K. Dash. "Microarray medical data classification using kernel ridge regression and modified cat swarm optimization-based gene selection system." Swarm and Evolutionary Computation 28 (2016): 144-160.

[14] Rani, M. Janji, and Durairaj Devaraj. "Two-stage hybrid gene selection using mutual information and genetic algorithm for cancer data classification." Journal of medical systems 43.8 (2019): 1-11.

[15] N. Cristianini, J. Shawe-Taylor, An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods, Cambridge University Press, 2000.

[16] Ibrahim, Hadeel Tariq, et al. "A grasshopper optimizer approach for feature selection and optimizing SVM parameters utilizing real biomedical data sets." Neural Computing and Applications 31.10 (2019): 5965-5974.

[17] Malathi, V., N. S. Marimuthu, and S. Baskar. "Intelligent approaches using support vector machine and extreme learning machine for transmission line protection." Neurocomputing 73.10-12 (2010): 2160-2167.

[18] Zarada, Jake M. Introduction to artificial neural systems. Vol. 8. St. Paul: West, 1992.

[19] E.K. Aydogan, I. Karaoglan, P.M. Pardalos, HGA: hybrid genetic algorithm in fuzzy rule-based classification of high-dimensional problems, Appl. Soft Comput. 12 (2) (2012) 800–806.

[20] Anagaw, Amare, and Yang-Lang Chang. "A new complement naïve Bayesian approach for biomedical data classification." Journal of Ambient Intelligence and Humanized Computing 10.10 (2019): 3889-3897.

[21] Naik, Bighnara, et al. "A harmony search based gradient descent learning-FLANN (HS-GDL-FLANN) for classification." Computational Intelligence in Data Mining-Volume 2. Springer, New Delhi, 2015. 525-539.

[22] Heermmann, Philip Dale, and Nahid Khazenie. "Classification of multispectral remote sensing data using a back-propagation neural network." IEEE Transactions on Geoscience and Remote Sensing 30.2 (1992): 81-88.

[23] Al-Shargabi, Bassam, Feda Alshami, and Rami Alkhawaledeh. "Enhancing multi-layer prediction for breast cancer prediction." International Journal of Advanced Science and Technology (2019).

[24] Fernández-Navarro, Francisco, et al. "Evolutionary generalized radial basis function neural networks for improving prediction accuracy in gene classification using feature selection." Applied Soft Computing 12.6 (2012): 1787-1800.

[25] Gulyiev, Namig J., and Vugar E. Ismailov. "A single hidden layer feedforward network with only one neuron in the hidden layer can approximate any univariate function." Neural computation 28.7 (2016): 1289-1304.

[26] Arulampalam, Ganesh, and Abdesselam Bouzerdoum. "A generalized feedforward neural network architecture for classification and regression." Neural networks 16.5-6 (2003): 561-568.

[27] Pao, Y-H., and Yoshiyasu Takefuji. "Functional-link net computing: system architecture, and functionalities." Computer 25.5 (1992): 76-79.

[28] Samanta, Sourav, et al. "Haralick features based automated glaucoma classification using back propagation neural network." Proceedings of the 3rd International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA) 2014. Springer, Cham, 2015.

[29] Igelthuk, Boris, and Yoh-Han Pao. "Stochastic choice of basis functions in adaptive function approximators and the functional-link net." IEEE transactions on Neural Networks 6.6 (1995): 1320-1329.
[30] Huang, Guang-Bin, Yan-Qiu Chen, and Haroon A. Babri. "Classification ability of single hidden layer feedforward neural networks." *IEEE Transactions on Neural Networks* 11.3 (2000): 799-801.

[31] LeCun, Yann, et al. "Handwritten digit recognition with a back-propagation network." *Advances in neural information processing systems* (1989).

[32] Denker, John S., et al. "Neural network recognizer for handwritten zip code digits." *Advances in neural information processing systems*. 1989.

[33] Shen, Lu, et al. "Multiple empirical kernel mapping based broad learning system for classification of Parkinson's disease with transcranial sonography." 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2018.

[34] Pao, Yoh-Han, Gwang-Hoon Park, and Dejan J. Sobajic. "Learning and generalization characteristics of the random vector functional-link net." *Neurocomputing* 6.2 (1994): 163-180.

[35] Chen, CL Philip, and Chun-Yang Zhang. "Data-intensive applications, challenges, techniques and technologies: A survey on Big Data." *Information sciences* 275 (2014): 314-347.

[36] Chen, CL Philip, and Zhulin Liu. "Broad learning system: An effective and efficient incremental learning system without the need for deep architecture." *IEEE transactions on neural networks and learning systems* 29.1 (2017): 10-24.

[37] Lu, Huijuan, et al. "A kernel extreme learning machine algorithm based on improved particle swarm optimization." *Memenetic Computing* 9.2 (2017): 121-128.

[38] Ayyad, Hossein, Abbas Khakhar Manshad, and Habib Rostami. "Application of hybrid neural particle swarm optimization algorithm for prediction of MMP." *FUEL* 116 (2014): 625-633.

[39] Mansour, Iman Ben, Ines Aalaya, and Moncef Tagina. "A gradual weight-based ant colony approach for solving the multiobjective multidimensional knapsack problem." *Evolutionary Intelligence* 12.2 (2019): 253-272.

[40] Zhang, Hong, et al. "Developing a novel artificial intelligence model to estimate the capital cost of mining projects using deep neural network-based ant colony optimization algorithm." *Resources Policy* 66 (2020): 101604.

[41] Mohapatra, P., Sreejit Chakravarty, and Pradipeta K. Dash. "An improved cuckoo search based extreme learning machine for medical data classification." *Swarm and Evolutionary Computation* 24 (2015): 25-49.

[42] Yamany, Waleed, et al. "Moth-flame optimization for training multi-layer perceptions." 2015 11th International computer engineering Conference (ICENCICO). IEEE, 2015.

[43] Majhi, Santosh Kumar. "How effective is the moth-flame optimization in diabetes data classification." *Recent Developments in Machine Learning and Data Analytics*. Springer, Singapore, 2019. 79-87.

[44] Chang, Yu-Tzu, et al. "Optimization the initial weights of artificial neural networks via genetic algorithm applied to hip bone fracture prediction." *Advances in Fuzzy Systems* 2012 (2012).

[45] Li, Hongjiang, et al. "Genetic algorithm for the optimization of features and neural networks in ECG signals classification." *Scientific reports* 7.1 (2017): 1-12.

[46] Albadr, Musafat Abbas Abbool, et al. "Spoken language identification based on optimised genetic algorithm–extreme learning machine approach." *International Journal of Speech Technology* 22.3 (2019): 711-727.

[47] Wang, Gai-Ge, Suash Deb, and Zhihua Cui. "Monarch butterfly optimization." *Neural computing and applications* 31.7 (2019): 1995-2014.

[48] Feng, Yanhong, et al. "Monarch butterfly optimization: A comprehensive review." *Expert Systems with Applications* (2020): 114418.

[49] Kira, Kenji, and Larry A. Rendell. "A practical approach to feature selection." *Machine learning proceedings* 1992. Morgan Kaufmann, 1992. 249-256.

[50] Cai, Hongmin, et al. "Feature weight estimation for gene selection: a local hyperlinear learning approach." *BMC bioinformatics* 15.1 (2014): 1-13.

[51] T.R. Golub, D.K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J.P. Mesirov, H. Coller, M. L. Loh, J. R. Downing, M. A. Caligiuri, et al.,"Molecular classification of cancer: class discovery and class prediction by gene expression monitoring," *Science*286(5439)(1999)531–537.

[52] U. Alon, N. Barkai, D.A. Nottermann, K. Gish, S. Ybarra, D. Mack, A.J. Levine, Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays, Proc. Natl. Acad. Sci.96(12) (1999)6745–6750.

[53] E.F. Petricoin III, A.M. Ardekani, B.A. Hitt, P.J. Levine, V.A. Fusaro, S.M. Steinberg, G.B. Mills, C. Simone, D.A. Fishman, E.C. Kohn, L. LA, Use of proteomic patterns in serum to identify ovarian cancer, Lancet 359 (9306) (2002) 572–577.

[54] Z. Zhu, Y.-S. Ong, M. Dash, Markov blanket-embedded genetic algorithm for gene selection, Pattern Recognit.40(11) (2007)3236–3248.

[55] Wang, Aiguo, et al. "Incremental wrapper based gene selection with Markov blanket." 2014 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE, 2014.

[56] V. Bolon-Canedo, N. Sánchez-Marocho, A. Alonso-Betanzos, Distributed feature selection: An application to microarray data classification, Appl. Soft Comput. 30(2015)136–150.

[57] Chinnaswamy, Arunkumar, and Ramakrishnan Srinivasan. "Hybrid feature selection using correlation coefficient and particle swarm optimization on microarray gene expression data." *Innovations in bio-inspired computing and applications*. Springer, Cham, 2016. 229-239.

[58] J. García-Nieto, E. Alba, Parallel multi-swarm optimizer for gene selection in DNA microarrays, Appl.Intell.37(2) (2012)255–266.

[59] Apolloni, Javier, Guillermo Leguizamón, and Enrique Alba. "Two hybrid wrapper-filter feature selection algorithms applied to high-dimensional microarray experiments." *Applied Soft Computing* 38 (2016): 922-932.

POURNAMASI PARHI, e-mail: pournamasi.parhi@gmail.com, Mobile No. +91 8249844227, ORCID iD: 0000-0002-6486-4911.

Pournamasi Parhi completed her Masters in Computer Application (M.C. A) degree from Biju Patnaik University of Technology (BPUT), India in 2008 and received the M.Tech degree in Computer Science from Siksha ‘O’ Anusandhan deemed to be University, India in 2012. She is currently pursuing her Ph.D. degree with the Department of Computer Science at Siksha ‘O’ Anusandhan deemed to be University Bhubaneswar, India. She has published two research articles in IEEE conferences. Her main research interests are in the areas of multi-objective optimization, evolutionary computing, medical data mining, and machine learning.
Dr. Ranjeeta Bisoi, e-mail: ranjeeta.bisoi@gmail.com, ORCID iD: 0000-0002-2651-2130.

Dr. Ranjeeta Bisoi received the M.C.A. (Masters in Computer Application) degree from North Orissa University, India in 2011, and the Ph. D. degree in Computer Engineering from Siksha “O” Anusandhan University, Odisha, India in 2015. She is currently working as a Research Assistant Professor in the Multidisciplinary Research Cell of Siksha “O” Anusandhan University, India. She has published 36 papers in international journals like Elsevier Science, Springer, Taylor-Francis, etc. and in IEEE conferences. Her research interests include soft computing, data mining, machine intelligence, renewable energy, and bioinformatics.

Pradipta Kishore Dash (Corresponding Author), e-mail: pkdash.india@gmail.com, Mobile No. +91-674-2727336, ORCID iD: 0000-0002-8950-7136.

Pradipta Kishore Dash (Life Senior Member, IEEE) received the B.E. and M.E. degrees in electrical engineering from the Indian Institute of Science, Bengaluru, India, in 1964, the Ph.D. degree in electrical engineering from the University of Calgary, Calgary, AB, Canada, in 1976, and the D.Sc. degree in electrical engineering from Utkal University, Bhubaneswar, India, with a focus on AI and soft computing applications in power engineering, in 2003. He is currently the Director of Research and Consultancy, Multidisciplinary Research Cell, Siksha ‘O’ Anusandhan University, Bhubaneswar, India. He has authored more than 500 research papers in reputed international journals and conferences in power systems, signal processing, renewable energy, computational intelligence, and data mining.