scientific reports

OPEN

A gene selection algorithm for microarray cancer classification using an improved particle swarm optimization

Arfan Ali Nagra[1], Ali Haider Khan[1], Muhammad Abubakar[1]✉, Muhammad Faheem[2], Adil Rasool[3]✉, Khalid Masood[1] & Muzammil Hussain[4]

Gene selection is an essential step for the classification of microarray cancer data. Gene expression cancer data (deoxyribonucleic acid microarray] facilitates in computing the robust and concurrent expression of various genes. Particle swarm optimization (PSO) requires simple operators and less number of parameters for tuning the model in gene selection. The selection of a prognostic gene with small redundancy is a great challenge for the researcher as there are a few complications in PSO based selection method. In this research, a new variant of PSO (Self-inertia weight adaptive PSO) has been proposed. In the proposed algorithm, SIW-APSO-ELM is explored to achieve gene selection prediction accuracies. This novel algorithm establishes a balance between the exploitation and exploration capabilities of the improved inertia weight adaptive particle swarm optimization. The self-inertia weight adaptive particle swarm optimization (SIW-APSO) algorithm is employed for solution explorations. Each particle in the SIW-APSO increases its position and velocity iteratively through an evolutionary process. The extreme learning machine (ELM) has been designed for the selection procedure. The proposed method has been employed to identify several genes in the cancer dataset. The classification algorithm contains ELM, K-centroid nearest neighbor, and support vector machine to attain high forecast accuracy as compared to the start-of-the-art methods on microarray cancer datasets that show the effectiveness of the proposed method.

Keywords Microarray cancer, Improved PSO, ELM, SVM, Evolutionary algorithms

The rapid growth of microarray knowledge in the previous decade has allowed researchers to analyze many genes instantaneously and find biological data for several determinations, particularly cancer classification. Microarray is used in clinics and research laboratories to understand the molecular procedures and effective treatment of many diseases. Cancer-related microarray genes have been determined through a gene selection algorithm, which is very useful in the biological cancer system. Several different microarray genes contain information about many disease samples. The number of genes has redundant data in other microarray data, so the current methods are not accurate for prediction[1]. In microarray datasets, the matrix is a challenging problem. Gene expression plays a vibrant role in this regard. Many computational methods are unsuccessful in recognizing a small subset of significant genes in microarray data, which eventually raises the challenge of microarray analysis. Apart from the proposed approach, alternatively, there is another solution that unifies multiple models into a uniform model for the evolution of cellular functionality phenomena[2].

An improved whale optimization method is proposed, incorporating a pooling mechanism and three efficient search strategies: migrating, preferential selection, and enriched encircling prey. An evaluation is conducted to assess the performance of enhanced whale optimization algorithm (E-WOA), which demonstrates that E-WOA surpasses the versions of WOA. Following the successful performance of E-WOA, a binary version called BE-WOA was developed to specifically identify effective features, notably in medical datasets. The validation of the binary whale optimization algorithm (BE-WOA) algorithm involves the use of medical ailments datasets and a comparison with the most advanced optimization algorithms. The evaluation is based on criteria such as fitness,

[1]Faculty of Computer Science, Lahore Garrison University, Lahore 54000, Pakistan. [2]School of Technology and Innovations, University of Vaasa, Vaasa, Finland. [3]Department of Computer, Bakhtar University Kabul, Kabul, Afghanistan. [4]Department of Computer Science, University of Central Punjab Pakistan, Lahore 54000, Pakistan. ✉email: abubakarqazi@lgu.edu.pk; adilrasool@bakhtar.edu.af

accuracy, sensitivity, precision, and the number of features.[3]. Another optimization algorithm, the quantum-based avian navigation optimizer algorithm (QANA) proposes a self-adaptive quantum orientation and quantum-based navigation with mutation strategies. It proposes a fast communication topology to share the information flow through the flocks. Also, the introduced long-term and short-term memories provide meaningful knowledge for partial landscape analysis, and the qubit-crossover operator produces the next generation. The experimental results prove that QANA is effective and scalable on test functions and real-world applications such as applied engineering problems[4].

Different gene selection methods are categorized as filter, wrapper and embedded methods[5]. The filter methods exploit a gene's general properties and estimate every gene without the classification algorithm. But in the wrapper method, a pool of genes is created, and with the help of classification algorithms, we can attain the accurate gene pool sunset for solving the classification problems. Embedded methods achieved the properties of wrapper and filter methods. However many embedded methods plan with genes one by one, which takes a long time for microarray data[6]. Evolutionary methods are more appropriate and refined than wrapper gene selection because of their capability to refine solutions on complex spaces of potential solutions. These days, a lot of approaches use hybrid techniques such as genetic algorithm (GA)[7], PSO[8,9] have been used progressively and proved to be good for microarray datasets[10–16]. A binary PSO hybrid with a filter method is used to attain accurate gene subsets[17]. Genetic algorithm techniques have been used to improve the classification accuracy[13,14]. The hybrid method of genetic algorithm (GA) and k- nearest neighbors (KNN) identifies the different classes of genes[15]. The deep learning method microarray optimized gene selection method for cancer classification has been introduced in[16]. Swarm intelligence algorithms in gene selection profiles based on the classification of microarray data were introduced[17]. In[18], authors propose the Hybrid leader selection strategy for many objective particle swarm optimization to encounter the Mult objective problem effectively.

In terms of identifying both breast cancer subtypes and other cancer types, the IMS and FES employing the core gene set outperformed the other techniques. Additionally, the IMS may be repeated even with different gene expression data (i.e. RNA-seq and microarray)[19]. The study, conceptual discussion, and demonstration of the algorithmic behaviors—including convergence trends, mutation and crossover rate variations, and running time—showed that they were consistent with the results of the literature[20]. Microarrays are motionless and one of the major methods employed to study cancer biology[21]. The effectiveness of minimum redundancy maximum relevance with imperialist competition algorithm (mRMR-ICA) is evaluated using ten benchmark microarray gene expression datasets. Comparing mRMR-ICA to the original ICA and other evolutionary algorithms, experimental results show an improvement in the precision of cancer classification and the number of useful genes[22]. The gene set that deep gene selection (DGS) chose has demonstrated greater effectiveness in the categorization of cancer. The number of genes in the original microarray datasets can be significantly decreased using DGS[23]. Recognizing vigorous prognostic biomarkers to stratify colorectal cancer (CRC) has been proposed in[24]. In DNA microarray technology, gene expression data have a high dimension with a small sample size. Consequently, the growth of well-organized and healthy feature selection approaches is crucial that classify a minor set of genes to attain better classification performance[25]. An advanced bio-inspired multi-objective procedure is suggested for gene selection in microarray data classification specifically in the binary domain of feature selection[26].

This manuscript proposes a gene selection method based on a new variant of PSO. Microarray facilitates researchers to measure the expression levels of several genes concurrently. For the classification of microarray data gene selection provides a vital role. Microarray facilitates to determination of the expression ranks of numerous genes concurrently. The selection of a prognostic gene with a small redundancy is a great challenge for the researcher. A PSO-based gene selection technique has limitations which is why a new method of PSO is proposed in this manuscript. In the proposed algorithm, self-inertia weight adaptive particle swarm optimization with extreme learning machine (SIW-APSO-ELM) is utilized to achieve gene selection prediction accuracies. The extreme learning machine (ELM) had prearranged for selection procedure. The proposed method with the few discriminative genes as selected by the algorithm, attains high forecast accuracy on microarray datasets and shows the usefulness of the proposed method. The primary contribution of the aforementioned method is to achieve an appropriate balance between the exploration capabilities of the increased inertia weight adaptive particle swarm optimization and the exploitation. The SIW-APSO algorithm is employed for solution exploration. The SIW-APSO conducts an evolutionary process where each particle iteratively enhances its velocities and positions. The extreme learning machine (ELM) has been designed for the selection procedure. The proposed method has been used to identify several genes in the cancer dataset. The classification algorithm contains ELM, K- centroid nearest neighbor (KCNN), and support vector machine (SVM) to attain the high forecast accuracy as compared to the start-of-the-art methods on microarray cancer datasets.

The article's arrangement is as follows: Related work has been presented in section "Related work", which contains a study of ELM, SVM, and KCNN. Section "Proposed method" provides a methodology, which describes the development of the proposed improved PSO. Section "Experimental results and discussion" introduces the experimental work and the results and discussion, which includes comparative analysis studies. Section "Conclusions" concludes the article by providing its findings.

Related Work

Extreme learning machine

The gradient-based learning algorithms have faced different problems. For hidden feedforward neural network's extreme learning algorithm has been used[19]. This algorithm selects the input weights arbitrarily; the hidden layer logically controls the output weights of SLFN. This algorithm significantly improved the enactment of a faster learning speed than gradient-based algorithms[20,21]. ELM has minor training bugs and minimal weights. The almost negligible weight in ELM shows the best performance. The parameters of SLFN are not tuned to obtain the analytical method. ELM converges very fast as compared to gradient-based algorithms.

Linear support vector machine

Support vector machine (SVM) is a machine learning algorithm[22]. The hypersphere around the part of feeding data is calculating the class problem which is handled by one class SVM. Inside the hypersphere are known as specific data points, and outside the hypersphere are called anomalous data points. For the one class SVM, γ and ν are the hyper-parameters, wherever $\gamma = \frac{1}{2\sigma^2}$ is used for the variance of the input. For the input data, variance is calculated for γ. A lower γ value maps input data points near to each other. A higher ν value accepts many outliers, whereas a low value shows data points outside the hyperspace. One class SVM provides the best result and the linear kernel achieves high accuracies[23,24].

K-nearest centroid neighbor (KNCN) Classifier

KNCN is a very popular classifier[25]. It has been observed that this classifier is a valuable approach in the finite sample size. Let $TS = \{\mathrm{p_i} \in \mathbb{R}^{\mathrm{d}}\}_{\mathrm{i=1}}^{\mathrm{N}}$ be a training sample and $c_i \in \{\mathrm{c_1, c_2, \ldots, c_M}\}$ means the class label for $\mathrm{p_i}$ with M number of classes. The centroid of an assumed set of points $Z = (z1, z2, \ldots zq)$ can be calculated as:

$$Z_q^c = \frac{1}{q}\sum_{i=0}^{q} Z_i \tag{1}$$

For a known query sample x, its unidentified class c can be projected by KNCN using the steps below:

1. With the help of nearest centroid neighbor (NCN) search the nearest neighbor of X from Ts.

$$TR_k^{NCN}(x) = \{p_{ij}^{NCN} \in \mathbb{R}^{\mathrm{d}}\}_{j=1}^{k} \tag{2}$$

2. Assign x to the class c, which the centroid neighbors in the set most frequently represent $TR_k^{NCN}(\mathrm{x})$.

$$\underset{c_j}{arg\max} \sum_{p_{ij}^{NCN} \in TR_k^{NCN}(x)} \delta(c_{jc=}c_n^{NCN}) \tag{3}$$

where c_n^{NCN} is the centroid neighbor p_{in}^{NCN}, and $\delta\left(cj = c_n^{NCN}\right)$.

Proposed method

Hybrid gene selection algorithm

Gene selection is typically a two-step procedure. One must recognize related genes and be alike to choose the minor subsets from the related genes. SIWAPSO-ELM has been used to simplify the gene selection procedure. It has been used to select the smallest gene subset pool from the primary gene pool. An improved inertia weight PSO for the updating global combined with extreme learning has been proposed for the best selection of dense gene subsets from the accurate gene pool. The steps in the proposed method are shown in Fig. 1.

In the first step, the gene, and the dataset are separated into training and testing sets. Furthermore, in the first phase 150 to 300 genes were selected from the total dataset. Create a second-level initial gene pool. After training and testing 5–5-fold cross-validation is applied to each gene subset. After that, every gene subset perdition is determined with the help of fivefold cross-validation and with the help of ELM on the training sets classification accuracy obtained. An improved inertia weight PSO has been used to select the optimal gene after the validation pool. The *ith* particle $Xi = (xi1, xi2, ..., xiD)$

Signifies an applicant gene subset. The measurement of the particles corresponds to the number of certain genes selected from the original gene pool. The Extreme Learning Machine (ELM) achieves higher accuracy when trained on specific gene subsets indicated by the *ith* particle. The proposed SIWAPSO is very accurate as compared to the traditional PSO.

It is far more effective for updating the best swarm position. The proposed method is also beneficial to decrease premature convergence of the swarm; SIW-APSO is implemented as an optimizer to revise the optimal location of the swarm. In the $(i+1)th$ generation, the velocity and position of the swarm, Pg, are updated by Eqs. (6 and 7).

Fitness function

The value of X is assessed by a fitness function f. Fitness is determined by the genetic aptitude acquired within this subset. The fitness value is computed in the following manner:

$$\boldsymbol{f(X) = w_1 * classifier(X)(1 - w_1) * (1 - \frac{No.genes}{p})} \tag{4}$$

wherever $\boldsymbol{w_1}$ is a weight coefficient [0, 1] which handles the combination of both intentions, the number of genes the number of selected genes in X, and p the dimension of the vector X. The additional period promises the detail of favoring subsets with less number of genes. The goal is to exploit the function f. As the main objective is to find a subset of genes with maximum accuracy, we choose $\boldsymbol{w_1} = 0.95$. Where accuracy is calculated as follows:

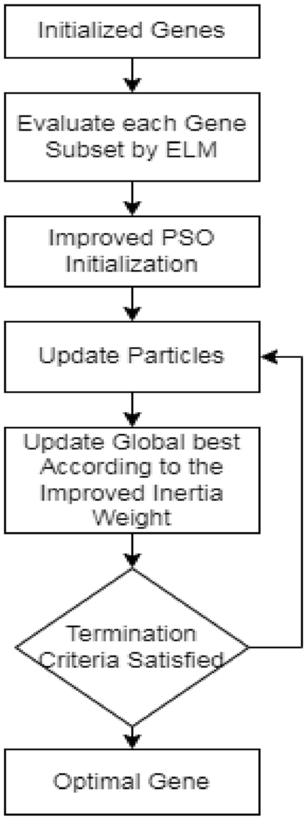

Figure 1. The framework of the proposed hybrid gene selection method.

$$accuracy = \frac{No.of\,correctlypredictedsamples}{No.of\,samples} \tag{5}$$

An improved inertia weight PSO for the updating global combined with extreme learning has been proposed for best selecting compact gene subsets from the refined gene pool. The pseudocode of a proposed method is as follows:

Step 1: In the first step, training and testing have been performed on the datasets.
And from the total datasets, 150 to 300 genes were selected for testing and validation.
Step 2: Determine the pbest Pb and gbest Pg.
Step 3: Find the fitness function of every particle.
Step 4: with the help of the fitness function, adjust the best Pb and gbest Pg.
Step 5: Update the position of all particles accordingly.
Step 6: Evaluate the fitness function and calculate accuracy with classifiers.
Step 7: The process is recurrent till the aim is encountered or the maximum optimization is achieved, or else, go to Step 3.

Self-inertia weight adaptive particle swarm optimization (SIW-APSO)

Since the evaluation of particle swarm optimization (PSO) numerous researchers tried to explore the solution for the precise value of inertia weights. A self-adaptive inertia weight PSO is proposed for the optimization of the exploration and exploitation of the particles. SIW helps to improve the premature convergence in PSO. The proposed method is able to improve the premature convergence of the PSO. It is vital to find the position of the particle in the population for all set of given epochs. When using this approach, it is critical to determine the particle's position in the population for all iterations. SIW-APSO initializes the random particles in the multidimensional search. A predetermine fitness function helps to evaluate the relative fitness function and for each iteration every particle in a randomized swarm indicated a point in feature space. However, SIW managed to attain the best solution. The Eqs. (4 and 5) determine the updates of velocity and position for each particle that moves around the dimensional search space. The Eqs. (6 and 7) provide the information about the iterative adaptive inertia weight as a function of iteration number, for each iteration the inertia weights update itself accordance to the improvement in the best fitness.

SIW-APSO improves the precocious convergence issue of PSO. The self-inertia weight is an adapted iteration of the PSO method. In this method, a random swarm of particles initializes. Every particle corresponds to a certain location inside the search space. A fitness function is derived from the pre-evaluated process, which greatly aids in achieving the desired optimal solution. The particles navigate inside a multidimensional solution space as a collective.

$$S_{id}^{t+1} = W_i^{t+1} S_{id}^t + k_1 r_1 \left(\frac{(\beta f \left(Z_i^{t+1} \right) - (\beta f \left(Z_i^t \right)}{(\beta f \left(Z_i^{t+1} \right)} \right) + k_2 r_2 (p_g^t - Z_i^t) \tag{6}$$

$$Z_{id}^{t+1} = S_{id}^{t+1} + Z_{id}^t \tag{7}$$

where Z_i^d is the position of particle i,W_i^{t+1} is the inertia weight of ith at the iteration (t + 1)th, k_1&k_2 are the constants, r_1&r_2 the random functions and "p_b^d" is the dth best position particle i,p_g^d is the dth global best position founded by the whole swarm[8]?

According to the Eq. (6), The repetitive nature of inertia weight is highly beneficial for determining actual fitness.

$$W_i^{t+1} = 0.9 if t = 0 \tag{8}$$

$$W_i^{t+1} = f\left(Z_i^t\right) - f(Z_i^{t-1}) if > 0 \tag{9}$$

where W_i^{t+1} and $f\left(Z_i^t\right)$ represent the inertia weight and best fitness value at the t^{th} and $t+1^{th}$iteration, respectively. $\frac{(\beta f \left(Z_i^{t+1} \right) - (\beta f \left(Z_i^t \right)}{(\beta f \left(Z_i^{t+1} \right)}$ Illustrate the improvement in the optimal fitness function that relies on the suggested inertia weight. The Eq. (6) shows the $f(Z_i^t)$ and $f(Z_i^{t-1} The)$, fitness function of the swarms at the t^{th} and $t-1^{th}$ repetitions, correspondingly. If $W_i^{t+1} = W_i^t$, the particle is unable to find an optimal solution due to its limitations as $f(Z_i^{t-1}) = f\left(Z_i^t\right)$ and there is no change in inertia weight. In the case of $t > 0$, as we can use in the above Eq. (6), then The variability in inertia weight is advantageous, leading to an improvement in fitness. The fitness function improves when the most optimal position is close to the other particles in the swarm.W_i^t Will grow to W_i^{t+1} at $t-1^{th}$ loop, it is advantageous for both exploration and exploitation. When $W_i^{t+1} = 0.9$ as described in Eq. (6) The inertia weight remains constant and functions similarly to a typical Particle Swarm Optimization (PSO). The oscillations exhibited a decline as the iterations progressed, an outcome of permissible modifications in the intended solution. SIW-APSO is capable of effectively augmenting the precision of PSO. The study reveals that the inertia weight becomes zero at certain phases and remains constant for multiple consecutive iterations, indicating that the swarm becomes stuck and unable to escape a local optimum. To adjust the range of inertia weight a linear function is used:

$$W(t) = W_l - (W_l - W_h)t/t_h \tag{10}$$

According to Eq. (6) The inertia weight for each swarm is modified autonomously at every iteration, utilizing the improvement in its own best fitness.[25]. An indication of a change in fitness is present. When the individual fitness of a particle improves at any iteration. A particle undergoes a direction change; otherwise, its moment of inertia remains constant at zero. The particle begins its search locally and continues its global exploration until it acquires additional inertia and an improvement in inertia. According to the mentioned method, balancing global and local research improves diversity.

Experimental results and discussion
Datasets
In this manuscript, eight microarray datasets are utilized for various types of cancer: Leukemia, Brain Cancer, Colon Cancer, SRBCT (Small Round Blue Cell Tumors), Lung Cancer, brain cancer, Lymphoma, 11_Tumors, and Diffuse large B cell lymphoma (DLBCL). These datasets are publicly available and were sourced from the GEMS system (http://www.gems-system.org). which are listed in Table 2. The comprehensive explanation of the datasets is recorded in Table 2. Table 1 has defined the parameter settings. The selection of these datasets includes a diverse range of cancers, which helps in demonstrating the generalizability and robustness of the gene selection technique across different types of cancer. Each of these cancers has distinct gene expression profiles, making them ideal for evaluating the effectiveness of gene selection methods in identifying relevant biomarkers.

Parameters	Value
Swarm size	30
Number of generations	100
C1, C2	2
W	0.9
V_{max}	4
v_{min}	– 4

Table 1. Parameters used for the proposed method.

Data	Total samples	Training samples	Testing samples	Number of classes	Number of genes
Leukemia	72	38	34	2	7129
Brain Cancer	60	30	30	2	7129
Colon Cancer	62	40	22	2	2000
SRBCT	83	63	20	4	2308
Lung Cancer	203	103	100	5	3312
Lymphoma	58	29	29	2	7129
11_Tumors	73	45	41	3	4200
DLBCL	67	42	39	2	3812

Table 2. Eight microarray dataset genes.

fivefold cross-validation is used in this manuscript which is a good compromise between the bias and variance of the model. It helps in ensuring that the results are not overly optimistic (low variance) while still providing a sufficient number of training samples to maintain generalizability (low bias). Compared to leave-one-out cross-validation or higher k-fold values, fivefold cross-validation is computationally less intensive, making it suitable for large microarray datasets. fivefold cross-validation is a widely accepted method in bioinformatics and machine learning literature, providing a balance between reliability and computational demand.

The parameter tuning process for SIW-APSO involves setting and optimizing the values of key parameters to achieve optimal performance of the algorithm. Here is a detailed discussion of the parameter tuning process, how the parameters were optimized, and their impact on the performance of SIW-APSO which is shown and highlighted blue in Table 1. A swarm size of 30 is a common choice, balancing diversity and computational efficiency. For iteration, setting it to 100 aims to balance solution quality and computational time. C1 and C2 values are typically set to the same value (2) to balance exploration (C1) and exploitation (C2). If C1 is much larger than C2, the particles may explore too much without converging. Conversely, if C2 is much larger than C1, the particles may converge prematurely without exploring the search space adequately. An inertia weight of 0.9 favors exploration in the initial stages of the optimization process. Properly tuned velocity limits Vmax is 4 and Vmin is − 4 ensure that particles explore the search space effectively without oscillating too much.

The experimental study has been conducted on eight microarray datasets, including SRBCT, Lung cancer, Colon cancer, brain cancer, Leukemia, and Lymphoma, shown in Table 2.

The proposed method of self-inertia weight PSO with ELM is to classify the eight Microarray datasets to verify selected gene subsets. Every experiment is made 500 times and means accuracies and standard deviation are listed in Table 3.

The prediction ability of the selected gene subsets

The predictive accuracy of certain groups of genes is confirmed using the proposed SIW-APSO-ELM method. The experiments have been conducted 500 times, Std., and the accuracy is shown in Table 3. The proposed method has obtained high accuracy for brain cancer. The proposed method obtains 100%. SIW-APSO-ELM method on the colon obtains excellent accuracies. These results indicate that the SIW-APSO-ELM can select highly valuable prognostic genes. Table 4 shows the best accuracy of the proposed method SIW-APSO-ELM> using the 5-fold validation on the eight cancer datasets.

In Table 5. KCNN classification accuracy with different parameters is given. We achieved 97% on Colon and 100% on SRBCT which shows good results in terms of parameters used. We achieved 100% accuracy on Lymphoma while it was slightly reduced in 11_Tumors which is 97% approximately.

In Table 6. SVM classifier shows the classification accuracy with different parameters. We achieved 97% on Colon and 100% on SRBCT which shows good results when we applied BPSO-GCSELM. We achieved 100% accuracy on Lymphoma while it was slightly reduced in 11_Tumors which is 99% approximately when we applied SIW-APSOELM.

In Table 7. Different datasets provided different values and it shows the accuracy, sensitivity, and specificity with different parameters. We achieved 96%, 93%, and 92% on Colon accuracy sensitivity and specificity.

In Table 8. A comparison of the proposed method with a modified version of the Moth Flame algorithm is given. It shows Lung accuracy of 94%using MMFA while 97% using the SIW-APSO-ELM approach.

In Table 9. The proposed method with selected genes on the Colon dataset is given that shows the description of each selected gene.

In Table 10. The proposed method with selected genes on the Leukemia dataset is given that shows the description of each selected gene.

In Table 11. The proposed method with selected genes on the Lymphoma dataset is given that shows the description of each selected gene.

In Table 12. The proposed method with selected genes on the SRBCT dataset is given that shows the description of each selected gene. It is linked to sample classes. Table 4 presents the experimental results with the latest gene selection methods. It demonstrates that the proposed technique has outclassed the other PSO[26] variants and other standard gene selection approaches such as IBPSO, SVM[27], IG-GA[28], EPSO[29], BPSO-GCS-ELM[30], mABC[31], because SIWAPSO-ELM has been used to simplify the genes selection procedure. It has been used to select the smallest gene subset pool from the primary gene pool which has been updating global position and

Selected gene subsets	Fivefold CV accuracy (%) ± std	Friedman test accuracy (%) ± std
SRBCT		
42,1003,1954,430,2050,123	100 ± 0.00	100 ± 0.00
545,1955,1434,509,971,255	100 ± 0.00	100 ± 0.00
1003,545,1911,153,123,1489,2161	100 ± 0.00	100 ± 0.00
1955,2050,545,2144,2045,123,1489	100 ± 0.00	100 ± 0.00
Leukemia		
4050,2642,2121	100 ± 0.00	100 ± 0.00
4050,2642,1882	100 ± 0.00	100 ± 0.00
4050,2642,3258	100 ± 0.00	100 ± 0.00
42,335,2642,1843,4050	100 ± 0.00	100 ± 0.00
Lymphoma		
152,2347,2650,5679,438,1855,5863	90.60 ± 0.023	86.11 ± 0.120
1855,2828,152,2437,806,530,1102	92.36 ± 0.027	89.33 ± 0.019
5279,4687,4940,5449,1133,1855,4519	93.51 ± 0.022	90.47 ± 0.029
152,2437,4829,2828,6441,806,2508	93.79 ± 0.020	90.45 ± 0.023
LUNG		
1765,2779,2841,1474,2045,3191,2763,2817,525,1630	98.27 ± 0.014	93.33 ± 0.011
525,1493,607,2763,792,580,867,368,3279,2158,1225	98.39 ± 0.023	93.47 ± 0.012
1765,883,2763,792,580,867,985,3279,2988,2045,814	98.67 ± 0.021	93.60 ± 0.019
1765,525,2763,2841,1474,2583,867,985,2045,814,918	98.67 ± 0.019	94.01 ± 0.024
Brain cancer		
1091,798,337	90.14 ± 0.036	89.62 ± 0.025
3052,973,3041,3692,4796	92.00 ± 0.023	91.78 ± 0.046
4628,7129,7045,4413,798	92.29 ± 0.020	90.22 ± 0.022
7129,2881,3052,865,1970,2935,4871	92.78 ± 0.012	91.88 ± 0.019
Colon		
1976,1325,1993,1870,1892,653,1917,187,22,1209,1060	93.63 ± 0.025	97.27 ± 0.013
377,792,14,1976,765,187,251,1110,175,53,1293,1740,200	93.00 ± 0.035	98.06 ± 0.013
792,1423,14,1976,1909,1110,1589,102,107,1916,175,1151	93.73 ± 0.031	98.71 ± 0.013
792,14,1976,765,1909,1524,1110,175,43,53,1293,1740,251	96.86 ± 0.033	99.05 ± 0.011
11_Tumors		
2131,1291,1798,1337	95.63 ± 0.025	95.27 ± 0.012
3052,973,3041,3692,4796	95.00 ± 0.035	97.66 ± 0.015
4628,7129,7045,4413,798	95.73 ± 0.031	95.71 ± 0.016
7129,2881,3052,865,1970,2935,4871	95.86 ± 0.033	94.05 ± 0.004
DLBCL		
1291,798,437,2345	100 ± 0.00	100 ± 0.00
2052,2973,3041,3692,4796	100 ± 0.00	100 ± 0.00
4628,7129,7245,7413,8798	100 ± 0.00	100 ± 0.00
7139,4831,3052,865,1970,3932,5871	100 ± 0.00	100 ± 0.00
SRBCT		
42,1003,1954,430,2050,123	100 ± 0.00	100 ± 0.00
545,1955,1434,509,971,255	100 ± 0.00	100 ± 0.00
1003,545,1911,153,123,1489,2161	100 ± 0.00	100 ± 0.00
1955,2050,545,2144,2045,123,1489	100 ± 0.00	100 ± 0.00
Leukemia		
4050,2642,2121	100 ± 0.00	100 ± 0.00
4050,2642,1882	100 ± 0.00	100 ± 0.00
4050,2642,3258	100 ± 0.00	100 ± 0.00
42,335,2642,1843,4050	100 ± 0.00	100 ± 0.00
Lymphoma		
152,2347,2650,5679,438,1855,5863	90.60 ± 0.023	85.11 ± 0.020
1855,2828,152,2437,806,530,1102	92.36 ± 0.027	89.33 ± 0.019
5279,4687,4940,5449,1133,1855,4519	93.51 ± 0.022	90.47 ± 0.029
152,2437,4829,2828,6441,806,2508	93.79 ± 0.020	90.45 ± 0.023
LUNG		
Continued		

Selected gene subsets	Fivefold CV accuracy (%) ± std	Friedman test accuracy (%) ± std
1765,2779,2841,1474,2045,3191,2763,2817,525,1630	98.27 ± 0.014	93.33 ± 0.011
525,1493,607,2763,792,580,867,368,3279,2158,1225	98.39 ± 0.023	93.47 ± 0.012
1765,883,2763,792,580,867,985,3279,2988,2045,814	98.67 ± 0.021	93.60 ± 0.019
1765,525,2763,2841,1474,2583,867,985,2045,814,918	98.67 ± 0.019	94.01 ± 0.024
Brain cancer		
1091,798,337	90.14 ± 0.036	89.62 ± 0.025
3052,973,3041,3692,4796	92.00 ± 0.023	91.78 ± 0.046
4628,7129,7045,4413,798	92.29 ± 0.020	90.22 ± 0.022
7129,2881,3052,865,1970,2935,4871	92.78 ± 0.012	91.88 ± 0.019
Colon		
1976,1325,1993,1870,1892,653,1917,187,22,1209,1060	93.63 ± 0.025	97.27 ± 0.013
377,792,14,1976,765,187,251,1110,175,53,1293,1740,200	93.00 ± 0.035	98.06 ± 0.013
792,1423,14,1976,1909,1110,1589,102,107,1916,175,1151	93.73 ± 0.031	98.71 ± 0.013
792,14,1976,765,1909,1524,1110,175,43,53,1293,1740,251	96.86 ± 0.033	99.05 ± 0.011
Tumors		
2131,1291,1798,1337	95.63 ± 0.025	96.27 ± 0.012
3052,973,3041,3692,4796	95.00 ± 0.035	96.06 ± 0.015
4628,7129,7045,4413,798	95.73 ± 0.031	96.71 ± 0.016
7129,2881,3052,865,1970,2935,4871	95.86 ± 0.033	95.05 ± 0.013
DLBCL		
1291,798,437,2345	100 ± 0.00	100 ± 0.00
2052,2973,3041,3692,4796	100 ± 0.00	100 ± 0.00
4628,7129,7245,7413,8798	100 ± 0.00	100 ± 0.00
7139,4831,3052,865,1970,3932,5871	100 ± 0.00	100 ± 0.00

Table 3. Classification accuracy using various gene subsets.

		IBPSO	IG_GA	EPSO	mABC	BPSO-GCS-ELM	SIW-APSO-ELM
Leukemia	Accuracy-Avg	98.34 ± 0.02	97.56 ± 0.32	95.11 ± 0.12	99.3 ± 0.01	100 ± 0.00	**100 ± 0.00**
Brain Cancer	Accuracy-Avg	77.34 ± 0.35	77.23 ± 0.45	74.65 ± 0.32	79.34 ± 0.03	88.3 ± 0.02	**90.21 ± 0.04**
Colon	Accuracy-Avg	93.45 ± 0.02	94.31 ±	93.23 ± 0.06	97.34 ± 0.09	97.38 ± 0.01	**98.45 ± 0.01**
SRBCT	Accuracy-Avg	**100 ± 0.00**	**100 ± 0.00**	**100 ± 0.00**	**100 ± 0.00**	**100 ± 0.00**	**100 ± 0.00**
Lung	Accuracy-Avg	95.86 ± 0.89	95.57 ± 0.23	95.67 ± 0.96	96.65 ± 0.56	94.78 ± 0.02	**97.62 ± 0.01**
Lymphoma	Accuracy-Avg	**100 ± 0.00**	**100 ± 0.00**	**100 ± 0.00**	**100 ± 0.00**	**100 ± 0.00**	**100 ± 0.00**
11_Tumors	Accuracy-Avg	95.06 ± 0.04	92.53 ± 0.07	95.92 ± 0.54	97.78 ± 0.23	97.88 ± 0.02	**99.87 ± 0.10**
DLBCL	Accuracy-Avg	**100 ± 0.00**	**100 ± 0.00**	**100 ± 0.00**	**100 ± 0.00**	**100 ± 0.00**	**100 ± 0.00**

Table 4. Comparison of the gene selection algorithms using particular datasets with fivefold validation. Bold values in Table 4 indicate the best results. (IBPSO = Improved Binary Particle Swarm Optimization, IG-GA = Information Gain Genetic Algorithm, EPSO = Enhance Binary Particle Swarm. Optimization, mABC = Artificial Bee Colony Algorithm, BPSO-GCS-ELM = Binary Particle Swarm Optimization-Gene to Class Sensitivity-Extreme Learning Machine, SIW-APSO-ELM = Self Inertia Weight Adaptive Particle Swarm Optimization-Extreme Learning Machine).

extreme learning for the best selection of dense gene subsets from the accurate gene pool. From Table 3, the SIW-APSO-ELM chooses a nearly similar number of genes as the former approaches on Leukemia, SRBCT, and DLBCL. At the same time, it determines the maximum number of genes in the colon, brain, and lung data, amongst others. ELM attains 100% accuracy on the Leukemia, DLBCL, and SRBCT datasets compared to other state-of-the-art methods.

From Table 3, the SIW-APSO-ELM chooses a nearly similar number of genes as the former approaches on Leukemia, SRBCT, and DLBCL. At the same time, it determines the maximum number of genes in the colon, brain, and lung data, amongst others. ELM attains 100% accuracy on the Leukemia, DLBCL, and SRBCT datasets compared to other state-of-the-art methods. Figure 2 shows the fivefold CV accuracy on the training data versus the iteration number of a proposed algorithm. With the help, we verified that it improves the premature convergence of the proposed method. The graph in Fig. 2. further shows that the tendency of improving convergence

		BPSO-GCS-ELM	SIW-APSO-ELM
Leukemia	Accuracy-Avg	100 ± 0.00	100 ± 0.00
Brain Cancer	Accuracy-Avg	88.30 ± 0.65	90.21 ± 0.23
Colon	Accuracy-Avg	97.38 ± 0.03	98.45 ±
SRBCT	Accuracy-Avg	100 ± 0.00	100 ± 0.00
Lung	Accuracy-Avg	94.78 ± 0.02	97.62 ±
Lymphoma	Accuracy-Avg	100 ± 0.00	100 ± 0.00
11_Tumors	Accuracy-Avg	97.88 ± 0.05	99.87 ± 0.04
DLBCL	Accuracy-Avg	100 ± 0.00	100 ± 0.00

Table 5. KCNN Classification Accuracy.

		BPSO-GCS-ELM	SIW-APSO-ELM
Leukemia	Accuracy-Avg	100 ± 0.00	100 ± 0.00
Brain Cancer	Accuracy-Avg	88.3 ± 0.65	90.21 ± 0.23
Colon	Accuracy-Avg	97.38 ± 0.01	98.45 ± 0.20
SRBCT	Accuracy-Avg	100 ± 0.00	100 ± 0.00
Lung	Accuracy-Avg	94.78 ± 0.04	97.62 ± 0.01
Lymphoma	Accuracy-Avg	100 ± 0.00	100 ± 0.00
11_Tumors	Accuracy-Avg	97.88 ± 0.01	99.87 ± 0.02
DLBCL	Accuracy-Avg	100 ± 0.00	100 ± 0.00

Table 6. SVM classification accuracy.

Number of selected genes	Dataset names	Accuracy (%)	Sensitivity (%)	Specificity (%)
7	SRBCT	100 ± 0.00	98.06 ± 0.09	100 ± 0.00
4	Leukemia	100 ± 0.00	97 ± 0.07	99 ± 0.13
7	Lymphoma	93.79 ± 0.020	90.23 ± 0.040	91.42 ± 0.10
11	LUNG	98.67 ± 0.019	95.34 ± 0.12	96.02 ± 0.11
6	Brain cancer	92.78 ± 0.012	94.16 ± 0.01	90.02 ± 0.11
13	Colon	96.86 ± 0.033	93.33 ± 0.30	92.12 ± 0.21
7	11_Tumors	95.86 ± 0.033	90.34 ± 0.20	91.16 ± 0.40
5	DLBCL	100 ± 0.00	97.56 ± 0.32	98.67 ± 0.03

Table 7. Classification sensitivity and specificity with different gene subsets.

		mMFA	SIW-APSO-ELM
Leukemia	Accuracy	99.6 ± 0.26	100 ± 0.00
SRBCT	Accuracy	98.0 ± 0.20	100 ± 0.00
Lung	Accuracy	94.78 ± 0.028	97.62 ± 0.011
DLBCL	Accuracy	100 ± 0.00	100 ± 0.00

Table 8. Comparison of the proposed algorithm with modified Moth Flame Algorithm.

rate for the four genes Brain Cancer, Lymphoma, 11_Tumors, and Colon is greatly impacted by the training examples and it improves over time as the model is trained.

Comparison with other classification models

The proposed approach has been evaluated against several state-of-the-art models, including IBPSO[32], SVM[33], IG_GA[34], EPSO[35], BPSO-GCS-ELM[36], mABC[37], and intelligent models[38]. This judgment was founded on the classification outcome and the number of genes irrespective of data dispensation and classification methods. The assessment outcomes on eight datasets are obtained in Table 4. From Table 3, the SIW-APSO-ELM method

No	Name	Description
65	H06524	Gelsolin precursor, plasma (human)
14	H20709	Sin alkali, smooth isoform (human)
1213	T92451	fibroblast and epithelial muscle
1976	K03474	Mullerian inhibiting substance gene,
179	T94579	Human chitotriosidase precursor mRNA

Table 9. The proposed method with selected genes on Colon datasets.

No	Name	Description
276	X62654	ME491 gene extracted from Sapiens gene antigen
1882	M27891	0.4549 CST3 Cystatin C
2288	M84526	DF D component
3985	X03934	GB DEF = T cell antigen T3-delta
1834	M23197	CD33 CD33 antigen

Table 10. Proposed method with selected genes on leukemia datasets.

No	Name	Description
1855	L17328_at	Fasciculation and elongation
2045	1641X	Cathepsin B; clone = 261,517
1467	1636X	Osteonectin = SPARC;
2356	263X	Similar to HuEMAP = (EMAP)
4940	U66559_at	Anaplastic lymphoma

Table 11. The proposed method with selected genes on the Lymphoma dataset.

No	Name	Description
1003	796,258	Sarcoglycan, alpha
1495	624,360	proteasome
1911	898,219	mesoderm-specific transcript (mouse) homolog
246	377,461	Caveolin 1, 22kD
1775	768,246	glucose-6

Table 12. The proposed method with selected genes on SRBCT dataset.

achieves 100% accuracy on the Leukemia, DLBCL, and SRBCT data with the genes selected by the other methods. Table 8 shows that the proposed algorithm performs much better as compared to the latest algorithm because the proposed algorithm has been used to simplify the gene selection procedure[39]. It has been used to select the smallest gene subset pool from the primary gene pool which has been updating global position and extreme learning for the best selection of dense gene subsets from the accurate gene pool.

Comparison SIW-APSO- ELM method with the KCNN and SVM

The proposed method achieved stunning performance with the other gene selection methods, such as BPSO-GCS-ELM on public microarray data. The SIW-APSO- ELM is matched with the BPSO-GCS-ELM process on the eight data employing ELM and KCNN consistent results are recorded in Tables 5 and 6. From Tables 5 and 6, KCNN and SVM were obtained with a fivefold CV Accuracy of 100% on the Leukemia and SRBCT data. The accuracy of the KCNN and SVM models is higher when using the genes identified by the SIW-APSO-ELM technique compared to the BPSO-GCS-ELM method for both colon cancer and lung cancer data. In the case of brain cancer, the SIW-APSO-ELM technique achieves more accuracy than the BPSO-GCS-ELM method for ELM. Similarly, for KCNN, the SIW-APSO-ELM method also yields higher accuracy compared to the BPSO-GCS-ELM method. The findings presented in Tables 4 and 5 demonstrate that the suggested method surpasses the BPSO-GCS-ELM method, as well as other PSO variants, in terms of performance. In the context of brain cancer, the SIW-APSO-ELM technique achieves more accuracy than the BPSO-GCS-ELM method in ELM. Similarly,

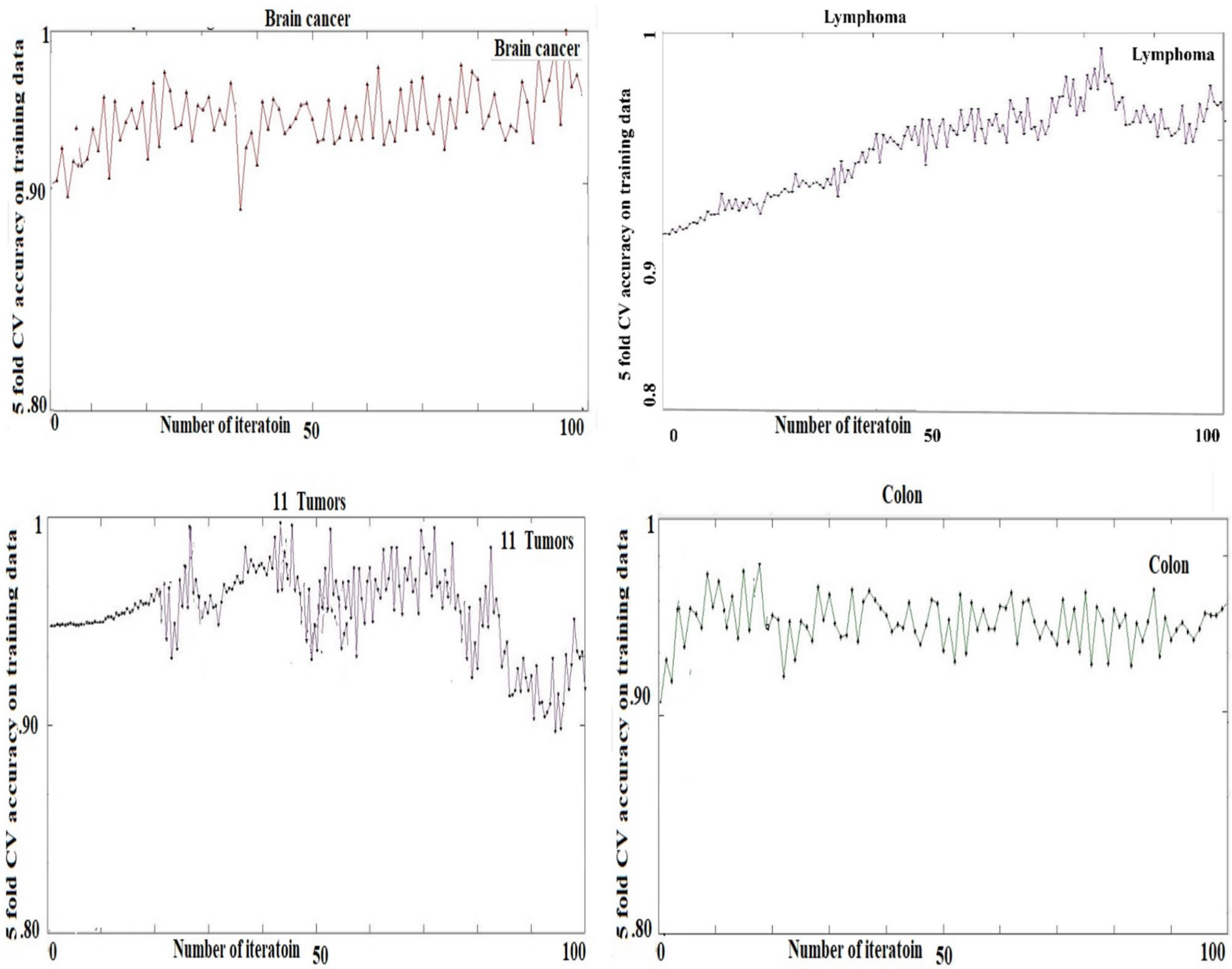

Figure 2. Fold CV accuracy on the training data versus the iteration number of a proposed algorithm.

in KCNN, the SIW-APSO-ELM method also outperforms the BPSO-GCS-ELM method in terms of accuracy. The BPSO-GCS-ELM technique chooses a smaller number of genes compared to the proposed method when applied to the Colon, Brain cancer, and Lung, including 11_Tumors and DLBC datasets.

Table 7 shows the accuracy, sensitivity, and specificity of the eight datasets. From Tables 5 and 6, we can observe that the proposed method performs an outstanding performance compared to the BPSO-GCS-ELM. From Fig. 3, the classification accuracy of Leukemia, DLBCL, and SRBCT achieved 100%. Figure 4 shows the ranks of genes from five independent runs on the two data sets, DLBCL and 11-Tumor, to assess the proposed method's reproducibility.

Figure 3 shows the classification accuracy for different datasets of the proposed method. It shows that some of the selected genes are better converged than others. However, it also shows that these variations are due to different parameters for different datasets that can impact the convergence rate of the selected genes.

Biological analysis of the selected gene subsets

Biological experiments on different data have been conducted. The highest five often selected genes have been listed in Tables 9, 10, 11 and 12. Table 12 shows the top five frequently selected genes with the proposed method on SRBCT dataset. For classification, gene X03934 is very critical which is listed in Table 10. From Table 9, the genes H06524, H20709, T94579, T92451, and K03474 were also selected. From Table 11, the gene U66559_at is also an essential gene for anaplastic lymphoma kinase. Figure 4. Shows the ranks of selected genes by using 5 independent runs on the two different datasets.

Conclusions

In this research, the gene subset classification is performed using a new variant of PSO along with an extreme learning machine (ELM). The algorithm is based on the SIWAPSO variant of swarm optimization to perform gene selection. The selection of prognostic genes with a small redundancy is a great challenge as the main drawback of a PSO-based gene selection method is its complexity. Gene selection is typically a two-step procedure. In the first step, data is initialized whereas in the second step, an improved inertia weight PSO is used to select the

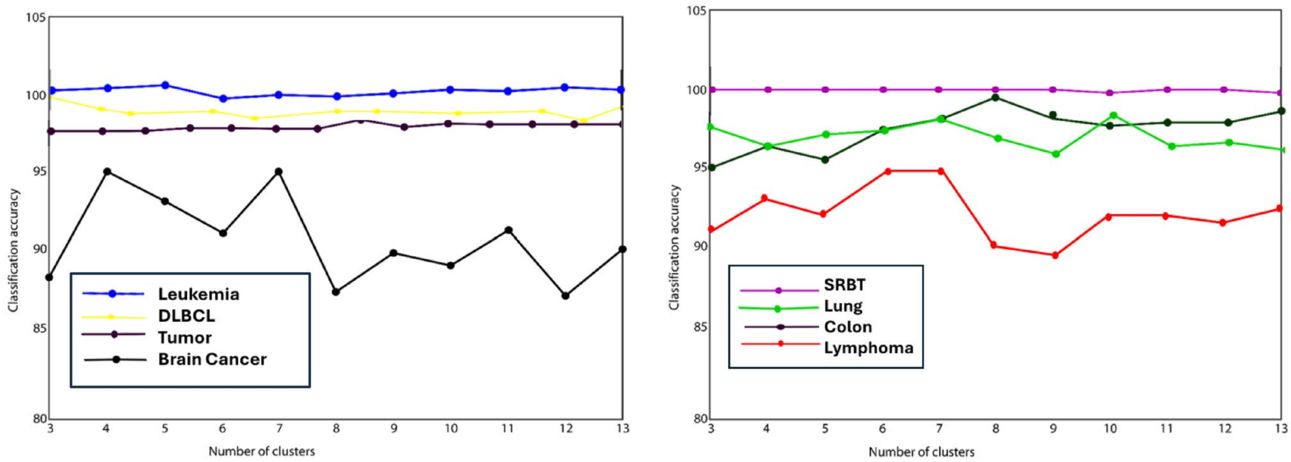

Figure 3. Classification Accuracy of proposed datasets (**a**) leukemia, DLBCL, tumor, brain cancer (**b**) SRBT, lung, colon, lymphoma.

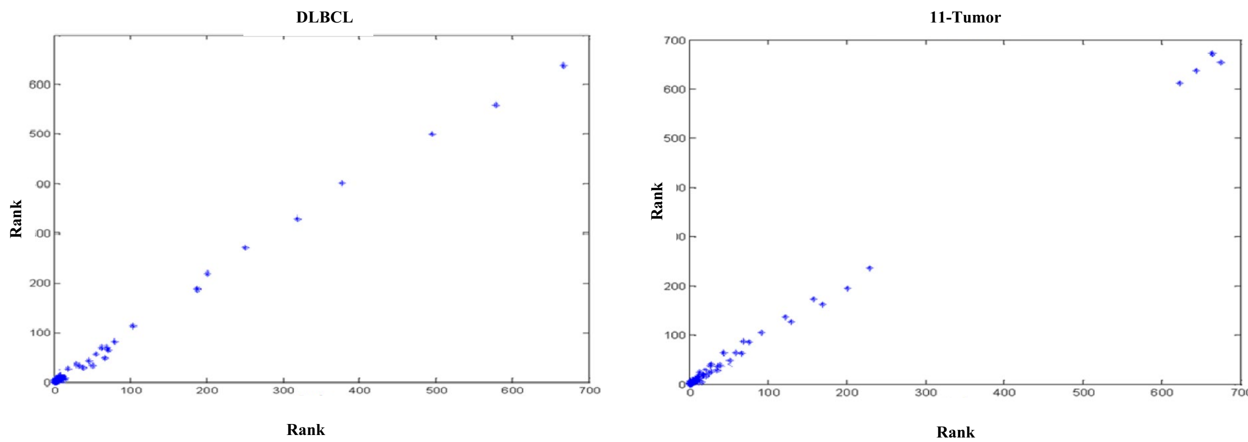

Figure 4. Ranks of genes from five independent runs on the two data sets.

optimal gene. In the proposed method, SIW-APSO-ELM is used to achieve the gene selection task. The proposed technique is more accurate in comparison to the traditional PSO. It updates the optimum position of the swarm. It is also an efficient technique as it decreases premature convergence of the swarm. The proposed method using the modified swarm optimization performs better than its counterparts with minimal error rates. The main algorithm contains an adaptive swarm optimization and an extreme machine learning method, which are more accurate and less complex than current state of the art methods. In the future, the ensemble learning and shallow deep learning frameworks with a few hidden layers may be employed to perform effective gene selection to reduce the processing time. The dataset may also be expanded to test the robustness of the proposed algorithm.

Data availability

All data supporting this study are openly available within the paper and online as open source, without restrictions".

Received: 3 May 2024; Accepted: 26 July 2024
Published online: 23 August 2024

References

1. Alba, E., García-Nieto, J., Jourdan, L. & Talbi, E.-G. Gene selection in cancer classification using PSO/SVM and GA/SVM hybrid algorithms. In *Evolutionary Computation, 2007. CEC 2007. IEEE Congress on, IEEE* pp. 284–290 (2007).
2. Kasperski, A. Life entrapped in a network of atavistic attractors: How to find a rescue. *Int. J. Mol. Sci.* **23**(7), 4017 (2022).
3. Nadimi-Shahraki, M. H., Zamani, H. & Mirjalili, S. Enhanced whale optimization algorithm for medical feature selection: A COVID-19 case study. *Comput. Biol. Med.* **148**, 105858 (2022).
4. Zamani, H., Nadimi-Shahraki, M. H. & Gandomi, A. H. QANA: Quantum-based avian navigation optimizer algorithm. *Eng. Appl. Artif. Intell.* **104**, 104314 (2021).
5. Mundra, P. A. & Rajapakse, J. C. Gene and sample selection for cancer classification with support vectors based t-statistic. *Neurocomputing* **73**, 2353–2362 (2010).

6. Jin, C., Jin, S.-W. & Qin, L.-N. Attribute selection method based on a hybrid BPNN and PSO algorithms. *Appl. Soft Comput.* **12**, 2147–2155 (2012).
7. Goldberg, D. E. *Genetic Algorithms in Search, Optimization and Machine Learning* (Addison-Wesley Longman Publishing Co., Inc., Boston, 1989).
8. Kennedy, J. & Eberhart, R. C. Particle swarm optimization. In *Proceedings of IEEE International Conference on Neural Networks*, 1942–1948 (1995).
9. Shi, Y. & Eberhart, R. C. A modified particle swarm optimizer. In *Proceedings of IEEE World Congress on Computational Intelligence*, 69–73 (1998).
10. O'Hagan, S., Knowles, J. & Kell, D. B. Exploiting genomic knowledge in optimising molecular breeding programmes: Algorithms from evolutionary computing. *PLoS ONE* **7**, e48862 (2012).
11. Chuang, L. Y., Lin, Y. D., Chang, H. W. & Yang, C. H. An Improved PSO algorithm for generating protective SNP barcodes in breast cancer. *PLoS ONE* **7**, e37018 (2012).
12. Chuang, L. Y., Huang, H. C., Lin, M. C. & Yang, C. H. Particle swarm optimization with reinforcement learning for the prediction of CpG islands in the human genome. *PLoS ONE* **6**, e21036 (2011).
13. Yang, C. S., Chuang, L. Y., Ke, C. H. & Yang, C. H. A hybrid feature selection method for microarray classification. *Int. J. Comput. Sci.* **35**, 285–290 (2008).
14. Shen, Q., Shi, W. M., Kong, W. & Ye, B. X. A combination of modified particle optimization algorithm and support vector machine for gene selection and tumor classification. *Talanta* **71**, 1679–1683 (2007).
15. Li, L. P., Weinberg, C. R., Darden, T. A. & Pedersen, L. G. Gene selection for sample classification based on gene expression data: Study of sensitivity to choice of parameters of the GA/KNN method. *Bioinformatics* **17**, 1131–1142 (2001).
16. Abdulqader, D. M., Abdulazeez, A. M. & Zeebaree, D. Q. Machine learning supervised algorithms of gene selection: A review. *Mach. Learn.* **62**(03), 51 (2020).
17. Pashaei, E. & Pashaei E. Gene selection using intelligent dynamic genetic algorithm and random forest. In *2019 11th International Conference on Electrical and Electronics Engineering (ELECO)* 470–474 (IEEE, 2019).
18. Leung, M.-F., Coello, C. A. C., Cheung, C.-C., Ng, S.-C. & Lui, A.-F. A hybrid leader selection strategy for many-objective particle swarm optimization. *IEEE Access* **8**, 189527–189545 (2020).
19. Lin, C. Y. *et al.* Deep learning with evolutionary and genomic profiles for identifying cancer subtypes. *J. Bioinform. Comput. Biol.* **17**(3), 1940005. https://doi.org/10.1142/S0219720019400055 (2019).
20. Dashtban, M. & Balafar, M. Gene selection for microarray cancer classification using a new evolutionary method employing artificial intelligence concepts. *Genomics* **109**(2), 91–107. https://doi.org/10.1016/j.ygeno.2017.01.004 (2017).
21. Grisci, B. I., Feltes, B. C. & Dorn, M. Neuroevolution as a tool for microarray gene expression pattern identification in cancer research. *J. Biomed. Inform.* **89**, 122–133. https://doi.org/10.1016/j.jbi.2018.11.013 (2019).
22. Wang, S. *et al.* Hybrid feature selection algorithm mRMR-ICA for cancer classification from microarray gene expression data. *Comb. Chem. High Throughput Screen* **21**(6), 420–430. https://doi.org/10.2174/1386207321666180601074349 (2018).
23. Alanni, R., Hou, J., Azzawi, H. & Xiang, Y. Deep gene selection method to select genes from microarray datasets for cancer classification. *BMC Bioinform.* **20**(1), 608. https://doi.org/10.1186/s12859-019-3161-2.PMID:31775613;PMCID:PMC6880643 (2019).
24. Alderdice, M. *et al.* Evolutionary genetic algorithm identifies *IL2RB* as a potential predictive biomarker for immune-checkpoint therapy in colorectal cancer. *NAR Genom. Bioinform.* **3**(2), lqab016. https://doi.org/10.1093/nargab/lqab016 (2021).
25. Ahmed, S. *et al.* An integrated feature selection algorithm for cancer classification using gene expression data. *Comb. Chem. High Throughput Screen.* **21**(9), 631–645. https://doi.org/10.2174/1386207322666181220124756 (2018).
26. Asif, M., Khan, M. A., Abbas, S. & Saleem, M. Analysis of space & time complexity with PSO based synchronous MC-CDMA system. In *2019 2nd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*, 1–5 (IEEE, 2019).
27. Huang, G. B., Zhu, Q. Y. & Siew, C. K. Extreme learning machine: a new learning scheme of FNNs. In *Proceedings of 2004 International Joint Conference on Neural Networks*, 985–990 (2004).
28. Abualkishik, A., Saleem, M., Farooq, U., Asif, M., Hassan, M. & Malik, J. A. Genetic algorithm based adaptive FSO communication Link. In *2023 International Conference on Business Analytics for Technology and Security (ICBATS)*, 1–4 (IEEE, 2023).
29. Soria-Olivas, E. *et al.* BELM: Bayesian extreme learning machine. *IEEE Trans. Neural Netw.* **22**, 505–509 (2011).
30. Cortes, C. & Vapnik, V. Support-vector networks. *Mach. Learn.* **20**, 273–297 (1995).
31. Han, F. & Huang, D. S. Improved extreme learning machine for function approximation by encoding a priori information. *Neurocomputing* **69**, 2369–2373 (2006).
32. Sánchez, J. S., Pla, F. & Ferri, F. J. On the use of neighbourhood-based non-parametric classifiers1. *Pattern Recognit. Lett.* **18**, 1179–1186 (1997).
33. Mohamad, M. S., Omatu, S., Deris, S. & Yoshioka, M. A modified binary particle swarm optimization for selecting the small subset of informative genes from gene expression data. *IEEE Trans. Inf. Technol. Biomed.* **15**, 813–822 (2011).
34. Mohamad, M. S. *et al.* An enhancement of binary particle swarm optimization for gene selection in classifying cancer classes. *Algorithms Mol Biol.* **8**, 1 (2013).
35. Moosa, J. M., Shakur, R., Kaykobad, M. & Rahman, M. S. Gene selection for cancer classification with the help of bees. *BMC Med Genet.* **9**, 2–47 (2016).
36. Lai, C.-M., Yeh, W.-C. & Chang, C.-Y. Gene selection using information gain and improved simplified swarm optimization. *Neurocomputing.* **19**(218), 331–338 (2016).
37. Han, F. *et al.* A gene selection method for microarray data based on binary PSO encoding gene-to-class sensitivity information. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **14**(1), 85–96 (2015).
38. Hussain, D. *et al.* Enabling smart cities with cognition based intelligent route decision in vehicles empowered with deep extreme learning machine. *CMC Comput. Mater. Contin.* **66**(1), 141–156 (2021).
39. Han, F., Sun, W. & Ling, Q.-H. A novel strategy for gene selection of microarray data based on gene-to-class sensitivity information. *PLoS ONE* https://doi.org/10.1371/journal.pone.0097530 (2014).

Author contributions

Author contributions: The authors confirm contribution to the paper as follows: study conception and design: A. A.N. and A.H.K.; data collection: A. A.N., M. A., ; analysis and interpretation of results: A. H. K., M.F., A.R.; draft manuscript preparation: M.A., K.M. and M.H.; supervision: A.A.N, K.M and M.F. All authors reviewed the results and approved the final version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to M.A. or A.R.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.