A Hybridized Feature Selection Approach in Molecular Classification using CSO and GA

Ahmed Elsawy[1], Mazen M. Selim[2,] Mahmoud Sobhy[3]
Computer Science dept.
Faculty of Computers and Informatics, Benha University
Benha, Qalyubia Governorate, Egypt

{ahmed.el_sawy, selimm, mahmoud.hassan}@fci.bu.edu.eg

Abstract — Feature selection in molecular classification is a basic area of research in chemoinformatics field. This paper introduces a hybrid approach that investigates the performances of chicken swarm optimization (CSO) algorithm with genetic algorithms (GA) for feature selection and support vector machine (SVM) for classification. The purpose of this paper is to test the effect of elimination of the inconsequential and redundant features in chemical datasets to realize the success of the classification. The proposed algorithm was applied to four chemical datasets and proved superiority in achieving minimum classification error rate in comparison with different feature selection algorithms for molecular classification.

Keywords— molecular classification; chicken swarm optimization; genetic algorithms; support vector machines; feature selection.

1. INTRODUCTION

Developing a new chemical entity as a drug is still a challenging, time-consuming and cost-intensive process [28]. Decreasing costs and speeding up the discovery process considered as primary objectives of drug discovery, both in pharma and biotech sector [24]. Improvements in computational techniques suggest an alternative to medical chemistry techniques for studying the structure and foretelling the biological activity of drug candidates and that way highly minimizing classical resource requirements [14].

One of the key problems in the structure-activity relationship research focuses on the characterization the quantification of chemical structures, as an appropriate correlation can only be developed if both the biological activity and chemical structure are quantified [23]. One significant method of numerical characterization of molecular structures was accomplished by applying the basics of graph theory to molecular structure and this resulted in many graph-theoretical invariants based on molecular graphs [2].

Several attempts are being devoted in the medicinal industry for evolving new drugs [6]. The drug discovery process comprises of seven steps, disease selection, target hypothesis, lead identification, lead optimization, pre-clinical trial, clinical trial and pharmacogenomic optimization [18]. Traditionally, these steps are carried out sequentially, and if one of these steps is delayed, it surely slows down the entire process [22]. Considering both, the potential benefits to human health and the immense cost in time and money of drug discovery, any tool that can enhance the performance of any stage included in drug discovery process will be highly prized [1]. A viable solution to this issue lies in the estimation of necessary properties of molecules directly from their structure without the input of any other experimental data through quantitative structure-activity relationship (QSAR) models [27].

The main supposition in the QSAR model is that all physicochemical and biological properties of a chemical item are statistically concerned with its molecular structure [26]. Quantitative relations generated from such studies help in predicting the leading contributions of specific structural aspects or chemical interactions in order to change physicochemical properties and biological activities and also in predicting properties and activities of untested and not yet synthesized compounds [6].

Several parameters are vital in the prediction ability of a QSAR model. On one hand, different techniques might be applied to check the linear or nonlinear behavior of a data set. On the other hand, feature selection techniques are applied to decrease the model complexity to diminish the overfitting/overtraining hazard and to choose the most important descriptors from an expansive number of descriptors [26].

The process of design, creation, organization, management, analysis, retrieval, spreading, visualization and use of chemical information is called Chemoinformatics [9]. Examination and manipulation of chemical structural information are made conceivable using molecular descriptors. These are numerical values that describe properties of molecules. For instance, they may represent the physicochemical properties of a molecule or they might be values that are determined by applying algorithmic techniques to the molecular structures.

The chosen descriptors are then connected to a biological activity of the corresponding compound by means of a mathematical model. Diverse modeling methods can be applied, some of which explicitly require a feature selection [7].

Molecular classification is carried out in three steps; the initial step is feature extraction, in this step all features of molecules are separated and represented in feature vectors [18]. The second step is called feature selection (or reduction), in which a subset of features is chosen from a larger set of features, which prompts to the diminishment of the dimensionality of features space for a successful classification task. Feature Selection helps in data comprehension, diminishing calculation necessity, lessening the effect of a curse of dimensionality and enhancing the predictor performance. The third and last step is classification, in which molecules are classified to their optimal class [7].

There are three developed types of feature selection: filter, wrapper, and embedded methods [30]. The first scheme (filter-based) uses statistical properties of the features to filter out poorly informative ones. This is done before applying any classification algorithm. A second approach (wrapper-based) is more computational, but often provides more accurate results than filter methods. A wrapper algorithm searches the feature space to score feature subsets according to their predictive power, optimizing the subsequent induction algorithm that uses the respective subset for classification [30]. The third approach (embedded methods) implements the feature selection in the process of a model building which adds an extra term that penalizes the size of the selected features to the standard cost function of SVM and optimizes the new objective function to select feature subset [17].

A lot of feature selection techniques have been developed and widely used such as genetic algorithms, forward selection, backward elimination, stepwise regression, and simulated annealing. Also, some swarm intelligence techniques have been applied to feature selection which is comprised of a population of artificial agents and inspired by the social behavior of animals (fish, birds, fireflies, and so forth) from the real world. A case of such techniques is ant colony optimization [15], [17], bat algorithm [21], and Cuckoo Search [24].

In this study, we developed a new wrapper feature selection algorithm using a combination between CSO algorithm and GA for obtaining the minimum features subset in chemical datasets and achieving better classification accuracy using SVM classifier.

This paper is organized as follows: Section II presents some of the related work that applied in feature selection for chemical compound classification. Section III accentuates the CSO algorithm while GA algorithm accentuated in section IV. In section V, the proposed algorithm for feature selection using CSO algorithm and GA is described. The experimental results and the datasets which used for validating and testing our proposed methods are discussed in section VI. Finally, conclusion is stated in section VII.

2. RELATED WORK

Feature selection reduces the dimensionality of dataset by selecting the prominent features from a set of features and eliminating the irrelevant and redundant ones. This causes both reduced processing time and increased classification accuracy.

One of the applied techniques that applied for wrapper Feature Selection in chemical compound classification is based on Chicken swarm optimization algorithm (FS-CSO). The CSO algorithm was used to find a set of features that minimize the classification error with a minor number of selected features. In this approach, the chicken swarm optimization is applied in the feature selection step and well-known k-nearest neighbor (KNN) classifier is applied in the classification step with three k values 3, 4, and 5. First, there is several feature vectors (molecular descriptor), Each molecule descriptor is an individual dimension and the values of each dimension range from 0 to 1. The fitness function for FS-CSO is to maximize classification performance given the training data while keeping a minimum number of features selected [5].

Particle swarm optimization (PSO) is another technique that was used for wrapper feature selection for molecular classification. PSO is a stochastic population-based optimization approach proposed by Kennedy and Eberhart in 1995 [4]. PSO is propelled by social behaviors, for example, bird flocking and fish schooling. The hidden wonder of PSO is that knowledge is optimized by social association in the population where thinking is not only personal but also social. It's additionally related, notwithstanding, to evolutionary Computation, and has binds to both genetic algorithms and evolution strategies. (FS-PSO). In the selection step, the particle swarm optimization is used and KNN classifier [11] in the classification step with k values 3, 4, and 5. The feature space with each feature outlined in an individual dimension and where each dimension ranges from 0 to 1. The fitness function for the PSO aims to reduce the classification error given the training data while keeping a minimum number of features selected.

Binary Particle Swarm optimization and neural networks are used together in two different methods in feature selection [20]. The first method called (BPSO-BP) used PSO as a first step and NNs as the second step in molecular classification. This approach consists of two stages. The first stage, BPSO was applied to feature selection and then a neural network is used to build a QSAR model based on the features selected in the first stage referred to as BPSO-BP. This approach depended on back-propagation to train the neural network in the second stage [20].

BPSO-BP comprises of two nested loops; BPSO is the outer loop. Every cycle of this loop produces a set of selected features. The neural network with back propagation is the inner loop. NN takes the selected features as input and is

trained for a predefined number of iterations. The model fitness is sent back to the BPSO stage to manage the feature selection in the outer loop. The second method called (BPSO-PSO) applied Binary PSO in both two stages. This approach re-builds up the results of BPSO-BP approach by addressing the restriction of back propagation. It utilizes particle swarm optimization (PSO) in the second stage for training and bootstrap aggregation (Bagging) keeping in mind the end goal to overcome the instability of PSO. BPSO-PSO method causes strong QSAR models while decreasing the changeability because of the decision of the back-propagation parameters [20].

Also, there are many artificial intelligence techniques that are used in feature selection in molecular classification [5].

3. CHICKEN SWARM OPTIMIZATION (CSO)

Chicken Swarm Optimization (CSO) was developed based on the chicken behavior which proposed by Meng, X.B.et al [19]. CSO is an optimization method based on chicken swarm foraging behaviors. Each kind of chickens follows various laws of movements. The predominant chickens in a flock will dominate the feeble ones. The most dominant hens stay close to the head roosters and the most submissive hens and roosters who remain at the periphery of the group. Each kind of chickens follows various laws of movements [9]. A hierarchal order assumes a huge part in the social lives of chickens. The predominant chickens in a flock will dominate the feeble ones.

According to the algorithm of CSO described in [19], there are at least four basics in the chicken behavior, as follows:

1) In the chicken swarm, there exist several groups. Each group has a dominant rooster and a block of hens and chicks.

2) The fitness value of the chickens is evaluated. The individuals with the best fitness will be the roosters which will be group leaders, and the individuals with the worst fitness values will be considered as chicks. The others would be the hens. Hens randomly decide which group to live in. The mother-child relationship between the hens and the chicks is also randomly established.

3) The hierarchal order, dominance relationship and mother-child relationship in each group will stay unchanged. These cases are updated every several (G) time steps.

4) The swarm consists of Nc virtual chickens divided as follow: Cnr, Cnh, Cnc, and Cnm which are the number of roosters, the hens, the chicks, and the mother hens, respectively. Each individual is represented by their positions in a D-dimensional space by $x_{i,j}$ (i ϵ[1; :::;Nc]; j ϵ [1; :::;D]).

Movement of Roosters:

Roosters with higher fitness values can search for food in a wider range of places and have a superiority for food access than those with worse fitness. Such movement described in equations (1) and (2).

$$RX_{i,j}^{t+1} = RX_{i,j}^{t} * (1 + RandN\ (0, \sigma^2)), \quad (1)$$

$$\sigma^2 = \begin{cases} 1, & if\ Rf_i \leq Rf_k \\ \exp\left(\frac{Rf_k - Rf_i}{|Rf_i + \epsilon|}\right), & otherwise \end{cases} \quad k \in [1, Nc], k \neq i, \quad (2)$$

where $Rx_{i,j}$ is the selected rooster with index i, RandN (0; σ^2) is a Gaussian distribution with zero mean and standard deviation σ^2, ϵ is the smallest constant in the computer used to avoid zero-division-error, is, k is roosters index that a randomly chosen from the roosters group, Rf_i is the fitness value of the corresponding rooster Rx_i.

Movement of Hens:

Hens follow their group-mate roosters to search for food. Besides, they would also randomly steal the decent food that found by other chickens; however, they would be quelled by the other chickens. The more dominant hens would have better chance in competing for food. These phenomena can be formulated mathematically as follows:

$$HX_{i,j}^{t+1} = HX_{i,j}^{t} + HS1 * Rand * \left(Rx_{r1,j}^{t} - Hx_{i,j}^{t}\right) + HS2 * Rand * \left(Rx_{r2,j}^{t} - Rx_{i,j}^{t}\right) \quad (3)$$

where, $HS1 = \exp(\ (Hf_i - Rf_{r1})/(abs\ (Hf_i) + \epsilon))$, (4)

and $HS2 = exp\ (Rf_{r2} - Hf_i)$ (5)

where Rand is a uniform random number over [0, 1]. $r_1 \epsilon$ [1, … , Nc] is an index of the rooster, which is the i^{th} hen's group-mate, while $r_2 \in$ [1, … , N] is randomly chosen index of a chicken (rooster or hen) which is randomly chosen from the swarm where r1 not equal to r2.

Movement of Chick:

The chicks can only move around their mother to search for food. This is formulated as in equation (6).

$$Cx_{i,j}^{t+1} = Cx_{i,j}^{t} + \text{Fc} * Cx_{m,j}^{t} - Cx_{i,j}^{t}, \quad (6)$$

Where $Cx_{m,j}^{t}$ is the position of the i^{th} chick's mother such that m ∈ [1, N], Fc is a parameter that represents how much speed a chick would follow its mother, to consider the differences between each chick Fc is chosen randomly in the range [0, 2].

4. GENETIC ALGORITHMS (GA)

GA is a meta-heuristics algorithm that mimics the long-term optimization process of biological growth for solving mathematical optimization problems [25]. GA is based on the selection and survival of the fittest. Problem solutions are abstract individuals in a population. Each solution is

computed with a fitness function. The fitness value represents the survivability of a solution, i.e. the probability to be a member in the next population and generating children with similar characteristics by handing down genetic information via evolutionary mechanisms like reproduction, variation, and selection, respectively. Reproduction and variation are achieved by mutation of genes and crossover [28].

GA can be defined as an intelligent probabilistic search algorithm which can be used in several combinatorial optimization techniques [25]. The theoretical essentials of GA were originally developed by Holland [13]. Individuals that adapt to their environment successfully will have a better chance of surviving and reproducing, while the individuals which are less fit will be eliminated. Meaning that the genes from the highly fit individuals will propagate to a numerous number of individuals in each successive generation. The combination of good characteristics from highly adapted ancestors may produce even more fit offspring. In this way, species evolve to adapt more and more to their environment [28].

GA algorithm mimics these processes by taking a foremost population of individuals and applying genetic operators such as crossover and mutation operators to each reproduction. Simple crossover operators are based on generating one or more crossover points randomly and then swapping the bits of the two parent strings to produce two child strings. After applying crossover operator, the mutation operator is applied to each child by inverting each bit in the solution with some small amount of random search [25]. It also helps to prevent loss of valuable genetic information by reintroducing information that is lost due to premature convergence and thereby expanding the search space. At an initial stage of GA, the crossover operator is mainly responsible for the search and so the mutation rate is set to a low value to allow minimal disruption.

The mutation rate at which the GA converges depends on the population replacement method [28]. Mutation is a divergence operator. It is meant with break one or more members of a population out of a local minimum or maximum space and potentially discover a better space occasionally. Since the end goal is to lead the population to convergence, crossover happen more frequently (typically at each generation). Mutation, being a divergence operation, it should happen less frequently than crossover, and typically affects only a few members of a population in any given generation [13].

The classification method that used in the fitness function in equation (7) is the support vector machine (SVM) classifier with Gaussian Radial Basis kernel function. Support vector machine (SVM) is an effective supervised learning algorithm that considered as one of the most popular machine learning methods that have been used in classification and regression applications like pattern recognition, data mining and machine learning application. SVM was developed in 1995 by Cortes and Vapnik [31]. The basic SVM takes a set of input data and predicts the corresponding class for each one. This process is known as the binary linear classification [3]. A SVM model represents these examples as points in space. Given a set of training examples, each of them belongs to one of two classes. For the linearly separable case, SVM determines the desired hyperplane that separates the training patterns. This hyperplane maximizes the sum of its distances to the most similar positive and negative training patterns, respectively. This sum is called margin [8]. Then, the new examples are mapped into the same space and emphasized to belong to a class based on which side they fall on.

5. HYBRID FEATURE SELECTION FOR CHEMICAL COMPOUND CLASSIFICATION (FSCSO-GA)

Firstly, data is divided into two categories randomly, training set and testing set. In the training phase, machine learning learns and selects the best features and in the testing phase, machine learning tests the machine learning knowledge and tests the features which were selected in the training phase, then classifies the data into active and in-active classes. Chicken swarm optimization (CSO) algorithm is used here to find a set of features that minimize the classification error with a minor number of selected features.

First, we have some of molecular descriptors that we called feature vectors. Each molecule descriptor is an individual dimension and the values of each dimension range from 0 to 1. There are some redundant or unwanted features in the descriptor that make it very huge, hence it requires an intelligent searching method to find optimal point in the search space that maximizes the given fitness function. The fitness function for the CSO is to maximize classification performance given the training data, while keeping a minimum number of features selected, as shown in equation (7)

$$f_\theta = w * E + (1 - w) * \sum_i \frac{\theta_i}{N} \ , \qquad (7)$$

Where f_θ is the fitness function given a vector θ sized N with 0 or 1 elements representing unselected and selected features, N is the total number of features in the dataset, E is the classification error rate and w is a constant that controls the importance of classification accuracy to the number of features selected.

The used variables are the same as the number of features in the given dataset. Variables are limited in the range [0; 1], where the variable value approaches 1; its corresponding feature is a candidate to be selected in classification. In the individual fitness calculation, the variable is a threshold to decide the exact features to be evaluated as shown in equation (8).

$$f_{i,j} = \begin{cases} 1 & if \ \ X_{i,j} > 0.5 \\ 0 & otherwise \end{cases}, \qquad (8)$$

Where $X_{i,j}$ is the dimension value for search agent i at dimension j. A simple truncation rule was used to ensure variable limits as the updated value can violate the limiting constraints; [0, 1].

After applying equation (8), the result is a vector containing 1 or 0 values that will passed to SVM classifier that computes the classification error rate to compute the fitness values for train individuals via equation (7).

But before we use the SVM classifier, may be one or more members of a population may be limited in a local minimum or maximum space, so we need a divergence operator which is intended to occasionally break one or more members of the population out of this problem and potentially discover a better minimum/maximum space.

One of the important operators that avoid premature convergence on a local maximum or minimum is GA operators especially (Mutation operator). Mutation operator maintains genetic diversity in the subsequent generations which avoid premature convergence on a local maximum or minimum. The probability of mutation, Pm, is the probability of modifying an integer of the array. In this algorithm, the chosen value for Pm is 0.01.

After we apply the mutation step for the vector of equation (8), this vector is passed to SVM classifier to compute the classification error rate and then compute the fitness value for this individual. SVM is based on the theory of statistical learning. In 1995, Vapnik introduced the SVM to solve two class of positive and negative pattern reorganization problem.

In this paper, SVM is used to find the lowest classification error and ensure the goodness of the selected features.

Algorithm 1 shows the proposed algorithms that combines CSO with GA for chemical compounds feature selection. This algorithm modifies FS-CSO to obtain better classification accuracy with minimum number of features.

Algorithm 1: Hybrid Feature Selection for Chemical Compound Classification using CSO and GA (FSCSO-GA)

1. Read feature vectors (D) from a chemical dataset, each feature vector regards one molecule
2. Initialize matrix of positions CX_i for chickens randomly (number of columns equal to length of the descriptor)
3. Transform CX_i values into zeros and ones using equation (8) forming binary matrix, A
4. (S) = {features that Corresponding to 1 in matrix A}
5. Apply SVM classifier between all descriptors using S features and form new vector (V) for new classes
6. Match classes in V with classes in A and calculate the error (E), and then calculate the fitness value in (7)
7. Initialize number of Roosters (Cnr), Hens (Cnh), Chicks (Cnc), Mother-Hens (Cnm) in the swarm, and G
8. Initialize maximum number of iterations, Max

9-**While** (T < Max)
  10- **if** (T % G == 0) **then**
    11-Rank the chickens' fitness values and establish a hierarchal order in the swarm;
    12- Separate the swarm into groups, and determine the relationship between the chicks and mother hens in a group;
  13- **end**
  14- **for** i = 1: ***Nc***
    15- **if** i == rooster **then**
      16 - Update RXi's location using equation 1;
    17- **end**
    18- **if** i == hen **then**
      19- Update HXi's location using equation 3;
    20- **end**
    21- **if** i == chick **then**
      22- Update CXi's location using equation 6;
    23- **end**
    24- Mutate the elements of matrix V according to mutation probability Pm forming new binary matrix B
    25- Evaluate the new solution using equation 7;
    26-If the new solution is better than its previous one, update it;
  27- **end for**
28- **end while**

6. EXPERIMENTAL RESULTS

In this paper, a hybrid algorithm FSCSO- GA applied to feature selection in molecular classification to maximize classification performance and minimize the number of selected features. The proposed technique is tested on four chemical datasets consisting of molecular descriptors with some physical properties that are considered as a target to each descriptor.

The proposed algorithm was applied on three public available standard chemical datasets C8, PAH, and Phenet. These datasets are available on the website of molecular descriptors that administrated by Milano Chemometrics and QSAR Research Group [29]. The list of datasets described in Table (I).

In these datasets, the values of the physical properties belong to descriptors described by decimal values, but the target values are only two, so we aim to convert these values into two binary values. Here, there is no a specific way for converting, the suggested value was the mean value of all values so that the values that are greater than or equal the mean value converted to 1 and the values that are smaller than the mean value will be 0.

The proposed algorithm FSCSO-GA is applied to the three datasets above and compared to two algorithms FS-CSO and FS-PSO and proved an advance in two datasets (C8 and Phenet) scoring minimum error rate.

Also, we applied FSCSO-GA algorithm on (Selwood) dataset after converting values in the same manner and compare results with four algorithms FS-CSO, FS-PSO and another two algorithms BPSO-BP and BPSO-PSO and the results show that FSCSO-GA obtain the minimum error rate in four results from all seven as shown in Table IV.

For each dataset, the instances are randomly divided into three sets namely training, validation, and testing sets in a cross-validation manner.

FSCSO-GA is randomly initialized with solutions in the feature space and is applied to minimize the fitness function in equation (7), a solution with the features selected is forced to be one of the initial solutions. The global parameter set for all the optimizers are decided by experiment experience as shown in Table (II).

Table I

DATASETS DESCRIPTION

Dataset	No. of Molecules	No. of Features/Molecules
PHENET	22	110
SELWOOD	31	53
C8	18	102
PAH	82	112

FSCSO-GA algorithm used to evaluate the classification performance with parameters indicated in Table (II).

The best value decided for the mutation probability Pm is 0.01 as shown in Table II.

In each dataset, molecules were classified according some physical properties. In C8 dataset, three properties BP, LogP, and DHForm were used where LogP has achieved minimum classification error rate. In PAH dataset, LogP, BP, MP properties are used, and minimum classification error ratio obtained using MP property. When using Phenethylamines dataset, only LogP property is applied to compute the classification accuracy.

Concerning the related algorithms FS-CSO and FS-PSO, the classification error ratio values were taken from the results obtained in [5]. Also, in the other two previous algorithms, BPSO-PSO and BPSO-BP, values of classification error ratio computed for Selwood dataset were taken from obtained results in [20].

We can see that FSCSO-GA obtains much-enhanced fitness values over FS-CSO and FS-PSO on the average fitness values obtained during 70 runs. The advance in the obtained fitness value can be interpreted by the clever capability of FSCSO-GA to search the feature space adaptively and distributed searching capability of FSCSO-GA that always avoid algorithm stagnation in addition to the Gaussian Radial Basis kernel function used in SVM classifier, which tends to yield good performance under general smoothness assumptions.

Also, we can remark that the output of FSCSO-GA, fitness even better than using the whole feature set while it keeps less number of features.

The comparison between FSCSO-GA algorithm and FS-CSO algorithm is showed in figures [1 - 8]. Each figure contains two charts; each chart accentuates the changes in fitness value in all iterations; the left chart represents the result when applying FS-CSO algorithm on four datasets with different physical properties, and the right chart represents the result when applying FSCSO-GA algorithm on the same datasets. We can remark that FSCSO-GA selects a minimum number of features in three datasets compared to other algorithms while it keeps better classification performance as outlined in Table (III) and (IV).

Finally, results of this paper proved using the mutation operator that we combined before applying SVM classifier preserves the genetic diversity in the subsequent generations which avoid premature convergence on a local maximum or minimum and so that improves the classification accuracy.

7. CONCLUSION

In this paper, we presented a hybrid wrapper feature selection algorithm for chemical compound classification based on Chicken Swarm Optimization (CSO) with Genetic Algorithms (GA) in the selection step and support vector machine (SVM) in the classification step to achieve better classification accuracy with a minor number of features. This study proved that our proposed algorithm FSCSO-GA is better than four previous algorithms, FS-CSO, FS-PSO, BPSO-PSO, and BPSO-BP for most of these datasets. The proposed algorithm proved an advance in both features reduction and classification accuracy.

Table II

INDIVIDUAL OPTIMIZER PARAMETER SETTING

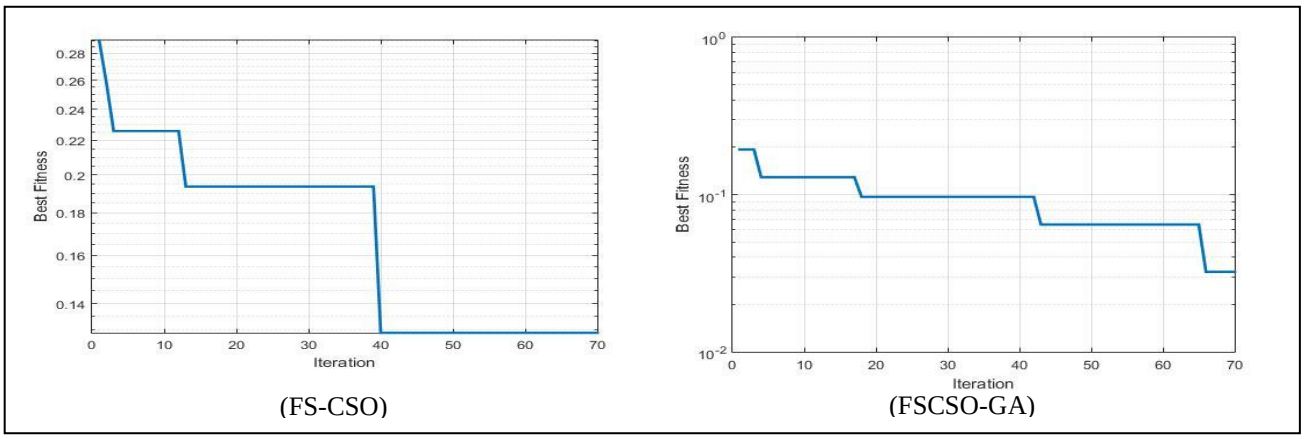
Algorithm	Parameter	Value
FSCSO-GA	R	0.15
	H	0.5
	M	0.7
	Pm	0.01

Table III
CLASSIFICATION ERROR ON TEST DATA FOR DIFFERENT OPTIMIZERS IN COMPARISON WITH THE DATA WITH ALL FEATURES

Dataset	All Features	Proposed Algorithm	FS-CSO [5]			FS-PSO [5]		
			K=3	K=4	K=5	K=3	K=4	K=5
C8-BP	0.3333	**0.0556**	0.1667	0.1667	0.1667	0.2222	0.1667	0.2222
C8 – LogP	0.5000	**0**	0.0556	0.0556	0.1667	0.1111	0.1111	0.1111
C8-DHForm	0.5000	**0.1111**	0.1667	**0.1111**	**0.1111**	0.2777	0.2222	0.1667
PAH-LogP	0.4444	0.2439	**0.2317**	0.2439	0.2439	0.2683	0.2805	0.3171
PAH-BP	0.2778	0.1220	**0.0976**	0.1098	0.1098	0.1220	0.1098	0.1098
PAH-MP	0.3333	**0.0845**	0.0976	**0.0845**	0.0976	0.0976	0.0845	0.0976
Phenet-LogP	0.3889	**0**	0.0999	0.0455	0.1364	0.1364	0.1366	0.1819

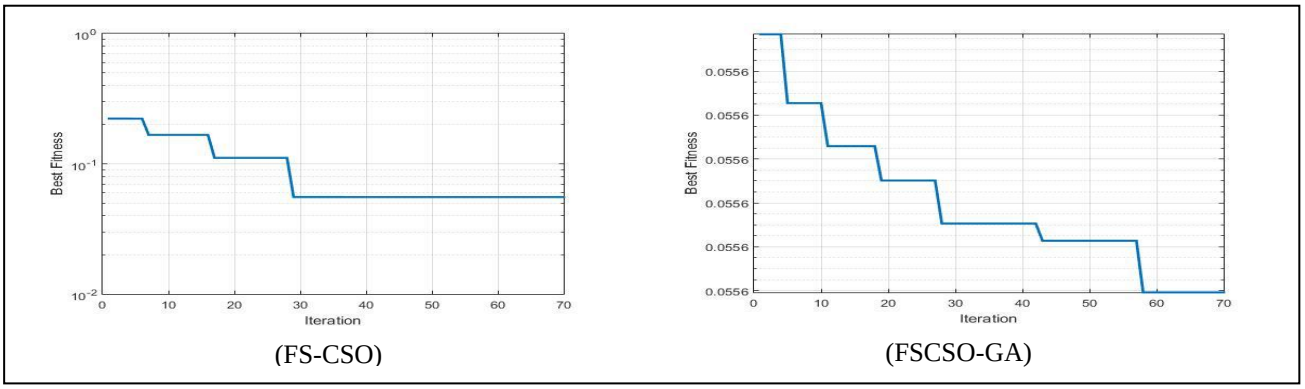
Table IV
CLASSIFICATION ERROR WHEN APPLYING FSCSO-GA, FS-CSO, FS-PSO, BPSO-PSO, AND BPSO-BP ON SELWOOD DATASET

Dataset	All features	Proposed Algorithm	BPSO-PSO [20]	BPSO-BP [20]	FS-CSO [5]			FS-PSO [5]		
					K=3	K=4	K=5	K=3	K=4	K=5
Selwood	0.4194	**0.0323**	0.1032	0.0981	0.0968	0.0645	0.1613	0.1613	0.1935	0.2258



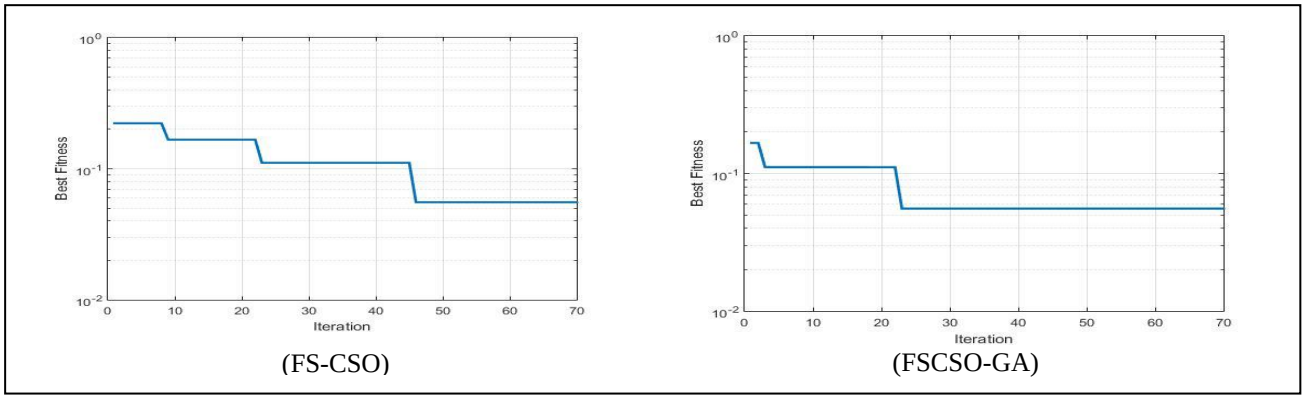
(FS-CSO) (FSCSO-GA)

Fig.1 Applying FS-CSO and FSCSO-GA on Selwood Dataset



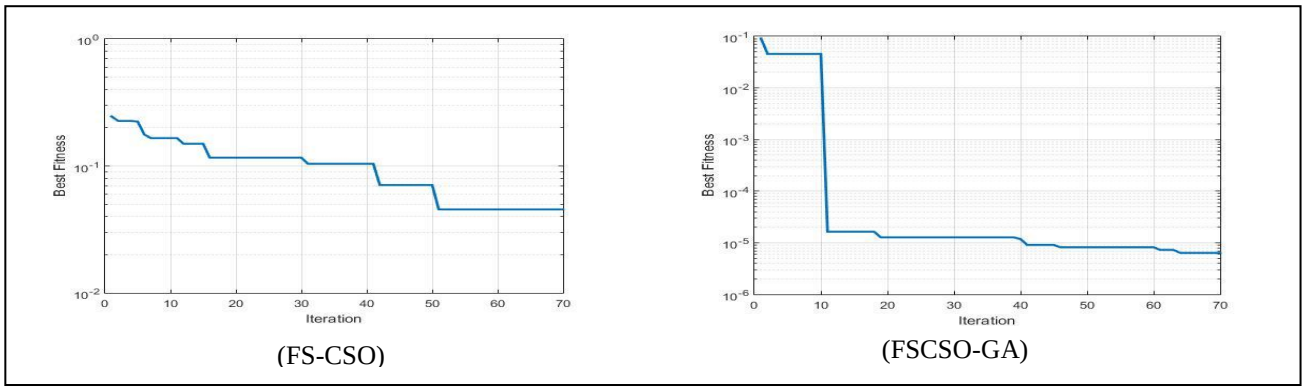
(FS-CSO) (FSCSO-GA)

Fig.2 Applying FS-CSO and FSCSO-GA on C8 Dataset based on LogP property



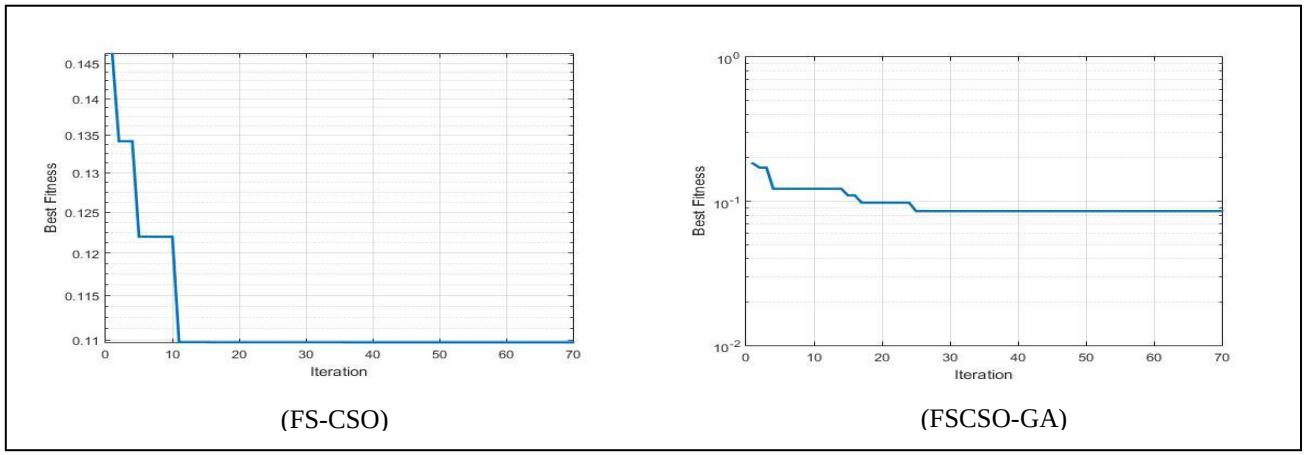
(FS-CSO) (FSCSO-GA)

Fig.3 Applying FS-CSO and FSCSO-GA on C8 Dataset based on BP property



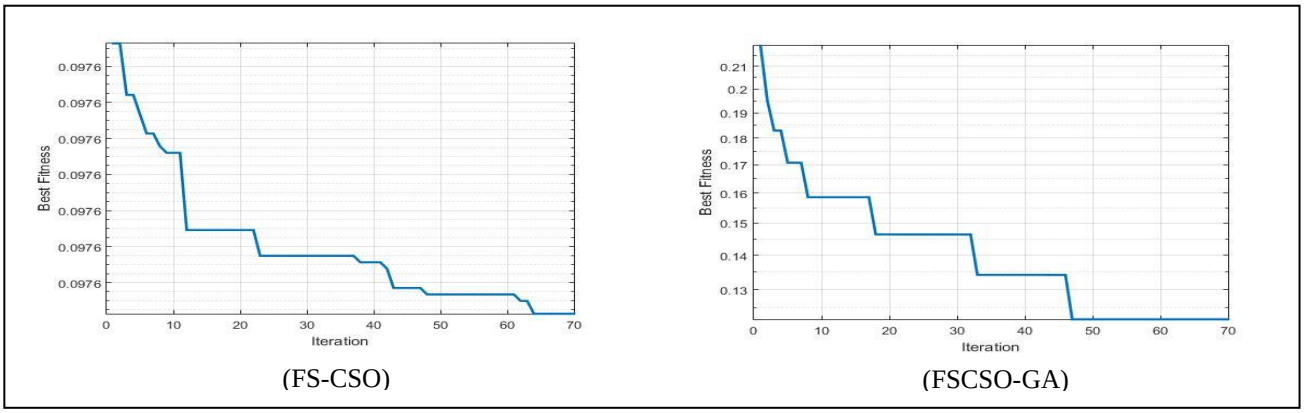
(FS-CSO) (FSCSO-GA)

Fig.4 Applying FSCSO and FSCSO-GA on Phenet Dataset based on LogP property



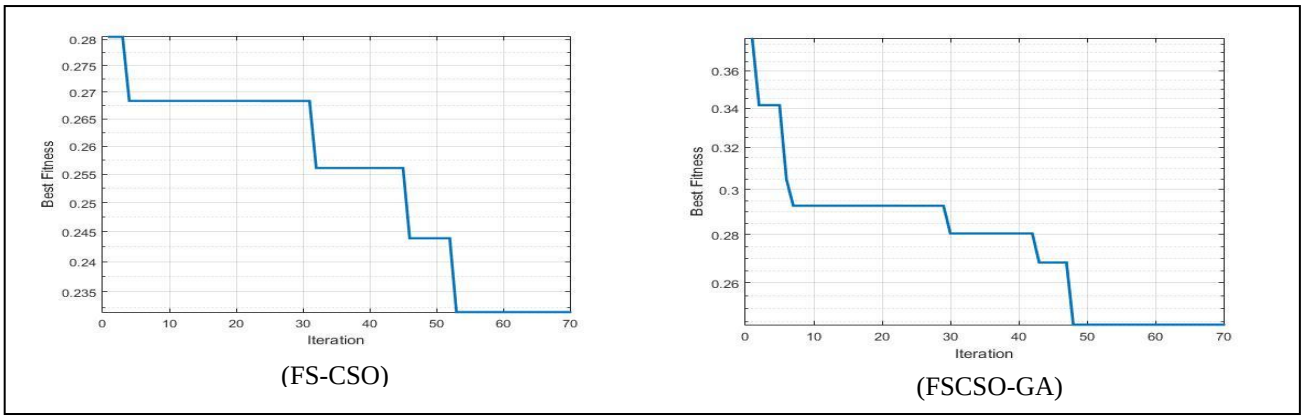
(FS-CSO) (FSCSO-GA)

Fig.5 Applying FS-CSO and FSCSO-GA on PAH Dataset based on MP property



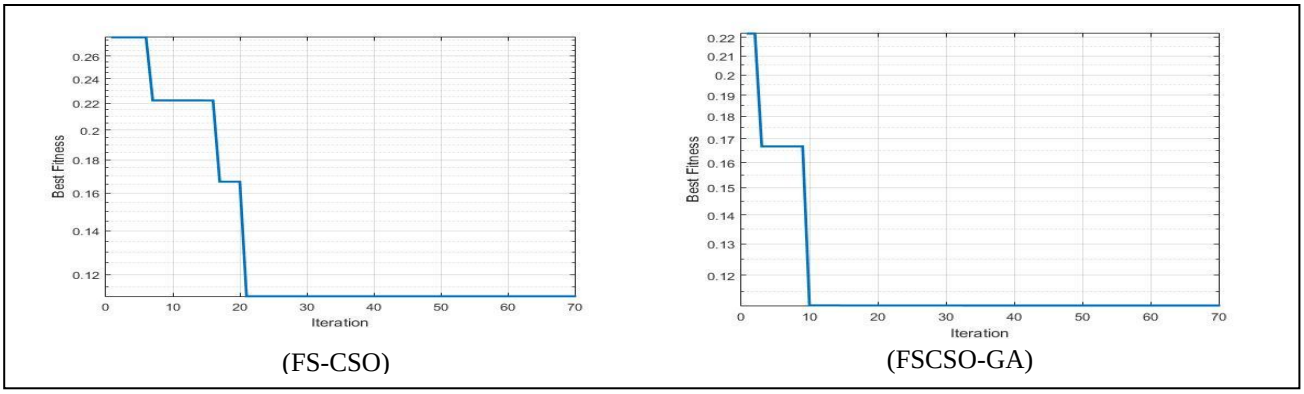
(FS-CSO) (FSCSO-GA)

Fig.6 Applying FS-CSO and FSCSO-GA on PAH Dataset based on BP property



(FS-CSO) (FSCSO-GA)

Fig.7 Applying FS-CSO and FSCSO-GA on PAH Dataset based on LogP property



(FS-CSO) (FSCSO-GA)

Fig.8 Applying FS-CSO and FSCSO-GA on C8 Dataset based on DH-Form property

REFERENCES

[1] Basak SC, Mills D, Mumtaz M, Balasubramanian K. Use of topological indices in predicting aryl hydrocarbon receptor binding potency of dibenzofurans: A hierarchical QSAR approach. Indian J Chem. 2003; 42A: 1385–1391.

[2] S.C.Basak,R.Natarajan,D.Mills,D.M.Hawkins,J.J.Kraker,Quantitative structure-activity relationship modeling of juvenile hormone mimetic compounds for culex pipiens larvae, with adiscussion of descriptor thinning methods, J.Chem.Inf.Model46(2006)65–77.

[3] Corinna Cortes and Vladimir Vapnik, "Support-Vector Networks", MachineLearning,20(3), 1995, pp.273-297 .

[4] R. Eberhart, and J. Kennedy, "A new optimizer using particle swarm theory". Proceedings of the Sixth International Symposium on Micromachine and Human Science, Nagoya, Japan, 1995, pp. 39-43.

[5] A. Elsawy, M. Mousa, M. Sobhy, "Feature Selection approach for Chemical Compound Classification based on CSO and PSO", Journal of Convergence Information Technology, accepted at March 15, 2017.

[6] Goyal, R. K., Dureja, H., Singh, G., & Madan, A. K. (2010). Models for Antitubercular Activity of 5′-O-[(N-Acyl)sulfamoyl]adenosines. Scientia Pharmaceutica, 78(4), 791–820.

[7] I. Guyon, A. Elisseeff, An introduction to variable and feature selection, Journal of Machine Learning Research 3 (2003) 1157–1182.

[8] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene Selection for Cancer Classification using Support Vector Machines," *Machine Learning*, vol. 46, pp. 389-422, 2002.

[9] A. Hafez, H. Zawbaa, E. Emary, and A. Mahmoud, "An innovative approach for feature selection based on chicken swarm optimization", 7th IEEE International Conference of Soft Computing and Pattern Recognition, , Kyushu University, Fukuoka, Japan,, November 13 - 1, 2015.

[10] Xu J, Hagler A. Chemoinformatics and Drug Discovery. Molecules. 2002; 7: 566–600. doi:10.3390/70800566

[11] R.Harikumar, M.Manjusha, "Performance Analysis of KNN Classifier and K-Means Clustering for Robust Classification of Epilepsy from EEG Signals", International Journal of Advanced

[12] Research Trends in Engineering and Technology (IJARTET), Jan.2016, vol. 3, Special Issue 7.

[13] J.H. Holland, Adaption in Natural and Artificial Systems, MIT Press, Cambridge, MA, 1975.

[14] H.-J. Huang, H.W. Yu, C.-Y. Chen, C.-H. Hsu, H.-Y. Chen, K.-J. Lee, F.-J. Tsai, C.Y.-C. Chen, Current developments of computer-aided drug design, J. Taiwan Inst. Chem. Eng. 41 (2010) 623–635.

[15] S. Kashef, H. Nezamabadi-pour, "An advanced ACO algorithm for feature subset selection", Neurocomputing, 5 January 2015, Vol.147, PP 271-279, ISSN 0925-2312.

[16] J. Kennedy, R. Eberhart, "Particle swarm optimization", Proceedings of IEEE International Conference on Neural Networks, Piscataway, 1995, NJ. pp. 1942-1948.

[17] R. Kohavi, G. John, Wrappers for feature subset selection, Artificial Intelligence - Special issue on relevance, dec. 1997, Volume 97 Issue 1-2, pp 273-324.

[18] R.A. Lewis, A general method for exploiting QSAR models in lead optimization, J. Med. Chem. 48 (2005) 1638–1648.

[19] X. Meng, Y. Liu, X. Gao, and H. Zhang, "A new bioinspired algorithm: chicken swarm optimization," in Advances in swarm intelligence, Springer, 2014, pp. 86–94.

[20] Z. Miled, D. Boyd, Z. Wang, G. Durst, R. Eberhart, "Particle Swarm Optimization and Neural Network Application for QSAR", , vol. 10, no. , pp. 194, 2004.

[21] R. Nakamura, L. Pereira, K. Costa, D. Rodrigues, J. Papa, X. Yang, "BBA: A binary bat algorithm for feature selection", SIBGRAPI - Conference on Graphics, Patterns and Images, 2012, PDB: 6581, 291-297.

[22] Kumar V, Madan AK. Topological Models for the Prediction of Cyclin-Dependent Kinase 2 Inhibitory Activity of Aminothiazoles. MATCH Commun Math Comput Chem. 2004; 51: 59–78.

[23] M.D.Parenti,G.Rastelli,Advances and applications of binding affinity prediction methods in drug discovery, Biotechnol. Adv.30 (2011).

[24] L. Pereira, D. Rodrigues, T. Almeida, C. Ramos, A. Souza, X. Yang and J. Papa, "A Binary Cuckoo Search and Its Application for Feature Selection", Cuckoo Search and Firefly Algorithm (book), Springer International Publishing, 2013, vol. 516, pp 141-154.

[25] C.R. Reeves, Modern Heuristic Techniques for Combinatorial Problems, Blackwell Scientific, 1993.

[26] Roy K, Saha A. Comparative QSPR Studies with Molecular Connectivity, Molecular Negentropy and TAU Indices. Part 2. Lipid-Water Partition Coefficient of Diverse Functional Acyclic Compounds. Internet Electron J Mol Des. 2003; 2; 288–305.

[27] Tarko L, Ivanciuc O. QSAR Modeling of the Anticonvulsant Activity of Phenylacetanilides with PRECLAV (Property Evaluation By Class Variables). MATCH Commun Math Comput Chem. 2001; 44: 201–214.

[28] L. Terfloth, J. Gasteiger, Neural networks and genetic algorithms in drug design, Drug Discov. Today 6 (2001) S102–S108.

[29] R.Todeschini, "molecular descriptors", Milano Chemometrics Milano Chemometrics and QSAR Research Group, Department of Environmental Sciences Department of Environmental Sciences University of University of Milano - Bicocca Bicocca, 2007, Website: http://www.moleculardescriptors.eu/dataset/dataset.htm.

[30] Ö. Uncu, I.B. Türksen, A novel feature selection approach: combining feature wrappers and filters, Information Sciences 177 (2007).

[31] V. Vapnik, The Nature of Statistical Learning Theory, Springer, New York, 1995.