Feature Selection approach for Chemical Compound Classification based on CSO and PSO

Ahmed Elsawy[1], Mahmoud Mousa[2], Mahmoud Sobhy[3]
[1,2,3] *Computer Science dept.*
Faculty of Computers and Informatics, Benha University
Benha, Alqualuobia governorate, Egypt
{ahmed.el_sawy[1], mahmoud.mossa[2], mahmoud.hassan[3]}@fci.bu.edu.eg

Abstract

with the improvement of profoundly efficient chemoinformatics data collection technology, classification of chemical data emerges as a vital topic in chemoinformatics. Towards building highly accurate predictive models for chemical data, here we introduce two feature selection algorithms. The first algorithm based on Chicken swarm optimization (FS-CSO) and the second algorithm based on Particle swarm optimization (FS-PSO). The proposed algorithms were applied to four datasets and FS-CSO proves advance over FS-PSO. Also, the two algorithms compared against two previous algorithms, BPSO-BP and BPSO-PSO, that used in feature selection for molecular classification and FS-CSO proves advance over them as well.

***Keywords**: Molecular Classification; Chicken Swarm Optimization; Particle Swarm Optimization; Feature Selection.*

1. INTRODUCTION

Chemoinformatics is concerned with the utilization of computational methods to handle chemical issues, with specific accentuation on the manipulation of chemical structural information. The term was introduced in the late 1990s [11]. Several endeavors have been made to define chemoinformatics; among the more generally cited the following:
The mixing of information resources to transform data into information, and information into knowledge, for the intended purpose of making better decisions faster in the arena of drug lead identification and optimization [1].

Chemoinformatics is a generic term that encompasses the design, creation, organization, management, retrieval, analysis, dissemination, visualization and use of chemical information [14]. The manipulation and examination 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.
Chemical and biological research is facilitated by enormous data repositories of chemical compounds from ultra-high- throughput screen techniques, where a large number of molecules are tested and classified based on their activities against given targets

A quantitative structure-activity relationship (QSAR) relates quantitative chemical structure attributes (molecular descriptors) to a biological activity [4]. QSAR studies get to be distinctly in drug discovery and development because their application can save substantial time and human resources. Several parameters are vital in the prediction ability of a QSAR model. On one hand, different measurable 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. The chose 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 [5].

Molecular classification is done 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 provides an approach to distinguishing the important features and expelling superfluous or redundant features from a dataset [10]. Feature Selection helps in comprehension data, diminishing calculation necessity, lessening the effect of a curse of dimensionality and enhancing the predictor performance. The third and final step is classification, in which molecules are classified to their optimal class that chose if each molecule is active or not active.

There are two types of feature selection approaches, i.e. wrapper and filter approaches. Their fundamental difference is that wrappers utilize a classification algorithm to evaluate the goodness of the features during the feature selection process while filters are independent of any classification algorithm [4].

Feature selection techniques, such as genetic algorithms, forward selection, backward elimination, stepwise regression, and simulated annealing have been used widely. Swarm intelligence is a computational knowledge based approach 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 [5], [8], bat algorithm [15], and Cuckoo Search [17]. In this paper, we will concentrate on the second step (Feature Selection) and attempt to solve it by using swarm intelligence techniques CSO and PSO as a wrapper-based approach.

This paper is organized as the following: section II presents some of the related work that used in feature selection for chemical compound classification. Section III accentuates CSO algorithm while PSO algorithm accentuated in section IV. In section V, The proposed algorithm for feature selection using chicken swarm optimization algorithm is described. Feature selection using particle swarm optimization algorithm is described in section VI. Section VII describes the datasets we used for validating and testing our proposed methods. The experimental results are discussed in section VIII. Finally, conclusions is stated in section IX.

2. RELATED WORK

Feature selection in chemical compound classification has been well studied. There are many techniques are used in feature selection for molecular classification. Binary Particle swarm optimization and neural networks are used together in two methods in feature selection [14]. 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. In the first stage, BPSO is used for feature selection and in the second stage a neural network is used to generate a QSAR model based on the features selected in the first stage referred to as BPSO-BP. This approach uses back propagation to train the neural network in the second stage [13].

The trouble in selecting a satisfactory learning rate for back propagation is illustrated. 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. The neural network takes the selected features as input, and is trained for a predefined number of iterations. The model fitness is fed 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 approach yields strong QSAR models, while diminishing the changeability due to the decision of the back-propagation parameters [13].

Also, there are many artificial intelligence techniques that are used in feature selection for molecular classification such as GA, ACO, and GSA [4].

3. CHICKEN SWARM OPTIMIZATION (CSO)

Chicken swarm optimization is bio-inspire metaheuristic optimization algorithm proposed by Meng, X.B.et al. [12]. The algorithm mimics the hierarchal order of a chicken swarm and the behaviors of its individual chickens. The hierarchal order of a chicken swarm is divided into several groups; each group comprises of one rooster and numerous hens and chicks. Each kind of chickens follows various laws of movements. A hierarchal order assumes a huge part in the social lives of chickens. The predominant chickens in a flock will dominate the feeble ones. There exist the most dominant hens that stay close to the head roosters and also the most submissive hens and roosters who remain at the periphery of the group.

The algorithm of CSO described in [12] as follows:

1) The chicken swarm is divided into several groups. In each group, there is a prevailing rooster, tailing it some hens and chicks.
2) The fitness value of the chickens is evaluated. The individuals with the best fitness will be the roosters each one will be a group leader, and the individuals with the worst fitness values will be considered as chicks. The others would be the hens.
3) The swarm consists of N virtual chickens divided as follow: Rn, Hn, Cn, and Mn 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; :::;N]; j ϵ [1; :::;D]).

Movement of Rooster: Roosters that have better fitness values can search for food in a wider range of place than those with worse fitness values; such movement described in equations (1) and (2).

$$X_{i,j}^{t+1} = X_{i,j}^{t} * (1 + Randn\,(0, \sigma^2)), \qquad (1)$$

$$\sigma^2 = \begin{cases} 1, & if\ f_i \le f_k \\ \exp\left(\frac{f_k - f_i}{|f_i + \epsilon|}\right), & otherwise \end{cases} \quad k \epsilon [1, N], k \neq i, \quad (2)$$

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

Movement of Hens: Hens follow their group-mate roosters to search for food. Besides, they would also randomly steal the good food found by other chickens; however, they would be quelled by the other chickens. The more dominant hens would have an advantage in competing for food than the more submissive ones. (3).

$$X_{i,j}^{t+1} = X_{i,j}^{t} + S1 * Rand * \left(x_{r1,j}^{t} - x_{i,j}^{t}\right) + S2 * Rand * \left(x_{r2,j}^{t} - x_{i,j}^{t}\right) \quad (3)$$

$$S1 = \exp(\,(f_i - f_{r1})/(abs\,(f_i) + \epsilon)), \qquad (4)$$

$$S2 = \exp\,(f_{r2} - f_i) \qquad (5)$$

where Rand is a uniform random number over [0, 1]. $r_1 \epsilon$ [1, … , N] is an index of the rooster, which is the i^{th} hen's group-mate, while $r_1 \in$ [1, … , N] is randomly chosen index of a chicken (rooster or hen) from the swarm.

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

$$x_{i,j}^{t+1} = x_{i,j}^{t} + \mathrm{FL} * x_{m,j}^{t} - x_{i,j}^{t}, \qquad (6)$$

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

4. PARTICLE SWARM OPTIMIZATION (PSO)

Particle swarm optimization (PSO) is a stochastic population-based optimization approach proposed by Kennedy and Eberhart in 1995 [2], [9]. 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.

PSO model comprises of a swarm of particles moving in an n-dimensional, real-valued search space of conceivable problem solutions. For the search space, as a rule, a specific quality measure, the fitness, is defined making it feasible for particles to analyze different problem solutions. Every particle i at the time t has the following characteristics: X_i^t is the position vector; V_i^t is the velocity vector; P_i^t is a small memory storing its own best position seen so far; G_i^t is the global best position obtained through communication with its fellow neighbor particles. This information flow is obtained by defining a neighborhood topology on the swarm. The instinct behind the PSO model is that by giving data information about good solutions a chance to spread out through the swarm, the particles will tend to move to great areas in the search space. At each time step (t) the velocity is updated and the particle is moved to a new position. This new position is simply calculated as the sum of the previous position and the new velocity:

$$X_i^{t+1} = X_i^t + X_i^{t+1} \qquad (7)$$

The update of the velocity from the previous velocity to the new velocity is determined by:

$$V_k^{i+1} = V_k + \text{c1} \cdot r_1 \left(P_k^i - X_k^i \right) + \text{c2} \cdot r_2 \left(G_i - X_k^i \right) \quad (8)$$

The parameters c1 and c2 are real numbers chosen uniformly and at random in a given interval, usually [0, 1]. These values determine the significance of $P_{i,t}$ and $G_{i,t}$ respectively.

The classification that used in the fitness function in equation (9) is the well-known K-nearest neighbor (KNN) classifier.

K-nearest neighbor is a supervised learning algorithm that classifies an obscure sample instance based on the majority of the K-nearest neighbor category.

The KNN classifier is a non-parametric instance-based classifier [7]. This algorithm depends on the nearest neighborhood estimation. The new cases are classified on the basis of similarity measure which is the distance metric. Most commonly used is Euclidean distance.

In KNN classifier, the class of x is found by the following procedure:

a) Determine the k instances which are nearest to the class x based on the distance measure.
b) The next step is to allow this k instances to vote to discover the class of x.

5. FEATURE SELECTION FOR CHEMICAL COMPOUND CLASSIFICATION USING CHICKEN SWARM OPTIMIZATION (FS-CSO)

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 a number 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, as shown in equation (9) while keeping a minimum number of features selected.

$$f_\theta = w * E + (1 - w) * \sum_i \frac{\theta_i}{N} \quad , \qquad (9)$$

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 classifier error rate and w is a constant that control the classification performance to the number of features selected.

The used variables are the same as the number of features in the given dataset. All 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 (10).

$$f_{i,j} = \begin{cases} 1 & if \quad X_{i,j} > 0.5 \\ 0 & otherwise \end{cases}, \qquad (10)$$

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

Algorithm 1: Feature Selection for chemical compound classification using Chicken swarm optimization algorithm (FS-CSO)

1. Read molecular descriptors (D) from a file, each descriptor regards one molecule
2. Selected features (S) = {⌀}
3. Initialize matrix of positions X_i for chickens randomly (number of columns equal to length of the descriptor)
4. Convert X_i values into zeros and ones using equation (10) forming binary matrix B
5. (S) = {features that Corresponding to 1 in matrix B}
6. Apply KNN classifier between all descriptors using S features and form new vector (V) for new classes
7. Compare classes in V with classes in B and calculate the error (E), then calculate the fitness in (9)
8. Initialize number of Roosters (Rn), Hens (Hn), Chicks (Cn), Mother-Hens (Mn) in the swarm, and G
9. Initialize maximum number of iterations M
10-**While** M is not reached
11- **if** T % G equals 0 **then**
12-Rank the fitness of all chickens and establish a hierarchal order in the swarm;
13- Divide the swarm into different groups, and
14- Determine the relationship between the chicks and mother hens in a group;
15- **end**
16- **foreach** chicken Xi in the swarm **do**
17- **if** Xi is a rooster **then**
18 - Update Xi's location using equation 1;
19- **end**
20- **if** Xi is a hen **then**
21- Update Xi's location using equation 3;
22- **end**
23- **if** Xi is a chick **then**
24- Update Xi's location using equation 6;
25- **end**
26- Evaluate the new solution using equation 9;
27-If the new solution is better than its previous one, update it;
28- **end**
29- **end**

6. FEATURE SELECTION FOR CHEMICAL COMPOUND CLASSIFICATION USING PARTICLE SWARM OPTIMIZATION (FS-PSO)

PSO algorithm is also applied to find combinations of features that minimize the classification error with a minor number of selected features. The feature space with each feature represented in an individual dimension and the span of each dimension ranges from 0 to 1. The fitness function for the PSO is to maximize classification performance given the training data, as shown in equation (9) while keeping a minimum number of features selected. The number of used variables is the same as the number of features in the given dataset. All 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 individual fitness calculation, the variable is a threshold to decide the exact features to be evaluated as shown in equation (10).

Algorithm 2: Feature Selection for chemical compound classification using Particle swarm optimization algorithm (FS-PSO)

1. Read molecular descriptors (D) from a file, each descriptor regards one molecule
2. Selected features (S) = {⌀}
3. Initialize matrix of positions X_i for particles randomly (number of columns equal to length of the descriptor)
4. Convert X_i values into zeros and ones using equation (10) forming binary matrix B
5. (S) = {features that Corresponding to 1 in matrix B}
6. Apply KNN classifier between all descriptors using S features and form new vector (V) for new classes
7. Compare classes in V with classes in B and calculate the error (E), then calculate the fitness in (9)
8. Initialize the position and velocity of each particle in the swarm
9. Initialize maximum number of iterations M
10- **While** M is not reached, for each iteration **do**
11- **for each** particle
12- Calculate fitness value
13- If the fitness value is better than its personal best
set current value as the new pbest
14- **end**
15- Choose the particle with the best fitness value of all as gbest
16- **for each** particle
17- Calculate particle velocity according (8)
18- Update particle position according (7)
19- **end**
20- Evaluate the new solution using equation 9;
21- If the new solution is better than its previous one, update it;
22- **end**

K-nearest neighbor is based on the minimum distance from the query instance to the training samples. KNN is utilized in the experiments based on trial and error basis where the selected values of K are (3, 4, and 5) as in FS-CSO and FS-PSO. The KNN is used as a classification to ensure the goodness of the selected features. The classifier is evaluated on a validation set inside the fitness function. In addition, the used fitness function incorporates both classification accuracy and reduction size [8].

7. DATA

We apply methodologies of CSO and PSO described in the previous sections to the problem of feature selection in molecular classification based on four public available standard chemical datasets respectively (C8, PAH, Phenet, and WineEW). These datasets are available on the website of molecular descriptors that administrated by Milano Chemometrics and QSAR Research Group [19]. The list of datasets described in table (I).
In the first three datasets, the values of the physical properties belong to descriptors have more than two values, so we convert these values into two values by computing the mean value of all values and the values that are greater than or equal the mean value will be 1 and the values that are smaller than the mean value will be 0.
Also, we applied FS-CSO and FS-PSO algorithms on (Selwood) dataset after converting values in the same manner and compare results with another two algorithms (BPSO-BP) and (BPSO-PSO) and the results show that FS-CSO is the best one with minimum error as showed in table IV.

Table I

DATASETS DESCRIPTION

Dataset	No. of Molecules	No. of Features
C8	18	102
PAH	82	112
PHENET	22	105
WINEEW	178	110
SELWOOD	31	13

8. EXPERIMENTAL RESULTS

In this paper the chicken swarm optimization (CSO) and particle swarm optimization (PSO) algorithms used in the feature selection step in molecular classification in order to maximize classification performance and minimize the number of selected features. The proposed technique will be tested on four chemical datasets consisting of molecular descriptors and some physical properties that are considered as a target to each descriptor.

For each dataset, the instances are randomly divided into three sets namely training, validation, and testing sets in a cross validation manner. The training set is used to train the used classifier while the validation set is used to evaluate the classifier performance and is applied inside the optimization fitness. The test data are used for the final evaluation of the whole feature selection and classification algorithm.

Individual solutions in the CSO and PSO are points in the feature space; d-dimensional space, where d is the number of features in the original dataset in the range [0; 1]. The well-known K-nearest neighbor (KNN) classifier was used in the fitness function. CSO is randomly initialized with solutions in the feature space and is applied to minimize the fitness function in equation (9) but a solution with all 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).

Chicken swarm optimizer (CSO) and Particle swarm optimizer (PSO) are used in the same manner to be compared to evaluate its classification performance with parameters indicated in table (III).
We can see that CSO obtains much-enhanced fitness values over PSO on the average fitness values obtained from the different number of runs. The advance in the obtained fitness value can be interpreted by the clever capability of CSO to search the feature space adaptively and distributed searching capability of CSO that always avoid algorithm stagnation.
We can remark that the output of CSO and PSO fitness even is better than using the whole feature set while it keeps less number of features.

In addition, The KNN classifier is used with a different number of neighbors (k=3, k=4, k=5), and the results show that some results are better at k=4 than the others as shown in table (II).

Figure 1 represents comparison between CSO algorithm and PSO algorithms for feature selection on Selwood dataset, this chart accentuates the changes in fitness value in all iterations; the left chart represents the result when applying FS-CSO algorithm and the right chart represents the result when applying FS-PSO algorithm. The figure shows that the FS-CSO gives result better than FS-PSO which is better than BPSO-PSO and BPSO-BP

Values in table (II) are the average ratios of features selected to the total number of features for different optimizers over different datasets. We can remark that CSO selects a minimum number of features in comparison with PSO while it keeps better classification performance as outlined in table (VII).

9. CONCLUSION

In this paper, we presented two feature selection algorithms for chemical compound classification, Feature Selection based on Chicken Swarm Optimization (FS-CSO) and Feature Selection based on Particle Swarm Optimization (FS-PSO) to obtain better classification accuracy with minor number of features. This study proved that FS-CSO is better than FS-PSO for many datasets. The two algorithms are also compared against two previous algorithms, BPSO-BP and BPSO-PSO, and FS-CSO algorithm proved an advance in both features reduction and classification accuracy.

Table II. INDIVIDUAL OPTIMIZER PARAMETER SETTING

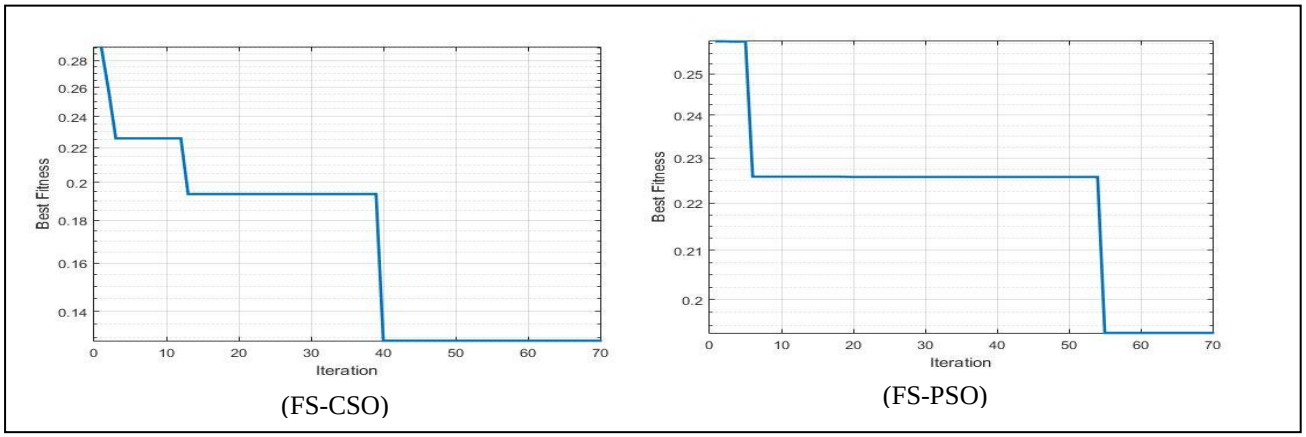
Algorithm	Parameter	Value
CSO	r	0.15
	h	0.5
	m	0.7
PSO	w	1
	c_1	1.5
	c_2	2.0

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

Dataset	All Features	CSO			PSO		
		K=3	K=4	K=5	K=3	K=4	K=5
C8 – LogP	0.5000	**0.0556**	**0.0556**	0.1667	0.1111	0.1111	0.1111
C8-BP	0.3333	**0.1667**	**0.1667**	**0.1667**	0.2222	**0.1667**	0.2222
C8-DHForm	0.5000	0.1667	**0.1111**	**0.1111**	0.2777	0.2222	0.1667
Phenet-LogP	0.3889	0.0999	**0.0455**	0.1364	0.1364	0.1366	0.1819
PAH-LogP	0.4444	**0.2317**	0.2439	0.2439	0.2683	0.2805	0.3171
PAH-BP	0.2778	**0.0976**	0.1098	0.1098	0.1220	0.1098	0.1098
PAH-MP	0.3333	0.0976	**0.0845**	0.0976	0.0976	0.0845	0.0976
WineEW	0.0449	**0.0112**	**0.0112**	**0.0112**	0.0169	0.0169	**0.0112**

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

Dataset	All features	BPSO-PSO	BPSO-BP	CSO			PSO		
				K=3	K=4	K=5	K=3	K=4	K=5
Selwood	0.4194	0.1032	0.0981	0.0968	**0.0645**	0.1613	0.1613	0.1935	0.2258



(FS-CSO) (FS-PSO)

Fig.1 Applying FS-CSO and FS-PSO on Selwood Dataset

10. REFERENCES

[1] N. Brown, "Chemoinformatics—an introduction for computer scientists", ACM Computing Surveys (CSUR), 2009, Vol. 41, Issue. 2, Article No. 8.

[2] 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.

[3] H. Gao, M. Lajiness, J. Drie, "Enhancement of binary QSAR analysis by a GA-based variable selection method", J Mol Graph Model, 2002, pp 259–268.

[4] M. Goodarzi, B. Dejaegher, Y. Heyden, "Feature Selection Method in QSAR Studies", Journal of AOAC International, May/Jun2012, Vol. 95 Issue 3, p636.

[5] M. Goodarzi,M. Freitas,R. Jensen, Ant colony optimization as a feature selection method in the QSAR modeling of anti-HIV-1 activities of 3-(3,5-dimethylbenzyl)uracil derivatives using MLR, PLS and SVM regressions, Chemometrics and Intelligent Laboratory Systems, Volume 98, Issue 2, 15 October 2009, pp 123-129, ISSN 0169-7439.

[6] 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.

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

[8] 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.

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

[10] 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.

[11] A. Leach,V. Gillet, "An Introduction to Chemoinformatics", Springer Netherlands, 2007, pp. 75-89.

[12] 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.

[13] 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.

[14] J. Mitchell, "Machine learning methods in chemoinformatics", WIREs Comput Mol Sci, 2014, 4: 468–481. doi:10.1002/wcms.1183.

[15] 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.

[16] C. Paris (1998). "Structure Databases. In Schleyer P von R, N L Allinger, T Clark, J Gasteiger, P A Kollman, H F Schaefer and P R Shreiner (Editors), The Encyclopedia of Computational Chemistry, Volume 4. Chichester, Wiley, pp 2771–2785.

[17] 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.

[18] A. Smalter, J. Huan, Y. Jia, G. Lushington. "GPD: A Graph Pattern Diffusion Kernel for Accurate Graph Classification with Applications in Cheminformatics". *IEEE/ACM transactions on computational biology and bioinformatics / IEEE, ACM*. 2010,7.2:197-207. doi:10.1109/TCBB.2009.80.

[19] 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.