International Journal of Engineering & Technology, 7 (2.27) (2018) 7-11

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper



EMOPS: an enhanced multi-objective pswarm based classifier for poorly understood cancer patterns

S. Subasree [1] *, N. P. Gopalan [2], N .K. Sakthivel [3]

[1] Research Scholar, Department of Computer Science and Engineering Bharath University, Chennai – 600 073, Tamil Nadu, India
[2] Department of Computer ApplicationsNational Institute of Technology, Thiruchirappalli, Tamil Nadu, India
[3] Department of Computer Science and EngineeringNehru College of Engineering and Research Centre, Thrissur, Tamil Nadu, India
**Corresponding author E-mail: drssubasree@gmail.com*

Abstract

Microarray based Cancer Pattern Classification is one of the popular techniques in Bioinformatics Research. This Research Work is noticed that for studying the expression levels through the Gene Expression profiling experiments, thousands of Genes have to be simultaneously studied to understand the patterns of the Gene Expression or Cancer Pattern. This research work proposed an efficient Cancer Pattern Classifier called An Enhanced Multi-Objective Pswarm (EMOPS) and it is studied thoroughly in terms of Memory Utilization, Execution Time (Processing Time), Sensitivity, Specificity, Classification Accuracy and FScore. The results were compared with the recently proposed classifiers namely Hybrid Ant Bee Algorithm (HABA), Kernelized Fuzzy Rough Set Based Semi Supervised Support Vector Machine (KFRS-S3VM) and Multi-objective Particle Swarm Optimization (MPSO). For analyzing the performances of the proposed model, this work considered a few cancer patterns namely Bladder, Breast, Colon, Endometrial, Kidney, Leukemia, Lung, Melanoma, Mom-Hodgkin, Pancreatic, Prostate and Thyroid. From our experimental results, it was noticed that the proposed model outperforms the identified three classifiers in terms of Memory Utilization, Execution Time (Processing Time), Sensitivity, Specificity, Classification Accuracy and FScore. To improve the performance of the system further in term of Processing Time, the proposed model Enhanced Multi-Objective Pswarm (EMOPS) is implemented under Parallel Framework and evaluated. That is the model is tested with Two, Four, Eight and Sixteen Parallel Processors and from the results, it is established that the Processing Time decreases considerably which will improve the performance of the Proposed Model.

Keywords: *Cancer Pattern Classifications; Gene Expression; Microarray, Multi-Objective Pswarm; Parallel Framework; Support Vector Machine.*

1. Introduction

This Microarray is a significant technology which facilitating to study various gene expressions. The microarray data, in general, are images and these microarray images could be converted into various gene expression. These Gene Expressions have been usually used for Gene Pattern Classifications. From the available literature survey [1-6], it was noticed that the Data Mining Techniques are facilitating to classify and predict various Cancer Gene Patterns.
The Classifiers are used to classify microarray samples for pattern classification. ie the normal microarray sample data set and cancer pattern samples can be classified with the help of Classifiers [12-16]. If the samples had a few subtypes of cancer pattern, then we needed multiclass cancer pattern classifiers [1-4]. From the literature survey, it was noticed that the Multi-Class Cancer Pattern Classifier can be employed to improve the classification accuracy [17].
This research work identified a few popular Multi-Class Classifiers which are recently proposed for Cancer Patter Prediction/Classification and all those Classifiers were discussed below.
The proposed model Enhanced Multi-Objective Pswarm (EMOPS) was implemented with Uni-Processor [1] and Parallel Processors as well. The detailed procedure of the Parallel Framework was discussed in the following section.
This Research paper is arranged and written as follows. The Section 2 briefly described the recently proposed Data Mining Classifiers. The proposed model, Enhanced Multi-Objective Pswarm Based Classifier (EMOPS) is implemented in Uni-Processing and Parallel Framework as well is described in Section 3. The results and strengths of the proposed model in Uni-Processing as well as Parallel Processing is discussed at Section 4 and Conclusion was given in Section 5

2. Recently proposed data mining classifiers

The characteristics and procedures of the three identified Classifiers namely i. Hybrid Ant Bee Algorithm (HABA) [4], ii. Kernelized Fuzzy Rough Set Based Semi Supervised Support Vector Machine (KFRS-S3VM) [1] and iii. Multi-objective Particle Swarm Optimization (MPSO) [6], [21-24] have been discussed in the following subsections.

2.1. Hybrid ant bee algorithm (HABA)

Ant Colony Optimization [1], [4], [10], [26] does maintain a colony of ants and make possible Permissible Ranges (PRs) in association with values proposed for a design model. Here, each and every ant is permitted to select a Permissible Range which will represent the path.
When all ants have chosen their paths, then the discrete value associated with the selected path is taken and for all ants, this is considered as candidate value. Then, the system evaluates the Artificial Bee Colony Approach by combining the candidate values of all the

Copyright © 2018 S. Subasree et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ants and this initializes the food source and the objective function can be evaluated with three phases and those phases named as i. Employed Bee Phase where Food sources assigned to Bees, ii. Onlooker Bee Phase, where a decision is taken by Bees and iii. Scout Bee Phase, where ants making out the random search. The proposed Ant Bee Algorithm combines the strength of Artificial Bee Colony (ABC) and Ant Colony Optimization (ACO). The procedure of Ant Bee Algorithm is described below.

Generate Initial solution space
Evaluate the Fitness of Objective function
if (Fitness Function Converged)
{ declare best solution
stop() }
Spilt the Database as Clusters
ACO()
//probabilistic based optimization
{ Set Parameters, Initialize Pheromone Trails
Construct path
Select and Construct Ant Solution
Update Pheromones }
ABC()
// Optimizes through ABC Algorithm
// Cluster based optimization based on intelligent foraging behaviour of bee
{ // No. of Parameters D; //Function fn;//No. of Bees NB;
// Lower Bound lb; //Upeer Bound ub;
Declare par, fn,D,NP,lb,ub,limit;
Initialization of parameter par=0
If(NP<limit)
{ abc_optim(par, fn, D=length(par) } }
Combine the results of ABC() and ACO()
Construct Solution

2.2. Kernelized fuzzy rough set based semi supervised support vector machine (KFRS-S3VM)

The Kernelized Fuzzy Rough Set (KFRS) [5] is used to classify Cancer Patterns and used to classify Gene Expressions from the Microarray datasets [5], [7-8]. The KFRS-S3VM has two popular feature selection techniques, namely i. Fuzzy Preference Based Rough Set (FPRS) and ii. Consistency Based Feature Selection (CBFS).
Gene Expressions based validations have done in this Scheme, which shown in the detailed procedure [5], [25]. The Forward Greedy Search Algorithm based Gaussian Kernel Approximation [4], [18-20] was designed as follows.

Input: Sample set $U = \{Z_1, Z_2, ...Z_m\}$, feature set A, decision F and stopping threshold ε
Output: Reduct red
Step 1: Initialize red to an empty set and β to 0
Step 2: For each attribute $a_i \varepsilon A - red$, red,
Compute $\beta_i = \beta_{\{a_i\}} \bigcup red$
Step 3: Find the maximal βi and the corresponding attribute ai
Step 4: Add attribute ai to red if it satisfies $\beta_i - \beta_{red}(F) > \varepsilon$
Step 5: Assign βi_to βred
Step 6: Repeat steps 2 to 5 while red ≠A
Step 7: Return red

The above procedure of Gaussian Kernel Approximation is initially starting with a null set of attribute and it is evaluating the all other remaining attributes in iterations and also it is selecting various features identifying by the Maximal Fuzzy Dependency [5], [8-9]. The fuzzy dependency (F) is calculated as follows [1], [3].

Input: Sample set $U = \{Z_1, Z_2, ...Z_m\}$, feature set A, decision F and parameters δ
Output: dependency β of F to A
Step 1: βA (f) ← 0
Step 2: i = 1 to m
Step 3: Find the nearest sample xi to zi with different class

$$Step\ 4: \ \beta_A(F) \leftarrow \beta_A(F) + \sqrt{1 - \left[exp\left(-\frac{\|zi - xi\|^2}{\delta}\right)\right]^2}$$

Step 5: Return $\beta_A(F)$
The algorithm will remove low dependency values those features that received from the data sets.

2.3. Multi-objective particle swarm optimization (MPSO)

The Particle Swarm Optimization[6] is one of the popular existing population based optimization techniques. The various candidate solutions are named as Particle and the population of these Particles is termed as Swarm.
Let us consider that there were N Particles in Swarm to achieve optimal fitness. The Particle Best Position pbest and Global Best Position gbest need to update to attain and compute fitness [27-30]. The MPSO was developed[5] by the authors Anirban Mukhopadhyay and et. al. as follows.

1) Input i. Data Matrix ii. Cluster Center C, iii. Particles N, iv. Samples S, v. Assign thr = 0.5, Sample Velocity SV
2) Output A
   a) Initialize Random Sample Locations and SVs as well
      i) Genes x_n, Samples Gene Set G_n, and Fitness P_n
   b) Initialize Random Sample Locations and SV as well
      ii) Calculate CellBoundary(xnd) for all cluster Centres till xnd $\geq$ Threshold
   c) Calculate CellBoundary and average Velocity Vnd
   d) Select Centres by evaluating and combining
   e) Take Average Calculation by crowding distance sorting for all derived solutions Select the best Sample Gene Gn

2.4. Identified problem

This research work has implemented the above discussed three Classifiers and studied thoroughly with a few Cancer Patterns in terms of Memory Utilization, Execution Time (Processing Time), Sensitivity, Specificity, Classification Accuracy and FScore. From our experimental results, it was noticed that the performances of these three classifiers are strongly depend on the patterns of the Gene/Cancer pattern. It was also noted that the Multi-objective Particle Swarm Optimization (MPSO) is relative outperforming other two classifiers. To improve the performance of the Multi-objective Particle Swarm Optimization (MPSO), this paper enhanced Multi-objective Particle Swarm Optimization (MPSO) and named as an Enhanced Multi-Objective Pswarm Based Classifier (EMOPS) and described in the following section.

3. EMOPS : an enhanced multi-objective pswarm based classifier

It was also noted that the Multi-objective Particle Swarm Optimization (MPSO) is relative outperforming other two classifiers. To improve the performance of the Multi-objective Particle Swarm Optimization (MPSO), this paper enhanced Multi-objective Particle Swarm Optimization (MPSO) and named as an Enhanced Multi-Objective Pswarm Based Classifier (EMOPS) and described in the following section.

3.1. Procedure of enhanced multi-objective pswarm based classifier (EMOPS)

As discussed in the previous section, the Multiobjective Particle Swarm Optimization (MPSO) considers the total number of particles to achieve optimal fitness. The Particle Best Position pbest and Global Best Position gbest will update to attain and compute fitness.
This research work noticed that the position and parameter values need to optimize in such a way to achieve a high level of Classification Accuracy. ie need to determine optimized centre

values to improve and achieve higher classification accuracy. To achieve higher classification accuracy, this work proposed an efficient model called an Enhanced Multi-Objective Pswarm Based Classifier (EMOPS). The procedure of this work will consider multiple competing solutions to find Global Best Position gbest, which will improve Classification and Prediction accuracy. The procedure for the Enhanced Multi-Objective Pswarm Based Classifier (EMOPS) is given below.

1) Input i. Data Matrix ii. Cluster Center C, iii. Particles N, iv. Samples S, v. Assign thr = 0.5, Sample Velocity SV
2) Output A
   a) Initialize Random Sample Locations and SVs as well
      i) Genes xn, Samples Gene Set Gn, and Fitness Pn
   b) Initialize Random Sample Locations and SVs as well
      i) Calculate CellBoundary(xnd) for all cluster Centres till xnd $\geq$ Threshold
   c) Calculate CellBoundary and average Velocity Vnd
   d) Calculate
      i) Strong-dominance updating strategy
   a) Compute Crowding Distance and Refresh for next Iteration
   b) Estimates the largest rectangle size
   c) Calculate the average distance of its two neighbouring solutions
   d) Select Centres by evaluating and combining
   e) Take Average Calculation by crowding distance sorting for all derived solutions
      i) Select the best Sample Gene Gn
   f) Select the Global Best Position gbest

3.2. Parallel computing framework

The proposed Enhanced Multi-Objective Pswarm Based Classifier (EMOPS) was implemented under Parallel Computing Framework to improve the performance of the proposed model in term of Execution Time. The Parallel Architecture was illustrated in the Fig. 1. The Model is designed by Parallel Framework to predict the Cancer Pattern. That is this work has implemented with One Processor, Two Processors, Four Processors, Eight Processors and 16 Processors.

As shown in the Fig. 1, the Parallel-Enhanced Multi-Objective Pswarm Based Classifier (EMOPS) has Multiple Populations, Objectives and Data Sets, which is created as Particle Swarm Decomposition. The decomposed Multiple Data Sets are allocated to Parallel Computing Resources for Cancer Pattern Classification/Prediction. These divided Data Sets will provide partial solution and it facilitates to find Global Solution which will provide final Classification / Prediction Pattern.

4. Performance analysis

This Research Work conducts Simulations to study the performances and classification abilities of the proposed model, Enhanced Multi-Objective Pswarm Based Classifier (EMOPS). The Cancer Genome Sequence Data Sets[11] namely NCBI.CGS.MER and NCBI.CS.MER are used to analysis the proposed model. The Simulation was performed as shown in Fig 1. For the Simulations, the various cancer patterns' are considered and the name of those patterns are i. Bladder, ii. Breast, iii. Colon. iv. Endometrial, v. Kidney, vi. Leukemia, vii. Lung, viii. Melanoma, ix. Mom-Hodgkin, x. Pancreatic, xi. Prostate and xii. Thyroid.

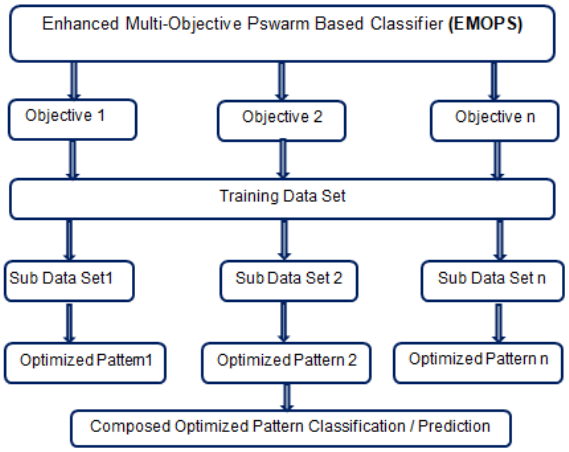


Fig. 1: Parallel Architecture of the Enhanced Multi-Objective Pswarm Based Classifier (EMOPS).

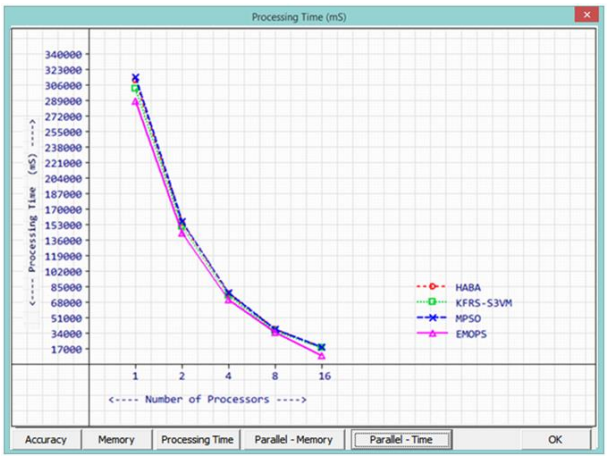


Fig. 2: Processing Time in MS (Parallel Processing) vs. Classifiers

The performance of the proposed Classifier was tested in terms of Execution Time (Processing Time), Sensitivity, Specificity, Classification Accuracy, FScore, and Memory Utilization. This work is developed an Interfacing Tool with the VC++ Programming Language to extract and validate the Gene Expressions which are downloaded from NCBI. The validated data is fed into BioWeka Simulation Tool for analyzing the performances of the proposed Classifier in terms of Execution Time (Processing Time), Sensitivity, Specificity, Classification Accuracy, FScore, and Memory Utilization.

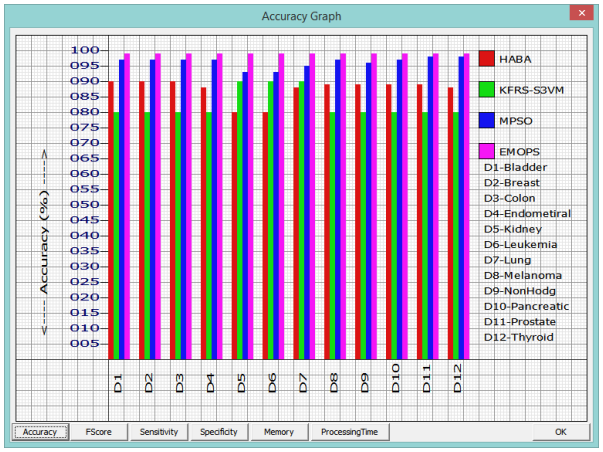


Fig. 3: Accuracy vs. Classifiers.

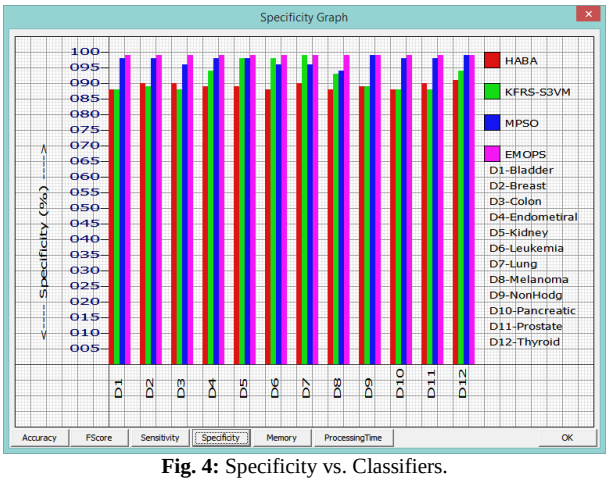


Fig. 4: Specificity vs. Classifiers.

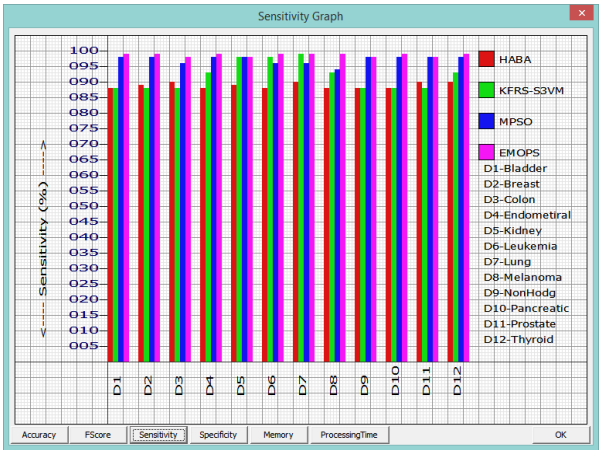


Fig. 5: Sensitivity vs. Classifiers.

The proposed Classifier EMOPS was implemented and studied thoroughly. The results were compared with the performances of the existing classifiers namely Hybrid Ant Bee Algorithm (HABA), Kernelized Fuzzy Rough Set Based Semi Supervised Support Vector Machine (KFRS-S3VM) and Multiobjective Particle Swarm Optimization (MPSO) which are illustrated from the Fig. 1 to Fig. 7. From the results, it was noticed that the proposed model outperforms the existing identified models in terms of Execution Time (Processing Time), Sensitivity, Specificity, Classification Accuracy, FScore, and Memory Utilization.

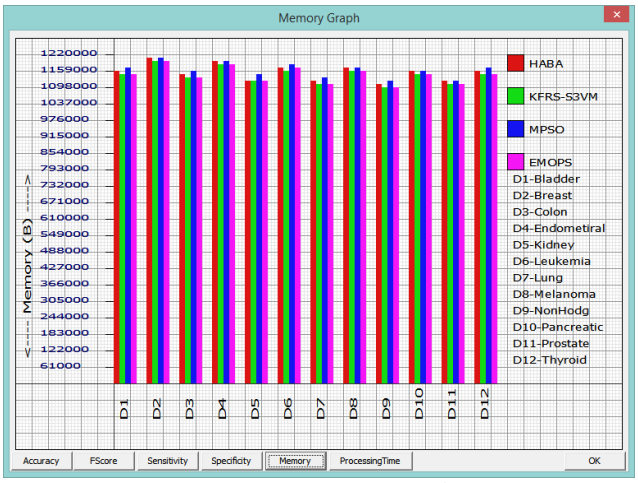


Fig. 6: Memory Usage vs. Classifiers.

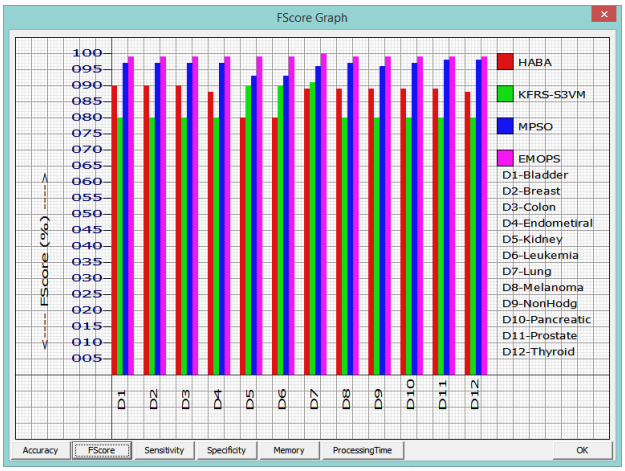


Fig. 7: FSCORE vs. Classifiers.

The experiment with Multi-Processors say 2, 4, 8 and 16 Processors was repeated number of times and average probabilities for predicting possible Cancer Patterns were recorded. It was noted that the Execution Time was reduced as number of processors involved were increased for Classification / Prediction.

5. Conclusion

This research work proposed an efficient Cancer Pattern Classifier called An Enhanced Multi-Objective Pswarm (EMOPS) and studied thoroughly. From our experimental results, it was noticed that the proposed model outperforms the identified three classifiers namely Hybrid Ant Bee Algorithm (HABA), Kernelized Fuzzy Rough Set Based Semi Supervised Support Vector Machine (KFRS-S3VM) and Multiobjective Particle Swarm Optimization (MPSO) in terms of Memory Utilization, Execution Time (Processing Time), Sensitivity, Specificity, Classification Accuracy and FScore. It is further observed that the execution time executed under Parallel Architecture is relatively lesser than that of Execution Time by Uni Processing.

References

[1] S. Subasree, N. P. Gopalan and N. K. Sakthivel, “A Comparative Study and Analysis of Data Mining Classifiers for Microarray based Cancer Pattern Diagnostics”, ACM International Conference on Informatics and Analytics (ICIA-16) (2016). https://doi.org/10.1145/2980258.2982039.

[2] Xu Gu, “A Multi-State Optimisation Framework for Parameter Estimation in Biological Systems”, IEEE/ACM Transactions on Computational Biology and Bioinformatics, (2016). https://doi.org/10.1109/TCBB.2015.2459686.

[3] Iman Behravan, Seyed Hamid Zahiri, Oveis Dehghantanha, “An Optimal SVM with Feature Selection using Multi- Objective PSO”, 1st Conference on Swarm Intelligence and Evolutionary Computation (CSIEC2016), Higher Education Complex of Bam, Iran, (2016).

[4] Pugalendhi Ganesh Kumar, et.al, “Hybrid Ant Bee Algorithm for Fuzzy Expert System Based Sample Classification”, IEEE Transactions on Computational Biology and Bioinformatics. Vol. 11, No. 2, Pp. 347-360, (2014). https://doi.org/10.1109/TCBB.2014.2307325.

[5] Debasis Chakraborty and Ujjwal Maulik, “Identifying Cancer Biomarkers From Microarray Data Using Feature Selection and Semi-supervised Learning”, IEEE Journal of Translational Engineering on Health and Medicine. Vol 2. (2014). https://doi.org/10.1109/JTEHM.2014.2375820.

[6] Anirban Mukhopadhyay and Monalisa Mandal, “Identifying Non-Redundant Gene Markers from Microarray Data: A Multiobjective Variable Length PSO-Based Approach”, IEEE/ACM Transactions on Computational Biology and Bioinformatics. Vol. 11, No. 6. (2014). https://doi.org/10.1109/TCBB.2014.2323065.

[7] Y. Yoon, S. Bien, and S. Park, “Microarray Data Classifier Consisting of k-Top-Scoring Rank-Comparison Decision Rules with a Variable Number of Genes”, IEEE Trans. Systems, Man, and Cybernetics Rev.Vol. 40, no. 2, Pp. 216-226, (2010). https://doi.org/10.1109/TSMCC.2009.2036594.

[8] Y. Hu, "Fuzzy Integral-Based Perceptron for Two-Class Pattern Classification Problems", Information Sciences. vol. 177, no. 7, Pp. 1673-1686, (2007). https://doi.org/10.1016/j.ins.2006.09.009.

[9] Z. Wang and V. Palade, "A Comprehensive Fuzzy-Based Framework for Cancer Microarray Data Gene Expression Analysis", Proceedings of IEEE International Conference of Bioinformatics and Bioengineering. Pp. 1003-1010, (2007). https://doi.org/10.1109/BIBE.2007.4375680.

[10] D. Devaraj and B. Yegnanarayana, "Genetic Algorithm-Based Optimal Power Flow for Security Enhancement", IEEE Proc. Generation, Transmission and Distribution, vol. 152, no. 6, Pp. 899-905, (2005). https://doi.org/10.1049/ip-gtd:20045234.

[11] Online http://www.biolab.si/supp/bi-cancer/projections/

[12] Witten, Ian H., and Eibe Frank, "Data Mining: Practical Machine Learning Tools and Techniques", Morgan Kaufmann, (2005).

[13] Zaki, Mohammed J., and Wagner Meira Jr, "Data Mining and Analysis: Fundamental Concepts and Algorithms", Cambridge University Press, (2014).

[14] Leskovec, Jure, Anand Rajaraman, and Jeffrey David Ullman, "Mining of Massive Datasets," Cambridge University Press, (2014). https://doi.org/10.1017/CBO9781139924801.

[15] Nikam, Sagar S. "A Comparative Study of Classification Techniques in Data Mining Algorithms," Oriental Journal of Computer Science & Technology, Pp.13-19, (2015).

[16] Han, Jiawei, Jian Pei, and Micheline Kamber, "Data Mining: Concepts and Techniques," Elsevier, (2011).

[17] Delveen Luqman Abd Al.Nabi, Shereen Shukri Ahmed, "Survey on Classification Algorithms for Data Mining (Comparison and Evaluation)," (ISSN 2222-2863) 4(8), (2013).

[18] R. S. Gunn, "Support Vector Machines fr Classification and Regression," ISIS technical report, vol. 14, (1998).

[19] Riaan Smit, "An Overview of Support Vector Machines," 2011.

[20] R. Srinivas, "Managing Large Data Sets Using Support Vector Machines," University of Nebraska at Lincoln, (2010).

[21] Y. Zhang, S. Wang, and G. Ji, "A comprehensive survey on Particle Swarm Optimization Algorithm and its applications," Mathematical Problems in Engineering, (2015). https://doi.org/10.1155/2015/931256.

[22] Trivedi, D. Srinivasan, K. Sanyal, and A. Ghosh, "A survey of Multi-Objective Evolutionary Algorithms based on decomposition," IEEE Transactions on Evolutionary Computation, (2016). https://doi.org/10.1109/TEVC.2016.2608507.

[23] Li, J. Li, K. Tang, and X. Yao, "Many-Objective Evolutionary Algorithms: A survey," ACM Computing Surveys (CSUR), Vol. 48, No. 1, (2015). https://doi.org/10.1145/2792984.

[24] J. Wang, W. Zhang, and J. Zhang, "Cooperative differential evolution with multiple populations for Multiobjective Optimization," IEEE Xplore: IEEE Transactions on Cybernetics, Vol. 46, No. 12, Pp. 2848-2861, (2016). https://doi.org/10.1109/TCYB.2015.2490669.

[25] Y.-J. Gong et al., "Distributed evolutionary algorithms and their models: A survey of the state-of-the-art," Applied Soft Computing, Vol. 34, Pp. 286-300, (2015). https://doi.org/10.1016/j.asoc.2015.04.061.

[26] Y. Xue, J. Jiang, B. Zhao, and T. Ma, "A Self-Adaptive Artificial Bee Colony Algorithm based on Global Best for Global Optimization," Soft Computing, Pp. 1-18, (2017).

[27] F. Qiu, L. Mo, B. Jiang, and L. Wang, "Multi-Objective Particle Swarm Optimization Algorithm Using Large Scale Variable Decomposition," Chinese Journal of Computers - Institute Of Computing Technology, Vol. 39, No. 12, Pp. 2598-2613, (2016).

[28] C. A. C. Coello and M. S. Lechuga, "MOPSO: A proposal for Multiple Objective Particle Swarm Optimization," IEEE Congress on Evolutionary Computation (CEC), Vol. 2. Pp. 1051-1056, (2016).

[29] B. Chen, W. Zeng, Y. Lin, and D. Zhang, "A New Local Search-Based Multi-Objective Optimization Algorithm," IEEE Xplore: IEEE Transactions on Evolutionary Computation, Vol. 19, No. 1, Pp. 50-73, (2015). https://doi.org/10.1109/TEVC.2014.2301794.

[30] X. Ma et al., "A Multi-Objective Evolutionary Algorithm based on Decision Variable Analyses for Multi-Objective Optimization Problems with Largescale Variables," IEEE Xplore: IEEE Transactions on Evolutionary Computation, Vol. 20, No. 2, Pp. 275-298, (2016). https://doi.org/10.1109/TEVC.2015.2455812.