Journal of Engineering and Applied Sciences 13 (3): 580-587, 2018
ISSN: 1816-949X
© Medwell Journals, 2018

EMOPS: An Enhanced Multi-Objective Particle Swarm Based Classifier for Poorly Understood Cancer Patterns

[1]S. Subasree, [2]N.P. Gopalan and [3]N.K. Sakthivel
[1]Department of Computer Science and Engineering, Bharath University,
600073 Chennai, Tamil Nadu, India
[2]Department of Computer Applications, National Institute of Technology,
Thiruchirappalli, Tamil Nadu, India
[3]Department of Computer Science and Engineering,
Nehru College of Engineering and Research Centre, Thrissur, Tamil Nadu, India

Abstract: Microarray based cancer pattern classification is one of the popular techniques in bioinformatics research. At the same time, it was 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 proposed an efficient cancer pattern classifier called an Enhanced Multi-Objective Particle Swarm (EMOPS) and it is studied thoroughly in terms of memory utilization, execution time (processing time), sensitivity, specificity, classification accuracy and F-score. The results were compared with that of 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 research 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 F-score.

Key words: Cancer pattern classifications, microarray, multi-objective Particle Swarm, SVM, gene expression, classification

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 (Gu, 2016; Behravan *et al*., 2016; Kumar *et al*., 2014; Chakraborty and Maulik, 2014; Mukhopadhyay and Mandal, 2014; Yoon *et al*., 2010), 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, i.e., the normal samples of the microarray data set and cancer pattern samples can be classified with the help of classifiers. If the samples had a few subtypes of cancer pattern, then we needed multiclass cancer pattern classifiers (Gu, 2016; Behravan *et al*., 2016; Kumar *et al*., 2014; Chakraborty and Maulik, 2014). From the literature survey, it was noticed that the multi-class cancer pattern classifier can be employed to improve the classification accuracy.

This research identified a few popular multi-class classifiers which are recently proposed for cancer patter prediction/classification and all those classifiers were discussed.

Recently proposed data mining classifiers: The characteristics and procedures of the three identified Classifiers namely Hybrid Ant Bee Algorithm (HABA) (Chakraborty and Maulik, 2014), Kernelized Fuzzy Rough Set based semi supervised Support Vector Machine (KFRS-S3VM) (Gu, 2016) and Multiobjective Particle Swarm Optimization (MPSO) (Yoon *et al*., 2010) have been discussed in the following study.

Corresponding Author: N.P. Gopalan, Department of Computer Applications, National Institute of Technology, Thiruchirappalli, Tamil Nadu, India

Hybrid Ant Bee Algorithm (HABA): Ant colony optimization (Gu, 2016; Chakraborty and Maulik, 2014) 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 ants and this initializes the food source and the objective function can be evaluated with three phases and those phases named as employed bee phase where food sources assigned to bees, onlooker bee phase where a decision is taken by bees and 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.

Algorithm 1:

```
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
```

Kernelized Fuzzy Rough Set based semi supervised Support Vector Machine (KFRS-S3VM): The Kernelized Fuzzy Rough Set (KFRS) (Mukhopadhyay and Mandal, 2014) is used to classify cancer patterns and used to classify gene expressions from the microarray datasets (Mukhopadhyay and Mandal, 2014; Hu, 2007; Wang and Palade, 2007). The KFRS-S3VM has two popular feature selection techniques, namely Fuzzy Preference based Rough Set (FPRS) and Consistency Based Feature Selection (CBFS).

Gene expressions based validations have done in this scheme which shown in the detailed procedure (Mukhopadhyay and Mandal, 2014). The forward greedy search algorithm based Gaussian Kernel approximation (Chakraborty and Maulik, 2014) was designed as follows.

Algorithm 2:

```
Input: Sample set U = {$Z_1$, $Z_2$, ..., $Z_m$} feature set A, decision F and stopping threshold
Output: Reduct red
Step 1: Initialize red to an empty set and $\beta$ to 0
Step 2: For each $a_i \varepsilon$A-red, red, compute $\beta_i = \beta\{a_1\} \cup$red
Step 3: Find the maximal $\beta$i and the corresponding attribute $a_i$
Step 4: Add attribute $a_i$ to red if it satisfies $\beta_i$-$\beta_{red}$(F)>$\varepsilon$
Step 5: Assign $\beta_{i\_}$to $\beta_{red}$
Step 6: Repeat steps 2-5 while red $\neq$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 (Mukhopadhyay and Mandal, 2014; Wang and Palade, 2007; Devaraj and Yegnanarayana, 2005). The Fuzzy dependency (F) is calculated as follows (Gu, 2016; Kumar *et al*., 2014):

Algorithm 3:

```
Input: Sample set U = {$Z_1$, $Z_2$, ..., $Z_m$} feature set A, decision F and parameters $\delta$
Output: dependency $\beta$ of F-A
Step 1: $\beta_A$(f)$\leftarrow$0
Step 2: i = 1-m
Step 3: Find the nearest sample $x_i$ to $z_i$ with different class:
Step 4: $\beta_A(F) \leftarrow \beta_A(F) + \sqrt{1 - \left[\exp\left(-\frac{\|z_i - x_i\|^2}{\delta}\right)\right]}$
Step 5: Return $\beta_A$(F)
```

The algorithm will remove low dependency values those features that received from the data sets.

Multi-Objective Particle Swarm Optimization (MPSO): The particle swarm optimization (Yoon *et al*., 2010) 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. The MPSO was developed by the researchers Mukhopadhyay and Mandal (2014) as follows.

Algorithm 4:

```
1. Input, Data matrix, Cluster center C, Particles N, Samples S,
   Assign thr = 0.5
2. Output A
   a. Initialize random locations and velocities as well
      i. Genes $x_n$, samples gene set $G_n$ and fitness $P_n$
   b. Initialize random locations and velocities as well
      i. Calculate cell boundary (xnd) for all cluster centres till
   xnd≥Threshold
   c. Calculate cell boundary and average velocity Vnd
   d. Select centres by evaluating and combining
   e. Take average calculation by crowding distance sorting for all derived
      solutions
3. Select the best sample gene Gn
```

Identified problem: This research 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 F-score. 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 study enhanced Multi-objective Particle Swarm Optimization (MPSO) and named as an Enhanced Multi-Objective Particle Swarm based classifier (EMOPS) and described in the following study.

MATERIALS AND METHODS

EMOPS: An Enhanced Multi-Objective Particle Swarm based classifier: As discussed in the previous study, 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 noticed that the position and parameter values need to optimize in such a way to achieve a high level of classification accuracy, i.e., need to determine optimized centre values to improve and achieve higher classification accuracy. To achieve higher classification accuracy this research proposed an efficient model called an Enhanced Multi-Objective Particle Swarm based classifier (EMOPS). The procedure of this research 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 Particle Swarm based classifier (EMOPS) is given.

Algorithm 5:

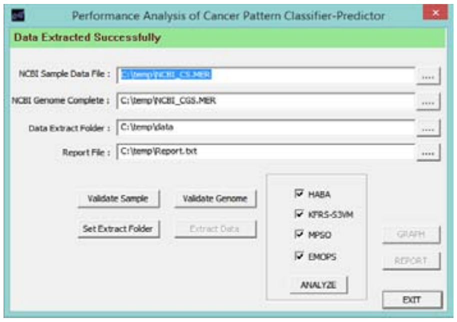
```
1. Input i. Data matrix ii. Cluster center C, iii. Particles N, iv. Samples S,
   v. Assign thr = 0.5
2. Output A
   a. Initialize random locations and velocities as well
      i. Genes $x_n$, samples gene set $G_n$ and fitness $P_n$
   b. Initialize random locations and velocities as well
      I. Calculate cellboundary(xnd) for all cluster centres till xnd≥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 size of the largest rectangle
         c. Takes 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
      ii. Select the best sample gene Gn
   e. Select the global best position gbest
```

RESULTS AND DISCUSSION

Performance analysis: This research conducts Simulations to study the performances and classification abilities of the proposed model, Enhanced Multi-Objective Particle Swarm based classifier (EMOPS). The cancer genome sequence data sets namely NCBI.CGS.MER and NCBI.CS.MER are used to analysis the proposed model. The simulation interface is shown in Fig. 1. For the simulations, the various cancer pattern's are considered and the name of those patterns are bladder, breast, colon, endometrial, kidney, leukemia, lung, melanoma, mom-hodgkin, pancreatic, prostate and thyroid.

The performance of the proposed classifier was tested in terms of execution time (processing time), sensitivity, specificity, classification accuracy, F-score and memory utilization. This research is developed an interfacing tool with the VC++ programming language to extract and validate the gene expressions which are downloaded from NCBI that is shown in Fig. 1. 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, F-score and memory utilization.

The proposed classifier EMOPS was implemented and studied thoroughly. The results were compared with


Performance Analysis of Cancer Pattern Classifier-Predictor

Data Extracted Successfully

NCBI Sample Data File : C:\temp\NCBI_CS.MER

NCBI Genome Complete : C:\temp\NCBI_CGS.MER

Data Extract Folder : C:\temp\data

Report File : C:\temp\Report.txt

☑ HABA
☑ KFRS-S3VM
☑ MPSO
☑ EMOPS

Fig. 1: VC++ based interface tool

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). The results are presented in Table 1-4. The outputs were plotted and shown in Fig. 2-7. 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, F-score and memory utilization.

Table 1: Performance analysis of EMOPS

Cancer type	Accuracy	F-score	Memory (B)	Processing time (msec)	Sensitivity	Specificity
Procedure: enhanced muti-objective particle swarm optimization						
Bladder	0.9921	0.9991	1158010	318818	0.9905	0.9917
Breast	0.9927	0.9920	1199450	322010	0.9934	0.9970
Colon	0.9934	0.9917	1145858	318314	0.9895	0.9924
Endometerial	0.9940	0.9945	1187298	321506	0.9956	0.9977
Kidney	0.9947	0.9974	1133692	317810	0.9885	0.9931
Leukemia	0.9953	0.9971	1175146	321026	0.9946	0.9916
Lung	0.9960	1	1121540	317306	0.9976	0.9970
Melanoma	0.9966	0.9929	1162980	320522	0.9937	0.9923
Mon-Hodg	0.9973	0.9926	1109388	316826	0.9898	0.9977
Pancreatic	0.9980	0.9954	1150828	320018	0.9927	0.9962
Prostate	0.9986	0.9983	1113700	316322	0.9888	0.9916
Thyroid	0.9993	0.9980	1155154	319514	0.9949	0.9916

Table 2: Performance analysis of MOPS

Cancer type	Accuracy	F-score	Memory (B)	Processing time (msec)	Sensitivity	Specificity
Procedure: enhanced muti-objective particle swarm optimization						
Bladder	0.9721	0.9791	1188210	331818	0.9805	0.9817
Breast	0.9727	0.9720	1219650	335010	0.9834	0.9870
Colon	0.9734	0.9717	1166058	331314	0.9617	0.9646
Endometerial	0.9740	0.9745	1207498	334506	0.9847	0.9868
Kidney	0.9335	0.9374	1153892	330810	0.9805	0.9851
Leukemia	0.9353	0.9371	1195346	334026	0.9636	0.9606
Lung	0.9560	0.9600	1141740	330306	0.9666	0.9660
Melanoma	0.9766	0.9729	1183180	333522	0.9479	0.9465
Mon-Hodg	0.9673	0.9626	1129588	319826	0.9832	0.9911
Pancreatic	0.9780	0.9754	1171028	323018	0.9861	0.9896
Prostate	0.9886	0.9883	1133900	319322	0.9814	0.9842
Thyroid	0.9893	0.9880	1175354	322514	0.9883	0.9842

Table 3: Performance analysis of HABA

Cancer type	Accuracy	F-score	Memory (B)	Processing time (msec)	Sensitivity	Specificity
Procedure: enhanced muti-objective particle swarm optimization						
Bladder	0.9021	0.9091	1168210	331818	0.8837	0.8849
Breast	0.9027	0.9020	1209650	335010	0.8966	0.9002
Colon	0.9034	0.9017	1156058	331314	0.9027	0.9056
Endometerial	0.8840	0.8845	1197498	334506	0.8888	0.8909
Kidney	0.8047	0.8074	1133892	330810	0.8917	0.8963
Leukemia	0.8053	0.8071	1175346	334026	0.8878	0.8848
Lung	0.8860	0.8900	1131740	330306	0.9008	0.9002
Melanoma	0.8966	0.8929	1173180	333522	0.8869	0.8855
Mon-Hodg	0.8973	0.8926	1119588	329826	0.8830	0.8909
Pancreatic	0.8980	0.8954	1161028	333018	0.8859	0.8894
Prostate	0.8986	0.8983	1123900	329322	0.9020	0.9048
Thyroid	0.8893	0.8880	1165354	332514	0.9081	0.9048

Table 4: Performance analysis of KFRS-S^3VM

Cancer type	Accuracy	F-score	Memory (B)	Processing time (msec)	Sensitivity	Specificity
Procedure: enhanced muti-objective particle swarm optimization						
Bladder	0.8021	0.8091	1158210	321818	0.8837	0.8849
Breast	0.8027	0.8020	1199650	325010	0.8966	0.8902
Colon	0.8034	0.8017	1146058	321314	0.8827	0.8856
Endometerial	0.8040	0.8045	1187498	324506	0.9388	0.9409
Kidney	0.9047	0.9074	1128892	310810	0.9817	0.9863
Leukemia	0.9053	0.9071	1170346	314026	0.9878	0.9848
Lung	0.9060	0.9100	1116740	310306	0.9908	0.9902
Melanoma	0.8066	0.8029	1163180	323513	0.9369	0.9355
Mon-Hodg	0.8073	0.8026	1109588	319826	0.8830	0.8909
Pancreatic	0.8080	0.8054	1151028	323018	0.8859	0.8894
Prostate	0.8086	0.8083	1113900	319322	0.8820	0.8848
Thyroid	0.8093	0.8080	1155354	322514	0.9381	0.8848

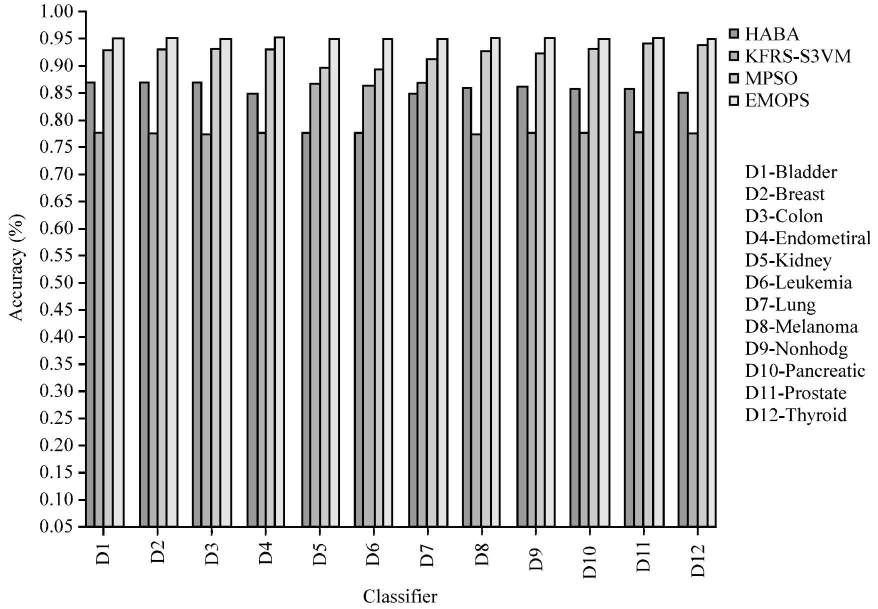


Fig. 2: Accuracy vs. classifiers

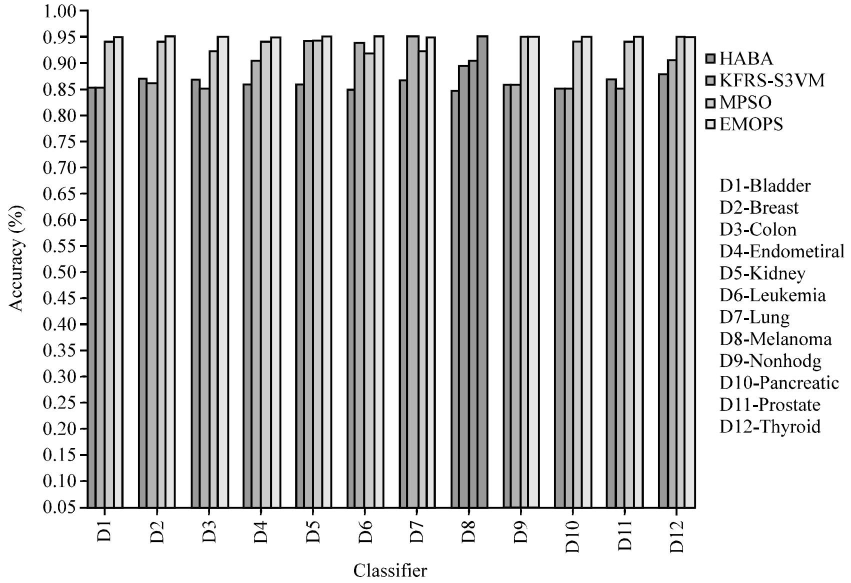


Fig. 3: Specificity vs. classifiers

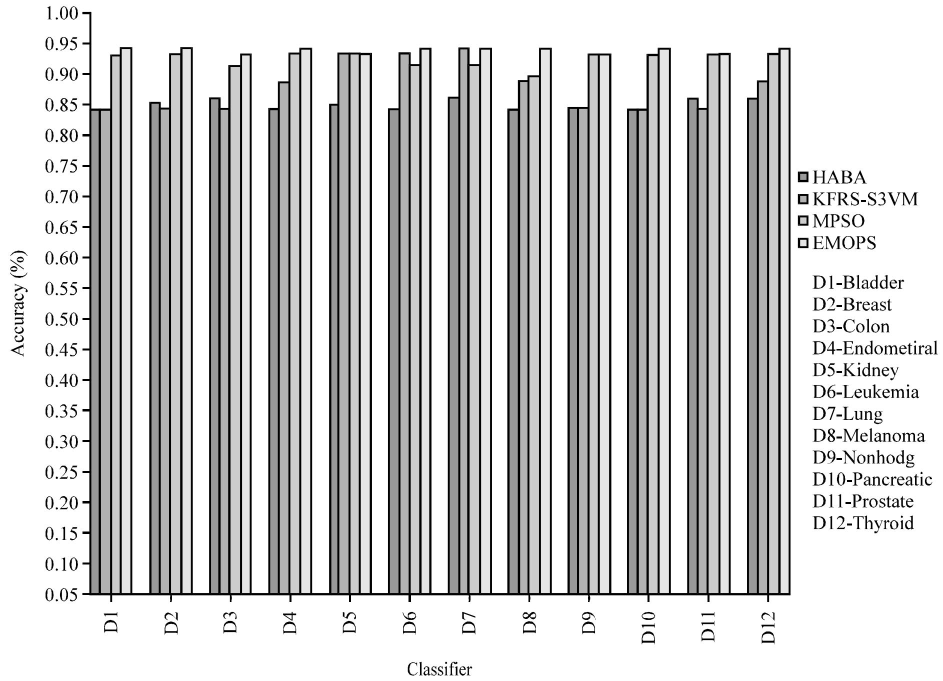


Fig. 4: Sensitivity vs. classifiers

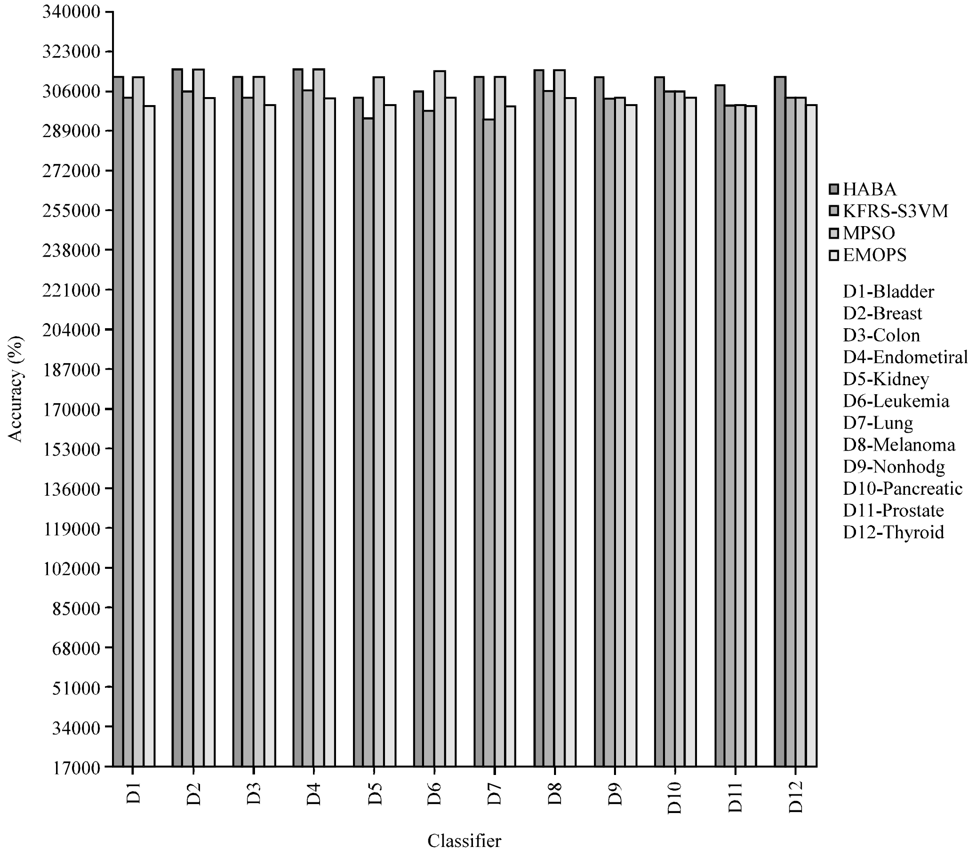


Fig. 5: Processing time vs. classifiers

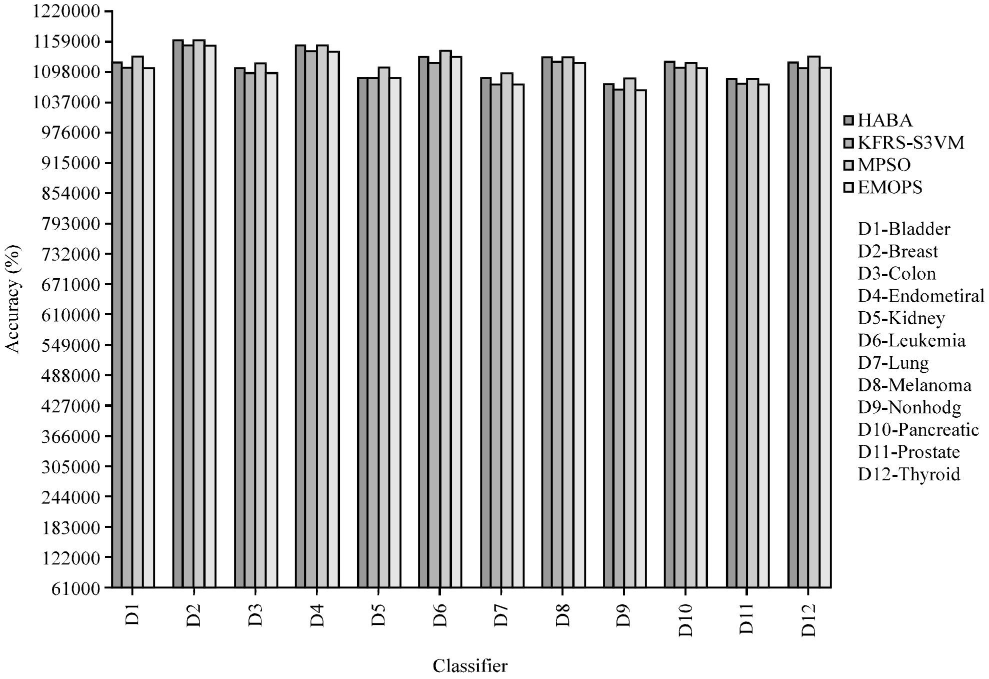


Fig. 6: Memory usage vs. classifiers

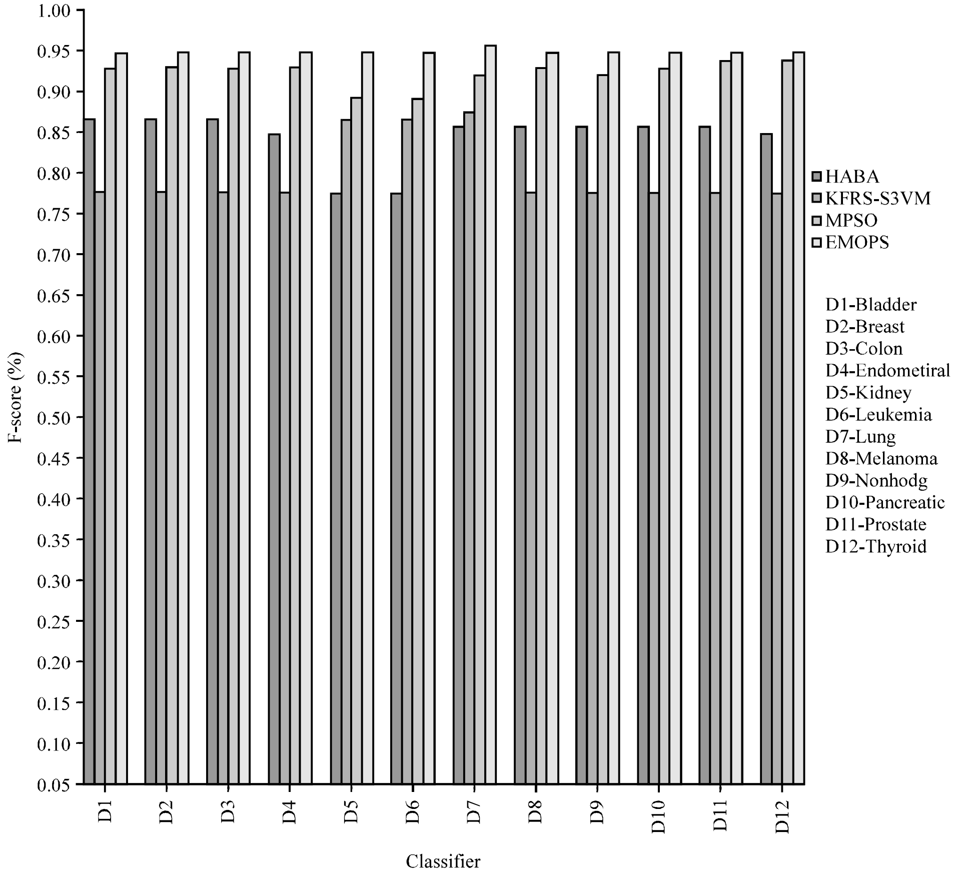


Fig. 7: F-score vs. classifiers

CONCLUSION

This research proposed an efficient cancer pattern classifier called an Enhanced Multi-Objective Particle Swarm (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 Multi-objective Particle Swarm Optimization (MPSO) in terms of memory utilization, execution time (processing time), sensitivity, specificity, classification accuracy and F-score.

REFERENCES

Behravan, I., O. Dehghantanha and S.H. Zahiri, 2016. An optimal SVM with feature selection using multi-objective PSO. Proceedings of the 1st IEEE Conference on Swarm Intelligence and Evolutionary Computation (CSIEC), March 9-11, 2016, IEEE, Bam, Iran, ISBN:978-1-4673-8737-8, pp: 76-81.

Chakraborty, D. and U. Maulik, 2014. Identifying cancer biomarkers from microarray data using feature selection and semisupervised learning. Transl. Eng. Health Med. IEEE. J., 2: 1-11.

Devaraj, D. and B. Yegnanarayana, 2005. Genetic-algorithm-based optimal power flow for security enhancement. Proc. IEE. Generat. Transmission Distribut., 152: 899-905.

Gu, X., 2016. A multi-state optimization framework for parameter estimation in biological systems. IEEE. ACM. Trans. Comput. Biol. Bioinf., 13: 472-482.

Hu, Y.C., 2007. Fuzzy integral-based perceptron for two-class pattern classification problems. Inf. Sci., 177: 1673-1673.

Kumar, P.G., C. Rani, D. Devaraj and T.A.A. Victoire, 2014. Hybrid ant bee algorithm for fuzzy expert system based sample classification. IEEE. ACM. Trans. Comput. Biol. Bioinf., 11: 347-360.

Mukhopadhyay, A. and M. Mandal, 2014. Identifying non-redundant gene markers from microarray data: A multiobjective variable length pso-based approach. IEEE. ACM. Trans. Comput. Biol. Bioinf., 11: 1170-1183.

Wang, Z. and V. Palade, 2007. A comprehensive fuzzy-based framework for cancer microarray data gene expression analysis. Proceedings of the 7th IEEE International Conference on Bioinformatics and Bioengineering, October 14-17, 2007, IEEE, Boston, Massachusetts, ISBN:1-4244-1509-8, pp: 1003-1010.

Yoon, Y., S. Bien and S. Park, 2010. Microarray data classifier consisting of K-top-scoring rank-comparison decision rules with a variable number of genes. IEEE. Trans. Syst. Man Cybern. Part C. Appl. Rev., 40: 216-226.