International Journal of Engineering Research and Technology.
ISSN 0974-3154 Volume 11, Number 4 (2018), pp. 661-673

http://www.irphouse.com

Smart Multi-Objective Particle Swarm Optimizer for Cancer Patterns Classification and Prediction

S. Subasree [1]
Research Scholar, Bharath University, Chennai – 600 073, Tamil Nadu, India.

N.P.Gopalan [2]
Professor, Department of Computer Applications, National Institute of Technology, Tiruchirappalli, Tamil Nadu, TN, India.

N.K. Sakthivel [3]
Professor and Head, Department of Computer Science and Engineering, Nehru College of Engineering and Research Centre, Pampady – 680 588, Kerala, India.

Abstract

The Cancer Pattern Classification and Prediction Technique through Microarray is the one of the demanding research techniques in the research field of Bioinformatics. A lot of Multi-Objective based methods and Multi-Objective Optimization Methods have been developed recently to identify and predict Cancer Patterns. At the same time, for predicting and classifying cancer patterns, hundreds of thousands of Genes Expressions Pattern needed to study to recognize those Gene Expression patterns or Cancer Patterns. As Gene Patterns are complex, we need to employ an efficient Multi Objective Particle Swarm Optimization (MOPSO) Technique to predict Cancer pattern population effectively and efficiently. Recently an Enhanced Multi-Objective Pswarm (EMOPS) was developed by the authors and its classification accuracy was improved. However, we revealed through this work that the current Classifier EMOPS fails to update entire population during each and every iteration to predict the gene pattern (cancer pattern) and it is depending on Global Best Particles that is the major issue which needed to focus for improving the classification accuracy by considering Optimization quality and convergence speed as well. To address this issue, we developed a Smart Multi-Objective Particle Swarm Optimizer (SMOPSO), where Particles are updated based on Pairwise Competitions instead of updating Global best particles. This model is simulated and analyzed thoroughly. The simulated work helps this research work to establish that the developed SMOPSO outperforms the EMOPS classifiers with regard to Convergence and Diversity

on bi-objective ZDT 1 and ZDT 3, Execution Time (Processing Time), and Classification Accuracy.

Keywords: Cancer Pattern Classification, Competitive Swarm Optimizer, Convergence Analysis, Gene Expression, Microarray, Multi-Objective Optimization.

I. INTRODUCTION

In the Research Field of Bioinformatics, it is observed that the Microarray Technology to create Data Sets is the most prominent and major technology that facilitates for studying different expressions of Genes and Cancer. This Microarray Technology will hold Image Data and these images can be transformed into different expressions of genes[1,2,3,4,5,6]. These transformed expressions of Gene and Cancer Patterns generally used for predicting and classifying the various patterns of Genes and Cancers. The survey was carried out by this work[1,2,6,7], it is observing that the existing Classifiers and Particle Swarm Optimization Techniques are helping for classifying and predicting different Patterns of Gene and Cancer as well effectively.

As discussed earlier, the Gene Patterns can be classified or predicted with hundreds of thousands of samples of microarray. The samples based on patterns can be classified as various subtypes and multiclass patterns could be built by various available multiclass classifiers [1,6,8,9]. From the literature survey, it is observed that the Multi-Class Cancer Pattern Classifiers were used for maximizing the classification and prediction accuracy[1,6,10]. However, from the literature survey, it is revealed that the classification accuracy could be improved further by Multi-Class Cancer Pattern Classifier if its entire patterns were populated.

This research work reviewed its previous classifier named Enhanced Multi-Objective Pswarm (EMOPS)[1,2] thoroughly. As this model is functioning under the framework of Multi-Class Cancer Pattern Classifier, it is classifying and populating particles with the idea of Global Best Position gbest[1,6,7]. It was observed that this model unable to employ entire particles for population which causes classification accuracy degradation. To address this issue, the Enhanced Multi-Objective Pswarm (EMOPS) was modified by introducing Pairwise Particles Population in this model to improve classification accuracy.

The developed model, Smart Multi-Objective Particle Swarm Optimizer (SMOPSO) was implemented and the detailed procedure was described at the below section.

This paper is organized as follows. The Multi-Objective Particle Swarm Optimization Technique and Enhanced Multi-Objective PSwarm based Classifier was discussed in Section 2. The functionalities and operations of the developed model, Smart Multi-Objective Particle Swarm Optimizer (SMOPSO) is described in Section 3. The experimental setup was narrated in Section 4 and results and comparative analysis of the developed model are discussed at Section 5. In Section 6, the conclusion of the developed model is described.

II. RECENTLY PROPOSED MULTI-OBJECTIVE PARTICLE SWARM OPTIMIZATION

The Particle Swarm Optimization approaches were developed for particle population and it deals with both unconstrained and continuous nonlinear optimization problems. In the following section, this work describes i. Enhanced Multi-Objective PSwarm based Classifier (EMOPS)[1,6] and ii. Multi-Objective Particle Swarm Optimization (MPSO) [1,6,7].

A. MULTI-OBJECTIVE PARTICLE SWARM OPTIMIZATION (MPSO)

The Particle Swarm Optimization[1,6,7] is predicted as the current demand and research optimization techniques which is working with the idea of population scenario. Here, the population is denoted as different group of candidate solutions. It is also understand that the particles population is named as Swarm.

That is the Optimal Fitness could be achieved by group of N Swarm Particles. It is also understand that the pbest and gbest needed to update to achieve fitness. Here the pbest is called as Particle Best Position and gbest is representing as Global Best Position.

The authors Anirban Mukhopadhyay and et. al. have developed the MPSO [6] by as follows.

1. *Input i. Cluster Center C, ii. Data Matrix, iii. Samples S, iv. Particles N, v. Assign thr = 0.5*
2. *Output A*
   a. *Assign initial default value to Velocity and Random Locations*
      i. *Samples Gene Set G_n, Genes x_n, and Fitness P_n*
   b. *Assign initial default value to Velocity and Random Locations*
      i. *Evaluate CellBoundary(xnd) entire Centres' Clusters till xnd $\geq$ Threshold*
   c. *Compute average Velocity Vnd and CellBoundary as well*
   d. *Evaluate and combine to Pick optimal Centres*
   e. *Average Calculation $\leftarrow$ Grouping derived solutions and Pick Optimal Gene Gn*

B. EMOPS : AN ENHANCED MULTI-OBJECTIVE PSWARM BASED CLASSIFIER

The author Subasree and et.al. developed an Enhanced Multi-Objective Pswarm Based Classifier (EMOPS)[1,5] that will improve the Patterns of Cancer and Gene Classification and Prediction as well. The detailed procedure was discussed in the below section.

Procedure of Enhanced Multi-Objective PSwarm based Classifier (EMOPS)

The Multiobjective Particle Swarm Optimization (MPSO), as discussed in the previous section, it groups hundreds of particles to attain fitness that should be optimal and optimized one and to achieve the same, this procedure used pbest and gbest. Here the pbest and gbest are represented as Particle Best Position and Global Best Position correspondingly. It is found that these parameters and positions of particles as well ought to optimize to attain better Classification Accuracy. That is it is needed to find and predict optimized centre values for maximizing better classification and prediction accuracy. For achieving better classification accuracy, this research work developed the EMOPS.

The operations of the EMOPS [1,2] is described elaborately below.

1. *Input i. Cluster Center C, ii. Data Matrix, iii. Samples S, iv. Particles N, v. Assign thr = 0.5*
2. *Output A*
   a. *Assign initial default value to Velocity and Random Locations*
      i. *Samples Gene Set Gn, Genes xn, and Fitness Pn*
   b. *Assign initial default value to Velocity and Random Locations*
      i. *Evaluate CellBoundary(xnd) entire Centres' Clusters till xnd* $\geq$ *Threshold*
   c. *Compute average Velocity Vnd and CellBoundary as well*
   d. *Evaluate*
      i. *High-dominance updating strategy*
         a. *Evaluate Distances of Growding in Clusters and Update for future Iteration*
         b. *Evaluate the Rectangle's Largest Size*
         c. *Takes the avg distance of its neighbouring solutions*
         d. *Evaluate and combine to Pick optimal Centres*
         e. *Average Calculation* $\leftarrow$ *Grouping derived solutions and Pick Optimal Gene Gn*
      ii. *Pick the optimized and Optimal (Gene Gn)*
   e. *Pick gbest*

III. SMART MULTI-OBJECTIVE PARTICLE SWARM OPTIMIZER (SMOPSO)

In this section, the developed Smart Multi-Objective Particle Swarm Optimizer (SMOPSO) is described.

A. Identifying Competing Particles Selection

The particle selection will have three important functions. First it will identify best Particles among competitions. In second step it will perform pairwise competition between particles. In the third step it will select the best particle based Learning process. The best particle will be selected for competition based on the Crowding distance.

The crowding distance is calculated by cluster centre with the given Threshold value. If the cluster centre is greater than Threshold value the Cell boundary and velocity will be calculated. The best particle will be identified based on Crowding distance and from that the Fronts Fi will be selected. It created swarm P ranging from F1, F2, F3 ……FK where K is representing as the Fronts Maximum Index. From list of Fronts, minimum number of Fronts t will be selected for competition. i.e |F1 U F2 U F3……FK| $\geq \gamma$ where γ is used for selecting best particles by optimizing its values. From each generation the best particles are selected for Multi-Objective PSO to find global best particle.

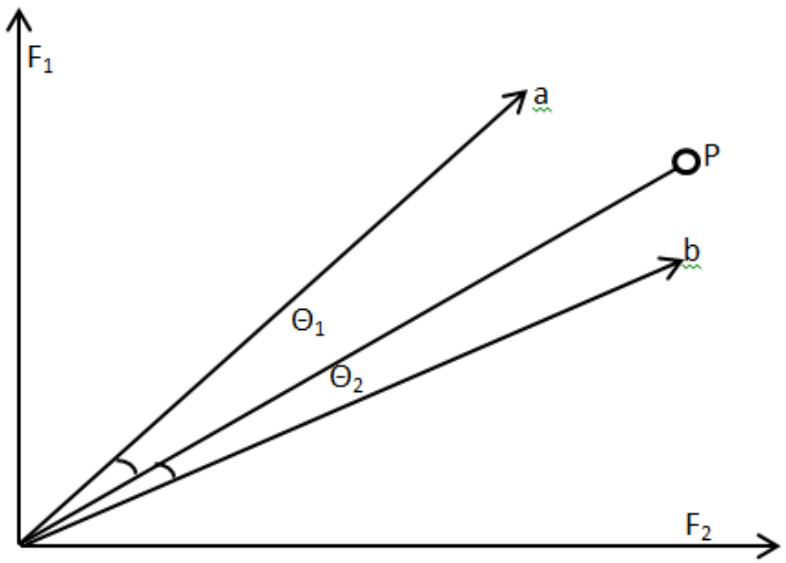

Fig 1: Pairwise Competition between best Particles

After creation of best particle selection, pairwise competition is conducted based on learning process, and the p, moving particles' directions can be found by using the winner. During the pairwise competitions, the randomly selected best particles a and b can be measured and selected in the moving particle p. It is needed to measure the

angles of a, b and p to conclude the best particle whose angle is smaller one that will be the best particle. This will be very closer to convergence direction. Figure 1 is representing competition pairs that is predicted as best Particles by the developed SMOPSO. From the figure Fig.1, it is predicted that the winner will be a θ2 because it has less angle compared to that of θ1. After finding the winner, the particle p velocity could be updated through learning process. Let p_i and v_i is the position and velocity of the ith particle swarm.

The updated position and velocity of the particle can be calculated by the below equations (1) and (2).

$$v_i' = R_1 v_i + R_2 (p_w - p_i) \quad (1)$$

$$p_i' = p_i + v_i' \quad (2)$$

Where R_1 and R_2 ranges from 0 to 1. R_1 and R_2 randomly generated vectors and P_w is the position of the winner. This mechanism will be continued until convergence and diversity take place. Finally the optimal convergence and diversity of gene pattern will be identified.

B. Procedure of SMOPSO

1. *Input i. Cluster Center C, ii. Data Matrix, iii. Samples S, iv. Particles N, v. Assign thr = 0.5*
2. *Output A*
    - i. *Assign initial default value to Velocity and Random Locations*
    - ii. *Samples Gene Set Gn, Genes xn, and Fitness Pn*
    - iii. *Assign initial default value to Velocity and Random Locations*
    - iv. *Calculate CellBoundary(xnd) entire Centres' Clusters till xnd ≥ Threshold*
    - v. *Compute average Velocity Vnd and CellBoundary as well Particle Selection*
    - vi. *Calculate Crowding Distance to select Fronts F_i*
    - vii. *Compare with γ to select best Particle*

    b. *Pairwise Ambitious Selection to choose best direction*

    - i. *Winner*
        1. *Update the Velocity*
        2. *Guide(Particle) → NextGeneration();*
        3. *Adjust γ to Optimize i. Convergence and ii. Diversity*
    - ii. *Looser*
        1. *Learn.winner()*
        2. *Update();*

c. Calculate and Update for Iterations

i. Strong-dominance updating Particle Selection from d.
ii. Update next Pairwise Particle
1. Learning by adjusting γ to Optimize i. Convergence and ii. Diversity

d. Pick the Optimal Gene Pattern Gn

IV. EXPERIMENTAL SETUP

This research work has presented Inverted Generational Distance(IGD) and the types of Zitzler – Deb – Thiele (ZDT) which used for evaluating and analysing our developed Classifier. It was understood from the Literature Survey that the Zitzler – Deb – Thiele (ZDT) family Models were sufficient to analyse and compare the Multi-Objective Particle Swarm Optimizer[1,3,4,5]. This Research Work adopts ZDT1 and ZDT 3 Test Problems[1,3] and results were demonstrated. The Inverted Generational Distance(IGD) Model[3] was employed for analysing and comparing the Multi-Objective Particle Swarm Optimizer, where 2000 sampled points were identified for analysis.

The Inverted Generational Distance(IGD) Model and the Zitzler – Deb – Thiele (ZDT) family Models were discussed below that demonstrated by the authors[1,3,12,13].

a. Inverted Generational Distance(IGD)

The distance is calculated based on Inverted Generational Distance(IGD) and the formula is given below.

A real pareto front and a set of candidate solutions

$$PF = \{y_1, y_2, \dots\dots\dots y_N\}$$

$$F = \{X_1, X_2, \dots\dots\dots X_k\}$$

$$IGD(F, PF) = \frac{1}{N}\left(\sum_{j=1}^{N} D_j^t\right)^{\frac{1}{t}}, 1 \leq t \leq \varphi$$

Where D_j^t is the minimal Eyclidean Distance from yj to F.

b. Zitzler – Deb – Thiele ZDT - 1

The Author[12,13] described the equation as below which didn't edit or didn't reframe by us.

"Decision space : $x \in [0,1]^{30}$

Objective Function : $f1(x) = f2(x) = g(x)\left(1 - \sqrt{x_1/g(x)}\right)$

$$g(x) = 1 + \frac{9}{n-1}\sum_{i=2}^{n} x_i$$

Optimal Solution : $00 \leq x_1 \leq 1, x_1 = 0$ where $i = 2 \ldots .30$

Characteristics : Convex Pareto Front"

c. Zitzler – Deb – Thiele ZDT – 2

The Author[12,13] described the equation as below which didn't edit or didn't reframe by us.

"Decision space : $x \in [0,1]^{30}$

Objective Function : $f1(x) = x_1$

$f2(x) = g(x) (1 - (x_1 / (g(x))^2))$

$g(x) = 1 + \frac{9}{n-1} \sum_{i=2}^{n} x_i$

Optimal Solution : $0 \leq x_1 \leq 1, x_1 = 0$ where $i = 2 \ldots .30$

Characteristics : Concave Pareto Front"

d. Zitzler – Deb – Thiele ZDT – 3

The Author[12,13] described the equation as below which didn't edit or didn't reframe by us.

"Decision space : $x \in [0,1]^{30}$

Objective Function : $f1(x) = x_1$

$f2(x) = g(x) \left(1 - \sqrt{x_1 / g(x)} - \frac{x_1}{g(x)} \sin(10\pi x_1)\right)$

$g(x) = 1 + \frac{9}{n-1} \sum_{i=2}^{n} x_i$

Optimal Solution : $0 \leq x_1 \leq 1, x_1 = 0$ where $i = 2 \ldots .30$

Characteristics : Discontinuous Pareto Front"

e. Zitzler – Deb – Thiele ZDT – 4

The Author[12,13] described the equation as below which didn't edit or didn't reframe by us.

"Decision space : $x \in [0,1]^{30} x [-5,5]^9$

Objective Function : $f1(x) = x_1$

$f2(x) = g(x) (1 - (1 - (x_1 / (g(x))^2))$

$g(x) = 1 + 10 (n - 1) + \sum_{i=2}^{n} \left(x_j^2 - 10 \cos\left(4\pi x_j\right)\right)$

Optimal Solution : $0 \leq x_1 \leq 1, x_1 = 0$ where $i = 2 \ldots .30$

Characteristics : Many Local Pareto Front"

f. Convergence Metrics

The Convergence Metric C_m is a measure as follows

$$C_m = \left[\sum_{i=1}^{|Q|} d_i\right]/|B| \quad (3)$$

Where d_i is the distance between the solution $i \epsilon B$ and the nearest member of pareto optimal solution.

g. Diversity Metrics

The second metric is the Diversity Metric D_m. This metric measures the diverse obtained non-dominated solutions are,

$$D_m = \frac{d_f + d_1 + \sum_{i=1}^{|Q|} d_i - d'}{d_f + d_1 + (|Q|-1)d'} \quad (4)$$

Where d is represented as distance among two consecutive solutions in the non-dominated solutions Q, where d is representing as average distances of the denoted solutions. Here where df and d1 are representing the distance among the extreme solution of pareto front and the Q which is the obtained boundary solutions.

V. EXPERIMENTAL RESULTS AND PERFORMANCE ANALYSIS

This Research Work is conducted Simulations for studying the performances and classification/prediction abilities of the developed model, Smart Multi-Objective Particle Swarm Optimizer (SMOPSO) for Cancer Patterns Classification and Prediction.

The NCBI.CGS.MER is the Cancer Pattern Data Sets[10] that is downloaded from NCBI for analysing the developed model. Simulation setup was created and for this purpose, there are different patterns of Cancer downloaded and a few important patterns that concern to most of the Patients are i. Breast, ii. Bladder, iii. Endometrial, iv. Colon. v. Kidney, vi. Lung, vii. Leukemia, viii. Melanoma, ix. Pancreatic, x. Mom-Hodgkin, xi. Thyroid xii. And Prostate . The procedure of the developed Smart Multi-Objective Particle Swarm Optimizer (SOPSO) was optimized to Tradeoff between Classification Accuracy and Computational Cost by adjusting γ. This developed model was executed to find the best Convergence and it was predicted that 2000 generations needed for Deep Analysis. The value of γ is very important to optimize convergence. There were 40 runs used for Inverted Generational Distance (IGD). The experimental results were shown in the Figures Fig. 2 to Fig. 6 to understand the Performances of the developed model.

From the Figures Fig.2 and Fig.3, it was established that the developed model obtained the Non-Dominated Solution Sets Associated with IGD values, Max. of 40 with the 2000 Generations. It is also noted that the developed model achieves relatively better Convergence to the Parento Front of ZDT 3 and ZDT 1.

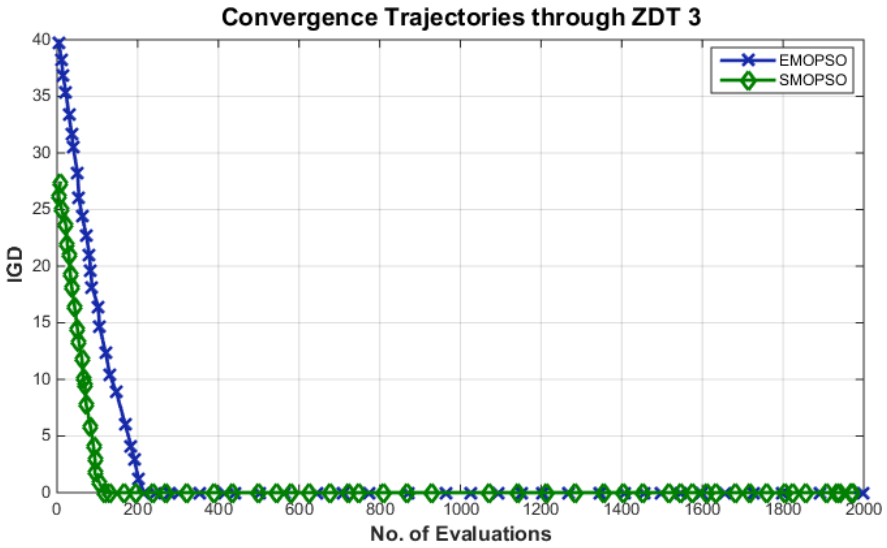

Fig 2: Convergence Trajectories on ZDT 3

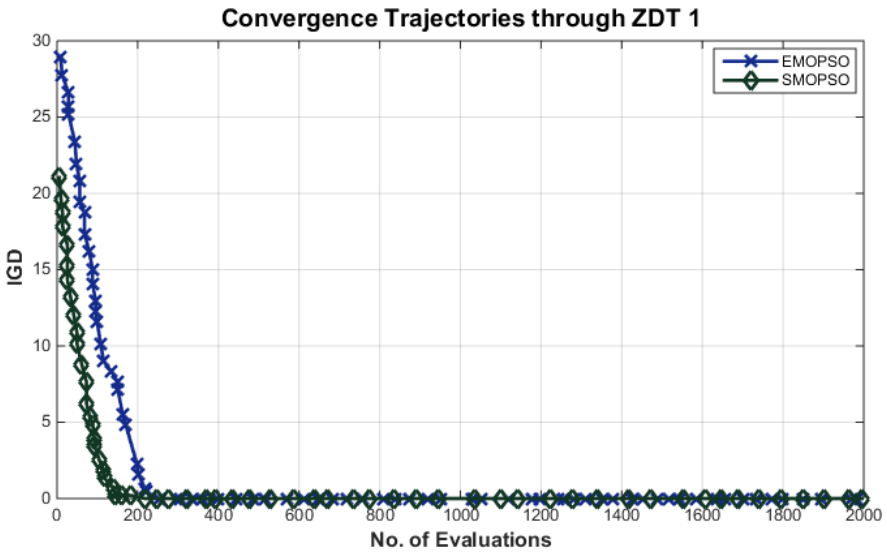

Fig 3: Convergence Trajectories on ZDT 1

From the Fig. 4, it is established that the developed model obtained the Non-Dominated Solution Sets Associated with IGD values for two sets of objectives. It is further revealed that the developed model achieves relatively higher values of Elite Particle Set on ZDT 3.

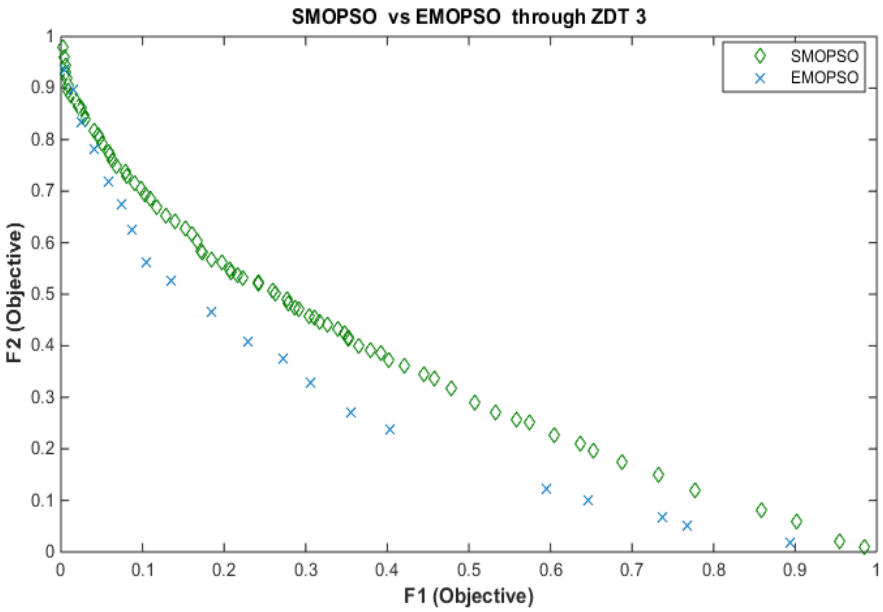

Fig 4: Non-Dominated Solution Set Obtained ZDT 3

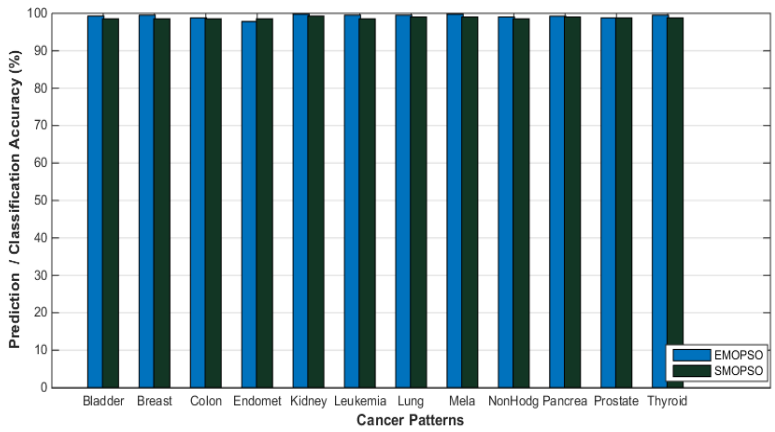

Fig 5: Classification Accuracy (Population Based)

The Classification/Prediction Accuracy of the developed model Smart Multi-Objective Particle Swarm Optimizer (SMOPSO) was shown in the Fig. 5. From the Results, it was revealed that the developed model SMOPSO performs better than that of our previous model Enhanced Multi-Objective Pswarm (EMOPS) for a few Cancer Patterns in term of Accuracy. It is revealed that the performance of a Classifier in term of Classification Accuracy and Pattern Prediction purely depends on the Cancer Pattern. ie it is noticed that the classification accuracy differs for the different cancer patterns.

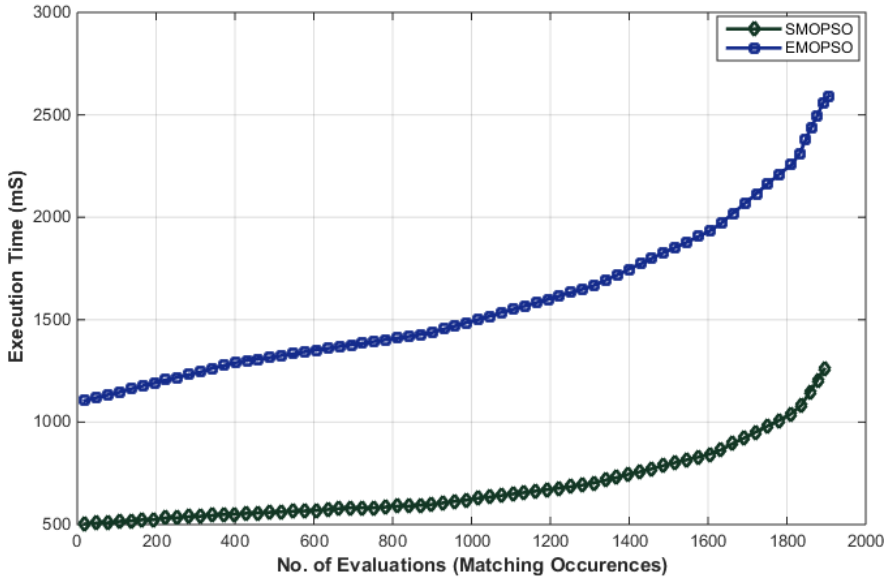

Fig 6: Execution Time of the Developed Model

From the Fig. 6, it is clearly established that that the execution time of the developed model Smart Multi-Objective Particle Swarm Optimizer (SMOPSO) is less than that of our previous Classifier, Enhanced Multi-Objective Pswarm Based Classifier (EMOPS) for predicting the various Cancer Patterns with different evaluations. It is also noted that the developed model SMOPSO achieves and provides fairness to the system as compared with our previous model, EMOPS.

VI. CONCLUSIONS

This research work has implemented both the developed classifier and predictor called Smart Multi-Objective Particle Swarm Optimizer (SMOPSO) and existing classifier Enhanced Multi-Objective Pswarm (EMOPS). The experiments were repeated thoroughly to study and to understand the classification/prediction efficiency of the developed model. From the experimental results, it was established that the developed SMOPSO outperforms our previous model, EMOPS classifiers in terms of Convergence and Diversity on bi-objective ZDT 1 and ZDT 3, Execution Time (Processing Time), and Classification Accuracy.

REFERENCES

[1] S. Subasree; N.P. Gopalan; and N. K. Sakthivel, EMOPS: An Enhanced Multi-Objective Pswarm based Classifier for Poorly Understood Cancer Patterns. International Journal of Engineering & Technology, Vol 7,Pp. 7-11, 2018.

[2] S. Subasree; N.P. Gopalan; and N. K. Sakthivel, An Enhanced Multi-Objective

Particle Pswarm Based Classifier for Poorly Understood Cancer Patterns. Journal of Engineering and Applied Sciences, Vol 13(3),580-587,2018.

[3] Xingyi Zhanga and et.al, A competitive mechanism based multi-objective particle swarm optimizer with fast convergence. Information Sciences, 63-76,2018.

[4] Prabhujit Mohapatra and et.al, A Modified Competitive Swarm Optimizer for Large Scale Optimization Problems. Applied Soft Computing, 340 – 362,2017.

[5] Zahra Pourbahman and Ali Hamzeh, Reducing the Computational Cost in Multi-objective Evolutionary Algorithms by Filtering Worthless Individuals. Neural and Evolutionary Computing, 2014.

[6] S. Subasree; N.P. Gopalan; and N. K. Sakthivel, A Comparative Study and Analysis of Data Mining Classifiers for Microarray based Cancer Pattern Diagnostics. Proceedings of ACM International Conference on Informatics and Analytics (ICIA-16), 2016.

[7] Xu Gu, A Multi-State Optimisation Framework for Parameter Estimation in Biological Systems. IEEE/ACM Transactions on Computational Biology and Bioinformatics,2016.

[8] Iman Behravan; Seyed Hamid Zahiri; and 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.

[9] Debasis Chakraborty; and Ujjwal Maulik, Identifying Cancer Biomarkers From Microarray Data Using Feature Selection and Semisupervised Learning. IEEE Journal of Translational Engineering on Health and Medicine. Vol 2,2014.

[10] 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.

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

[12] Pradyumn Kumar Shukla, "Exploiting second order information in computational multi-objective evolutionary optimization," Proceedings of the aritficial intelligence 13th Portuguese conference on Progress in artificial intelligence, Lecture Notes of Computer Science,2007.

[13] Carlos A Coello and et.al., Evolutionary Algorithms for solving Multi-Objective Problems, Second Edition, Springer, 2007.