ISSN: 0976-3104
SPECIAL ISSUE (ASCB)
Jayaraman and Velumani

THE IIOAB JOURNAL

ARTICLE OPEN ACCESS

A HYBRID PSO-GA APPROACH FOR BIOMARKER DISCOVERY

Ramsingh Jayaraman and Bhuvaneswari Velumani*

Department of Computer Applications, Bharathiar University, Coimbatore, INDIA

ABSTRACT

The advancement in genomic and proteomic has paved way for identifying methodologies to identify gene involved in life threatening diseases. Biomarker refers to specific gene and products with biochemical features to measure the progress of the diseases. Various soft computing techniques are applied to identify biomarkers from large micro array chips. In this paper a hybridized approach for biomarker discovery using Particle Swarm Optimization (PSO) and Genetic Algorithm (GA) is proposed. The proposed approach is tested with microarray lymphoma dataset. Based on the experimental run twelve gene were identified as significant gene and 3 gene were found in the disease pathway which can be verified with experts to name as biomarkers using the methodology proposed. The experimental results are validated using panther tool for biological significance and verified with related literature papers. From the results it is found the performance of the algorithm is good for identifying significant gene for bio marker discovery.

Received on: 18th-March-2015
Revised on: 20th-May-2015
Accepted on: 26th-June-2015
Published on: 20th-Aug-2015

KEY WORDS

Biomarker, PSO, GA, Gene Pathway, Significant gene Gene Ontology (GO)

***Corresponding author:** **Email**: bhuvanes_v@yahoo.co.in, j.ramsingh@hotmail.com, **Tel:** 9500897389

INTRODUCTION

DNA microarray is a rapid growing technology used for the study of gene expression profiles and it is a classic probe hybridization method which provides access to thousands of gene at once. Generally, microarray data are images, which are transformed into gene expression matrixes in which rows represent gene and columns represent the expression values of gene under various experimental conditions. Microarray technologies enable the simultaneous interrogation of the expression level of thousands of gene to obtain a quantitative assessment of their differential activity in a given tissue or cell [1]. Recently the microarray technology has a gained interest of their use in clinical trials and disease diagnosis for identifying genomic factors that are prognostic for prediction or identification of Biomarker.

Biomarkers (short for biological markers) are biological measures of a biological state. By definition, a biomarker is "a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes or pharmacological responses to a therapeutic intervention [2]. Biomarkers are the measures used to perform a clinical assessment such as blood pressure or cholesterol level and are used to monitor and predict health states in individuals or across populations so that appropriate therapeutic intervention can be planned [3-5]. Biomarker discovery is proved as one of the most broadly applicable and successful means of translating molecular and genomic data into clinical practice [6]. Evolutionary algorithm and soft computing techniques are applied in large for analysis of gene profiles for Biomarker prediction and identification. Genetic algorithms are hybridized with bio-inspired algorithms for microarray analysis to identify the significant gene. Identification of significant gene from large gene set becomes essential to identify the biomarkers of specific diseases [7]. The authors Zhang F [8], Hui-Huang Hsu [9] used data mining to identify biological markers for the diagnostic classification and prognostic assessment in the context of microarray and proteomic data A hybridized Particle Swarm Optimization and Genetic Algorithm PSO-GA is proposed in this paper to identify significant gene from DNA microarray for biomarker discovery. The proposed work is implemented using R-language for Lymphoma cancer dataset. The paper is organized as follows. Section II discusses the literature related to the work. Section III describes the hybridized approach for identifying biomarker using PSO with Genetic algorithm. In Section IV the experimental results are discussed followed by Conclusion in section V,

THE IIOAB JOURNAL
ISSN: 0976-3104

followed by Limitations and Future scope in section VI.

LITERATURE STUDY

This section discusses the literature related to the work. The literature are discussed in three sub sections namely Preprocessing techniques of microarray data for significant gene analysis, literature related to evolutionary algorithm and hybridized approaches.

Statistical techniques for microarray data

Statistical measures are widely used to identify the most and least selected attributes from large samples using various statistical tests as F-Test, T-Test, Chi-Square and Annova. In microarray dataset the gene are viewed as random samples and rank based statistical measures are widely used to identify the best expressed gene profiles from microarray data. The literature for ranking gene as part of pre-processing of microarray data is discussed in this section.

Statistical approaches like k Nearest Neighbor, Iterative regression imputation, Mean Squared Error, F-Test, T-test, chi-square, ANOVA and F-Test are used to pre-process the microarray dataset to impute missing values and rank gene. The authors Miguel Rocha and Isabel Rocha [4] suggested substituting of missing values with mean, median and mode using imputation method k nearest neighbor for microarray dataset. M. Templ et al., [10] used KNN methods for estimating missing values in compositional data. The authors also have discussed about T-test and ANOVA for identifying differentially expressed gene in microarray data. Matt Blackwell [11] analyzed Multiple Hypothesis Testing F-test in microarray data and S. N. Mukherjee, S. J. Roberts et al., [12] proposed a gene-ranking algorithm whose main novelty is the use of bootstrapped P-value for microarray datasets. The authors Guoqiang Yu, et.al, [13] used MSE (Mean Squared Error) to select candidate gene for classification. MSE is minimized to achieve the desired output (class target).Yoko Omura and Jun Sese [14], used Mean Square Error to select significant gene and generates pClusters from the selected gene. Mohd Sazli Saad [15], used MSE as objective function to calculate fitness value of each chromosome represented as gene in Genetic algorithm to pass the best chromosome for the consecutive generations.

PSO – GA

Particle swarm optimization is a computational method that optimizes a problem by iteratively trying to improve a candidate solution with regard to a given measure of quality [16]. PSO optimizes a problem by having a population of candidate solutions, here dubbed particles, and moving these particles around in the search-space according to simple mathematical formulae over the particle's position and velocity [17]. Each particle's movement is influenced by its local best known position but, is also guided toward the best known positions in the search-space, which are updated as better positions are found by other particles [18]. This is expected to move the swarm toward the best solutions. The algorithm is very simple but powerful.

The genetic algorithm is a model of machine learning which derives its behavior from a metaphor of the processes of evolution in nature Genetic algorithm are found to be widely used microarray data to improve the classification accuracy [19] . Algorithm is started with a set of solutions (represented by chromosomes) called population. Solutions from one population are taken and used to form a new population [20]. This is motivated by a hope, that the new population will be better than the old one. Solutions which are selected to form new solutions (offspring) are selected according to their fitness - the more suitable they are the more chances they have to reproduce [21]. This is repeated until some condition (for example number of populations or improvement of the best solution) is satisfied. Laetitia Jourdan et. al.[4] they have proposed a 2-phase approach using a specific genetic algorithm for the feature selection problem, they used a genetic algorithm (GA). Linyu Yang [19] and C.H. Ooi and Patrick Tan (2003) [22] used Genetic algorithm to design a gene-selection scheme predict the disease. J. Yu, et. al.,[23] and Edmundo Bonilla Huerta [6] applied Genetic Algorithm to find the gene subsets that produce high prognostic classification accuracy.

The authors Barnali Sahua et.al. [16] and Keisuke Kameyama [18] used Particle swarm Optimization (PSO) as a optimization technique for the classification of high dimensional cancer microarray data. Qinghai Bai [21],

THE IIOAB JOURNAL
ISSN: 0976-3104

concluded that performance of PSO is fast, easy and effective for optimization problem compared to other approaches and author strongly suggested that, PSO provides better result when combined with other techniques. Yuji Zhang et.al, [24] proposed a method of hybrid of genetic algorithm and particle swarm optimization (GA-PSO) to train the NN (Neural Network) models. The authors C.H. Ooi and Patrick Tan [8], Linyu Yang ,et.al., [19] applied genetic algorithm for multi-class prediction problem. The authors stated that the parallelized searching capability of GA helps to design a gene-selection scheme that determines the optimal set of gene in a multiclass dataset which classifies the samples within the dataset with minimal error and accurate prediction. Shital C. Shah [25] described that, GA based approach uniquely identified some gene/SNPs that could not be identified by the approaches like Information Gain and REG Regression approaches. Laetitia Jourdan et. al. [26] Proposed a 2-phase approach using a specific genetic algorithm with clustering algorithm for multi factorial disease prediction. The authors Ramsingh et.al. [27] Proposed genetic algorithm with K-Means clustering to identify significant gene from microarray dataset. The authors Vanitha et.al, [28] used evolutionary approaches such as optimization methods to generate best gene combinations to achieve higher level classification accuracy.

The biomarker gene are some specific gene which is highly significant causing gene mutation resulting in specific diseases. Indentifying biomarker gene can be viewed as an optimization problem to identify minimum number of gene which are highly significant bases on their gene expression profiles. From the literature study it is found that PSO, GA and other approaches are applied as standalone approaches for identifying significant gene from microarray dataset. These approaches when applied and iterated for more number of runs identify the best gene expression profiles which are highly significant. Biomarker identification requires accurate validation of the gene expression profiles to identify the gene in specific diseases. In this paper a novel methodology is proposed to filter the optimal gene using a two level optimization process using the hybrid PSO-GA approach to identify the significant gene to discover biomarkers for diseases.

BIOMARKER DISCOVERY USING PSO-GA

This section discusses the proposed methodology and framework to identify genomic biomarker for microarray dataset. The framework for Biomarker discovery using PSO-GA given in **Figure−1** is consists of three phases Pre Processing phase, phase II as PSO-GA hybrid approach and verification and validation phase as the third phase.

Phase I

Pre-processing phase is the first phase used for analyzing Gene expression features of microarray dataset using statistical techniques. Microarray data contains of noisy and inconsistent data. The Pre-processing techniques are used to reduce noisy and inconsistent data using imputation methods. k Nearest Neighbor (kNN) is a data mining technique to impute missing expression data in microarray in the proposed work to fill in the missing values. For each gene with missing values, it finds the k Nearest Neighbor using a distance metric, confined to the columns for which those gene are not missing. The microarray data is normalized using the statistical techniques.

Based on the detailed study and analysis of various statistical methods for microarray data through experimental result and comparison MSE is found to identify highly significant gene which are semantically relevant compared to other approaches from the literature study. MSE is used as fitness function to rank significant gene. The mean squared error is arguably the most important criterion used to evaluate the performance of a predictor or an estimator. (The subtle distinction between predictors and estimators is that random variables are predicted and constants are estimated.) The mean squared error is also useful to relay the concepts of bias, precision, and accuracy in statistical estimation.

$$MSE \quad \sum_{i=1}^{n}(x_i x)^2 \qquad \text{----Eq 1}$$

n = size (number of individuals)

$\mathbf{x}$ = mean value of individuals

$\mathbf{x}$ = value of individual

THE IIOAB JOURNAL
ISSN: 0976-3104

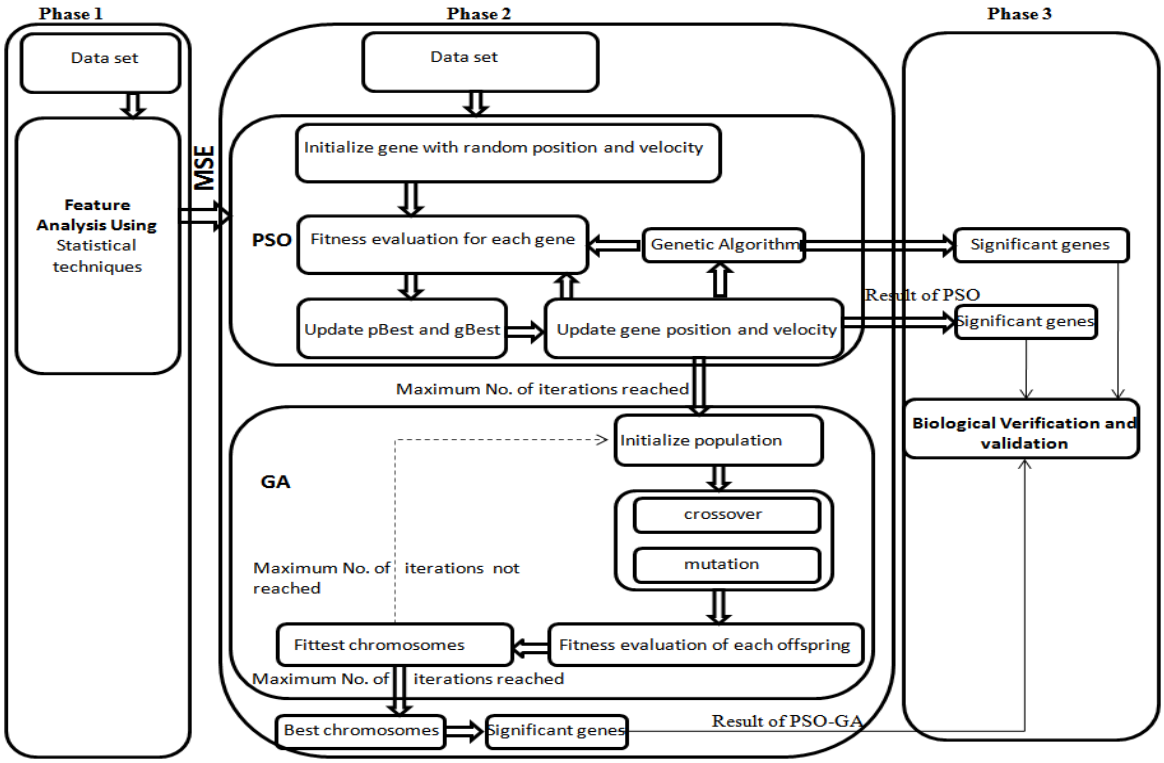

Fig: 1. Biomarker discovery using PSO-GA

Phase II: PSO – GA hybrid approach

The second phase of the framework is presented with the methodology to identify significant gene for biomarker discovery using hybridized PSO - GA algorithm. In this methodology the PSO algorithm is applied initially on the dataset and best populated gene are passed as input to GA process to identifying the best significant gene, as a two level optimization process the detailed methodology is presented in the below section.

PSO algorithm – parameters

In PSO algorithm the particles are initialized at random positions, to explore the search space to find better solutions. In each iteration each particle adjusts its velocity to follow two best solutions. The first is the cognitive part, where the particle follows its own best solution found so far. This is the solution that produces the lowest cost (has the highest fitness). This value is called pBest (particle best). The other best value is the current best solution of the swarm, i.e., the best solution by any particle in the swarm. This value is called gBest (global best).

In the proposed PSO algorithm gene expression values in microarray dataset are encoded as a particle. The fitness value of the particle is calculated using MSE given in Eq 1 gBest and pBest are calculated in each iteration based on the fitness value of gene, where gBest is the best fitness value of all gene in the entire population and pBest is the fitness value of gene in each population. the velocity and position of the particles are calculated using Eq 2 and Eq 3. The best ranked gene are passed to the GA process. The best ranked gene are removed from the PSO population for next iteration. This cycle is iterated until there is minimum number of gene in the population. The parameters of PSO algorithm is given in **Table–1**

$$\text{New velocity } v' = v + c1.r1.(pBest - x) + c2.r2.(gBest - x) \quad \text{----Eq 2}$$

$$\text{New position } x' = x + v' \quad \text{----Eq 3}$$

$$\text{New position } x' = x + v'$$

THE IIOAB JOURNAL
ISSN: 0976-3104

v = current velocity
x = current position
c = coefficient
r = random values, $0 >= r1 <= 1$ *and* $0 >= r2 <= 1, 0 <= c1 <= 2$, *and* $0 <= c2 <= 2$

The parameters for Hybridized PSO - GA algorithm for PSO Approach is given in **Table-1**.

Table: 1. Parameters of PSO

No. of particles	4026
Fitness function	MSE
Total no. of iterations	100

The below section describes the Genetic Algorithm parameters modeled for the work.

Genetic algorithm - parameters

Genetic Algorithms are efficient search methods based on the principles of natural selection and population genetics [29]. GA can search the solution space to find an optimal or near optimal solution by using evaluation and genetic operator functions to maintain the useful schemata of chromosomes in the population, in which chromosomes are evaluated using a fitness function to determine their fitness. According to the principle of survival of the fittest, Chromosome with a higher fitness will have higher probability of survival in each generation and thereby offspring with higher probability are generated. The genetic algorithm parameters modeled for identifying significant gene are given in **Table– 2**.

Table: 2. Parameters of GA

Chromosome - Size		33
Generations		100
Populations –Size		100
Selection		Random Selection
Fitness		MSE
GA operators	Crossover	Single point crossover
	Mutation	Substitution
	Mutation rate	0.8

The GA cycle stops the process if the genetic algorithm reaches local optimum or after the maximum number of solutions that already defined. The pseudo code for proposed work is given below:

Proposed pseudo code : Hybridized PSO and GA

```
Pop // input , population for PSO
// gene is a particle
Repeat
For each gene
   Initialize gene
End
do
   For each gene
      Calculate fitness value
      If the fitness value is better than the best fitness value (pBest) in history
         Update pBest = current best value
   End
   Choose the gene with the best fitness value of all the particles as the gBest
   For each gene
      Calculate gene velocity
Equation7  // equation to calculate velocity
      Update gene position
Equation 8  //equation to calculate position
   End
If (maximum iterations ==TRUE)
```

THE IIOAB JOURNAL
ISSN: 0976-3104

```
Population p=best 1000 gene from PSO
pop=(pop-best 1000 gene)
if(size(pop) < sufficient)
 break(PSO)
end if
//Genetic algorithm
Initialize population p
do
For i in 1:100   \\ Crossover
Randomly select two parents X_A and X_B from p
Generate X_C and X_D by one-point crossover to X_A and X_B
End for
For j in 1:size(p1)   \\ Fitness
Fitness value = sum_{j=1}^{n} (x_j - x̄)^2
End for

Select 100 chromosomes with minimum MSE value
 //Best chromosome selection
while there is no recurrence of chromosome in generation

If recurrence of chromosome = TRUE
For (k in 1:100)  //Mutation
Select duplicate chromosome Xk from p1
Mutate a gene of Xk
        If Xk is unfeasible
                Replace with best gene
        End if
End for
End if
Else
Stop // reaches local optimum or maximum no. of generation
Update p  // Update
p = best 100 fittest chromosomes
Returning best solution p //Return
PSO(pop)
```

Phase II: Verification and validation

The proposed work is validated biologically using PANTHER tool. PANTHER is an online biological tool used to analyze the biological significance of proteins and gene. PANTHER produces the output of list of gene which are involved in Gene Ontology functionalities such as Molecular function, biological process, cellular component and disease pathway. In this proposed work GO functionalities and pathway are analyzed to find out biological relevance and biological significance of 12 extracted gene from phase II. The experimental results of these approaches are discussed in result and discussion section. The results are also verified and compared with literature result.

Significant gene

The gene from the microarray dataset is identified as significant based on the biological validation using Gene Ontology. The gene is available in GO only when the gene has functionalities in Biological Process, Molecular Function and Cellular Component. The gene identified using the proposed approaches are termed as significant when the gene is available under GO taking part in all the three functionalities.

RESULTS AND DISCUSSION

Experimental results

The experimental results of proposed work are discussed in this section. Lymphoma microarray dataset is used for the proposed work. The dataset is downloaded from [http:/llmpp.nih.gov/lymphoma/data.shtml]. The microarray dataset consists of 4026 gene and 96 samples. In this section the experimental analysis of PSO-GA approaches are presented in detail. The experiments were conducted on the data set by executing the algorithms PSO and GA individually. The experimental analysis are presented below.

THE IIOAB JOURNAL
ISSN: 0976-3104

Normal PSO

The PSO algorithm is applied on the lymphoma dataset for identification of significant gene for biomarker discovery. On experimental run 33 gene are identified from 4026 gene using PSO approach in which 12 gene are found as significant based on the biological validation taking part in all the functionalities in this approach. The **Figure–1** presents the best significant gene based on fitness value using PSO approach. The fitness plot of the normal PSO approach for the dataset is given in **Figure–2**.

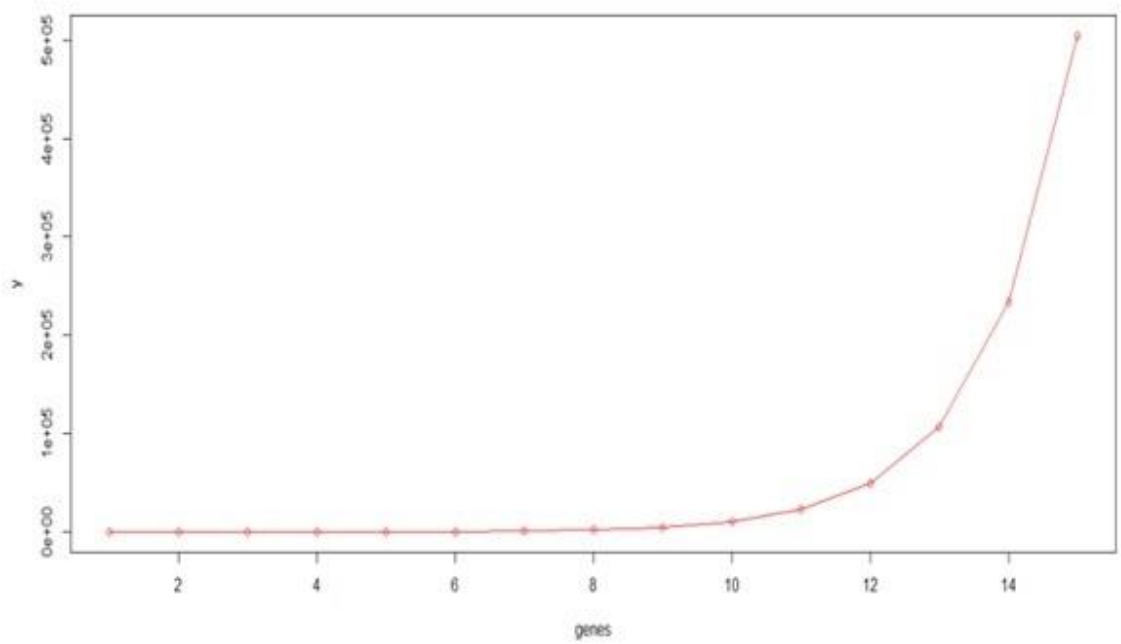

Fig: 2. significant gene from normal PSO

Genetic algorithm

The experimental result of Genetic Algorithm is presented below. The chromosome for GA approach is modeled with 33 gene of microarray data. The GA is run with the with a population size of 100 for 50 and 100 generations. 424 gene are identified using GA is given in **Table–3** with fitness score. **Figure–3** represents the significant gene for 50 and 100 generations. From the graph it is observed that the no of gene get optimum after 15 generations. 12 gene were found to be significant in GA approach based on the biological validation of GO.

Table: 3. Experimental Result - GA

Total No of Genes	4026
Significant Gene	424
Best Chromosome Fitness For first 10 generation	1230.21, 1046.75, 1009.72, 1140.30, 975.18, 975.72, 926.19, 911.82, 898.91, 911.82
Best Fit chromosome	688.31

THE IIOAB JOURNAL
ISSN: 0976-3104

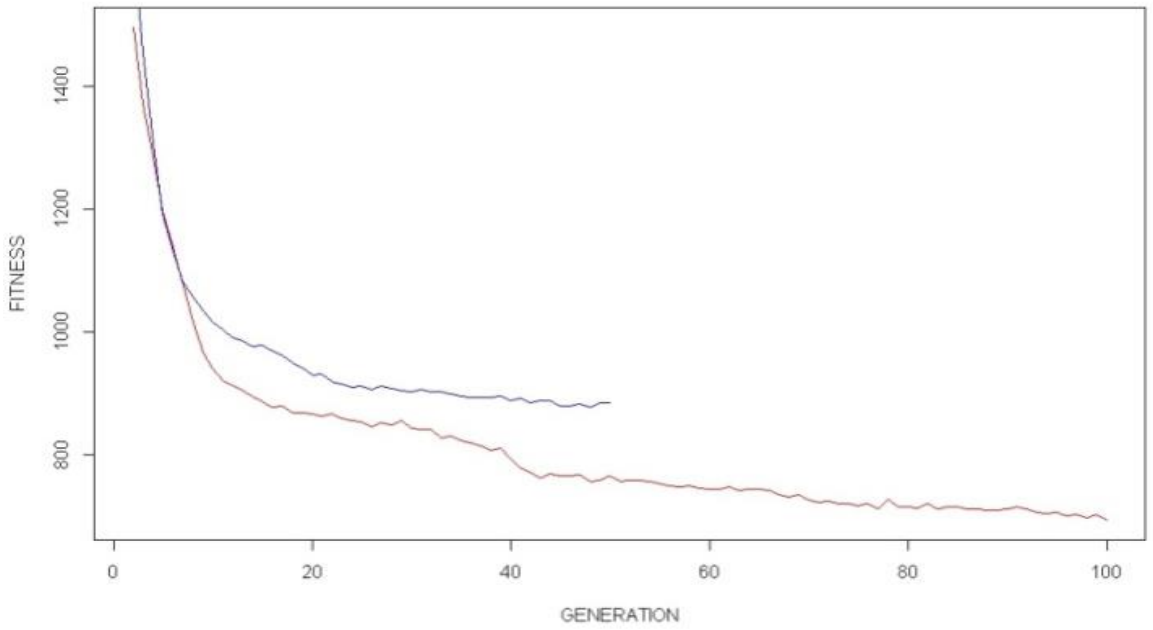

Fig: 3. Significant gene of 50, 100 generations

Hybrid PSO- GA

The experimental result of the proposed approach hybrid PSO- GA with comparison of Normal PSO and GA approach are discussed below. To identify the biomarker the proposed approach PSO-GA identifies the significant gene with a two level optimization process filtering the best gene in every population of the iterations.

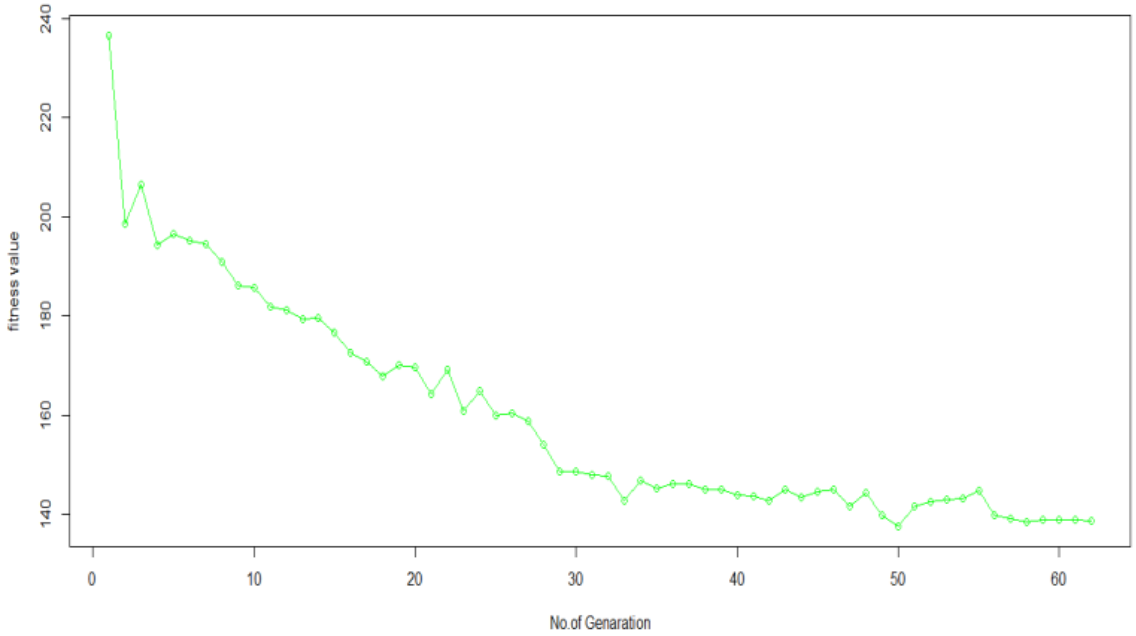

Fig: 4. Best Chromosomes from Hybrid PSO - GA

The **Figure–4** represents the value of 62 best chromosomes derived from hybrid PSO - GA approach. The comparison of gene for the proposed approach PSO-GA with PSO and GA is given in **Table–3** and **Figure–5**. On analysis of gene it is found that 12 common gene found in all the three approaches is termed as highly significant gene. 12 Gene identified from the dataset are:

THE IIOAB JOURNAL
ISSN: 0976-3104

GENE722X, GENE1634X, GENE699X, GENE3836X, GENE2485X, GENE707X, GENE1432X, GENE1806X, GENE185X, GENE2029X and GENE2551X.

Table: 4. Gene counts of PSO, GA & PSO-GA

	No. of iteration	No. of significant gene
Normal PSO	100	33
GA	100	424
Hybrid PSO - GA	100	62

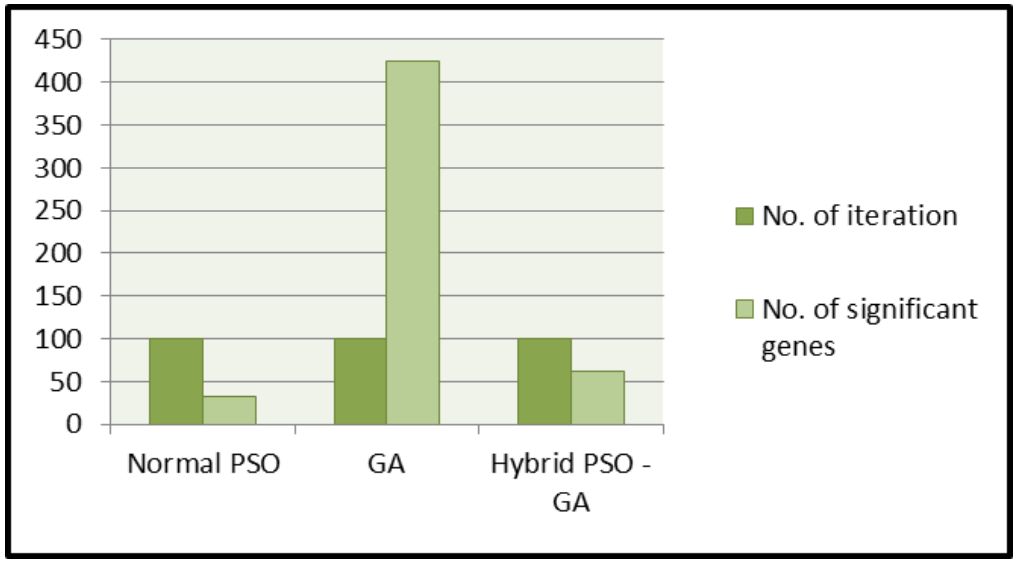

Fig: 5 Comparison of PSO approaches

Biological validation

The extracted gene using PSO-GA approaches are verified using PANTHER tool for validating biological significance. The biological validations of the gene are done using Gene Ontology (GO) functionalities and pathway.

GO functionalities

The identified gene are verified in Gene Ontology functionalities Biological Process, Molecular Function and Cellular Component. A biological process is a process of a living organism. Biological processes are made up of any number of chemical reactions or other events that result in a transformation. Molecular function describes activities, such as catalytic or binding activities, that occur at the molecular level. GO molecular function terms represent activities rather than the entities (molecules or complexes) that perform the actions, and do not specify where or when, or in what context, the action takes place. Cellular component refers to the unique, highly organized substances of which cells, and thus living organisms, are composed. Cells are the structural and functional units of life. The gene ontology found to be for the extracted twelve gene for BP,MF,CC is significant in **Figure– 6(a),6(b),6(c).**

THE IIOAB JOURNAL
ISSN: 0976-3104

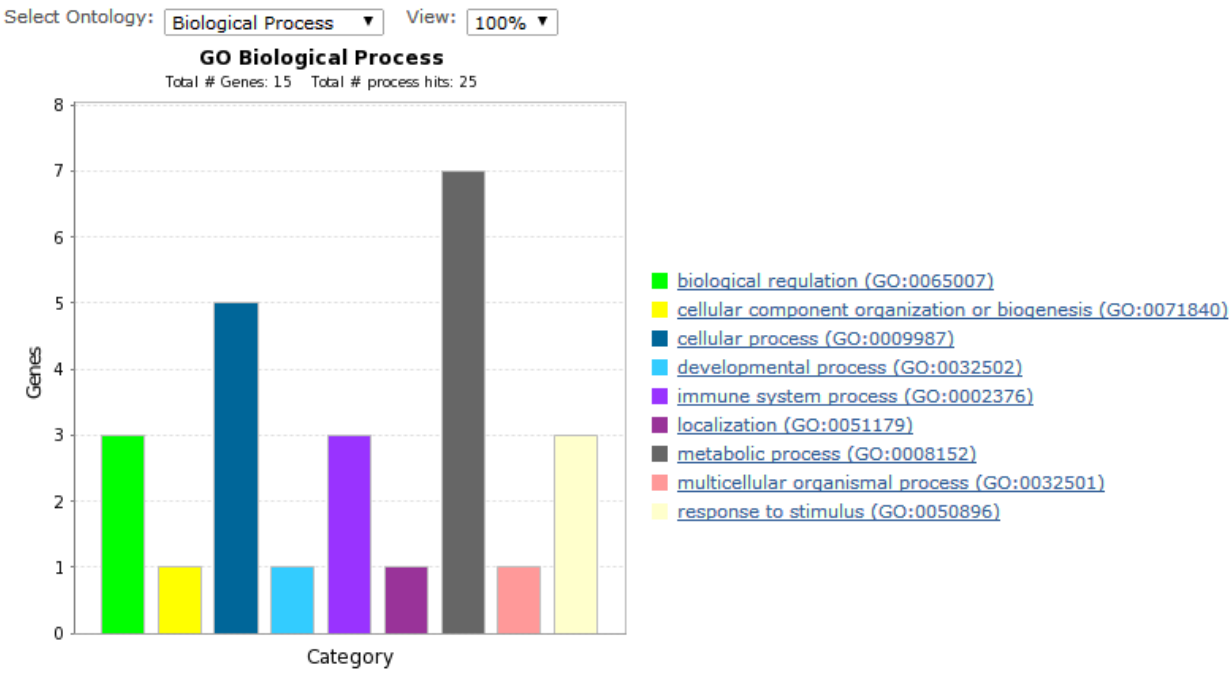

Fig: 6 (a) Biological processes

It can be inferred from **Figure–6(a)** identified 12 gene are involved in 9 biological process like biological regulation, biogeneis, cellular process, development process, immune system process, localisation, metabolic process, multi cellular organism process, response to stimulates and all the 12 gene are active in more than one process

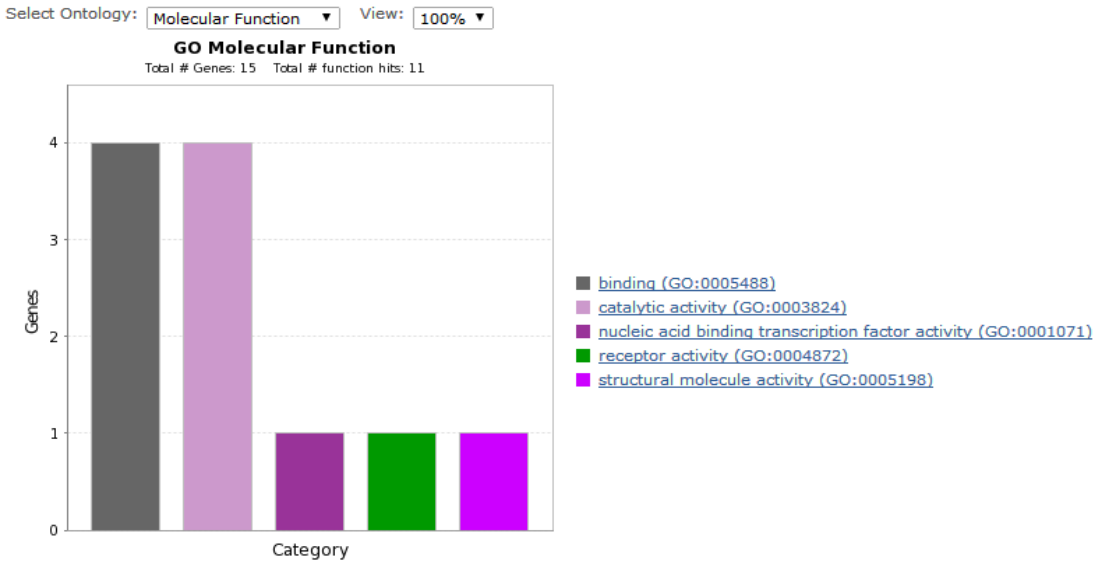

Fig: 6 (b). Molecular functions

THE IIOAB JOURNAL
ISSN: 0976-3104

From **Figure–6 (b)** it can be inferred that identified 12 gene are involved in 5 molecular function like binding, catalytic activity, nucleic acid binding, respirator activity, structural molecular activity and are active in more than one process. Each colour represent the molecular activity of the gene.

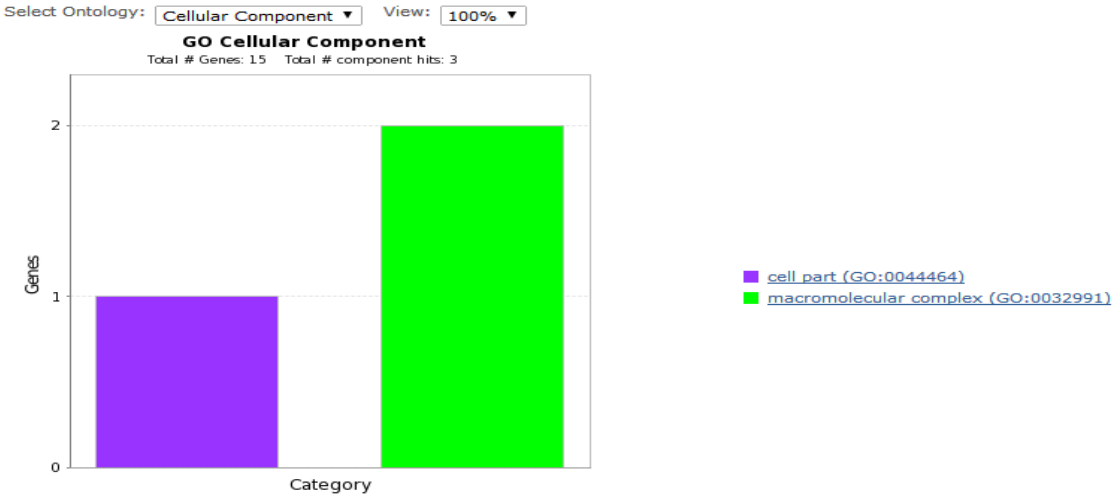

Fig: 6 (c). Cellular components

Figure–6 (c) It can be inferred that identified 12 gene are involved in 2 cellular components like cell part and macromolecular complex. Hence based on the result it is found that the gene are involved in the BP and there are chances to be marked as biomarker. where the five gene GENE1432X, GENE1806X, GENE185X, GENE2029X and GENE2551X identified in this work are also found in literatures [30][31][32].

Pathway validation

The experimental results are also verified with pathway database. The **Figure–7** shows gene involved in the diseases pathways using panther tool for the biological validation of the proposed approach.. The 12 gene identified from PSO-GA approach are found in more than 1 pathway. Out of 12 gene five gene GENE1432X, GENE1806X, GENE185X, GENE2029X and GENE2551X identified in this work is found already in literatures [30][31][32] .
From the experimental analysis the three gene GENE1432X [30], GENE1806X [31], GENE185X [32] are found in the diseases pathway. The three gene are identified in the proposed approach found in the diseases pathway which has to be further verified for biological significance by experts to mark as biomarkers.

In **[Figure–6]**, based on the forecast threshold, the number of data aggregation is larger when compared to the Kalman filter. Due to this data aggregation is reduced redundant transmission and communication consumption power.

CONCLUSION

Biomarker discovery has become important in many aspects of drug discovery and development. Biomarkers predict toxicity and help to save hundreds of millions of dollars with the early termination of costly clinical trials. Many biomarkers play important roles in diagnostic and prognostic applications, where they help to detect or predict diseases. In this work, a hybridized approach of Particle Swarm Optimization and Genetic Algorithm for retrieving highly significant gene is done as a two level optimization process. From the experimental results out of 4026 gene of Lymphoma microarray dataset 12 gene were found to be significant in all the three approaches PSO, GA, PSO-GA. Out of 12 gene 3 gene were found in the diseases pathway and GO functionalities using the biological validation approach of this work. 3 gene when verified with domain experts can be marked as biomarker.

The proposed approach is found to be better for identifying optimal gene using the two level process PSO-GA is verified based on biological validation. The proposed algorithm compared with the other approaches is found to be significant in performance.

THE IIOAB JOURNAL
ISSN: 0976-3104

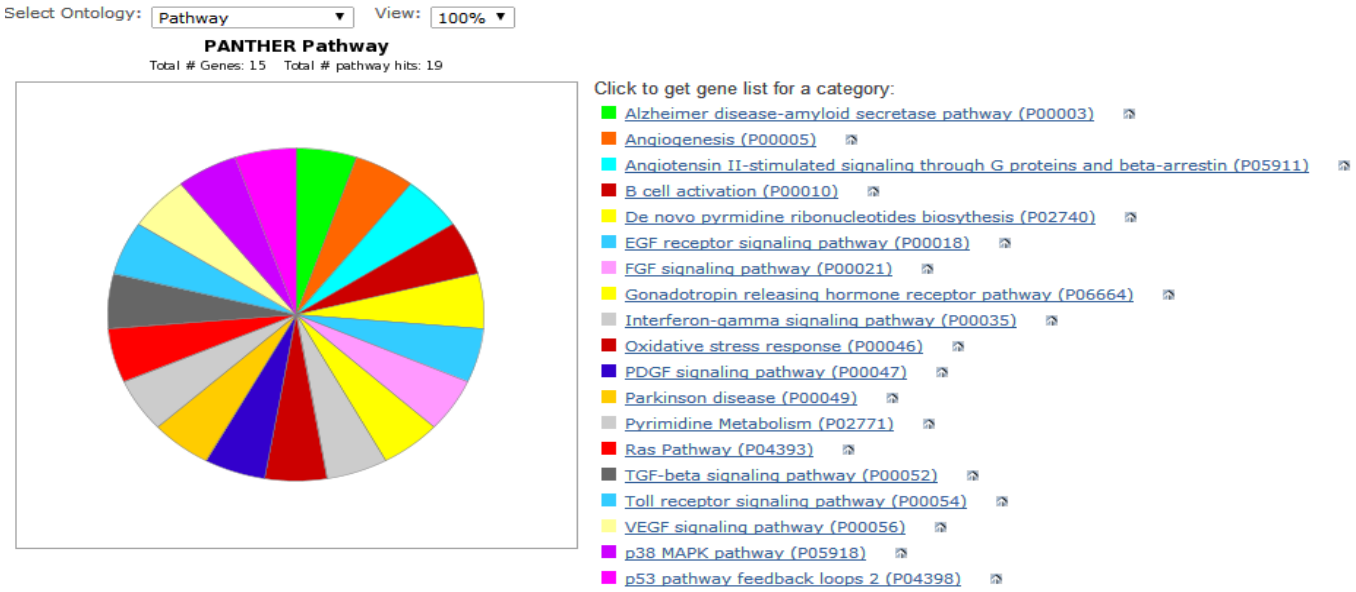

Fig: 7. Pathway of 12 Gene using PSO-GA

LIMITATION AND FUTURE SCOPE

The limitation of the proposed work is, that it is iterated only for little iteration due to the hardware constraints. The work is tested only on one dataset the proposed approach can be verified with other microarray dataset in future. The experimental run can be further extended by implementing using map reduce framework with the existing hardware. The approach will be further tried with other algorithmic approaches in future.

ACKNOWLEDGEMENT
None.

CONFLICT OF INTERESTS
Authors declare no conflict of interest.

FINANCIAL DISCLOSURE
No financial support was received to carry out this project.

REFERENCES

[1] Todd Walter [20.11] A multi-variance analysis in the time domain, 413−426.
[2] Ji-Hoon Cho, Dongkwon Lee et.al.[2004] Gene selection and classification from microarray data using kernel machine. *FEBS Letters* 571 :93–98.
[3] Laurent Gautier et,al. 2004] affy—analysis of Affymetrix GeneChip data at the probe level. *Oxford journals*, 20 (3): 307–315.
[4] Nicola Day et.al., [2012] Alzheimer's disease biomarker discovery using in silico literature mining and clinical validation. *Journal of Translational Medicine*, 10:217
[5] Edmundo Bonilla Huerta et.al., [2008] Gene Selection for Microarray Data by a LDA-Based Genetic Algorithm. *Springer Berlin Heidelberg*, ISSN 0302-9743, 5265: 250−261 .
[6] Ramachandran S Vasan. [2008] Biomarkers of Cardiovascular Disease : Molecular Basis and Practical Considerations. *American Heart Association* ,ISSN: 0009−7322.
[7] Margaret Sullivan Pepe et.al.,[2001] "Phases of Biomarker Development for Early Detection of Cancer", *Journal of the National Cancer Institute*, 93(14)
[8] Paul Yacci et.al. [2009] Feature Selection of Microarray Data Using Genetic Algorithms and Artificial Neural Networks ,RIT Scholar work.
[9] K Hron et al.[2010] Imputation of missing values for compositional data using classical and robust methods. *Computational Statistics & Data Analysis* 54(12):3095–3107.
[10] Matt Blackwell.[2008] Multiple Hypothesis Testing: *The F-test journals*, 2(3): 30–35.
[11] Shinn-Ying Hoa b et.al.[2006] Interpretable gene expression classifier with an accurate and compact fuzzy rule base for microarray data analysis. *Elsevier* 85(3):165−176. Epub.
[12] Guoqiang Yu, et.al.,[2010] Matched Gene Selection and Committee Classifier for Molecular Classification of

Heterogeneous Diseases. *Journal of Machine Learning Research* 11: 2141−2167.

[13] Shital C. Shah, et.al. [2004] "Data mining and genetic algorithm based gene/SNP selection", *Artificial Intelligence in Medicine* 31: 183—196, Elsevier B.V. doi:10.1016/j.artmed.2004.04.002.

[14] Mohd Sazli Saad,[2012] Implementation of PID controller tuning using differential evaluation and genetic algorithms, *ICIC international*, ISSN 1349−4198.

[15] Barnali Sahua et.al.[2012] A Novel Feature Selection Algorithm using Particle Swarm Optimization for Cancer Microarray Data, International Conference on Modeling Optimization and Computing (ICMOC-2012)

[16] Javed khan et.al, [2001] Classification and diagnostic prediction of cancer using gene expression profiling and artificial neural network. *Nature Medicine* **7**: 673 − 679 doi:10.1038/89044.

[17] Keisuke Kameyama, [2009] Particle swarm optimization- a survey, IEICE TRANS.INF. & SYST., Vol-E92-D,No.7 .

[18] Linyu Yang, et.al., [2001] An Entropy-based Adaptive Genetic Algorithm for Learning Classification Rules . 2 ,*IEEE* published in **:**Evolutionary Computation, Proceedings of the 2001 Congress.

[19] Miguel Rocha and Isabel Rocha,[2011] Data Mining Techniques for DNA Microarray Data. *Nature Medicine* **7**: 673 − 679.

[20] Russell d Wolfinger et.al.,[2001] Assessing Gene Signi. cance from cDNA Microarray Expression Data via Mixed Models. Journal of computational biology ,Vol. 8, Pages 625–637.

[21] C.H. Ooiand Patrick Tan,[2003] Genetic algorithms applied to multi-class prediction for the analysis of gene expression data . *oxfordjournals*, 19(1)

[22] JY Yeh,[2008] Applying Data Mining Techniques for Cancer Classification on Gene Expression Data. *Cybernetics & Systems*, 39(6): 583−602. Doi: 10.1080/01969720802188292.

[23] Yongliang Yang et.al.[2008] Integrative Genomic Data Mining for Discovery of Potential Blood-Borne Biomarkers for Early Diagnosis of Cancer, 3(11):3661

[24] Sandrine dudoit, et.al, [2002] Statistical methods for identifying differentially expressed gene in replicated cDNA microarray experiments, *Statistica Sinica* 12: 111−139.

[25] Luciano S´anchez, et.al.,A Multiobjective,[2006] Genetic Fuzzy System with Imprecise Probability Fitness for Vage Data. International Symposium on Evolving Fuzzy Systems, 0-7803-9719-3/06/$20.00 ©2006 IEEE.

[26] K Dheenathayalan, J Ramsingh,V Bhuvaneswari.[2014] Identifying significant gene from DNA microarray using Genetic Algorithm. published in IEEE Xplore – 2014 International Conference on Intelligent Computing Applications - 978-1-4799-3966-4/14 $31.00 © 2014 IEEE DOI 10.1109/ICICA.20120

[27] Vathany Kulasingam, Eleftherios P Diamandis et.al, [2008] Strategies for discovering novel cancer biomarkers through utilization of emerging technologies. *Nature Clinical Practice Oncology*, 5(10)

[28] Alex Kotlarchyk, [2011] Identification of microRNA biomarkers for cancer by Combining multiple feature selection techniques. *Journal of Computational Methods in Sciences and Engineering*, 11(5−6):283−298.

[29] Barbarino Julia M, Staatz Christine E, et.al. [2008] PharmGKB summary: cyclosporine and tacrolimus pathways. *Pharmacogenetics and genomics* .

[30] Maitland Michael L, Lou Xing Jian, et.al..[2010] Vascular endothelial growth factor pathway. *Pharmacogenetics and genomics*

[31] M. Whirl-Carrillo, E.M. McDonagh, et.al. [2012] Pharmacogenomics Knowledge for Personalized Medicine. *Clinical Pharmacology & Therapeutics.*

[32] Anuska M Glas et.al., [2005] Gene expression profiling in follicular lymphoma to access clinical aggressiveness and to guide the choice of treatment. *Blood Journal*;105(1):301−307.

[33] Ahmad m. Sarhan, [2009] Cancer Classification Based on Microarray Gene Expression Data Using DCT and ANN. JATIT.

[34] Christophe Ambroise, et.al,[2002] Selection bias in gene extraction on the basis of microarray gene-expression data. *PNAS*, 99(10).