International Journal of Enhanced Research in Science, Technology & Engineering
ISSN: 2319-7463, Vol. 4 Issue 12, December-2015

Comparision of GA based Approach for Gene Cancer Classification with FFBNN as classifier with PSO for Dimensionality Reduction and ANFIS as Classifier

Vaishali P Khobragade[1], G. Swamy[2]

[1,2]Vivekananda Institute of Technology & Science, Karimnagar, Telangana-505001, India

ABSTRACT

The field of gene cancer classification has seen a remarkable spurge in the current scenario. Many techniques have been devised to cater to the problem of gene cancer classification with microarray technology. These devise techniques mainly comprises of, dimensionality reduction, feature selection, and gene classification for the process of gene cancer classification.. In our work, microarray gene classification by GA with FFBNN was proposed for precise classification of genes to their corresponding gene types. To gauge the efficiency of any method it is better to compare it with existing methods. Thus, analysis is necessary for the techniques that are utilized in the gene classification process. Hence, in this study, we present a comparative analysis of familiar methods that are utilized in the microarray gene classification process. We compare the GA with FFBNN approach with that of PSO with ANFIS .The efficacy of the classification methods are assessed by the performance measures such as accuracy, positive predicate value, false discovery rate, specificity, and sensitivity.

Keywords: Microarray gene expression, Classification, Dimensionality Reduction, Feature Selection, Genetic Algorithm (GA), Feed Forward Back propagation Neural Network (FFBNN) Partial Swarm Optimizer (PSO), Adaptive Neural Fuzzy Inference system (ANFIS).

I. INTRODUCTION

DNA Microarray is one such technology which facilitates the researchers to examine and deal with issues which were some time ago considered to be non-traceable. Deoxyribonucleic Acid (DNA) microarray technology offer tools for examining the expression levels of bulk amount diverse genes at the same time [1]. With this technology at hand the biologists concentrating on the area of genome can investigate the expression level of huge number of genes simultaneously thereby increasing the efficacy of their findings. [2][3][4]. This technology endows a distinctive tool, which is at the moment used for medical diagnosis and gene investigations. Investigation into the expressiveness of the genes under various circumstances and their response can be examined. A microarray method also plays an important function in personalized medicine since it can be used to recognize the individual unique genetic vulnerability to treat the diseases [5]. Prediction, classification and clustering methods are employed in examining and understanding of the data [9]. Microarray gene expression data plays a vital role in categorizing of biological samples [10]. Gene expression data from DNA microarrays are interpret by several variables (genes) with only a small number of observations [7][8. Gene expression profiles appraised by microarray technology have given an exact, consistent and objective cancer classification than the standard checks .The cancer classification data comprises of few samples, but in conjunction with every sample it represents a huge number of genes[11][12]. To infer gene expression data apparently it is very difficult. It is impossible or inconvenient to delve into a huge numeral of genes using conventional methods. DNA Microarray is one such technology which facilitates the researchers to examine and deal with issues which were some time ago considered to be non-traceable. The classification of cancer is a vital topic of investigation in area of medicine. The classification is an imperative step in deducing the treatment and diagnosis of diseases in humans. Due to the nature of the disease(cancer) the techniques used for the classification is requires efficiency and accuracy of the

result for proper monitoring and the treatment. The capability to successfully discriminate among cancer classes which are already identified or yet to be revealed using gene expression data is significant facet for cancer classification.

Particularly, DNA microarray profiling technology is incredibly useful in disease diagnosis and foretelling, at the same time in subtype detection [13][14]. Huge number of expression level of genes corresponding to the available samples can be A standard microarray experiment dataset contains expression levels hundreds of genes in samples [15]. In Biomedical research the microarray technology is emerging as robust tool. The expression data is represented in a matrix form, where the row denotes genes and the columns are denoting the samples. This matrix is termed as gene expression matrix [16]. A robust model is indispensable for predicting the class membership of data, creating an exact label on training data, and predicting the label for any anonymous data correctly in order to achieve a high classification accuracy [8]. Classification analysis of microarray gene expression data has been carried out extensively to determine the biological features as well as to differentiate intimately related cell types that normally appear in the diagnosis of cancer [13]. Some of the classification techniques for gene expression data analysis are decision tree, k-nearest neighbor classifier (KNN), support vector machine (SVM), neural network, etc. Normally, the techniques used for the classification of microarray gene expression data are divided into two groups: one is based on clustering and the other is based on machine-learning approach [14]. There are many techniques developed for the microarray gene classification. In classification, the genes in the microarray dataset are classified into their corresponding class types. Normally, all microarray gene classification techniques perform three basic steps during the classification process, they are: dimensionality reduction, feature selection, and gene classification. In our prior work, microarray gene classification was performed by GA with FFBNN for precise classification of genes to their respective gene types. However, it is uncertain that the GA and FFBNN will perform their operations properly in gene classification process. Thus, an analysis is essential for the techniques that are utilized in the gene classification process. Hence, here we proposed a comparative analysis of well known methods that are used in the microarray gene classification process. The performances of well-known methods such as GA (Genetic Algorithm) are analyzed with the AI techniques namely, FFBNN. The methods that are employed for microarray gene classification process are GA with FFBNN All these methods separately perform the aforementioned three basic steps. The performances of the classification methods are evaluated by the performance measures such as accuracy, specificity, and sensitivity. The rest of the paper is organized as follows: Section 2 reviews the recent related works of the microarray gene classification process. The well known classification method such as GA with FFBNN and PSO with ANFIS is explained in Section 3. The experimental result and conclusion of this paper are given in Section 4 and 5, respectively

II. RELATED WORK

The data from the Microarray gene expression is mostly rendered in the area of cancer classification. Plenty of researchers were carried out for the booming microarray gene cancer classification. The important problem of cancer classification using gene expression data is addressed using various data classification techniques. A handful of recent works available in the literature are reviewed other than the techniques mentioned above are discussed in this section. Chhanda Ray [50] proposed a method with the contour for every graph associated with a DNA gene expression pattern to describe using eight-directional chain code sequence. Initiation of the cancer is identified by considering the variations of DNA gene expression model with concurrently examining the behaviour of thousands of genes. Moreover, the classification of cancer genes was carried out on the basis of probability distribution of codes of the eight-directional chain code sequences indicating DNA microarray gene expression patterns Mark A. Iwen [5] in their work discussed a heuristic approach for classification.

They presented a rule-based gene expression data classifier termed Boolean Structure Table Classification (BSTC). This method clearly relates to association rule-based methods. They have pointed out the significance of their results which guarantees to be in polynomial space or time. Danh V. Nguyen1 [9] have introduced an algorithm which rescales the data based on a threshold. They considered the leukemia microarray data set of Golub et al. 1999. On this data set they applied a Partial Least Squares (PLS) method. They discussed the importance of Principal Components Analysis (PCA), Quadratic Discriminant Analysis (QDA) and Logistic Discrimination (LD). Jian J. Dai_ Linh Lieu [10], in their research provides a review of three rescaling techniques that are Partial Least Squares (PLS), Sliced Inverse Regression (SIR) and Principal Component Analysis (PCA). They have discussed evaluation with respect to the performance of the mentioned techniques. They inferred a simple t-score based gene selection approach works fine with two-class problems. Alok Sharma et.al. [11] in their investigation focused the usage of LDA technique to decrease the size of the feature space for cancer classification considering microarray gene expression data.

They supported the use of the k-nearest neighbour classifier on the reduced data. Donald Geman [13] discussed a rank-based classifier, depending on expression assessment between selected pairs of genes. They carried out the comparisons

invariant to the majority of the transformations implicated in pre-processing and normalization. Cheng-San Yang [7] presented the application of two feature selection methods which belongs to the category of filter methods. The two specific methods demonstrated were information gain and correlation-based selection method. Along with the mentioned filters they used binary particle swarm optimization as a wrapper which performs feature selection on gene subsets which are selected. Thus they evaluated the performance of classification. Gad Getz [15] discussed a joined two-way clustering move towards gene microarray data investigation. The central idea is to recognize subset of the genes and samples. Their idea is to cluster one of the subset when then one the other subset appear to come out as constant and important one. The algorithm presented is based on the repetitive clustering approach that does a search. Andrej Kastrin [16] has discussed Latent variable modelling which offers a promising approach to deal with huge size microarray data. The latent variable model is based on a few latent variables that capture most of the gene expression information. A description is made in such a way to get a reduction in dimension by the use of a latent variable method.

This greatly reduced the number of features considered to characterize microarray data. They proposed a general latent variable structure for prediction of predefined classes of samples using gene expression profiles from microarray experiments, where all individual genes are used in the predictor set. Pengyi Yang and ZiliZhanag [12] proposed an ingrained two-layer approach. They concentrated on combining the best features of filter and wrapper algorithms and avoided their downside. The hybrid algorithm, is called GAEF (Genetic Algorithm with Embedded Filter). This algorithm consists of two steps for feature selection. The first step, uses a Genetic Algorithm (GA) to pre-select features and the considers a filter selector. This approach is used to detect a few feature subset for precise sample classification. For reducing the dimension and to retain the class discrimination information a feature selector is used. Generalized rule induction (GRI) was amalgamate with DQDA to establish relationships rules which give insight of the associations among cancer classes and the related genes. The gene expression data recorded on DNA microarrays has been classified by using this proposed technique. The proposed technique has been tested by using benchmark datasets and it has been found that the technique was faster than neural network and the classification performance was also high compared to neural network.

III. MICROARRAY GENE CANCER CLASSIFICATION

As discussed our previous work [30] the microarray gene classification technique involves three major steps namely (i) Dimensionality reduction, (ii) Feature selection, and (iii) Gene classification. The GA technique performs the dimensionality reduction process for obtaining the dataset with small size. The features like Standard Deviation, Probability of GA-indexed gene, and new statistical features are extracted from the dimensionality reduced dataset. After that, the gene classification is carried out by using the features extracted during the feature extraction process. Here we use FFBNN perform the gene classification process. The basic microarray gene classification process is explained in the following subsections.

DIMENSIONALITY REDUCTION

Initially [30], the dimensionality reduction process is carried out on the microarray cancer gene dataset for diminishing the complexity in the gene classification. This process is performed because the dataset size is high dimensional, which increases the processing time and does not produce accurate result for the classification process. Let, $M_{ij}\ ;\ 1 \le i \le S,\ 1 \le j \le G$ be the microarray cancer gene data, where, S indicates the number of samples and G indicates the number of genes. Dataset M_{ij} contains N number of cancer class types, which is represented as $D_c = \{l_1, l_2, \cdots l_N\}$. The gene dataset can be represented as,

$$M_{ij} = \begin{bmatrix} g_{(1,1)} & g_{(1,2)} & \cdots & g_{(1,G)} \\ g_{(2,1)} & g_{(2,2)} & \cdots & g_{(2,G)} \\ \vdots & \vdots & \vdots & \vdots \\ g_{(S,1)} & g_{(S,2)} & \cdots & g_{(S,G)} \end{bmatrix} \quad (1)$$

Each row and column of the gene expression dataset index values are represented as

$$R_i = \{r_1, r_2, \cdots r_G\}, C_j = \{o_1, o_2, \cdots o_S\} \quad (2)$$

DIMENSIONALITY REDUCTION BY GA

The dimensionality reduction by GA process is briefly explained in the prior work[30]. Initially in GA, the initial chromosome, $C_m = [r_{11}^{(m)}\ r_{22}^{(m)}\ r_{33}^{(m)}\ \cdots\ r_{nK}^{(m)}];\ 0 \le m \le N_p - 1$, where K is the value based on size of the chromosome and n represents the genes row index value in M_{ij} where $n \in r_G$. The fitness function is carried out to choose the best chromosomes among the generated chromosomes. The fitness function is given as,

$$f1 = \frac{S^{(C_m^{(s)})} * S^{(l_1)} * S^{(l_2)} * \cdots S^{(l_N)}}{E^{(C_m)} * T^{(C_m)} * t_1} \quad (3)$$

where, $S^{(C_m^{(s)})}$ is the standard deviation of the chromosome $C_m^{(s)}$ and $S^{(l_1)}, S^{(l_2)}, \cdots S^{(l_N)}$ are also the standard deviations of the genes cancer class types. All the generated chromosomes gene values are given to the networks such as FFBNN, ANFIS and Fuzzy ANN to obtain the error ($E^{(C_m)}$) and time ($T^{(C_m)}$) parameters of the chromosome C_m. $E^{(C_m)}$ is the error produced when the networks are trained by the chromosome C_m. Time parameter $T^{(C_m)}$ represents the time taken by the networks to train the C_m, and t_1 is the defined threshold value. The best $N_p / 2$ chromosomes containing minimum fitness values are selected. The selected chromosomes are involved in the crossover and mutation operations with the single point crossover at crossover rate C_R and mutation rate M_R, respectively. This process is repeated until it reaches the utmost number of iterations I. Once it reaches I, the $N_p / 2$ chromosomes having minimum fitness value are selected. The dimensionality reduced dataset from GA is represented as P_{uv}.

DIMENSIONALITY REDUCTION BY PSO

The dimensionality reduction process is performed over the microarray gene expression dataset M_{ij} by utilizing an optimization algorithm called PSO. The procedure of PSO is discussed below. PSO define each particle as a possible solution to a problem in D-dimensional space. We arbitrarily generate initial particles for genes and velocities for each particle. The randomly generated initial particles and velocity of each particle are represented as,

$$P = (p_1, p_2, p_3 \ldots\ldots p_n) \quad n = 1, 2, 3 \ldots\ldots X \quad (4)$$

$$V = (v_1, v_2, v_3 \ldots\ldots v_n) \quad n = 1, 2, 3 \ldots\ldots X \quad (5)$$

The generated particles and velocities are bounded between the minimum and maximum values i.e., all particles should be within the specified intervals. Before each iteration, the particles are checked to find whether those particles are within the intervals. The gene values of particles are randomly generated between the intervals $[1, r_G]$ in the dataset M_{ij} . The evaluation function values are calculated for each individual particle to determine the optimal solution. From the result of fitness values of all particles, the maximum fitness value is selected as an optimum value. Initially, the optimum value is considered as a pbest (flocal) value and then as a gbest (fglobal) value. The evaluation function can be calculated as,

$$F = \frac{S^{(p_n^{(s)})} * S^{(l_1)} * S^{(l_2)} * \cdots S^{(l_N)}}{E^{(p_n)} * T^{(p_n)} * t_1} \quad (6)$$

where, $S^{(p_n^{(s)})}$ is the standard deviation of the particle $p_n^{(s)}$ and $S^{(l_1)}, S^{(l_2)}, \cdots S^{(l_N)}$ are also the standard deviations of the genes cancer class types. All the generated particles gene values are given to the networks such as FFBNN, ANFIS and Fuzzy ANN to obtain the error ($E^{(p_n)}$) and time ($T^{(p_n)}$) parameters of the particle p_n. $E^{(p_n)}$ is the error produced when the networks are trained by the particle p_n. Time parameter $T^{(p_n)}$ is the time

taken by the networks to train the p_n, and t_1 is the defined threshold value. In initial iteration, the values of velocity are assigned as zero. Using the randomly generated and initial velocity of particles, the fitness values of these particles are determined. We define pbest and gbest values from this fitness result. The pbest value is called local best and gbest value is called global best. All particles having fitness values evaluated by the fitness function need to be optimized. The particles fly through the problem space by following the current optimum particles. After finding the best values, all particles try to change its position and velocity. To change the position, two data are used. First one is the distance between the current particle position and pbest, and second one is the distance between the current position and gbest. This modification can be represented by velocity. Velocity of each particle can be modified by using the following equations,

$$V_n^{(o+1)} = V_n^{(o)} + f_1 * r_1() * (flocal_n - x_n^{(o)}) + f_2 * r_2() * (fglobal_n - x_n^{(o)}) \quad (7)$$

$$x_n^{(o+1)} = x_n^{(o)} + V_n^{(o+1)} \quad (8)$$

Where, $V_n^{(o)}$ is the velocity of n[th] particle at iteration o, and f_1, f_2 are the learning factors. $flocal$ is the position of the best fitness value of the particle at current iteration, $fglobal$ is the position of the particle with the best fitness value in the swarm, r_1, r_2 are the random numbers generated in the range of [0, 1] and $x_n^{(o)}$ is the current position of the particle n at iteration o. Each particle knows its best value (pbest) and position. Also, each particle knows the best value in the group (gbest) among the pbest. Particles change their position and velocity for each iteration until it reaches the termination criteria. This process is repeated until the utmost number of iterations is reached. Once the maximum number of iterations is achieved, then the process gets terminated. The last solution pointing the particle is considered as the best possible particles. The dimensionality reduced dataset from PSO is represented as P'_{uv} .

FEATURE SELECTION

The features [30] are selected from the dimensionality reduced datasets P_{uv} and P'_{uv} . The features like Standard Deviation, Probability of GA-indexed gene, and new statistical features are selected from the dimensionality reduced dataset. The features that are selected from the dataset P_{uv} are briefly explained in the previous work. From this dataset, the extracted features are $F^{(d)}$, $F^{(c)}$, $F^{(A)}$, $F^{(D)}$, $F^{(p)}$ and $F^{(ss)}$. Also, the similar features are extracted from the dataset P'_{uv}, which is represented as $F^{(d)'}$, $F^{(c)'}$, $F^{(A)'}$, $F^{(D)'}$, $F^{(p)'}$ and $F^{(ss)'}$.

GENE CLASSIFICATION

Using the SD, Probability of GA-indexed gene, and new statistical features determined in the previous phase, the gene classification process is carried out. To perform the classification process, here we utilized three AI techniques such as Feed Forward Back Propagation Neural Network (FFBNN), ANFIS, and Fuzzy NN. Each technique is trained and tested with the features that are obtained from the dataset P_{uv} and P'_{uv} , individually.

CLASSIFICATION USING FFBNN

Classification by FFBNN using the features from P_{uv} is already explained in our previous work[30]. In this classification process, the FFBNN is designed with six input neurons, H_d hidden layers, and one output layer. The FFBNN training process is performed with the bias and activation functions of input and output layers, respectively. After that, the network learning error rate is calculated and the error gets minimized by allocating weights to the hidden layer and output layer neurons via back propagation algorithm. Testing process is done for the column gene values in the dimensionality reduced dataset P_{uv} . The well trained FFBNN classifies the column gene values into any one of the cancer class types by using the extracted features. The same FFBNN training and testing process is performed with the features from P'_{uv} .

CLASSIFICATION USING ANFIS

ANFIS architecture comprises five layers of nodes. Out of the five layers, the first and fourth layers are the adaptive nodes while the second, third and fifth layers are the fixed nodes. ANFIS has a five layered feed-forward neural network structure, which is shown in Figure 4.3.

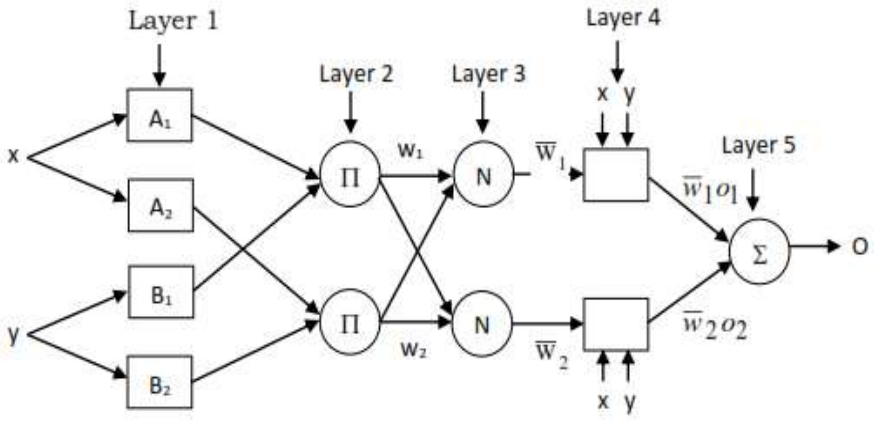


Figure: 1

ANN along with ANN are used here. Both ANN and FL, takes the input and the output as the input and output membership function respectively. The input Sugeno type model is used that makes use of if-then- rules to produce an output for each rule. Rule outputs consist of linear combination of the input variables plus a constant term, and the final output is the weighted average of each rule's output. Sugeno type if-then-rules are defined for both dimensionality reduced datasets by the input values i.e., extracted features. Each layer in the ANFIS system is processed with the input values that are briefly explained [28].

The ANFIS system is well trained by the features from P_{uv} and the same training process is performed in ANFIS with the features from P'_{uv} . Testing process is performed in ANFIS with the one column of gene values in the dataset and the remaining columns gene values are utilized in the training process.

IV. EXPERIMENTAL RESULT

The proposed classification technique is implemented in the MATLAB platform version 7.8 and evaluated using the microarray gene expression dataset. The dataset contains number of genes and samples i.e., 675x156. The high dimensional dataset is subjected to dimensionality reduction using a GA and PSO. The dimensionality reduced dataset from GA and PSO are the dimension of 10x156. Among these 156 samples, 1 to 139 samples are AD class type and 140 to 156 samples are NL class type. Then, the feature selection process is performed over the dimensionality reduced datasets and these selected features are given to the FFBNN, training and testing process. The FFBNN training and testing process is explained in our previous work[30].The abovementioned procedure is performed until all samples are involved in both training and testing process. The performance of proposed technique is evaluated by using the statistical measures. The statistical measures [27] are applied to determine the classification performance. The performance analysis has shown that the proposed technique has successfully classified the genes to their specified gene types. To analyze the performance of GA and PSO methods, the parameters in GA and PSO values are changed. The parameters of GA such as crossover, chromosome length, and mutation rate are changed as well as the PSO parameters such as population size and chromosome length are also changed.

PERFORMANCE OF GA WITH FFBNN

Performance of GA with FFBNN best, worst cases is as follows:

Table: 1 Performance of GA with FFBNN best, worst cases

Statistical	Number	Best	Worst
Sensitivity (%)	10	96.40	89.21
	20	99.28	89.93
	30	99.28	87.77
FPR (%)	10	52.94	58.82
	20	35.29	70.59
	30	41.18	70.59
Accuracy (%)	10	91.03	83.97
	20	95.51	83.33
	30	94.87	81.41
Specificity (%)	10	47.06	41.18
	20	64.71	29.41
	30	58.82	29.41
PPV (%)	10	93.71	92.54
	20	95.83	91.24
	30	95.17	91.04
NPV (%)	10	61.54	31.82
	20	91.67	26.32
	30	90.91	22.73
FDR (%)	10	6.29	7.46
	20	4.17	8.76
	30	4.83	8.96
MCC (%)	10	49.00	27.20
	20	74.83	18.43
	30	70.72	15.38

PERFORMANCE OF PSO WITH ANFIS

The PSO with ANFIS finest and worst cases (bold and italic formats) TP, TN, FP and FN values are listed in the following Table 5.11 and Table 5.12.

Table 2: PSO parameters modification results of TP, FP, TN and FN values

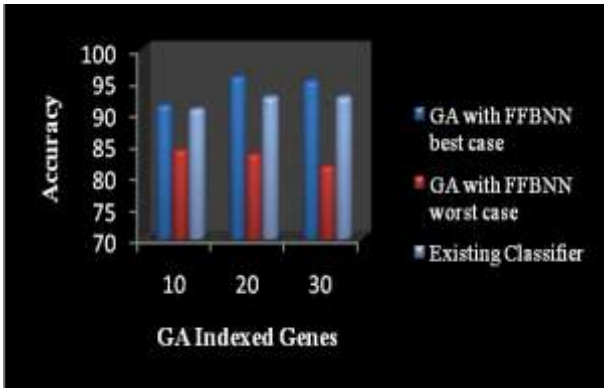
Population size	Particles Length	(TP)	(FP)	(TN)	(FN)	ACC
10	**10**	**133**	**6**	**11**	**6**	**92**
	20	134	9	8	5	91
	30	131	8	9	8	90
20	10	134	10	7	5	90
	20	130	8	9	9	89
	30	**135**	**7**	**10**	**4**	**93**
30	**10**	**139**	**2**	**15**	**0**	**99**
	20	131	5	12	8	92
	30	135	1	16	4	97

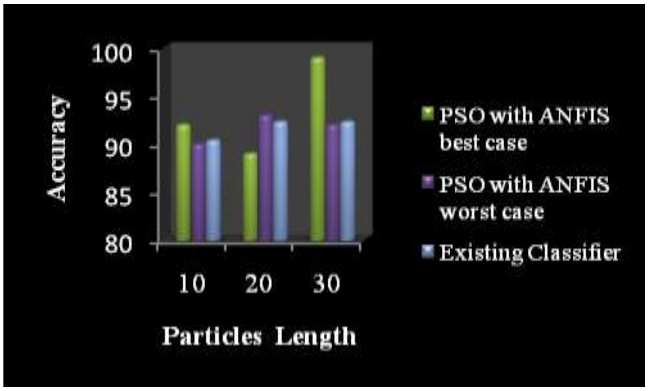
The best and worst case statistical values of the classification of PSO with ANIFS are tabulated & presented in Table 3

Table 3: The best and worst case statistical values of the PSO with ANIFS

Statistical Measures	Particles Length	Best Case	Worst Case	Classifier [26]
Sensitivity (%)	10	96	94	91.37
	20	97	94	94.24
	30	100	94	94.96
FPR (%)	10	35.3	47.1	17.65
	20	41.2	47.1	23.53
	30	11.8	29.4	29.41
Accuracy (%)	10	92	90	90.38
	20	89	93	92.31
	30	99	92	92.31
Specificity (%)	10	65	53	82.35
	20	59	53	76.47
	30	88	71	70.59
PPV (%)	10	96	94	97.69
	20	95	94	97.04
	30	99	96	96.35
NPV (%)	10	65	53	53.85
	20	71	50	61.90
	30	100	60	63.16
FDR (%)	10	4	6	2.31
	20	5	6	2.96
	30	1	4	3.65
MCC (%)	10	59.3	47.0	61.64
	20	60.4	45.2	64.56
	30	89.2	58.9	62.46

When compared with the proposed method PSO with ANIF provides Low accuracy when the GA index is 20, but high accuracy when GA is 30.The comparison of the methods GA with FFBNN and PSO with ANFIS as classifiers in depicted in Figure 5.8. The Figure also depicts the comparison with existing SVM**.**





REFERENCES

[1] Mark A. Iwen, Willis Lang and Jignesh M. Patel, "Scalable Rule-Based Gene Expression Data Classification", In Proceedings of the IEEE 24th International Conference on Data Engineering, 2008

[2] Nguyen and Rocke, Classification of Acute Leukemia based on DNA Micro array Gene Expressions using Partial Least Squares, Kluwer Academic, Dordrecht, 2001

[3] Jian J. Dai, Linh Lieu, and David Rocke, "Dimension Reduction for Classification with Gene Expression Micro array Data", Statistical Applications in Genetics and Molecular Biology: Vol. 5, No. 1, 2006

[4] Alok Sharma and Kuldip K. Paliwal, "Cancer classification by gradient LDA technique using micro array gene expression data", Data & Knowledge Engineering, Vol. 66, pp. 338-347, 2008

[5] Yuh-Jye Lee and Chia-Huang Chao, "A Data Mining Application to Leukemia Micro array Gene Expression Data Analysis", International Conference on Informatics, Cybernetics and Systems (ICICS), Kaohsiung, Taiwan, 2003

[6] Donald Geman, Bahman Afsari, Aik Choon Tan and Daniel Q. Naiman, "Micro array Classification from Several Two-Gene Expression Comparisons", In Proceedings of ICMLA, pp.583-585, 2008

[7] Chun-Hou Zheng, Bo Li, Lei Zhang and Hong-Qiang Wang, "Locally Linear Discriminant Embedding for Tumor Classification", In Proceedings of ICIC, pp.1093-1100, 2008

[8] Cheng-San Yang, Li-Yeh Chuang, Chao-Hsuan Ke and Cheng-Hong Yang, "A hybrid Feature Selection Method for Micro array Classification", International Journal of Computer Science, Vol. 35, No. 3, 2008

[9] James J. Chen and Chun-Houh Chen, "Micro array Gene Expression", Encyclopedia of Biopharmaceutical Statistics, 2nd Edition, Marcel Dekker, Inc., pp. 599-613, 2003

[10] Gad Getz, Erel Levine and Eytan Domany, "Coupled two-way clustering analysis of gene micro array data", In Proceedings of the National Academy of Sciences, Vol. 97, No. 22, pp. 12079-12084, 2000

[11] Andrej Kastrin, "Item Response Theory Modeling for Micro array Gene Expression Data", Advances in methodology and statistics, Vol. 6, No. 1, pp. 51-67, 2009

[12] Srinivas Mukkamala, Qingzhong Liu, Rajeev Veeraghattam and Andrew H. Sung, "Computational Intelligent Techniques for Tumor Classification (Using Micro array Gene Expression Data)", In Proceedings of World Congress on Lateral Computing, 2005

[13] Fei Pana, Baoying Wanga, Xin Hub and William Perrizoa, "Comprehensive vertical sample-based KNN/LSVM classification for gene expression analysis", Journal of Biomedical Informatics, Vol. 37, pp. 240–248, 2004

[14] Peng Qiu, Z. Jane Wang and Ray Liu, "Ensemble Dependence Model-based Cancer Classification using Gene Micro array Data", IEEE International Workshop on Genomic Signal Processing and Statistics (GENSIPS), 2005

[15] Seeja and Shweta, "Microarray Data Classification Using Support Vector Machine", International Journal of Biometrics and Bioinformatics (IJBB), Vol. 5, No. 1, pp. 10-15, 2011

[16] Manoj Kumar, "The Widely Used Diagnostics “DNA Microarray”-A Review", American Journal of Infectious Diseases, Vol. 5, No. 3, pp. 207-218, 2009
[17] Danh V. Nguyen, David M. Rocke, "Tumor Classification by Partial Least Squares Using Micro array Gene Expression Data", Bioinformatics, Vol. 18, No. 1, pp. 39-50, 2002
[18] Pengyi Yang and Zili Zhang, "An Embedded Two-Layer Feature Selection Approach for Microarray Data Analysis", IEEE Intelligent Informatics Bulletin, Vol.10, No.1, pp. 24-32, 2009
[19] Yee Hwa Yang and Natalie P. Thorne, "Normalization for Two-color cDNA Microarray Data", Science and Statistics: A Festschrift for Terry Speed, Vol. 40, pp. 403-418, 2003
[20] Ahmad M. Sarhan, "Cancer Classification Based on Micro array Gene Expression Data Using DCT and ANN", Journal of Theoretical and Applied Information Technology, Vol. 6, No. 2, pp. 208-216, 2009
[21] Bharathi and Natarajan, "Cancer Classification of Bioinformatics data using ANOVA", International Journal of Computer Theory and Engineering, Vol. 2, No. 3, pp. 369-373, June 2010
[22] Bo Li, Chun-Hou Zheng, De-Shuang Huang, Lei Zhang and Kyungsook Han, “"Gene expression data classification using locally linear discriminant embedding", Computers in Biology and Medicine, Vol. 40, pp. 802–810, 2010
[23] Xiaosheng Wang and Osamu Gotoh, "A Robust Gene Selection Method for Micro array-based Cancer Classification", Journal of Cancer Informatics, Vol. 9, pp. 15-30, 2010
[24] Mallika and Saravanan, "An SVM based Classification Method for Cancer Data using Minimum Micro array Gene Expressions", World Academy of Science, Engineering and Technology, Vol. 62, No. 99, pp. 543-547, 2010
[25] Chhanda Ray, "Cancer Identification and Gene Classification using DNA Micro array Gene Expression Patterns", International Journal of Computer Science Issues, Vol. 8, Issue 2, pp. 155-160, March 2011
[26] Seeja and Shweta, "Microarray Data Classification Using Support Vector Machine", International Journal of Biometrics and Bioinformatics (IJBB), Vol. 5, No. 1, pp. 10-15, 2011http://en.wikipedia.org/Sensitivity_and_specificity
[27] AnandaKumar and Punithavalli, "Efficient Cancer Classification using Fast Adaptive Neuro-Fuzzy Inference System (FANFIS) based on Statistical Techniques", International Journal of Advanced Computer ScienceandApplicationspp.132-1372011.
[28] Heiss, Held, Estévez, Perez,Holzmann and Perez, "Classification of Sleep Stages in Infants: A Neuro Fuzzy Approach", IEEE Engineering in Medicine and Biology Magazine, Vol. 21, No. 5, pp. 147-51., 2002
[29] P.K.Vaishali & Dr.A.VinayaBabu.”,A Classification of Microarray Gene Expression Data Using Hybrid Soft Computing Approach”, IJCSI Vol 9,Issue 6.

.