# Complexity-Reduced Tumor Classification System using Microarray Gene Expression Dataset

N.Gopala Krishna Murthy
Associate Professor
Department of CSE,
G.V.V.R.Institute of Technology,
Andhra Pradesh

O.NagaRaju
Head of the Department,
Department of Computer Science,
Government Degree College,
Andhra Pradesh

Allam Appa Rao
Director, CR Rao Advanced
Institute of Mathematics, Statistics
& Computer Science (AIMSCS),
University of Hyderabad Campus,
Andhra Pradesh

## ABSTRACT

The classification of cancer based on gene expression data is the advancements in DNA Microarray technology and genome sequencing. The important feature is to predict the genes for various diagnosis purposes using such micro-array gene expression dataset and also the gene expressions that are relevant to a particular type of genes. Lot of research works have been carried out to produce a better solution to improve the prediction accuracy of cancer gene prediction. But the analysis results are not up to the convincing level artificial intelligence is exploited to improve the prediction accuracy meanwhile state-of-the art insists necessary enhancements which are essential in the classification module instead in the features module. The enhanced classifier called Principal component analysis used in latter researches is used in this work for the performances comparison of the classifier as a conventional prediction methodology.

This work intends to apply the developed classifier and dominant gene prediction methodology to predict extensive set of cancer expression datasets. The experimental study will be carried out by considering the techniques on CNS tumor, colon tumor and ALL_AML Leukemia. The prediction performance of the proposed methodology will be compared against the conventional prediction methodologies and the results will be validated extensively. The method will be implemented in the working platform of MATLAB and the performance will be analysed.

## Keywords

Artificial Neural Network, Artificial Bee Colony algorithm, CNS tumor, colon tumor and ALL_AML Leukemia.

## 1. INTRODUCTION

In Microarray technology, vast quantities of data are collected because of the quantity of information brought forth from every experiment [1]. Microarray technology has become one of the indispensable tools that several biologists utilize to monitor genome broad expression levels of genes in a provided organism [2]. Generally, a microarray is a glass slide on to which DNA molecules are fixed in an orderly manner at particular locations named spots. By the presence of genetic variations [3] [4] nearly all human genetic diseases such as cancer and developmental abnormalities are classified. To explore significant normal and abnormal biological phenomena, the discovery and analysis of gene expression patterns of several model organisms shows a fascinating opportunity. Within a particular mRNA sample [5], DNA microarrays can simultaneously measure the expression level of thousands of genes. In order to: 1) identify diagnostic or prognostic biomarkers; 2) classify diseases; 3) monitor the response to therapy; and 4) understand the mechanisms involved in the genesis of disease processes [7], such high-throughput expression profiling can be used to compare the level of gene transcription in clinical conditions.

Many microarray technologies that supervise the level of expression of a multitude of genes have lately come forth [6]. In several ways to measure gene expression, Microarrays may be utilized, but one of the most popular applications is to compare expression of a set of genes from a cell maintained in a specific condition to the same set of genes from a reference cell maintained under normal conditions [8]. Provided DNA-microarray information for a set of cells classified by a rendered phenotype and for a set of control cells, an significant issue is to recognize "patterns" of gene expression that can be applied to predict cell phenotype. It is possible to examine the expression of thousands of genes at once with microarrays. Checking for elevated expression of sure genes can help in foretelling cancer [9].

As an imperative tool, Micro array technology has come forth in the tracking of genome-wide expression levels of gene [10]. In diverse organs applying micro array technologies [11], separate genes, gene ensembles, and the metabolic ways fundamental to the structurally workable organization of an organ and its physiological function are brought out by the analysis of the gene expression profiles. The application of micro array technology can automate the diagnostic task and improve the precision of conventional diagnostic techniques. By micro array technology [12], simultaneous analysis of thousands of gene expressions is alleviated. Gene expression microchip likely the most speedily developing tool of genome analysis allows supervising of the expression levels of tens of thousands of genes under various experimental conditions simultaneously. In their atmosphere, this shows a rich tool in the analysis of united gene reaction to variations, and renders suggestions related to the organization of the concerned gene networks [13]. The usage of fuzzy logic for interpretation of gene expression information has not been searched considerably. As a result of the fuzzification of measurements, initial investigations proposed that poor quality clusters are made. Estimating the generalizability of these experiments is not simple because calculating the quality of clusters goes forward to be an extremely subjective task, and various fuzzification systems have not been attempted [14].

To diminish the dimensionality of gene expression information [15] many methods have been suggested earlier. Applying micro array information, numerous machine learning techniques have been effectively used to cancer classification [16] [17]. But, categorization in micro array technology is considered to be extremely hard, because of the high dimensionality and insignificant sample size of the gene expression information. For the successful classification of gene expression data, lots of researches have been performed. In the following section [19] [20], a few recent works available in the literature are reviewed.

## 2. RELATED WORKS AND RESEARCHES

Three various classification methods have been focused on by Jayashree Dev *et al.* [21]: BPN, FLANN and PSO-FLANN and discovered that the integrated approach of Functional Link Artificial Neural Network (FLANN) and Particle Swarm Optimization (PSO) could foretell the disease as compared to other method. The nonlinearity of the classification issue is overcome by this suggested technique. In order to classify various kinds of cancer genes from vast amount of DNA microarray gene expression information, this suggested algorithm could be developed.

For gene expression information analysis, an algorithm has been suggested by Alok Sharma *et al.* [22]. The algorithm first separates genes into subsets, the sizes of which are comparatively small, then chooses informative smaller subsets of genes from a subset and merges the selected genes with another gene subset to modify the gene subset. Till all subsets were merged into one informative subset, they iterated this process. For all the test datasets, they depicted promising classification accuracy. By examining three distinct gene expression datasets, they illustrated the effectiveness of the suggested algorithm.

Classification method for microarray gene expression has been suggested by A. Castano F. *et al.* [23], obtained by the light reflection analysis of genomic material. This suggested algorithm was in two-stages, in the first stage; from thousands of genes, two filter algorithms recognize prominent expression genes. The suggested methodology was performed, in the second stage, utilizing chosen gene subsets as new input variables. The methodology was composed of a combination of Logistic Regression (LR) and Evolutionary Generalized Radial Basis Function (EGRBF) neural networks which have depicted to be extremely precise in previous research in the modeling of high-dimensional patterns. The results received were contrasted, at last, with nonparametric statistical tests and verify good synergy between EGRBF and LR models.

An algorithm has been suggested by Chhanda Ray *et al.* [24] to examine DNA microarray gene expression patterns efficiently for vast amount of DNA microarray information. On the basis of the variations of DNA microarray gene expression patterns of the same organism by concurrently supervising the expression of thousands of genes, this development method was identified. Finally, based on the distribution probability of codes, classification of cancer genes was also focused.

Microarray analysis or gene expression profiling has been discussed by Venkatesh *et al.* [25] that gives techniques to examine thousands of genes in a single sample. By providing large amount of data, Micro array analysis was providing challenges in various fields, which could be processed to get helpful data. In this study, the gene samples received from biopsy samples are gathered from colon cancer patients. They brought in a learning vector quantization method that determines arti-fact states and separate malignant genes from regular genes. Finally, organism was identified based on the variations of DNA microarray gene expression patterns of the same organism.

## 3. ARTIFICIAL BEE COLONY (ABC) ALGORITHM

Artificial Bee Colony (ABC) algorithm is based on the representation of the bees searching behavior. Generally the bee colony consists of a single queen bee that is responsible for laying eggs, thousands of male bees called drones and thousands of worker bees, which are the sterile bees, and the young bee larvae are called broods. The artificial bee colony algorithm consists of 3 types of bees namely the employed bee, onlooker bee and the scout bee. Scout bee is responsible for carrying out random searches in the environment. A bee who visits the source visited by him previously is called an employed bee and the bee that waits in the beehive for decision making is called the onlooker bee. In the ABC algorithm it is assumed that the colony consists of an equal number of employed bees and onlooker bees and for every solution there is an employed bee in the hive. An onlooker waits and decides on whether an optimal solution is acceptable or not. The fitness of the species is given by the profitability of the solution it describes. Information survives by continuing to circulate within the nest, and is capable of reproducing itself by recruiting new foragers who become informed of the food source and come back to the nest and share their information.

## 4. CLASSIFICATION OF TUMOR GENES

The tumor genes are classified in to subsets using the different classifiers. The different classifiers used here are Principal component analysis (PCA) and Artificial Neural network (ANN). This can be done to reduce the dimension of the datasets to make the analysis easier. Dimension reduction has the approaches to find a subset of the original variables and transforms the data in the high-dimensional space to a space of fewer dimensions.

### 4.1 Principal Component Analysis

PCA is a linear transformation technique to form a simplified data set retaining the characteristics of the original data set. The transformation is given by

$$Y = E^T .X \tag{1}$$

Consider the original data matrix contains $d$ dimensions and $n$ observations which have to be reduced to $k$ dimensional subspace. Here $E_{d \times k}$ is the projection matrix which contains $k$ Eigen vectors corresponds to the $k$ highest Eigen values, and where $X_{d \times n}$ is a mean centred data matrix.

### 4.2 Artificial Neural Network

The dimensionality reduced microarray gene dataset is utilized for training the feed forward Neutral network with back propagation. The network maps a set of input values to a set of output values and the graphical representation of a parametric function is supposed. The neural network to train the gene dataset is shown in the fig (1). The steps to be followed in this process is given below

*Step 1:* Set the input weights of every neuron network with $N_g^{''}$ input layers, a $N_g^{''}$ hidden layers and an output layer are designed.

*Step 2:* The designed NN is weighted and biased. The developed NN is shown in the Fig.1.

*Step 3:* The basis function and the activation function which is chosen for the designed NN are shown in Eqn (2) and Eqn (3) respectively whereas $\hat{M}_c$ is the dimensionality reduced microarray gene data, $w_{ij}$ is the weight of the neuron, $\alpha$ is the bias and y ranges [0,1].

$$Y_i = \alpha + \sum_{j=0}^{N_g''-1} w_{ij}\hat{M}_{cij}, \quad 0 \le i \le N_s''-1 \tag{2}$$

$$g(y) = \frac{1}{1+e^{-y}} \tag{3}$$

*Step 4:* The basis function given in Eq. (2) is commonly used in all the remaining layers (hidden and output layer, but with the number of hidden and output neurons, respectively). The output of the ANN is determined is determined by giving it $\hat{M}_c$ as the input.

*Step 5:* The learning error is determined for the NN as follows where $E$ is the error in the FF-ANN, $D$ is the desired output and $Y_b$ is the actual output.

$$E = \frac{1}{N_s''} \sum_{b=0}^{N_s''-1} D - Y_b \tag{4}$$
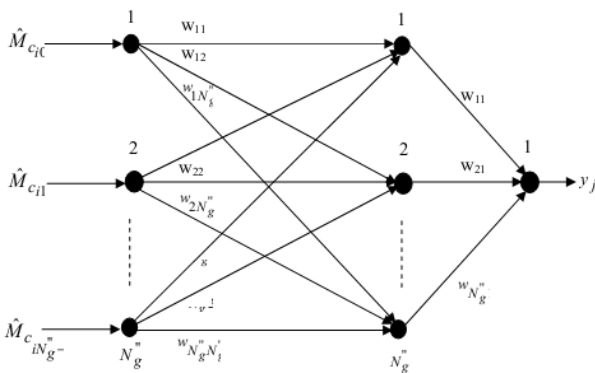


Fig 1 $n$ Inputs one output Neural Network to train the gene dataset

The pre-processing steps for predicting dominant genes are explained in the following steps in the fig (2).
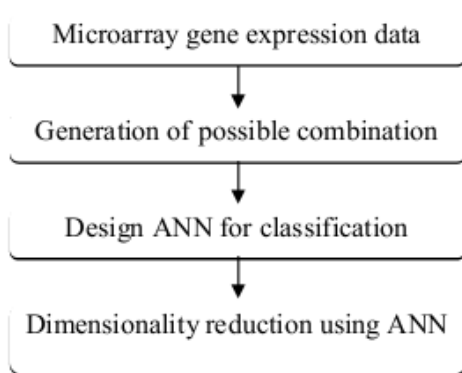


Fig 2 Pre-processing steps for dominant gene prediction

## 4.3 Minimization of Error by BP Algorithm
The Back Propagation algorithm is used to train the network and to minimize the total errors. The steps involved in Back Propagation Algorithm are discussed below:

*Step 1*: Assign random weights to the links range [0, 1] in the Artificial Neural network above

*Step 2*: In case of training perform output calculation based on two functions i.e. Basis function (the product of weights and inputs) and Activation function (non-linear).

*Step 3*: In order to determine the BP error using Eq. (2), the training gene data sequence is given to the NN. Eq. (2), Eq. (3) and Eq. (4) show the basis function and transfer function.

*Step 4*: The weights of all the neurons are adjusted when the BP error is determined as follows,

$$w_{ij} = w_{ij} + \Delta w_{ij} \tag{5}$$

The change in weight $\Delta w_{ij}$ given in Eq. (5) can be determined as $\Delta w_{ij} = \gamma . y_{ij} . E$ , where $E$ is the BP error and $\gamma$ is the learning rate; normally it ranges from 0.2 to 0.5.

*Step 5*: After adjusting the weights, steps (2) and (3) are repeated until the BP error gets minimized. Normally, it is repeated till the criterion, $E < 0.1$ is satisfied.

When the error gets minimized to a minimum value it is construed that the designed ANN is well trained for its further testing phase and the BP algorithm is terminated. Thus, the neural network is trained by using the samples.

# 5. DOMINANT GENE PREDICTION USING ABC ALGORITHM
The ABC algorithm is applied on the classified test sequence and then this test sequence is evaluated and the dominant genes are predicted. This can be done using the random selection of population using the search algorithm. The random populations are the indices of the test sequence which are classified as CNS tumor (or) colon tumor (or) ALL_AML Leukemia. The fitness of the selected population is evaluated using (7) where $N_{out}$ is the network output obtained from the ANN $N_{fit}$ in (6) is the fitness value of the initially generated population.

$$N_{fit} = \frac{1}{(1-\mu_{net})^c} \tag{6}$$

$$\mu_{net} = \frac{\sum_{k=0}^{SN-1} N_{out}}{|k|} \tag{7}$$

$$\begin{cases} c = 0 \ if \ test \ sequence \ is \ class \ 1 \\ c = 1 \ if \ test \ sequence \ is \ class \ 2 \end{cases} in \ case \ of \ CNS \ tumor$$

$$\begin{cases} c = 0 \ if \ test \ sequence \ is \ negative \\ c = 1 \ if \ test \ sequence \ is \ positive \end{cases} in \ case \ of \ colon \ tumor$$

$$\begin{cases} c = 0 \ if \ test \ sequence \ is \ ALL \\ c = 1 \ if \ test \ sequence \ is \ AML \end{cases} in \ case \ of \ ALL\_AML \ Leukemia$$

The steps involved in the ABC algorithm is discussed below

Step1: Initialize the population of solutions $x_{ij}$ , $i = 1 \ldots SN$ and $j = 1 \ldots D$ where $SN$: the number of food source and D is the number of optimization parameters.

Step2: Evaluate the population by the means of fitness

Step3: Produce new solutions $v_{ij}$ for the employed bees by using (8) and evaluate them

$$v_{ij} = x_{ij} + \phi_{ij}(x_{ij} - x_{kj}) \tag{8}$$

Step4: Apply greedy selection process using the probability values $p_i$ for the solutions $x_{ij}$ using (2) where $fit_i$ is the fitness value of the solution $i$ which is proportional to the nectar amount of the food source in the position $i$ and SN is the number of food sources which is equal to the number of employed bees.

$$p_i = \frac{fit_i}{\sum_{n=1}^{SN} fit_n} \tag{9}$$

Step5: Produce the new solutions $v_{ij}$ for the onlookers from the solutions $x_{ij}$ selected depending on $p_i$ and evaluate them

Step6: Apply the greedy selection process

Step7: Determine the abandoned solution for the scout, if exists, and replace it with a new randomly produced solution $x_{ij}$ by using (10)

$$x_i^j = x_{\min}^j + rand(0,1)(x_{\max}^j - x_{\min}^j) \tag{10}$$

Step8: Memorize the best solution achieved and $cycle = cycle + 1$ until the maximum no. of cycle reached.

## 6. EXPERIMENTAL RESULTS

The proposed classification technique is implemented in the MATLAB platform and it is evaluated using the microarray gene expression data. The training dataset is of dimension $N_g = 7129$ and $N_s = 60$ in case of CNS tumor, $N_g = 2000$ and $N_s = 62$ in case of colon tumor, $N_g = 7129$ and $N_s = 72$ in case of ALL_AML [26]. This high dimensional training dataset is subjected to dimensionality reduction using ANN and so a dataset of dimension $50\times60$, $50\times62$, $50\times72$ for CNS tumor, colon tumor and ALL_AML respectively is obtained.

This training dataset is utilized to design the ANN and then the test input sequence is tested through the ABC algorithm. Testing is done under the evaluation of sensitivity and specificity values. These values are among the terms true positive (TP), true negative (TN), false positive (FP), false negative (FN) [27] which are given in the Table (1) below

**Table 1. TP, TN, FP and FN values for four datasets for the classifiers PCA and ABC algorithm with ANN**

| Data sets | Dimension reduction using PPCA | | | | Dimension reduction using ABC with ANN | | | |
|---|---|---|---|---|---|---|---|---|
| | TP | TN | FP | FN | TP | TN | FP | FN |
| CNS tumor | 6 | 12 | 11 | 9 | 21 | 12 | 11 | 9 |
| Colon tumor | 0 | 25 | 0 | 13 | 21 | 22 | 1 | 18 |
| ALL_AML | 2 | 12 | 11 | 13 | 1 | 24 | 12 | 1 |

The graphical representation for the Sensitivity, Specificity, FPR, accuracy, PPV, NPV, FDR, MCC for the PPCA and GA with ANN classifiers under different datasets are shown below
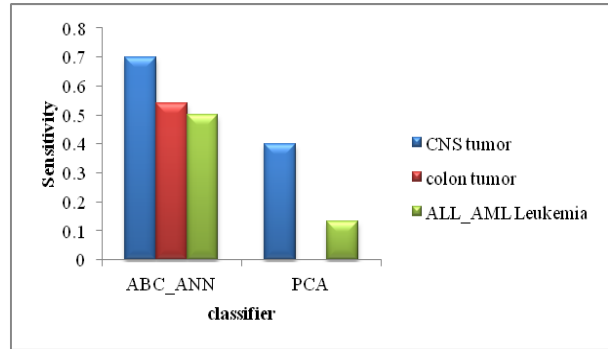


Fig 3 Graphical representation of Sensitivity value for PCA and ABC with ANN classifiers under different datasets
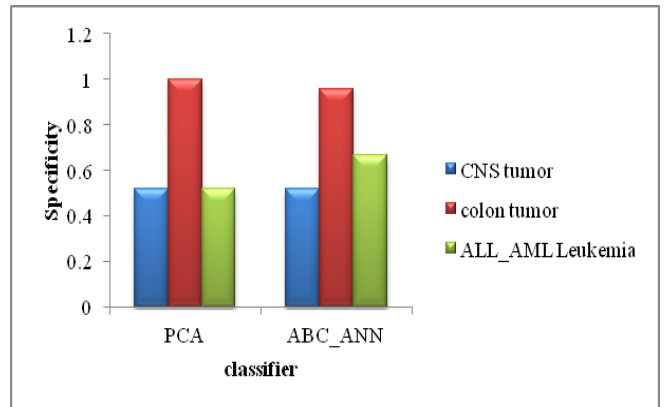


Fig 4 Graphical representation of Specificity value for PCA and ABC with ANN classifiers under different datasets
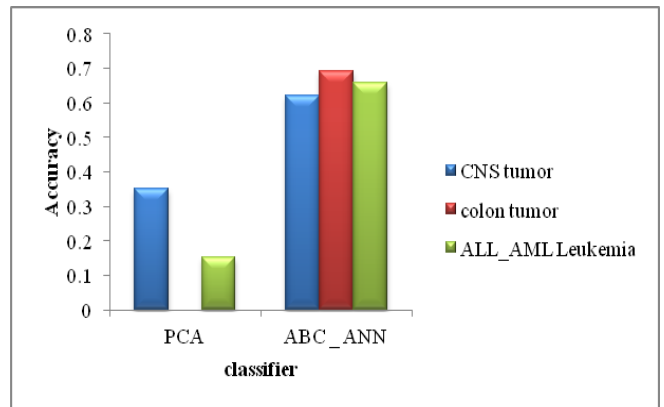


Fig 5 Graphical representation for accuracy of PCA and ABC with ANN classifiers under different datasets
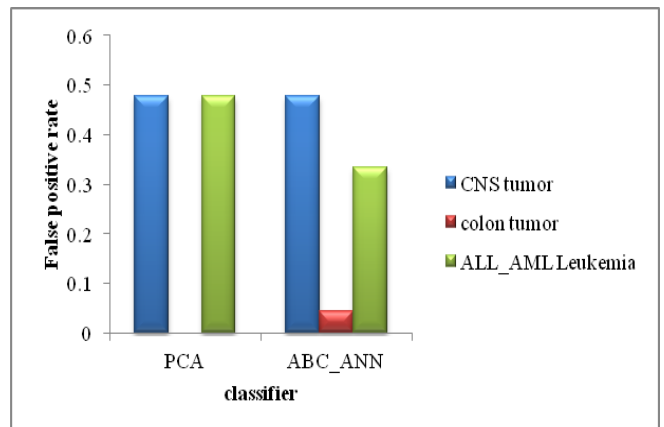
Fig 6 Graphical representation of false positive rate for PCA and GA with ANN classifiers under different datasets
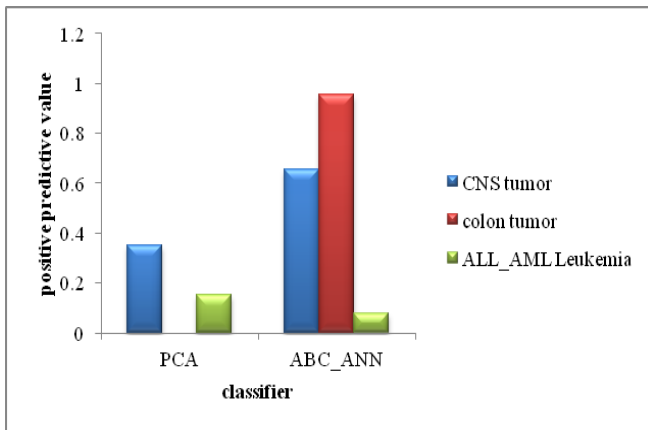


Fig 7 Graphical representation of positive prediction value for PPCA and GA with ANN classifiers under different datasets
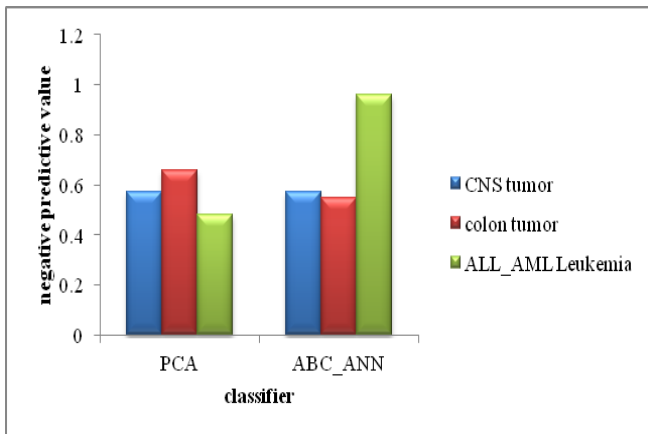


Fig 8 Graphical representation of negative prediction value for PCA and ABC with ANN classifiers under different datasets
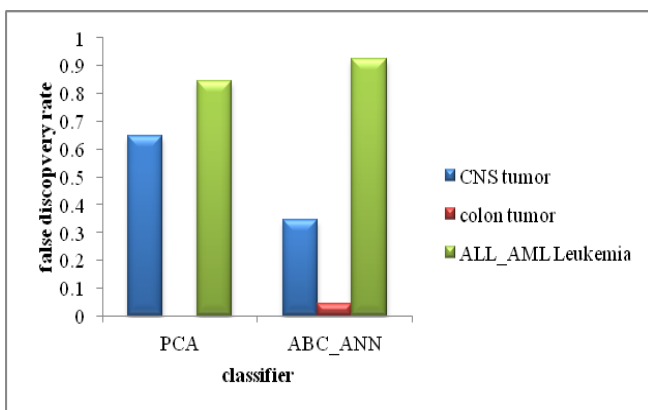


Fig 9 Graphical representation of false discovery rate for PCA and ABC with ANN classifiers under different datasets
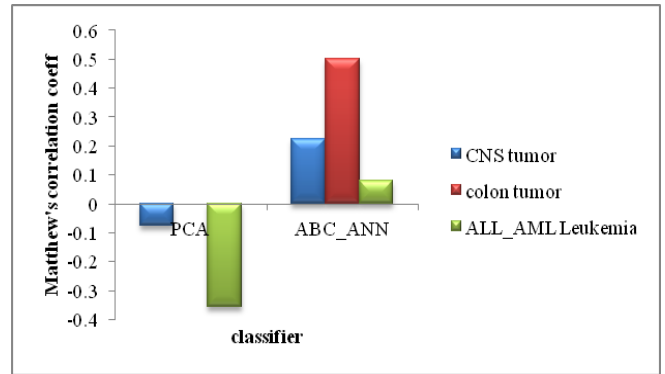


Fig 10 Graphical representation of Matthew's correlation coefficient for PCA and ABC with ANN classifiers under different datasets

## DISCUSSION

Considering the values in the Table (1), sensitivity, specificity, accuracy, positive prediction value, negative prediction value, false positive rate, false discovery rate and Matthews's correlation coefficient can be found out. The graphical representations of the testing values are plotted in the figures 3 – 10.

In case of sensitivity in fig (3), the cancer prediction rate using the classifier ABC with ANN is better than PCA whereas ABC with ANN performs with 50%-65% but PCA has only 40%. The specificity in fig (4) also yields a better outcome with 50% - 100% in case of ABC with ANN but PCA with 35% - 60%. The accuracy rate in fig (5) gives 35% in PCA and 70% in ABC with ANN. The false positive rate in fig (6), which is recognized as an error, is observed as average rate of 32% in PCA and 29% in ABC with ANN. In fig (7), positive predictive value which is the proportion of the positive results has been identified as maximum of 30% in PCA and 95% in ABC with ANN. In fig (8) the negative prediction value, the proportion of negative results, has 60% in PCA and 95% in ABC with ANN. The false discovery rate in fig (9) which is the identification of false results is analyzed as minimum of 46% in PCA and 43% in ABC with ANN. The Matthew's correlation coefficient has been used for the identification of the results, which hold its value from the range of 0 to 1. The value 1 concludes as a correct identification, while analyzing the fig (10) it is concluded that ABC with ANN has better results than PPCA. So the cancer prediction can be done easily using the proposed classifier ABC with ANN.

## 7. CONCLUSION

Most literary works had shown interest on developing better classifiers to improve the classification accuracy of microarray gene expression dataset. A mechanism to predict the cancer patients using artificial neural network is presented. The training and testing of the three datasets under consideration is undergone using classification function (sensitivity and specificity measures) respectively. Results presented in this paper shows that the cancer prediction is easily done using the proposed ABC with ANN mechanism while comparing to the conventional PCA method. Even the accuracy level of the proposed classifier is encouraging ANN needs some enhancement to improve the level. Some adaptive techniques with the ANN will be used for the better improvement in the future works.

## 8.REFERENCES

[1] Brian T. Wilhelm, Josette-Renée Landry, "RNA-Seq-quantitative measurement of expression through massively parallel RNA-sequencing", Methods, Vol. 48, No. 3, pp. 249–257, July 2009.

[2] Gad Getz, Erel Levine and Eytan Domany, "Coupled two-way clustering analysis of gene microarray data", Proceeding of the National Academy of Sciences of the United States of America, Vol. 97, No. 22, 2000.

[3] Adi L. Tarca, Roberto Romero and Sorin Draghici, "Analysis of microarray experiments of gene expression profiling", American Journal of Obstetrics and Gynecology, Vol. 195, No. 2, pp. 373–388, 2006

[4] M. Kathleen Kerr, Mitchell Martin, and Gary A. Churchill, "Analysis of Variance for Gene Expression Microarray Data", Journal of Computational Biology, Vol. 7, No. 6, pp. 819-837, December 2000.

[5] Laura J. vant Veer, Hongyue Dai, Marc J. van de Vijver, Yudong D. He, Augustinus A. M. Hart, Mao Mao, Hans L. Peterse, Karin van der Kooy, Matthew J. Marton, Anke T. Witteveen, George J. Schreiber, Ron M. Kerkhoven, Chris Roberts, Peter S. Linsley, René Bernards and Stephen H. Friend "Gene expression profiling predicts clinical outcome of breast cancer", Nature, Vol. 415, pp. 530-536, 2002.

[6] Javed Khan, Richard Simon, Michael Bittner, Yidong Chen, Stephen B. Leighton, Thomas Pohida, Paul D. Smith, Yuan Jiang, Gerald C. Gooden, Jeffrey M. Trent, and Paul S. Meltzer, "Gene Expression Profiling of Alveolar Rhabdomyosarcoma with cDNA Microarrays", Cancer Res , Vol. 58, pp. 5009-5013, 1998.

[7] Sridhar Ramaswamy and Todd R. Golub, "DNA Microarrays in Clinical Oncology", Journal of Clinical Oncology, Vol. 20, No. 7, pp. 1932-1941, 2002.

[8] Mark Schena, Renu A Heller, Thomas P Theriault, Ken Konrad, Eric Lachenmeierc and Ronald W Davis "Microarrays: biotechnology's discovery platform for functional genomics", Trends in Biotechnology, Vol. 16, No. 7, pp. 301–306, 1998.

[9] Jie Chen, Huai Li, Kaihua Sun, Kim B., "How will bioinformatics impact signal processing research?", IEEE Signal Processing Magazine, Vol. 20, No. 6, pp. 106-206, 2003.

[10] Ahmad M. Sarhan "Cancer Classification Based On Microarray Gene Expression Data Using DCT and ANN", Journal of Theoretical and Applied Information Technology, Vol. 6, No. 2, pp. 208-216, 2009.

[11] Yendrapalli, Basnet, Mukkamala and Sung, "Gene Selection for Tumor Classification Using Microarray Gene Expression Data", In Proceedings of the World Congress on Engineering, London, U.K., Vol. 1, 2007

[12] Sandrine Dudoit, Jane Fridlyand and Terence P. Speed, "Comparison of Discrimination Methods for the Classification of Tumors Using Gene Expression Data", Journal of the American Statistical Association, Vol. 97, pp. 77-87, 2002

[13] Peterson and Ringner, "Analyzing Tumor Gene Expression Profiles", Artificial Intelligence in Medicine, Vol. 28, No. 1, pp. 59-74, 2003.

[14] Almut Schulze and Julian Downward, "Navigating gene expression using microarrays - a technology review", Nature Cell Biology, Vol 3, pp. 190-195, August 2001.

[15] Chintanu K. Sarmah, Sandhya Samarasinghe, Don Kulasiri and Daniel Catchpoole, "A Simple Affymetrix Ratio-transformation Method Yields Comparable Expression Level Quantifications with cDNA Data", World Academy of Science, Engineering and Technology, Vol. 61, pp.78-83, 2010

[16] Khlopova, Glazko and Glazko, "Differentiation of Gene Expression Profiles Data for Liver and Kidney of Pigs", World Academy of Science, Engineering and Technology, Vol. 55, pp. 267-270, 2009

[17] Ahmad m. Sarhan, "Cancer classification based on microarray gene expression data using DCT and ANN", Journal of Theoretical and Applied Information Technology, Vol. 6, No. 2, pp. 207-216, 2009

[18] Ying Xu, Victor Olman and Dong Xu, "Minimum Spanning Trees for Gene Expression Data Clustering", Genome Informatics, Vol. 12, pp. 24–33, 2001

[19] Lucila Ohno-Machado, Staal Vinterbo and Griffin Weber, "Classification of Gene Expression Data Using Fuzzy Logic", Journal of Intelligent & Fuzzy Systems, Vol. 12, No. 1, pp. 19-24, January 2002

[20] Li-Yeh Chuang, Cheng-Hong Yang and Li-Cheng Jin, "Classification Of Multiple Cancer Types Using Fuzzy Support Vector Machines And Outlier Detection Methods", Biomedical Engineering applications, Basis and Communications, Vol. 17, No. 6, pp. 300-308, December 2005.

[21] Jayashree Dev, Sanjit Kumar Dash, Sweta Dash and Madhusmita Swain, "A Classification Technique for Microarray Gene Expression Data using PSO-FLANN", International Journal on Computer Science and Engineering, Vol. 4, No. 09, pp. 1534-1539, Sep 2012.

[22] Alok Sharma, Seiya Imoto, and Satoru Miyano, "A top-r Feature Selection Algorithm for Microarray Gene Expression Data", IEEE/ACM Transactions on Computational Biology And Bioinformatics, Vol. 9, Vol. 3, pp. 754-764, 2012.

[23] A. Castaño F., Fernández-Navarro, C. Hervás-Martínez and P. A. Gutierrez, "Neuro-logistic Models Based on Evolutionary Generalized Radial Basis Function for the Microarray Gene Expression Classification Problem", Neural Process Letter, Vol. 34, pp. 117–13, 2011.

[24] Chhanda Ray, "Cancer Identification and Gene Classification using DNA Microarray Gene Expression Patterns", International Journal of Computer Science Issues, Vol. 8, No. 2, pp. 155-160, 2011.

[25] Venkatesh and Thangaraj, "Investigation of Micro Array Gene Expression Using Linear Vector Quantization for Cancer", International Journal on Computer Science and Engineering, Vol. 02, No. 06, pp. 2114-2116, 2010

[26] CNS tumor, colon tumor, lung tumor and Diffuse Large B-Cell Lymphoma from http://levis.tongji.edu.cn/gzli/data/mirror-kentridge.html

[27] en.wikipedia.org/wiki/Sensitivity_and_specificity