Acute Leukemia Classification using Bayesian Networks

[1] **Abdel Nasser H. Zaied,** [2] **Mona G. Hebishy,** [3] **Mohamed A. Saleh**
[1] Vice Dean for Education and Students Affaires, [2] Teaching Assistant, [3] Dean,
College of Computers and Informatics, Zagazig University, Egypt
[1] nasserhr@zu.edu.eg, [1] nasserhr@gmail.com

ABSTRACT

In this study, two models for constructing acute leukemia classifiers using the Signal-to-Noise Ratio (SNR) gene selection method in conjunction with the Bayesian Networks (BNs) have been proposed. In the first model, genes of the acute leukemia training dataset are ranked using the SNR method and then top ranked genes are selected and used to construct the acute leukemia BN classifier. In the second model, genes of the acute leukemia training dataset are clustered using the*k*-means clustering and then genes of each cluster are ranked using the SNR method after that top ranked genes from gene clusters are selected and used to construct the acute leukemia BN classifier. From the experimental evaluation, the results showed that the classification accuracies achieved by the acute leukemia classifiers constructed according to either of these two models are good compared with the classification accuracies achieved in other studies. The results also indicated that the second model is better than the first model.

Keywords: *Acute leukemia, Bayesian networks, Signal-to-noise ratio, k-means clustering, Gene expression data, Microarray, Cancer classification.*

1. INTRODUCTION

The fact that different therapies should be used with distinct tumor types to maximize efficacy and minimize toxicity has made cancer classification a crucial aspect of cancer diagnosis and treatment. Cancer classification has been based primarily on morphological appearance of tumor [1]. But, this conventional method is unable to discriminate among tumors with similar histopathology features which vary in clinical course and in response to treatment [2]. This shortcoming led to increasing interest in changing the basis of tumor classification from morphologic to molecular.

Golub et al [3],presented a method for performing cancer classification based on gene expression monitoring by Deoxyribonucleic Acid (DNA) microarrays and their findings demonstrated the feasibility of using gene expression data produced by DNA microarrays for classifying tumors. Since this pioneering work, studies that use microarray gene expression data to classify cancer did not stop.

The classification methods used in these studies include Support Vector Machines (SVMs)[4, 5, 6 & 7],Artificial Neural Networks (ANNs) [8 & 9],Decision Trees (DTs)[10, 11 & 12],*K*-Nearest Neighbor (KNN)[5 & 6], Bayesian Networks (BNs)[13 & 14],Fuzzy Inference (FI) [15],Penalized Logistic Regression (PLR) [16]and discriminate analysis methods [10 & 17].But, the distinctive characteristics of the gene expression datasets make their classification difficult. The gene expression dataset is a two dimensional array with a large number of genes (rows) and a small number of samples (columns). Moreover, most genes in these datasets are not related to the performance of the classification and taking such genes into account during classification increases the dimension of the classification problem, poses computational difficulties and introduces unnecessary noise in the process [18]. Applying gene selection method to find the most informative genes that discriminate various classes before classification became increasingly popular to speed up and increase the accuracy of gene expression data classification. So, the process of classifying gene expression data merges two main steps[19 & 20]:*implementing an effective gene selection technique* and *choosing a powerful classifier*.

The main objective of this study is to build classifiers that classify acute leukemia into its two subtypes: Acute Lymphoblastic Leukemia (ALL) and Acute Myeloid Leukemia (AML) based on its gene expression profiling using DNA microarrays. To achieve this goal, two models were proposed by employing the two gene selection approaches that are based on the signal-to-noise ratio gene selection method that were proposed by Mishra &Sahu [6] in conjunction with the Bayesian networks.

The remainder of this paper is structured as follows: In the next section, a quick background about DNA microarrays and acute leukemia was presented. The followed section reviewed the methods used in our two models which are the signal-to-noise ratio gene selection method, the*k*-means clustering and the Bayesian networks. In the followed section, the proposed models were described. The evaluation and the results were later presented and discussion of the findings followed. The conclusions and recommendations finalized the paper.

2. BACKGROUNDS

2.1 DNA Microarrays

DNA microarrays provide a platform where one can measure the expression levels of tens of thousands of genes in a sample. There are three major steps involved in a typical DNA microarray experiment [21]:

a. *Preparation of microarrays*: DNA microarrays are available in two different formats: oligonucleotide arrays and

Complementary Deoxyribonucleic Acid (cDNA) microarrays. Oligonucleotide arrays are generated by synthesizing specific oligonucleotides in a predetermined spatial orientation on a solid surface using a technique called photolithography. cDNA arrays are generated by printing a double stranded cDNA onto a solid support, such as glass or nylon, using robotic pins.

b. *Preparation of fluorescently labeled cDNA probes and hybridization*: Ribonucleic Acid (RNA) from two different sources (e.g. normal tissue and tumor tissue) is prepared. The isolated RNA is converted to cDNA which in turn is labeled with fluorescent dyes. The most frequently used fluorescent dyes are Cyanine 3 (CY3) (green) for control samples and Cyanine 5 (CY5) (red) for test samples. After that, the hybridization takes place and then the array is washed.

c. *Slide scanning, image and data analysis*: Slides are scanned using a confocal laser scanner capable of interrogating both the CY3- and CY5- labeled probes to produce separate Tagged Image File Format (TIFF) images for each label. These images are subsequently analyzed to calculate the relative levels of expression of each gene. The raw data obtained after image analysis have to be further analyzed before one can identify the list of differentially regulated genes.

Since their invention, DNA microarrays have become commonplace in biomedical research especially in cancer biology[22].

2.2 Acute Leukemia

Leukemia is the cancer of the blood; Leukemia starts in the bone marrow. In most cases, the leukemia invades the blood fairly quickly. From there, it can go to other parts of the body such as the lymph nodes, spleen, liver, central nervous system (the brain and spinal cord), testicles or other organs [23]. Leukemia is the most common cancer in children and adolescents [24]. In 2000, approximately 256,000 children and adults around the world had a form of leukemia, and 209,000 died from it. This represents about 3% of the almost seven million deaths due to cancer that year and about 0.35% of all deaths from any cause [25]. At the local level, Leukemia accounts for about 33% of pediatric malignancies [26]. Acute leukemia is the most dangerous type of leukemia because it worsens quickly, and if not treated; would probably be fatal in a few months. If acute leukemia occurs in the lymphoid cells then it is called acute lymphoblastic leukemia. On the other hand, if it occurs in the myeloid cells then it is called acute myeloid leukemia.

3. METHODS

3.1 Signal-to-Noise Ratio (SNR) Gene Selection Method

In the SNR method, we start with a dataset S consisting of m expression vectors,$X^i = (x_1^i, \dots\,.., x_n^i), 1 \leq i \leq m,$ where m is the number of patient samples and n is the number of genes measured. Each patient sample is labeled with $Y \in \{+1, -1\}$(e.g. ALL versus AML). For each genex_j, we calculate the mean μ_j^+ (resp.μ_j^-) and standard deviation σ_j^+ (resp.σ_j^-) using only the patient samples labeled +1 (resp. -1). We want to find genes that will help discriminate between the two classes; therefore, we calculate the following score [27]:

$$F(x_j) = \left|\frac{\mu_j^+ - \mu_j^-}{\sigma_j^+ + \sigma_j^-}\right| \qquad (1)$$

We then simply take the genes with the highest $F(x_j)$score as our top genes.

Mishra &Sahu[6], proposed the following two gene selection approaches that are based on the signal-to-noise ratio method:

- *In the first approach*, the genes of microarray training dataset are clustered by the *k*-means clustering and then the SNR ranking is implemented to get top ranked genes from each cluster which are used to construct the classifier.
- *In the second approach*, the genes of microarray training dataset are ranked by implementing only the SNR ranking and then the top scored genes are selected and used to construct the classifier.

These two gene selection approaches succeeded to achieve high classification accuracies when they were used to construct support vector machine and *K*-nearest neighbor classifiers from the acute leukemia gene expression dataset, so, we chose to use them in our two models.

3.2 *K*-Means Clustering

The *k*-means algorithm gains its name from its method of operation; it clusters n data points into Kclusters, where K is provided as an input parameter. Let $X = \{x_i\}, i = 1, \dots\,., n$ be the set of $n$data points to be clustered into a set of K clusters, $C = \{c_k\}, k = 1, \dots, K$. K-means algorithm finds a partition such that the squared error between the empirical mean of a cluster and the points in the cluster is minimized. Let μ_k be the mean of clusterc_k. The squared error between μ_k and the points in cluster c_k is defined as:

$$J(c_k) = \sum_{x_i \in c_k} \|x_i - \mu_k\|^2 \qquad (2)$$

The goal of k-means is to minimize the sum of the squared error over all the K clusters:

$$J(C) = \sum_{k=1}^{K} \sum_{x_1 \in c_k} \| x_i - \mu_k \|^2 \qquad (3)$$

K-means starts with an initial partition with K clusters and assigns patterns to clusters so as to reduce the squared error.

3.3 Bayesian Networks

Although BNs are not widely used in cancer gene expression data classification studies, they have several characteristics that make them proper for this purpose [13 & 14].

BNs belong to the family of probabilistic graphical models which are used to represent knowledge about an uncertain domain. A BN represents the joint probability distribution for a set of random variables efficiently based on the concept of conditional independence. A BN assumes a form of directed acyclic graph (DAG). Figure (1) shows an example BN structure. Each node in the graph corresponds to a random variable and each edge represents the probabilistic dependency between variables. For example in Figure (1), an edge from node *A* to node *C* indicates that a value taken by variable *C* depends on the value taken by variable *A*. *A*is then referred to as a parent of *C* and, similarly, *C* is referred to as the child of *A*.

Descendants of a given node are the set of nodes that can be reached on a direct path from that node and ancestor nodes are the set of nodes from which the node can be reached on a direct path. The structure of the acyclic graph guarantees that there is no node that can be its own ancestor or its own descendent.

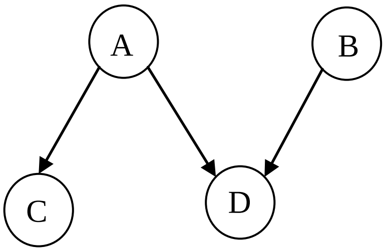

Fig 1: The directed acyclic graph (DAG) structure of an example BN

The BN reflects the following simple conditional independence statement, "each variable is independent of its non-descendents in the graph given the state of its parents". So, a BN which consists of n nodes (variables),$X = \{X_1, \ldots, X_n\}$, represents the joint probability distribution, as follows:

$$P(X) = \prod_{i=1}^{n} P(X_i / Pa_i) \qquad (4)$$

Where Pa_i is the set of parents of X_i in the network structure and $P(X_i \mid Pa_i)$ is the local probability distribution related to the nodeX_i which depends only on its parents. For discrete random variables, this local probability distribution is often represented by a table, listing the local probability that a child node takes on each of the feasible values -for each combination of values of its parents. The global structure of a BN encodes the conditional independence relationships among all variables and is called the qualitative part of the BN. Local probability distributions for all nodes constitute the quantitative part of the BN.

Because a BN encodes the joint probability distribution for a set of variables, the conditional probability of any interesting variable given observations of some of the other variables can be inferred efficiently. Therefore, once the BN whose nodes represent gene expression levels and the cancer class label is constructed from the gene expression data, the probability of the cancer class label given some gene expression levels for a new sample can be inferred. This is the BN classifier [13].

4. PROPOSED MODELS

In this study, two models for constructing acute leukemia classifiers were proposed. In the first model, genes of the acute leukemia training dataset are ranked using the SNR method after that the top ranked genes are selected and used to construct the acute leukemia BN classifier which in turn is evaluated using the acute leukemia test dataset, as shown in Figure (2).

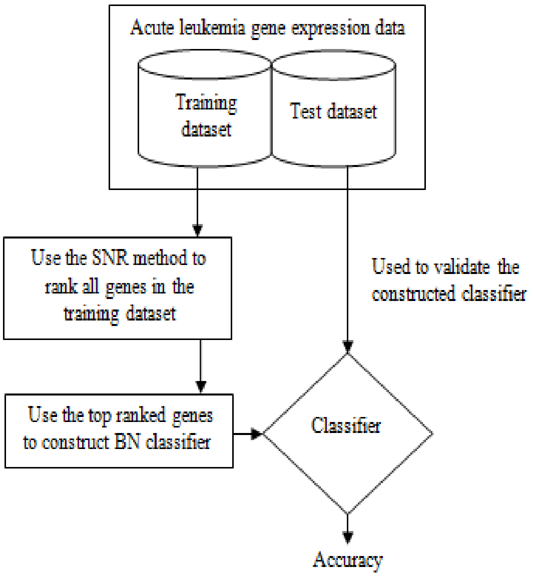
Acute leukemia gene expression data
Training dataset
Test dataset
Use the SNR method to rank all genes in the training dataset
Used to validate the constructed classifier
Use the top ranked genes to construct BN classifier
Classifier
Accuracy

Fig 2: The first model proposed for constructing acute leukemia classifiers

In the second model, genes of the acute leukemia training dataset are clustered using the *k*-means clustering then genes of each cluster are ranked using the SNR method after that the top ranked genes from different clusters are selected and used to construct the acute leukemia BN classifier which in turn is evaluated using the acute leukemia test dataset, as shown in Figure (3).

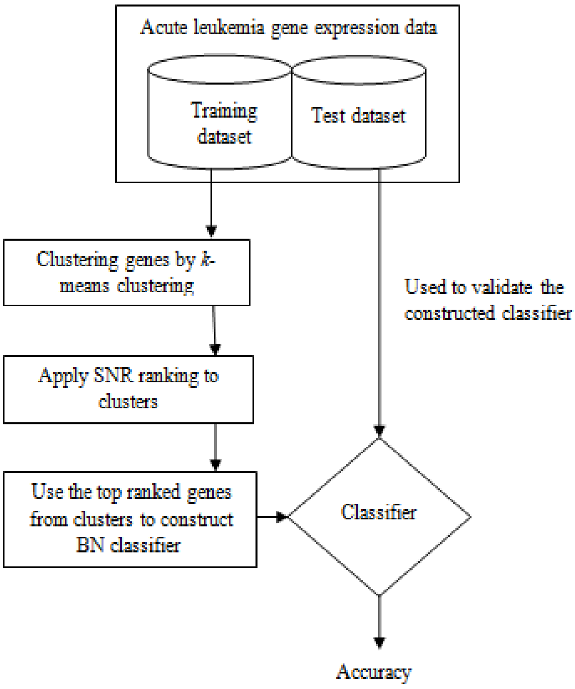
Acute leukemia gene expression data
Training dataset
Test dataset
Clustering genes by k-means clustering
Used to validate the constructed classifier
Apply SNR ranking to clusters
Use the top ranked genes from clusters to construct BN classifier
Classifier
Accuracy

Fig 3: The second proposed model for constructing acute leukemia classifiers

5. EXPERIMENTAL EVALUATION AND RESULTS

To evaluate the proposed models, seven acute leukemia classifiers have been constructed according to each model methodology using the top 5, 10, 20, 30, 50, 70 and 90 ranked genes and the classification accuracy of these classifiers was calculated as the percentage of test instances that were correctly classified, as follows:

$$\boldsymbol{Classificationaccuracy} = \frac{\|\boldsymbol{correct}\|}{\|\boldsymbol{test}\|}\% \qquad (5)$$

Where $\|correct\|$ and $\|test\|$ denote the number of test instances whose labels are correctly predicted and the total number of test instances, respectively.

During this experimental work, we have used the acute leukemia benchmark dataset downloaded from the Broad Institute’s website [28]. This dataset includes expression profiles of 7,129 human DNA probes (genes) spotted on Affymetrix Hu6800 microarrays of 72 patients with either acute myeloid leukemia or acute lymphoblastic leukemia. Tissue samples were collected at time of diagnosis before treatment, taken either from bone marrow (62 cases), or peripheral blood (10 cases) and reflect both childhood and adult leukemia. The gene expression profiles of the original dataset are represented as $\log_{10}$ normalized expression values. The dataset is divided into a training set containing 38 samples and a test (validation) set containing 34 samples.

Table (1) indicates the classification accuracies achieved by different acute leukemia BN classifiers constructed according to the first model methodology.

Table 1: Classification accuracies of acute leukemia BN classifiers constructed according to the first model

Acute leukemia BN classifier	No of correctly classified samples	Classification accuracy %
5-gene	29 out of 34	85.29
10-gene	30 out of 34	88.24
20-gene	28 out of 34	82.35
30-gene	31 out of 34	91.18
50-gene	29 out of 34	85.29
70-gene	30 out of 34	88.24
90-gene	32 out of 34	94.12

The 90-gene acute leukemia BN classifier achieved the highest classification accuracy (94.12%) followed by the 30-gene acute leukemia BN classifier (91.18%). Both the 10-gene and 70-gene acute leukemia BN classifiers came in the third position (88.24%) followed by both the 5-gene and 50-gene acute leukemia BN classifiers (85.29%) and the 20-gene acute leukemia BN classifier came in the last position (82.35%), as shown in Figure (4).

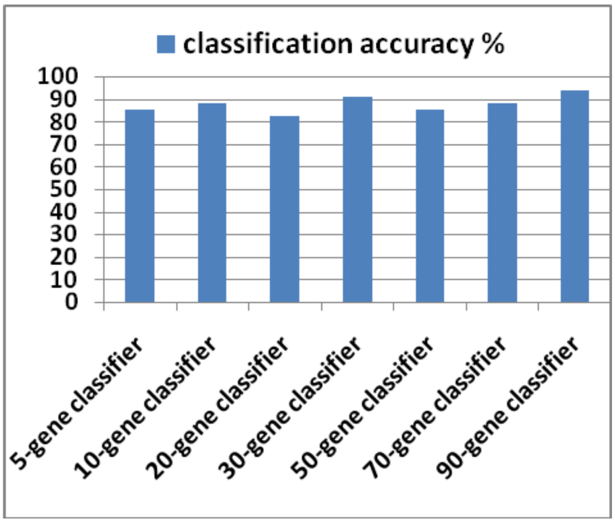

Fig 4: Classification accuracies of acute leukemia BN classifiers constructed according to the first model

When these acute leukemia BN classifiers were constructed according to the second model methodology, they achieved different classification accuracies, as shown in Table (2).

Table 2: Classification accuracies of acute leukemia BN classifiers constructed according to the second model

Acute leukemia BN classifier	No of correctly classified samples	Classification accuracy %
5-gene	30 out of 34	88.24
10-gene	30 out of 34	88.24
20-gene	30 out of 34	88.24
30-gene	32 out of 34	94.12
50-gene	31 out of 34	91.18
70-gene	32 out of 34	94.12
90-gene	32 out of 34	94.12

The 30-gene, 70-gene and 90-gene acute leukemia BN classifiers achieved the highest classification accuracy (94.12%) followed by the 50-gene acute leukemia BN classifier (91.18) and the 5-gene, 10-gene and 20-gene acute leukemia BN classifiers came in the last position (88.24%), as shown in Figure (5).

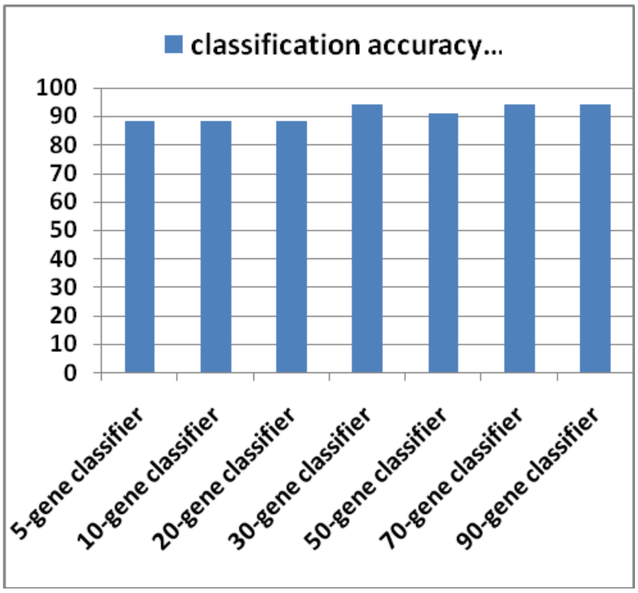

Fig 5: Classification accuracies of acute leukemia BN classifiers constructed according to the second model

6. DISCUSSION

6.1 General Discussion

The results, indicated in the previous section, show that increasing the number of top ranked genes used to construct the classifier does not always mean increasing the classification accuracy achieved by the classifier. These results are matched with Mishra & Sahu [6] and Wang & Makedon, [5] results.

When comparing the results achieved by applying the first model with the results achieved by applying the second model, the second model gives better results than the first model. In the second model, applying the *k*-means clustering before using the SNR method assures that the selected gene subsets have no redundancy and thus enhances the classification accuracy of most classifiers, as shown in Table (3) and Figure (6).

- The classification accuracy of the 5-gene, 20-gene, 30-gene, 50-gene and 70-geneacute leukemia BN classifiers were improved from 85.29% to 88.24%, from 82.35% to 88.24%, from 91.18% to 94.12%, from 85.29% to 91.18% and from 88.24% to 94.12% respectively.
- The classification accuracy of the 10-gene and 90-gene acute leukemia BN classifiers didn’t change due to using the same methodology in selecting the set of top ranked genes in the two models, in other words, applying the *k*-means clustering did not affect the set of top ranked genes.

6.2 Comparison with Previous Works

By comparing proposed models' results with methods present in literature, the following points are highlighted:

- Golub et al [3], constructed acute leukemia classifier using a weighted voting scheme. They constructed their classifier using the top 50 ranked genes selected according to the SNR method. When, they applied their classifier to the test dataset, 29 samples out of 34 were correctly classified which corresponds to 85.29% classification accuracy.
  In this study, the 5-gene and 50-gene acute leukemia BN classifiers constructed according to the first model achieved the same classification accuracy achieved by Golub et al [3]. Moreover, the 10-gene, 30-gene, 70-gene and 90-gene acute leukemia BN classifiers constructed according to the first model; and the 5-gene, 10-gene, 20-gene, 30-gene, 70-gene and 90-gene acute leukemia BN classifiers constructed according to the second model achieved higher classification accuracies than Golub et al classifier.

- Furey et al [27], used the top 25, 250, 500 and 1000 ranked genes selected according to the SNR method to construct acute leukemia SVM classifiers, their classifiers succeeded to classify 30 to 32 samples out of 34 correctly which corresponds to 88.24% to 94.12% classification accuracy.
  In this study, the 10-gene and 70-gene acute leukemia BN classifiers constructed according to the first model; and the 5-gene, 10-gene and 20-gene acute leukemia BN classifiers constructed according to the second model achieved 88.24% classification accuracy. The 30-gene acute leukemia BN classifier constructed according to the first model and the 50-gene acute leukemia BN

classifier constructed according to the second model achieved 91.18% classification accuracy. The 90-gene acute leukemia BN classifier constructed according to the first model; and the 30-gene, 70-gene and 90-gene acute leukemia BN classifiers constructed according to the second model achieved 94.12% classification accuracy.

- Hwang et al [13], constructed BN classifier for acute leukemia. Their BN classifier achieved 94.12% classification accuracy.
  In this study, the 90-gene acute leukemia BN classifier constructed according to the first model; and the 30-gene, 70-gene and 90-gene acute leukemia BN classifiers constructed according to the second model achieved the same classification accuracy achieved by Hwang et al classifier.

- Seeja&Shewta[7], used SVM trained using different kernels; they used the top 200 ranked genes selected according to the SNR method to construct their classifiers. When they applied their classifiers to the test dataset, classifier trained with the Radial Basis Function (RBF) kernel achieved 94.12% classification accuracy and the classifiers trained with the poly kernel and normalized poly kernel achieved 91.18% classification accuracy.

In this study, the 30-gene acute leukemia BN classifier constructed according to the first model and the 50-gene acute leukemia BN classifier constructed according to the second model achieved the same classification accuracy achieved by Seeja&Shewta classifiers trained with the poly kernel and normalized poly kernel. The 90-gene acute leukemia BN classifier constructed according to the first model; and the 30-gene, 70-gene and 90-gene acute leukemia BN classifiers constructed according to the second model achieved the same classification accuracy achieved by Seeja&Shewta classifier trained with the RBF kernel.

- Yap et al [14], introduced a Bayesian Network-based Noise Correction framework named BN-NC. They used the entropy-based discretization and gene selection technique. Their classifier surpassed our best classifiers (97.05% classification accuracy).

Table 3: Accuracy of classifiers constructed according to the first model against accuracy of classifiers constructed according to the second model

Acute leukemia BN classifier	First Model		Second Model		Improvement %
	No of correctly classified samples	Classification accuracy %	No of correctly classified samples	Classification accuracy %	
5-gene	29 out of 34	85.29	30 out of 34	88.24	03.45
10-gene	30 out of 34	88.24	30 out of 34	88.24	00.00
20-gene	28 out of 34	82.35	30 out of 34	88.24	07.15
30-gene	31 out of 34	91.18	32 out of 34	94.12	03.22
50-gene	29 out of 34	85.29	31 out of 34	91.18	06.91
70-gene	30 out of 34	88.24	32 out of 34	94.12	06.66
90-gene	32 out of 34	94.12	32 out of 34	94.12	00.00

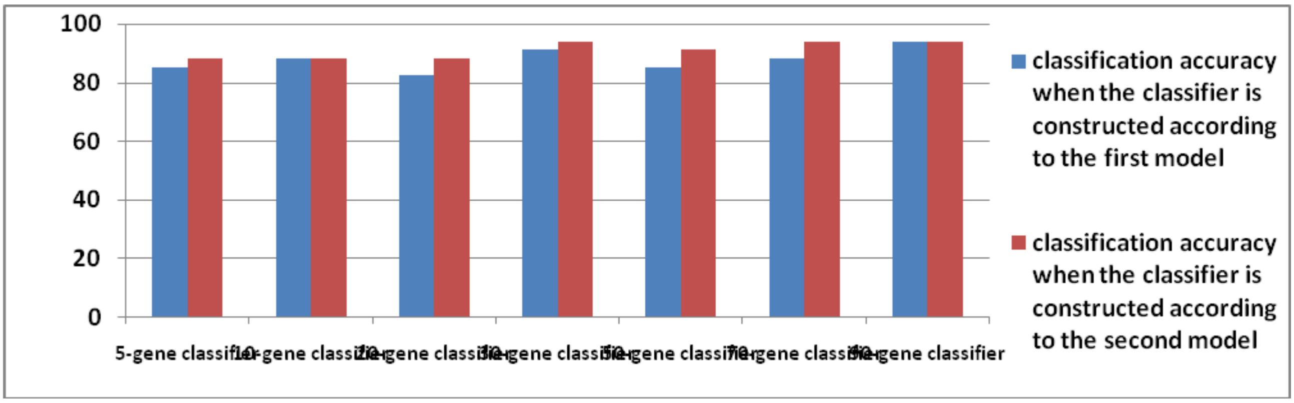

Fig 6: Accuracy of classifiers constructed according to the first model against accuracy of classifiers constructed according to the second model

7. CONCLUSION

In this study, two models for constructing acute leukemia classifiers using the signal-to-noise ratio gene selection method and Bayesian networks have been used. In the first model, genes of the acute leukemia training dataset are ranked using the SNR method and then top ranked genes are selected and used to construct the acute leukemia BN classifier. In the second model, genes of the acute leukemia training dataset are clustered by *k*-means clustering and then genes of each cluster are ranked using the SNR method after that the top ranked genes from gene clusters are selected and used to construct the acute leukemia BN classifier.The results indicated that the classification accuracies achieved by acute leukemia classifiers constructed according to the two models are comparable with the classification accuracies achieved in other studies. The results also indicated that the second model is better than the first model.

REFERENCES

[1] Zhang X., Ke H., ALL/AML Cancer Classification by Gene Expression Data Using SVM and CSVM Approach, Genome Informatics 11, 2000, pp. 237–239.

[2] Zhang H., Yu C., Singer B., Xiong M., Recursive Partitioning for Tumor Classification with Gene Expression Microarray Data, PNAS 98(12), 2001, pp. 6730-6735.

[3] Golub T., Slonim D., Tamayo P., Huard C., Gaasenbeek M., Mesirov J., Coller H., Loh M., Downing J., Caligiuri M., Bloomfield C., Lander E., Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring, Science 286, 1999, pp. 531-537.

[4] Jaeger J., Sengupta R., Ruzzo W., Improved Gene Selection for Classification of Microarrays, Pacific Symposium on Biocomputing, 2003, pp. 53–64.

[5] Wang Y., Makedon F., Application of Relief-F Feature Filtering Algorithm to Selecting Informative Genes for Cancer Classification using Microarray Data, Proceedings of the IEEE Conference on Computational Systems Bioinformatics (CSB’04), IEEE Press, Stanford, CA, USA, 2004, pp. 497–498.

[6] Mishra D., Sahu B., Feature Selection for Cancer Classification: A Signal-to-noise Ratio Approach, International Journal of Scientific & Engineering Research 2(4), 2011, pp. 1-7.

[7] Seeja K., Shweta, Microarray Data Classification Using Support Vector Machine, International Journal of Biometrics and Bioinformatics (IJBB) 5(1), 2011, pp. 10-15.

[8] Plagianakos V., Tasoulis D., Vrahatis M., Computational Intelligence Techniques for Acute Leukemia Gene Expression Data Classification, proceedings of 2005 IEEE international joint conference on neural networks 4, 2005, pp. 2469 – 2474.

[9] Kanth M., Kulkarni U., Giridhar B., Gene Expression Based Acute Leukemia Cancer Classification: a Neuro-Fuzzy Approach, International Journal of Biometrics and Bioinformatics (IJBB) 4(4), 2010, pp. 136-146.

[10] Dudoit S., Fridlyand J., Speed T., Comparison of Discrimination Methods for the Classification of Tumors using Gene Expression Data, Journal of the American Statistical Association 97(457), 2002, pp. 77-87.

[11] Dettling M., Buhlmann P., Boosting for Tumor Classification with Gene Expression Data, Bioinformatics 19(9), 2003, pp. 1061-1069.

[12] Netto O., Nozawa S., Mitrowsky R., Macedo A., Baranauskas A., Applying Decision Trees to Gene Expression Data from DNA Microarrays: A Leukemia Case Study, In: XXX Congress of the Brazilian Computer Society, X Workshop on Medical Informatics. P. 10. Belo Horizonte, MG, 2010, pp. 1489-1498.

[13] Hwang K., Cho D., Park S., Kim S., Zhang B., Applying Machine Learning Techniques to Analysis of Gene Expression Data: Cancer Diagnosis, in Proceedings of the first Conference on Critical Assessment of Microarray Data Analysis (CAMDA2000), 2000.

[14] Yap G., Tan A., Pang H., and Learning Feature Dependencies for Noise Correction in Biomedical Prediction, in proceedings of the Eleventh SIAM International Conference on Data Mining, 2011, pp. 71-82.

[15] Madeswaran T., Nawaz G., Classification of Micro Array Gene Expression Data using Statistical Analysis Approach with Personalized Fuzzy Inference System, International Journal of Computer Applications, 31(1), 2011, pp. 5-12.

[16] Mahmoodian H., Marhaban M., Rahim R., Rosli R., Saripan M.,A Combinatory Algorithm of Univariate and Multivariate Gene Selection, Journal of Theoretical and Applied Information Technology 5(2), 2009, pp. 113-118.

[17] Cho J., Lee D., Park J., Lee I., New Gene Selection Method for Classification of Cancer Subtypes Considering within-Class Variation, FEBS letters 551, 2003, pp. 3-7.

[18] Wang Z., Palade V., Xu Y., Neuro-Fuzzy Ensemble Approach for Microarray Cancer Gene Expression

Data Analysis, Proceedings of the Second International Symposium on Evolving Fuzzy System (EFS'06), IEEE Computational Intelligence Society, 2006, pp. 241-246.

[19] Kanth M., Kulkarni U., Giridhar B., Acute Leukemia Cancer Classification using Single Genes, Journal of Computing 3(2),2011, pp. 112-116.

[20] Salem D., AbulSeoud R., Ali H., DMCA: A Combined Data Mining Technique for Improving the Microarray Data Classification Accuracy, International Proceedings of Chemical, Biological and Environmental Engineering (IPCBEE) 21, 2011, pp. 36-41.

[21] Somasundaram K., Mungamuri S., Wajapeyee N., DNA Microarray Technology and its Applications in Cancer Biology, Applied Genomics and Proteomics 1(4), 2002.

[22] Ma S., Huang J., Regularized ROC Method for Disease Classification and Biomarker Selection with Microarray Data, Bioinformatics 21(24), 2005, pp. 4356-4362.

[23] Sewak M., Reddy N., Duan Z., Gene Expression Based Leukemia Sub-Classificationusing Committee Neural Networks, Bioinformatics and Biology Insights 3, 2009, pp. 89-98.

[24] American cancer society, Childhood leukemia, 2010.

[25] Mathers C., Murray C., Lopez A., Boschi-Pinto C., Cancer Incidence, Mortality and Survival by Site for 14 Regions of the World, Geneva, World Health Organization (GPE Discussion Paper No. 13), 2001.

[26] Labib N., Malek M., Data Mining for Cancer Management in Egypt Case Study: Childhood Acute Lymphoblastic Leukemia, World Academy of Science, Engineering and Technology 8, 2005, pp. 309-318.

[27] Furey T., Cristianini N., Duffy N., Bednarski D., Schummer M., Haussler D., Support Vector Machine Classification and Validation of Cancer Tissue Samples using Microarray Expression Data, Bioinformatics 16(10), 2000, pp. 906-914. http://www.broad.mit.edu/cgi-bin/cancer/publications/pub_paper.cgi?mode=view&paper_id=43