Proceedings of International Conference on Public Policy and Social Science, UiTM Melaka Malaysia, November 2012
ISBN 978-967-11354-5-7

COMPARATIVE STUDY OF FEATURE SELECTION METHOD OF MICROARRAY DATA FOR GENE CLASSIFICATION

Nurulhuda Ghazali[1], Noraini Hasan[2], Norliana Mohd Lip[3], Nur Hafizah Ghazali[4], Mohammad Faridun Naim Tajuddin[5], Puteh Saad[6]

[1,2,3]Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Melaka
nurulhudaghazali@melaka.uitm.edu.my, *norainihasan@melaka.uitm.edu.my*, *norliana7287@melaka.uitm.edu.my*

[4]School of Computer and Communication Engineering, Universiti Malaysia Perlis
nurhafizahghazali@yahoo.com

[5]School of Electrical Systems Engineering, Universiti Malaysia Perlis
faridun@unimap.edu.my

[6]Faculty of Computer Science and Information System, Universiti Teknologi Malaysia
puteh@utm.my

ABSTRACT

Recent advances in biotechnology such as microarray, offer the ability to measure the levels of expression of thousands of genes in parallel. Analysis of microarray data can provide understanding and insight into gene function and regulatory mechanisms. This analysis is crucial to identify and classify cancer diseases. Recent technology in cancer classification is based on gene expression profile rather than on morphological appearance of the tumor. However, this task is made more difficult due to the noisy nature of microarray data and the overwhelming number of genes. Thus, it is an important issue to select a small subset of genes to represent thousands of genes in microarray data which is referred as informative genes. These informative genes will then be classified according to its appropriate classes. To achieve the best solution to the classification issue, we proposed an approach of minimum Redundancy-Maximum Relevance feature selection method together with Probabilistic Neural Network classifier. The minimum Redundancy-Maximum Relevance feature selection method is used to select the informative genes while the Probabilistic Neural Network classifier acts as the classifier. This approach has been tested on a well-known cancer dataset which is Leukemia. The results achieved shows that the gene selected had given high classification accuracy. This reduction of genes helps take out some burdens from biologist and better classification accuracy can be used widely to detect cancer in early stage.

Keywords: Biotechnology, Classification, Feature Selection, Gene Expression Microarray

INTRODUCTION

Every living organism has discrete hereditary units known as genes. Each gene provides some function or mechanism either by itself or it will combine with other genes that will eventually producing some property of its organism. Genes in DNA is expressed by transferring its coded information into proteins that dwell in the cytoplasm. This process is called as gene expression [1]. There are several experimental techniques to measure gene expression such as expression vector, reporter gene, northern blot, fluorescent hybridization, and DNA microarray. DNA microarray technology allows the simultaneous measurement of the expression level of a great number of genes in tissue samples [2]. It yields a set of floating point and absolute values. Many explored on classification methods to recognize cancerous and normal tissues by analyzing microarray data. Traditionally cancer is diagnosed based on the microscopic examination of patients' tissue. This kind of diagnosis may fail when dealing with unusual or atypical tumors and it takes a long time and might however limit the finding of tumor cell especially in early tumor detection [3]. If tumor cell is found in its critical stage, then it might be too late to cure the patient.

The microarray technology typically produces large datasets with expression values for thousands of genes (2000-20000) in a cell mixture, but only few samples are available (20-80) [4]. This study is focused on gene selection and classification of DNA microarray data in order to identify tumor samples from normal samples. Gene selection is a process where a set of informative genes is selected from the gene expression data in a form of microarray dataset. This process helps improve the performance of the classifier. On the other hand, classification is a process to classify microarray data in several classes that have its own characteristics. There are several techniques that have been used in gene selection such as ReliefF Algorithm, Information Gain, minimum Redudancy Maximum Relevance (mRMR) and Chi Square. For classification of microarray data, a few techniques have been applied in the bioinformatics field to classify the highly dimensional data. These techniques include Random Forest, Naïve Bayes and Probabilistic Neural Network (PNN).

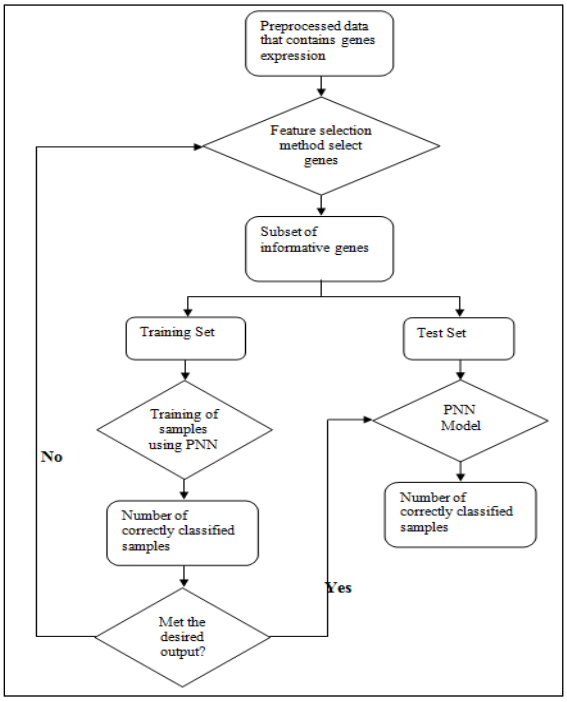
MRMR is chosen as the primary technique for gene selection since this technique is proposed originally for gene selection [5]. The advantage of this technique is it focuses on redundancy of genes together with the relevance of genes. Unlike other techniques; ReliefF [6], Information Gain [7] and Chi Square [8], they were firstly proposed only for general feature selection, rather than genes. For comparison, these four techniques are used to select genes in order to measure the performance. As for classification, the technique chosen in this research is Probabilistic Neural Network (PNN) classifier. PNN has been use in many studies of feature classification [9, 10]. These studies have proved that PNN yield better result in classification accuracy compared to other existing classifiers. Thus, this research combines a few feature selection methods together with PNN classifier to classify microarray data according to its classes.

The objective of this research is to compare the effectiveness of the selection of a set of meaningful genes using minimum Redundancy-Maximum Relevance (mRMR), Information Gain, ReliefF and Chi Square and classify them using Probabilistic Neural Network (PNN) classifier.This paper will brief on research method, followed by the implementation of the method, as well as the experimental settings. Lastly, the results will be discussed and concluded.

RESEARCH

The proposed method involved two stages where the first stage is the gene selection stage and the second one would be the classification stage. In gene selection method, the technique chosen is a technique called minimum Redundancy-Maximum Relevance (mRMR) feature selection and will be compared to three other methods namely ReliefF, Information Gain and Chi Square. mRMR is a feature selection framework that was introduced by Ding and Peng in 2005. They supplement the maximum relevance criteria along with minimum redundancy criteria to choose additional features that are maximally dissimilar to already identified ones. This can expand the representative power of the feature subset and help improves their generalization properties. The classification problem will be handled by Probabilistic Neural Network (PNN) technique. PNN has been widely used in solving classification problems. This is because it can categorize data accurately [11]. Both techniques will be assessed on a bench mark cancer dataset which is Leukemia [12].

FIGURE 1 OVERALL PROCESS OF FEATURE SELECTION AND CLASSIFICATION



IMPLEMENTATION

3.1 Preprocessed data

Here we chose to use the dataset leukemia which consists of 2 classes, AML and ALL. The data will be tested, evaluated and classified into ALL class and AML class. Before continue with experiment, firstly the data needs to be preprocessed in order to fit the feature selection methods and classifier. The original data is in form of ARFF format. Meanwhile, as for mRMR, the data need to be converted to CSV file format as shown below in Figure 2. Figure below is example of leukemia dataset in CSV format, viewed in Microsoft Excel. The first column indicates the class of leukemia (ALL/AML) while the rest of the table stores the values of genes.

FIGURE 2 CONTINUOUS DATA IN CSV FORMAT

class	att1	att2	att3	att4	att5	att6	att7	att8
1	-214	-153	-58	88	-295	-558	199	-176
1	-139	-73	-1	283	-264	-400	-330	-168
1	-76	-49	-307	309	-376	-650	33	-367
1	-135	-114	265	12	-419	-585	158	-253
1	-106	-125	-76	168	-230	-284	4	-122
1	-138	-85	215	71	-272	-558	67	-186
1	-72	-144	238	55	-399	-551	131	-179
1	-413	-260	7	-2	-541	-790	-275	-463
1	5	-127	106	268	-210	-535	0	-174
1	-88	-105	42	219	-178	-246	328	-148
1	-165	-155	-71	82	-163	-430	100	-109
1	-67	-93	84	25	-179	-323	-135	-127
1	-92	-119	-31	173	-233	-227	-49	-62
1	-113	-147	-118	243	-127	-398	-249	-228
1	-107	-72	-126	149	-205	-284	-166	-185
1	-117	-219	-50	257	-218	-402	228	-147
1	-476	-213	-18	301	-403	-394	-42	-144
1	-81	-150	-119	78	-152	-340	-36	-141
1	-44	-51	100	207	-146	-221	83	-198

3.2 Feature Selection Methods

As been brief earlier, this study involves the comparison of 4 feature selection methods, namely mRMR, ReliefF, Information Gain and Chi Square. The comparisons of these methods in terms of its performance are being measured using PNN classifier.

3.2.1 mRMR Feature Selection Method

mRMR was introduced by Ding and Peng in year 2003. mRMR stands for minimum Redundancy-Maximum Relevance feature selection. The purpose of this method is to select a feature subset that best characterizes the statistical property of a target classification variable. These features has to be mutually as dissimilar to each other as possible, but marginally as similar to the classification variable as possible. The owner of this method believes that combining a "very effective" gene with another "very effective" gene often does not form a better feature set. One of the reasons is
that the two genes could be highly correlated and leads to redundancy of feature set. In brief, the mRMR minimizes redundancy and used a series of intuitive measures of relevance and redundancy to select useful features for both continuous and discrete datasets. If a gene has expressions randomly or uniformly distributed in different classes, its mutual information with these classes is zero whereas if a gene is strongly differentially expressed for different classes, it should have large mutual information. Thus the mutual information is used as a measure of relevance of genes. The mutual information I of two variables x and y is defined based on their joint probabilistic distribution $p(x,y)$ and the respective marginal probabilities $p(x)$ and $p(y)$:

$$I(x,y) = \sum_{i,j} p(x_i, y_j) \log \frac{p(x_i, y_j)}{p(x_i)p(y_j)} \tag{1}$$

The measurements of the level of similarity between genes are based on their mutual information. The main idea of minimum redundancy is to select the genes such that they are mutually maximally dissimilar. Let S denote the subset of features that is the most relevant for classification. The minimum redundancy condition is

$$minW_1, \qquad W_1 = \frac{1}{|S|^2} \sum_{i,j \in S} I(i,j), \tag{2}$$

where the $I(i,j)$ is used to represent $I(g_i,g_j)$ for notational simplicity and $|S|$ is the number of features in S.

To measure the level of discriminant powers of genes when they are differentially expressed for different targeted classes, again mutual information $I(h,g_i)$ is used between targeted classes $h=\{h_1,h_2,\ldots,h_K\}$ and the gene expression g_i. Thus $I(h,g_i)$ quantifies the relevance of g_i for the classification task. Thus the maximum relevance condition is to maximize the total relevance of all genes in S:

$$maxV_1, \qquad V_1 = \frac{1}{|S|} \sum_{i \in S} I(h,i), \tag{3}$$

where $I(h,g_i)$ is referred as $I(h,i)$.

The minimum redundancy – maximum relevance feature set is obtained by optimizing the conditions in equations (2) and (3) simultaneously. Optimization of these two conditions requires combining them into a single criterion function. In this paper the two conditions are equally important, and consider two simplest combined criteria:

$$\max(V_1 - W_1), \tag{4}$$

$$\max(V_1/W_1),$$

(5)

An exact solution to the mRMR requirements requires O($N^{|S|}$) search to obtain (*N* is the number of genes in the whole gene set, Ω). To achieve optimal solution, a simple heuristic algorithm is used. The first feature is selected according to Eq. (3), i.e. the feature with the highest *I*(*h*,*i*). The rest features are selected in an incremental way: earlier selected features remain in the feature set. Suppose a set of *m* features is already selected for the set *S*, and additional features are selected from the set $\Omega_S = \Omega - S$ (i.e. all genes except those already selected). The following two conditions are optimized:

$$max_{i\in\Omega_S} I(h,i),$$

(6)

$$min_{i\in\Omega_S} \frac{1}{|S|} \sum_{j\in S} I(i,j).$$

(7)

The condition in Eq. (6) is equivalent to the condition in Eq. (3), while Eq. (7) is an approximation of the condition of Eq. (2). The two combinations of Eqs. (4) and (5) for relevance and redundancy lead to the selection criteria of a new feature:

(1) MID: Mutual Information Difference criterion,
(2) MIQ: Mutual Information Quotient criterion,

These optimizations can be computed efficiently in O(|*S*|•*N*) complexity.

3.2.2 ReliefF Algorithm

The main idea of the ReliefF algorithm that was proposed by Kononenko in 1994 is to estimate the quality of attributes that have weights greater than the threshold using the distinction of an attribute value between a given instance and the two nearest instances namely Hit and Miss. The algorithm of ReliefF is shown in the following figure.

FIGURE 3 RELIEFF ALGORITHM [13]

Input : *a vector space for training instances with the value of attributes and class values*
Output : *a vector space for training instances with the weight W of each attribute*
Set all weights W[A]=0.0
for *i=1* to *m* do begin
    randomly select instance R_i
    find k nearest hit H_j
for *each class* $C \neq class(R_i)$ do
    find k nearest miss Mj(C) from class C
for *A=1* to *all attribute* do

$$W(A) = W(A) - \sum_{j=1}^{k} \frac{diff(A,R_j,H_j)}{m,k} + \sum_{C\neq class(R_j)}^{k} [\frac{P(C)}{1} - P(class(R_j)) \sum_{j=1}^{k} diff(A,R_j,M_j(C))]/(mk)$$

According to the algorithm, the nearest hits H_j is finds k the nearest neighbors which is from the same class while the nearest misses $M_j(C)$ is k the nearest neighbors which is from a different class. W (A) which is the quality estimation for all attributes A is updated based on their values for R_i, hits H_j and $M(C)$. Each probability weight must be divided by 1-P (class (R_i)) due to the missing of the class of hits in the sum of *diff* (A, R, $M_j(C)$).

3.2.3 Information Gain

The next method of feature selection that will be used in the experiment is Information Gain. The equation of the information gain method applied is as following.

Let $\{c_i\}^m_{i=1}$ denote the of classes. Let V be the set of possible values for feature f. Thus, the information gain of a feature f is defined as:

$$G(f) = \sum_{i=1}^{m} P(c_j) \log P\left(c_j\right) + \sum_{v \in V} \sum_{i=1}^{m} P\,(f = v) P(c_j \,|f = v) log P(c_j | f = v) \tag{8}$$

In information gain, the numeric features need to be discretized. Therefore, an entropy-based discretization method is used and has been implemented in WEKA software.

3.2.4 Chi Square

Chi- Square or χ^2 –statistic is another feature selection method that will be used as comparison. The equation of this method is as follow:

$$X^2(f) = \sum_{v \in V} \sum_{i=1}^{m} \frac{[A_i(f = v) - E_j(f = v)]^2}{E_j(f = v)} \tag{9}$$

Where,

V is the set of possible values for feature f,
A_i ($f = v$) is the number of instances in class c_i with $f = v$
E_i($f = v$) is the expected value of A_i ($f = v$)
E_i($f = v$) is computed with:

$$E_i(\, f = v) = P\,(\, f = v) P(c_i) N \tag{10}$$

where *N* is the total number of instances.

Same as information gain, this method also requires numeric features to be discretized. The following is the example of the chi square algorithm which is consists of two phases. For discretization, in the first phase, it begins with a high significance level (sigLevel) for all numeric attributes. The process involved in phase 1 will be iterated with a decreased sigLevel until an inconsistency rate, δ is exceeded in the discretized data. In the Phase 2, begin with sigLevel0 determined in Phase 1, each attribute *i* is associated with a sigLevel [*i*] and takes turns for merging until no attribute's value can be merged. If an attribute is merged to only one value at the end of Phase 2, it means that this attribute is not relevant in representing the original dataset. Feature selection is accomplished when discretization ends.

FIGURE 4 CHI SQUARE ALGORITHM

```
Phase 1:
set sigLevel = .5;
do while (InConsistency (data) < δ) {
for each numeric attribute {
Sort(attribute, data);
chi-sq-calculation(attribute, data)
do {
chi-sq-calculation(attribute, data)
} while (Merge (data))
}
sigLevel0 = sigLevel;
sigLevel = decreSigLevel(sigLevel);
}
Phase 2:
set all sigLvl [i] = sigLevel0
do until no-attribute-can be-merged {
for each attribute i that can be merged {
Sort(attribute, data);
chi-sq-initialization(attribute, data);
do {
chi-sq-calculation(attribute, data)
} while (Merge(data))
if (InConsistency (data) < δ)
sigLvl [i] = decreSigLevel(sigLvl [i]);
else
attribute i cannot be merged
```

3.2.5 PNN Classifier

The probabilistic neural network was developed by Donald Specht. This network provides a general solution to pattern classification problems by following an approach developed in statistics, called Bayesian classifiers. Bayes theory, developed in the 1950's, takes into account the relative likelihood of events and uses a priori information to improve prediction.

The probabilistic neural network uses a supervised training set to develop distribution functions within a pattern layer. These functions are used to estimate the likelihood of an input feature vector being part of a learned category, or class. The learned patterns can also be combined, or weighted, with the a priori probability, also called the relative frequency, of each category to determine the most likely class for a given input vector. If the relative frequency of the categories is unknown, then all categories can be assumed to be equally likely and the determination of category is solely based on the closeness of the input feature vector to the distribution function of a class.

The probabilistic neural network has three layers. The network contains an input layer which has as many elements as there are separable parameters needed to describe the objects to be classified. It has a pattern layer, which organizes the training set such that each input vector is represented by an individual processing element. And finally, the network contains an output layer, called the summation layer, which has as many processing elements as there are classes to be recognized. Each element in this layer combines via processing elements within the pattern layer which relate to the same class and prepares that category for output. Sometimes a fourth layer is added to normalize the input vector, if the inputs are not already normalized before they enter the network.

In the pattern layer, there is a processing element for each input vector in the training set. Normally, there are equal amounts of processing elements for each output class. Otherwise, one or more classes may be skewed incorrectly and the network will generate poor results. Each processing element in the pattern layer is trained once. An element is trained to generate a high output value when an input vector matches the training vector. The training function may include a global smoothing factor to better generalize classification results. In any case, the training vectors do not have to be in any special order in the training set, since the category of a particular vector is specified by the desired output of the input. The learning function simply selects the first untrained processing element in the correct output class and modifies its weights to match the training vector.

The pattern layer operates competitively, where only the highest match to an input vector wins and generates an output. In this way, only one classification category is generated for any given input vector. If the input does not relate well to any patterns programmed into the pattern layer, no output is generated.

The Parzen estimation can be added to the pattern layer to fine tune the classification of objects. This is done by adding the frequency of occurrence for each training pattern built into a processing element. Basically, the probability distribution of occurrence for each example in a class is multiplied into its respective training node. In this way, a more accurate expectation of an object is added to the features which make it recognizable as a class member.

The first layer (input layer) of PNN accepts input in *d*-dimensional input vectors. The second layer calculates the Gaussian basis function (GBFs) as in the Eq. 11 below:

$$\rho_{m,k}(X) = \frac{1}{(2\pi\sigma_{m,k}^2)^{\frac{d}{2}}} exp - (\frac{\left\|X - v_{m,k}\right\|^2}{2\sigma_{m,k}^2})$$

(11)

where it specifies the GBF for *m*-th cluster in the *k*-th class where $\sigma_{m,k}^2$ is the variance $v_{m,k}$ is the cluster centroid and *d* represents the dimension of the input vector.

The third layer of PNN is where the class conditional probability density function is estimated, given by the formula

$$f_k(X) = \sum_{m=1}^{M_k} \beta_{m,k}\rho_{m,k}(X)$$

(12)

where M_k is the number of clusters for class *k* and $\beta_{m,k}$ is the intra-class mixing coefficient that can be defined as below.

$$\sum_{m=1}^{M_k} \beta_{m,k} = 1$$

The flow of PNN can be explained further by the following pseudo-code:

FIGURE 5 PNN ALGORITHM [14]

<1> input layer
Given an unknown pattern or feature vector **x**
<2> pattern layer: $\mathbf{x}^i$ is the i^{th} reference pattern vector
for i = 1:N
$y^i = \mathbf{x}^i\mathbf{x}^T - 0.5(\mathbf{x}\mathbf{x}^T + \mathbf{x}^i(\mathbf{x}^i)^T)$;
$y^i = \exp(y^i/h^2)$; % go through activation function
end
<3> summation layer
for j = 1:n
sum(j) = 0;
for all i in {1,..., N} % all instances in the same taxon
sum(j) += $y^{(j,i)}$;
end
sum(j) = $\frac{sum(j)}{2\pi^{m/2}h^m N_j}$ % go through activation
function
end
pattern **x** belongs to taxon j with some memberships as
membership(j) = $\frac{sum(j)}{\sum_{j-1}^{n} sum(j)}$ for all j in [1,n]
<4> output layer
assign pattern **x** to taxon *j* (s_j) with the highest membership such that
s_j* = argmax{*membership(j)*}
all j ∈{1,...,n}
Conclusion: assign **x** to taxon j with membership(j).

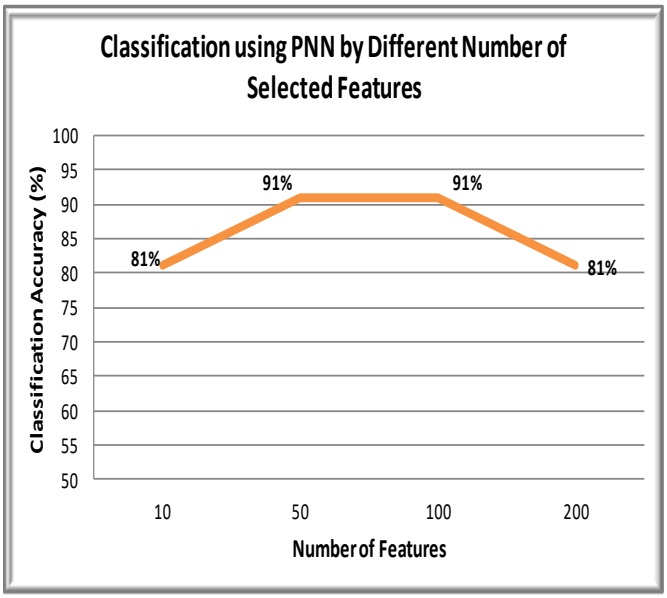
3.2.6 Experimental Settings

Generally, the experiment was divided into 2 phases; Feature Selection and Classification. For both phases, the dataset being tested was Leukemia dataset. This dataset consist of 7129 genes, 72 samples and 2 classes namely Acute myeloid leukemia (AML) and Acute lymphoblastic leukemia (ALL). Table 1 shows the description of leukemia dataset. This experiment was conducted on a platform of Microsoft Windows XP Professional Edition (Service Pack 3) using Intel Core 2 Duo processor and 2.5 Gigabyte of RAM.

TABLE 1 LEUKEMIA DATASET

Class	No. of samples	No. of genes per sample
ALL	47	7129
AML	25	7129
Total	72	

FIGURE 6 CLASSIFICATION USING PNN BY DIFFERENT NUMBER OF SELECTED FEATURES



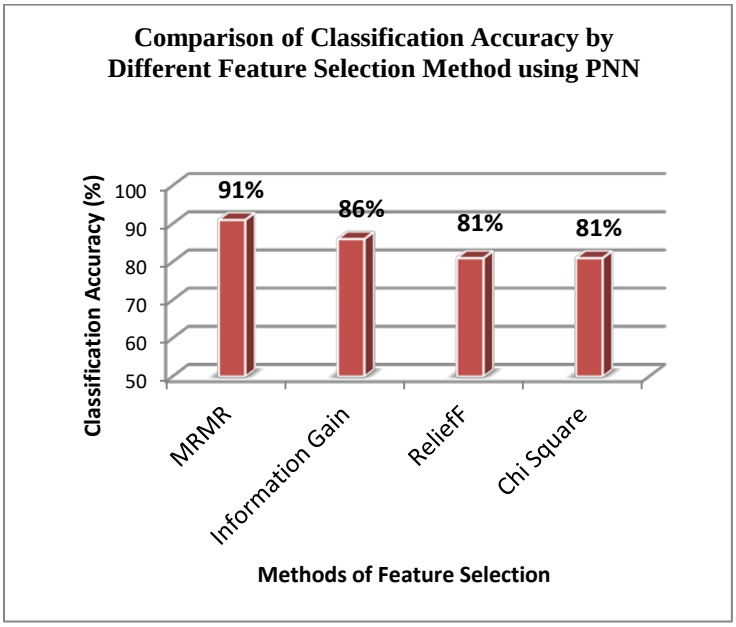
The above figure shows the classification using PNN by different number of selected features. The numbers of selected features being tested to evaluate the mRMR performance are 10, 50, 100 and 200 features. According to the result achieved,

classification accuracy produced by 50 and 100 selected genes has 91% of accuracy whereas for 10 and 200 selected genes, the accuracy was a bit lower (81%).

The low accuracy produced by 10 genes is due to its lack of informative genes to give information about a class. Lacking information gives the classifier a very small training data that is not enough to produce a great model. This lead to misclassification of test set using a poor model. For 200 selected genes, as been explained before; the poor performance is caused by redundancy among features. Let say, in 200 genes, only 100 genes that gives information about the class whereas the other 100 genes consist of noises and irrelevant genes. These genes will then being fit into classifier to be train and yield a poor model because of its irrelevant information. As for 50 and 100 genes selected, the information given was sufficient for the classifier to produce a very good model. These subset of genes does not have any redundant genes among them and produce a great performance of classification.

In feature selection, there are a few other techniques that have been used to select informative genes such as Chi Square, ReliefF and Information Gain. Thus, to ensure that mRMR technique gives promising results, a comparison of these techniques is conducted to see which technique gives better classification accuracy. Figure 7 displays the graph of the comparison among feature selection techniques.

FIGURE 7 COMPARISON OF CLASSIFICATION ACCURACY BY DIFFERENT FEATURE SELECTION METHOD USING PNN



Based on the graph above, it is very clear that mRMR achieve the best performance among other techniques. mRMR produces 91% classification accuracy compared to other feature selection techniques that yield lower accuracy, 86% for Information Gain, 81% for ReliefF and Chi Square. The lowest percentage in comparison

of feature selection techniques belongs to ReliefF and Chi Square techniques. The key of success in mRMR technique is this technique focuses more on redundant of genes rather than relevance only [15]. This technique has been proven in giving tremendous result based on previous work. By eliminating redundant genes, mRMR performs well by selecting only very informative genes that strongly contribute in determining its class. Compared to ReliefF, Information Gain and Chi Square, these techniques compute only the relevance of features/genes and ignoring the existence of redundancy in feature subsets. This leads to misclassification of genes due to its irrelevant features.

CONCLUSION

Based on overall results achieved, it is clear that mRMR and PNN classifier gives better results compared to other existing techniques. mRMR technique focuses on redundancy of genes and at the same time the maximum relevance also been taken into account. Unlike other feature selection techniques (Information Gain, ReliefF and Chi Square), where they only focus on the maximum relevance of features and ignore the existing of redundancy problem in features subset. Thus, this gives credit to mRMR where it concerns on both problem and that is the major reason why it produces better results.

The result obtained from several experiment has prove that mRMR techniques select very useful genes and reduce redundancy whereas PNN acts as a great classifiers that gives a better result. The selection of genes is important since it affect the efficiency of classifiers if the data given are huge. Furthermore, by selecting only the relevant genes, biologists do not have to waste time in investigating the wrong genes that causes cancer. They only have to rely on the selected genes to carry on their research. Thus, it can be said that combination of mRMR technique and PNN classifier gives great result in classification of microarray data.

However, during this research, there were some problems and limitations exist while implementing the feature selection techniques and PNN technique. One of the problems and limitations is the used of limited dataset. This research only uses one dataset which is leukemia. Hence it does not generalize the result obtained and there are no comparisons of results between other dataset which might have different type of diseases, different number of sizes and different number of classes.

Since the problems exists because of a limited dataset, thus, to overcome this situation, it is important to use different types of dataset which different in terms of sizes, diseases and classes. This will help getting a better result and comparison between the datasets can show the most suitable dataset to be used by feature selection techniques and PNN technique and further validate the performance of the designed technique. Examples of other datasets would be colon tumor, lung cancer, breast cancer, and prostate cancer.

REFERENCES

Russell, P.J., *Essential iGenetics,* San Francisco: Benjamin Cummings, 226-265, 2003.

Paul, T.K and Iba, H., Gene selection for classification of cancers using probabilistic model building genetic algorithm, *BioSystems,* Vol 82: 208-225, 2005.

Xu, R. and Wunsch, D. C., Probabilistic Neural Networks for Multi-class Tissue Discrimination with Gene Expression Data, *Proceedings of the International Joint Conference on Neural Network,*Vol 3: 1696-1701, 2003.

Huerta, E.B., Duval, B., Hao, J.K., A hybrid GA/SVM approach for gene selection and classification of microarray data.

Ding, C. and Peng, H., Minimum Redundancy Feature Selection from Microarray Gene Expression Data, *Proceedings of the Computational Systems Bioinformatics,* 2003.

Kononenko, I., Estimating Attributes: Analysis and Extensions of Relief. *Proceedings of the European Conference on Machine Learning,* Springer-Verlag New York, 171-182, 1994.

Cover, T., and Thomas, J., *Elements of Information Theory,* New York :John Wiley and Sons,
1991.

Zheng, Z., Srihari, R. and Srihari, S., A Feature Selection Framework Text Filtering, *Proceedings of the Third IEEE International Conference on Data Mining,* 2003.

Pastell, M.E. and Kujala, M., A Probabilistic Neural Network Model for Lameness Detection,
American Dairy Science Association, Vol 90: 2283- 2292, 2007.

Shan, Y., Zhao, R., Xu, G., Liebich, H.M. and Zhang, Y., Application of Probabilistic Neural Network in the Clinical Diagnosis of Cancers based on Clinical Chemistry Data, *Analytica Chimica Acta,* 77-86, 2002.

Nur Safawati binti Mahshos, *Pengecaman Imej Kapal Terbang Dengan Menggunakan Teknik Rangkaian Neural Radial Basis Fuction Dan Rambatan Balik,* Universiti Teknologi Malaysia: Tesis Sarjana Muda, 2008.

Golub, T.R., Slonim, D.K, Tamayo, P., *et al*., Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring, *Science,* Vol 286: 531-537, 1999.

Park, H., and Kwon, H-C., Extended Relief algorithms in instance-based Feature Filtering, *Sixth International conference on Advanced Language Processing and Web Information Technology,* 123-128, 2007.

Bi, C., Saunders, M. C. and McPheron, B. A., Wing Pattern-Based Classification of the *Rhagoletis pomonella* Species Complex Using Genetic Neural Networks, *International*

Journal of Computer Science & Applications, Vol 4: 1-14,2007.

Peng, H., Long, F. and Ding, C., Feature Selection Based on Mutual Imformation: Criteria of
Max-Dependency, Max-Relevance, and Min-Redundancy, *IEEE Transactions on Pettern Analysis and Machine Intelligence,* Vol 27: 1226-1238, 2005.