



# Gene Expression Data Classification with Kernel independent Component Analysis

Abdallah Bashir Musa

College of Mathematics and Computer Science, Hebei University, Baoding 071002, CHINA

Available online at: [www.isca.in](http://www.isca.in), [www.isca.me](http://www.isca.me)

Received 22<sup>nd</sup> March 2014, revised 15<sup>th</sup> April 2014, accepted 10<sup>th</sup> May 2014

## Abstract

The challenge of classifying the characteristics of gene expression data is that the size of the training data is significantly lower than the number of features. Logistic regression (LR) is standard statistical method that broadly used in medical, epidemiology and bioinformatics communities for classification task; however, in such situation of gene expression data, LR does not work efficiently due to multi- collinearly and over- fitting problems, therefore, modifying of LR to analysis the microarray data is required. For solving those problems, reduction dimension is usually used. Recently, kernel approaches have proven to be good for classification such type of data. Kernel independent component analysis (KICA) is the nonlinear form of independent component analysis (ICA). In this paper, LR is applied to classify the features that selected by KICA. To evaluate the classification performance of this technique, this method has compared to kernel principle component analysis (KPCA) and independent component analysis (ICA). Numerous performance metrics such as accuracy, sensitivity, specificity, precision, F-score, the area under receiver operating characteristic curve (AUC) and the receiver operating characteristic (ROC) analysis are used.

**Keywords:** Gene expression data. Kernel principal component analysis (KICA). Logistic regression (LR). Kernel principal component analysis (KPCA)..

## Introduction

Logistic Regression (LR)<sup>1-4</sup> is considered as a standard multivariate statistical classification technique that has been used broadly in a variety of applications including document classification<sup>5</sup> and bioinformatics<sup>6-8</sup>. The advantage of using LR is that it yields a-posteriori probabilities, In other words, besides predicting class labels LR provides a probabilistic interpretation about this labeling. LR has an additional advantage that the extension to the multi-class case is well described.

The fast increasing of microarray technology has generated a riche and huge of gene expression data sets. Usually, gene expression data sets include large number of features with small number of data size and also it includes a high level of noises. Statistically, it's known that applying of all types of regression including logistic regression requires the data size to be relatively bigger than then the number of features; this makes the classical LR classification method invalid to analyze these types of data. Therefore, constructing a logistic regression model using gene expression data is very difficult and considered as a serious technical challenge, which has attracted researchers' attention. Although gene expression data has big number of features, it has been observed that a few of them can be explaining most of data variation, so these features can be extracted to build reliable classifier. Accordingly, for analyzing gene expression data, initially we need to select or extract only these few relevant genes (features) from the original data.

Instead of classifying the whole expression data, feature selection methods aim to select the most essential features with low dimensions from the original data and then the classification method can be applied on these selected features. The previous studies shown that the choice of gene selection method has much effect on the performance of the classification methods, and thus the classification methods should be considered together with the gene selection criteria<sup>9</sup>. Feature selection represents an essential part in classification tasks. Generally, the main aim of feature selection methods is to decrease the computational complexity and produces a few meaningful features that include the maximum information in the data, which can be used as an input features to a classifier to gain maximum accuracy that can be obtained. The important advantage of features selection methods that it selects few uncorrelated or independent features (gene) consequently solves the problems of over fitting and co-linearity simultaneously, which result in improving the accuracy of classification when applying logistic regression to those selected features.

Several machine learning techniques have been used to classify the gene expression data, such as support vector machine (SVM); LORIS NANNI<sup>10</sup> who applied support vector machine (SVM) with different feature reduction methods, ISABELLE GUYON<sup>11</sup> who applied the SVM method of recursive feature elimination (RFE) to gene selection, Michael P. S. Brown<sup>12</sup> who used SVMs to classify genes based on gene expression,

Terrence S<sup>13</sup> who developed a method for analyzing microarray data using SVM, Hao Helen Zhang<sup>14</sup> who applied regularization method with support vector machines (SVMs) to classify cancer data, NIR FRIEDMAN<sup>15</sup> who analysis gene expression data based on Bayesian networks, Pierre Baldi<sup>16</sup> who develop a Bayesian probabilistic framework for microarray data analysis and Kyoungwha Bae<sup>17</sup> who proposed a Gene selection method by using a two-level hierarchical Bayesian model, Kyeong Eun Lee<sup>18</sup> who used Bayesian mixture prior to perform the gene expression data variable selection. Although SVM usually achieves low test error despite small sample sizes and it gives good performance regarding classifying of gene expression data<sup>19</sup>, it considered to be a black box predictor, it neither makes it prediction implicit nor gives incite in the rule governing its prediction which is not the case in LR<sup>20</sup>.

Regard to logistic regression, it has been successfully extended to classify microarray data; J.G. Liao (2007)<sup>21</sup> develops a parametric bootstrap method for building a logistic regression model for disease classification using microarray data, Fort and Lambert-Lacroix<sup>22</sup> employed partial least squares and logistic regression for classifying microarray data, Shen L. Tan EC<sup>23</sup> they used penalized logistic regression for cancer classification using microarray expression data, JI ZHU<sup>24</sup> proposed penalized logistic regression (PLR) as an for the microarray cancer diagnosis problem, Maureen A<sup>25</sup> who presented a logistic regression approach for identifying enriched biological groups in gene expression data and Danhv. Nguyen<sup>26</sup> applied PCA partial least squares (PLS) with LR for tumor classification using microarray gene expression data.

From features selection/extraction point of view, popular features section methods such as principle components analysis (PCA) and independent component analysis have been used for classification gene expression data; K.Y.Yeung and W.L.Ruzzo<sup>27</sup> who used PCA for clustering gene expression data and De-Shuang Huang<sup>28</sup> applied Independent component analysis-based penalized discriminant method for tumor classification using gene expression data.

Because applying of linear features selection methods may give inaccurate results for some real- world data sets, recently, kernel features selection methods have been developed and used, such as kernel principal component analysis; Zhenqiu Liu, Dechang Chen and Halima Bensmail<sup>29</sup>, they used kernel principal component analysis(KPCA) for dimension reduction and applied logistic regression for classification gene expression data and recently, Qingsong Gao<sup>30</sup> shows that KPCA with LR is a valid and powerful gene- or region-based method for the analysis of genome-wide association studies( GWAS) data set.

This paper, proposes the application of kernel independent component analysis (KICA)<sup>31</sup> with logistic regression for classifying gene expression data. Comprehensive comparison experiment between KICA with LR, ICA and KPCA, with LR is performed.

**Logistic regression:** Logistic Regression (LR)<sup>1-4, 32</sup> is a multivariate statistical classification method commonly used for modeling dichotomous (binary) data. Let  $x \in R^n$  denote a vector of explanatory or feature variables, and let  $y \in \{-1,+1\}$  denote the associated binary class label or outcome. The logistic model is described as:

$$\text{pr}(y/x) = \frac{1}{1 + \exp(-y(\beta^T x + \alpha))} = \frac{\exp(y(\beta^T x + \alpha))}{1 + \exp(y(\beta^T x + \alpha))} \quad (1)$$

Where  $\text{Pr}(y/x)$  is the conditional probability of  $y$  given  $x \in R^n$ . The logistic model has parameters  $\alpha \in R$  represent the intercept tem and  $\beta \in R^n$  represent the weight vector.  $\beta^T x + \alpha = 0$  defines a hyperplane in the feature space, on which  $P(y/x)=0.5$ . The conditional probability  $\text{Pr}(y/x)$  is larger than 0.5 if  $\beta^T x + \alpha$  has the same sign as  $y$ , and less than 0.5 otherwise.

Suppose we are given a set of  $m$  observed or training data  $\{x_i, y_i\}_{i=1}^m$ , where  $x_i \in R^n$  denote the  $i$ -th sample and  $y_i \in \{-1,+1\}$  denote the corresponding class label. These  $m$  samples are assumed to be independent samples. According to the logistic model, the vector of the conditional probabilities associated of these samples is:

$$\text{pr}(\alpha, \beta)_i = p(y_i/x_i) = \frac{\exp y_i(\beta^T x_i + \alpha_i)}{1 + \exp y_i(\beta^T x_i + \alpha_i)} \quad i = 1, \dots, m \quad (2)$$

The likelihood function associated with the samples is

$$\prod_{i=1}^m \text{pr}(\alpha, \beta)_i, \text{ and the log likelihood function is:} \quad \sum_{i=1}^m \log \text{pr}(\alpha, \beta)_i = - \sum_{i=1}^m f(\beta^T a_i + \alpha y_i) \quad (3)$$

Where  $a_i = x_i y_i \in R^n$  and  $f$  is the logistic loss function that is:

$$f(z) = \log(1 + \exp(-z)) \quad (4)$$

Using (4),(3) can be written as

$$\sum_{i=1}^m \log \text{pr}(\alpha, \beta)_i = - \sum_{i=1}^m \log(1 + \exp(-(\beta^T a_i + \alpha y_i))) \quad (5)$$

The negative of the log likelihood function is called the (empirical) logistic loss, and dividing by  $m$  we obtain the average logistic loss,

$$l_{avg}(\alpha, \beta)_i = \frac{1}{m} \sum_{i=1}^m \log(1 + \exp(-(\beta^T a_i + \alpha y_i))) \quad (6)$$

The model parameters  $\beta$  and  $\alpha$  can be determined by maximum likelihood estimation from the observed examples, by solving the convex optimization problem.

$$\text{minimize } l_{avg}(\alpha, \beta)_i \quad (7)$$

The problem (7) is called the logistic regression problem (LRP). This LRP is a smooth convex optimization problem, and can be solved by a wide variety of methods, such as gradient descent, steepest descent, Newton, quasi-Newton, or conjugate-gradients (CG) methods. In this paper the Newton method is used. Once we find maximum likelihood values of  $\alpha$  and  $\beta$ , that is, a solution of (7), we can predict the probability of the two possible outcomes. Given a new features vector  $x \in R^n$ , by using the associated logistic regression model, the logistic regression classifier is formed as:

$$\phi(x) = \text{sgn}(\beta^T x + \alpha)$$

$$\text{Where } \text{sgn}(z) = \begin{cases} +1 & z > 0 \\ -1 & z \leq 0 \end{cases} \quad (8)$$

**Independent Component Analysis (ICA):** ICA<sup>33-35</sup> is a relatively new statistical and computational technique for data analysis. ICA originated from the signal-processing community, where it was developed as a powerful procedure for blind source separation<sup>33</sup>. The task of ICA is to find representation of non-Gaussian data so those components are statistically independent or as independent as possible<sup>34</sup>. The basic ICA model for feature transformation can be described as:

$$S_i = u x_i \quad (9)$$

Where  $x_i$  is  $n \times p$  matrix represents the observed feature vectors,  $s_i$  is  $n \times p$  matrix represent the new independent estimated vectors for classification purpose,  $u$  is called the  $n \times n$  de-mixing matrix and is used to find an entirely new coordinate system of statistically independent non-Gaussian directions, with the first IC direction being the most non-Gaussian. The algorithm works iteratively and determines the most non-Gaussian direction first. Based on this direction it finds the next most non-Gaussian direction which is independent from the first, etc. For  $n \times p$  dimensional data vectors it determines up to  $n \times p$  dimensional independent vectors, so it projects the feature vectors representing the original data into independent components,  $u$  must be estimated from the data. Statistical independence has been defined as the joint probability density function of the component "s<sub>i</sub>" is equal to the product of marginal densities functions of the individual components. Many algorithms have been developed for performing ICA, the fixed-point algorithm is popular among them. Fixed point fast ICA presented by Hyvarinen and Oja<sup>36</sup> is used in this paper. In fast ICA, principle component analysis (PCA) is used to perform the

whitening before estimating the independent components vectors; the original input vectors will be transformed to a set of new uncorrelated vectors with zero means and unity variance. Using PCA to get the whitening reduced the dimension of  $x_i$  and consequently, reduced the number of  $s_i$  that will be computed. After the process of data whitening is finished, the fixed point-algorithm is performed to estimate the transformation matrix and independent components. Mutual information is described as a measure of the dependence between random variables. Minimizing the mutual information between the components is equivalent to maximizing their negentropy. The negentropy in the fast ICA can be approximately expressed as follows:

$$J_G(s_{i(i)}) \approx [E\{G(s_{i(i)})\} - E\{G(V)\}]^2 \quad (10)$$

Where  $G$  is practically any non-quadratic function,  $V$  is a Gaussian variable with zero mean and unit variance and  $\mu_i$  is  $n-u$  dimensional vector, comprising one of the rows of the matrix. There are many functions can be used as  $G$ <sup>33</sup>. Substituting in equation (10) obtaining the following optimization problem:

$$\text{Maximize } \sum J_G(\mu_i) = [E\{G(\mu_i^T x)\} - E\{G(V)\}]^2 \quad (11)$$

Subject to

$$E\{(\mu_i^T x)^2\} = 1 \quad i = 1, 2, \dots, n \quad (12)$$

One new independent component can be estimated by solving this optimization problem through the Fast ICA algorithm and based on this the whole reduced independent components  $s^*_i$  matrix can be estimated. In this paper, we used Fast ICA with skew to obtain the independent components.

**Kernel Independent Component Analysis (KICA):** Kernel principle component analysis (KICA)<sup>37</sup> is the kernel version the ICA. For a given training data  $x$ , suppose this training data is transforming to new feature space  $F$  through some nonlinear mapping. Where, is a Mercer's kernel that allows the  $\Phi(x), F = \Phi(x) \quad k(x_i, x_j) = \Phi^T(x_i)\Phi(x_j)$  calculation of the dot product in this space without explicitly knowing the nonlinear mapping. In this nonlinear space the centering and whitening that have been mentioned in the previous section of ICA is obtained as follow:

For the centering task the data where  $\Phi^T(x_i) \quad i=1, 2, \dots, k$ , should be transformed to

$$\Phi^*(x_i) = \Phi(x_i) - E(\Phi(x_i)) \quad (13)$$

Where: For the whitening  $E(\Phi^*(x_i)) = 0$  in this space, the task here is to find a transformation matrix  $Q$  satisfy that the

covariance matrix of the data is unit matrix  
 $(\Phi(x_i) = Q(\Phi^*(x_i))$

For arbitrary vector the KICA transformation can be obtained as: where  $W^*$  denotes the orthogonal transformation matrix that can  $Z \in x$  be obtained as described for ICA, while  $Q$  is the matrix obtained from kernel centering and whitening.  $Z^* = W^* Q\Phi(Z)$

Different kernel functions can be used, but choosing a suitable kernel function for a certain application is pertinent. Gaussian kernel is selected for this paper.

### Methodology

**The data sets:** The gene expression data sets used in this study are the most two famous gene expression data; Colon tumor and Leukemia cancer data sets. Table-1 give a numerical summary of the data sets.

**Table-1**  
**Summary of the gene expression data sets**

Data set	Data size	Number of features
Colon tumor	62	2000
Leukemia cancer	72	7129

**Experimental set up:** Kernel Independent component analysis (KICA) with Gaussian kernel, kernel principles components analysis (KPCA) with the Gaussian (RBF) kernel and independent component analysis (ICA) are used as feature selection methods for dimensionality reduction in logistic regression. The kernel ica algorithm with Gaussian kernel is used for applying KICA; for a given training data  $x$  it returns a matrix  $W$  where the kernel independent component is:  $s=W*x$ . The performance of KICA is compared to the performance of KPCA and ICA. For all the methods, the original dimension is reduced to explain 85 percent for Colon tumor and 75 percent leukemia cancer for the total variation of the data. The Gaussian (RBF) is used for the application of KPCA. For applying ICA, the reduction dimension and whitening is obtained by using the principle component analysis (PCA) method. Then for obtaining

the independent components from these reduced whitening data, fast ICA algorithm with skew is used.

Once the feature KICA, KPCA and ICA transformed the original features spaces into new lower dimension space Logistic regression with Newton' method with 10 fold cross-validation methods is applied to classify the features in this new space.

### Results and Discussion

For the application of KICA the kernel -ica version 1.2<sup>31</sup> available at <http://www.di.ens.fr/~fbach/kernel-ica/> is used. For the application of KPCA, the Statistical Pattern Recognition Toolbox for MATLAB (stprtool) version 2.11<sup>38</sup>. For the application of ICA, the fast-ICA-2.5<sup>39</sup> software package, is used. The ll\_logreg A large logistic regression problem package version 0.8.2<sup>40</sup> available at ([http://www.stanford.edu/~boyd/ll\\_logreg/](http://www.stanford.edu/~boyd/ll_logreg/)), with Newton method, under matlab (7.8.0347- R2009a) interface is used for estimation the logistic regression model parameters. The ROCs are obtained by using spss.16.0 (SPSS Inc, Chicago, IL, USA).

The classification results of LR application in the features that selected by the kernel independent component analysis (KICA), kernel principle component analysis (KPCA) and independent component analysis (ICA) are presented in table-2, table-3 and table-4 respectively. Each value in these tables represents the average associated with corresponding metric. The number of components for the new dimension selected by those methods for the two data sets is shown in table 2. The ROCs of the Colon tumor and Leukemia cancer for three methods are shown in figure-1 and figure-2 respectively. The performances measures of three methods for Colon tumor and Leukemia cancer is depicted in figure-3.

**Discussion:** From table- 2, table-3 and table-4, it can be observed that performance measures values of KICA are greater than corresponding measures values for KPCA and ICA for the two data sets, also it's vivid that ICA performs better than KPCA on Leukemia cancer data set. This can be confirmed from figure-3, where the figure shows that KICA has higher bar for all performance measures for the two data sets in comparison to KPCA and ICA. The ROCs curve for the two data sets also support this point.

**Table-2**  
**The results of the performance measures for K ICA**

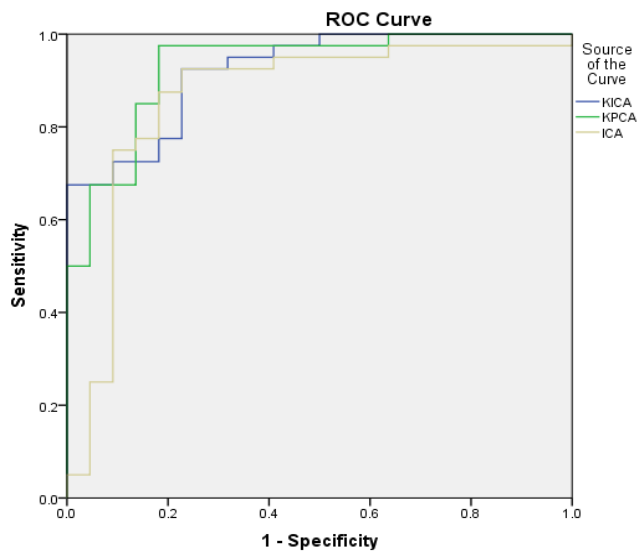
Data set	The performance measures						Number of components
	Accuracy	Sensitivity	Specificity	F-score	Precision	AUC	
Colon tumor	0.901	0.968	0.791	0.910	0.887	0.952	24
Leukemia cancer	0.886	0.834	0.944	0.888	0.954	0.943	26

**Table-3**  
**The results of the performance measures for KPCA**

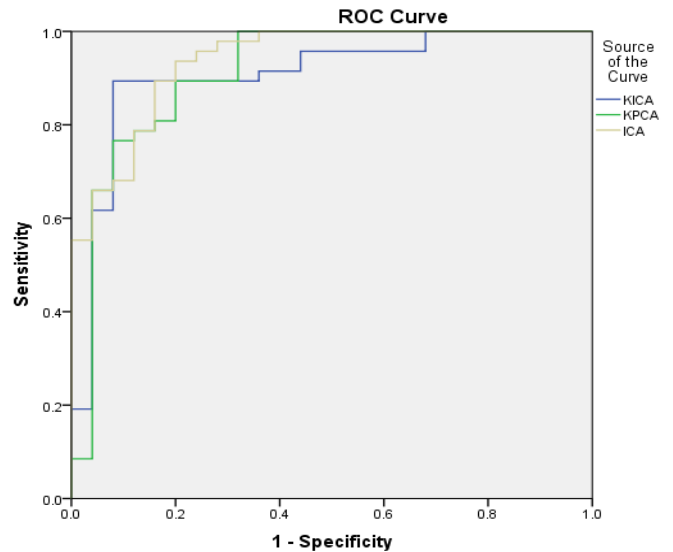
Data set	The performance measures					
	Accuracy	Sensitivity	Specificity	F-score	Precision	AUC
Colon tumor	0.887	0.925	0.818	0.914	0.902	0.930
Leukemia cancer	0.819	0.894	0.680	0.866	0.840	0.911

**Table-4**  
**The results of the performance measures for ICA**

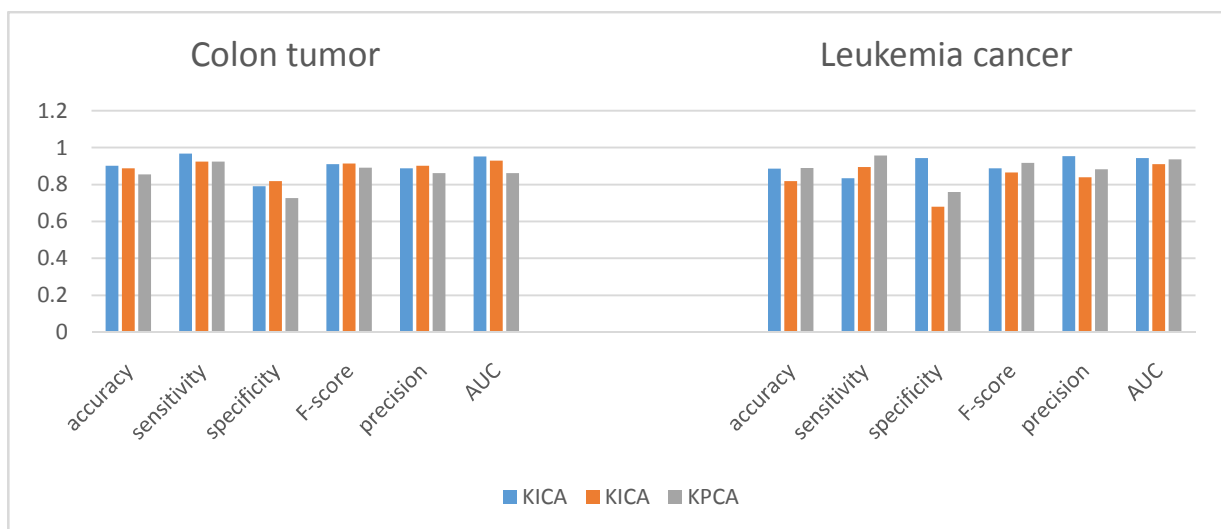
Data set	The performance measures					
	Accuracy	Sensitivity	Specificity	F-score	Precision	AUC
Colon tumor	0.855	0.925	0.727	0.892	0.861	0.861
Leukemia cancer	0.889	0.957	0.760	0.918	0.882	0.937



**Figure-1**  
**The ROCs for Colon**



**Figure-2**  
**The ROCs for Leukemia cancer**



**Figure-3**  
**The performance measures of KICA, KPCA and ICA for Colon and Leukemia cancer data sets**

## Conclusion

In this paper the kernel independent component analysis (KICA) is used as a features selection method for reduction dimension of gene expression data; Colon tumor and Leukemia cancer data sets. LR for classification is applied for the features that selected by KICA. KPCA and ICA with LR are used to assess the performance of KICA with LR. The comparison shows that KICA outperformed ICA and KPCA. That's why KICA can be used effectively as a features selection method with LR for classifying the gene expression data. From the same classification point of view, and from the fact that naïve Bayse classification method required the training features to be independent, future work may apply KICA technique with naïve Bayse.

## References

1. Hosmer D.W. and Lemeshow S., Applied logistic regression, 2nd edn. Wiley series in probability and statistics, Wiley, Inc, New York, (2000)
2. Menard S., Applied logistic regression analysis, 2<sup>nd</sup> edn. Sage publications Inc, (2002)
3. Neter J., Kutner M.H., Nachtsheim C.J. and Wasserman W., Applied linear statistical models, 4th edn. Irwin, Chicago, (1996)
4. Ryan T.P., Modern regression methods, 2nd edn. Wiley, New York, (2008)
5. Brzezinski J.R. and Knafl G.J., Logistic regression modeling for context-based classification. In: Proceedings tenth international workshop on database and expert systems applications, 755-759 (1999) doi:10.1109/DEXA.1999.795279
6. Liao J.G. and Chin K.V., Logistic regression for disease classification using microarray data: model selection in a large p and small n case, *Bioinformatics*, **23**(15), 1945-1951(2007)
7. Sartor M.A., Leikauf G.D., Medvedovic Lrpath M, A logistic regression approach for identifying enriched biological groups in gene expression data, *Bioinformatics*, **25**(2), 211-217 (2008)
8. Asgary M.P., Jahandideh S., Abdolmaleki P. and Kazemnejad A., Analysis and identification of b-turn types using multinomial logistic regression and artificial neural network, *Bioinformatics*, **23**(23), 3125-3130 (2007)
9. Lee Jae Won, et al., An extensive comparison of recent classification tools applied to microarray data, *Computational Statistics and Data Analysis*, **48**(4), 869-885(2005)
10. Nanni Loris, Sheryl Brahmam, and Alessandra Lumini, Combining multiple approaches for gene microarray classification, *Bioinformatics*, **28**(8), 1151-1157 (2012)
11. Guyon Isabelle, et al., Gene selection for cancer classification using support vector machines, *Machine learning*, **46**(1-3), 389-422 (2002)
12. Brown Michael P.S., et al., Knowledge-based analysis of microarray gene expression data by using support vector machines, *Proceedings of the National Academy of Sciences*, **97**(1), 262-267(2000)
13. Furey, Terrence S., et al., Support vector machine classification and validation of cancer tissue samples using microarray expression data, *Bioinformatics*, **16**(10), 906-914 (2000)
14. Zhang, Hao Helen, et al., Gene selection using support vector machines with non-convex penalty, *Bioinformatics*, **22**(1), 88-95 (2006)
15. Friedman, Nir, et al., Using Bayesian networks to analyze expression data, *Journal of computational biology*, **7**(3-4), 601-620 (2000)
16. Baldi, Pierre and Anthony D. Long, A Bayesian framework for the analysis of microarray expression data: regularized t-test and statistical inferences of gene changes, *Bioinformatics*, **17**(6), 509-519 (2001)
17. Bae, Kyoungwha and Bani K. Mallick, Gene selection using a two-level hierarchical Bayesian model, *Bioinformatics* **20**(18), 3423-3430 (2004)
18. Lee, Kyeong Eun, et al., Gene selection: a Bayesian variable selection approach, *Bioinformatics*, **19**(1), 90-97 (2003)
19. Bae, Kyoungwha and Bani K. Mallick, Gene selection using a two-level hierarchical Bayesian model, *Bioinformatics*, **20**(18), 3423-3430 (2004)
20. Musa A.B., Comparative study on classification performance between support vector machine and logistic regression, *Int J Mach Learn Cybern*, **4**(1), 13-24 (2013)
21. Liao J.G. and Khew-Voon Chin, Logistic regression for disease classification using microarray data: model selection in a large p and small n case, *Bioinformatics*, **23**(15), 1945-1951(2007)
22. Fort, Gersende, and Sophie Lambert-Lacroix, Classification using partial least squares with penalized logistic regression, *Bioinformatics*, **21**(7), 1104-1111(2005)
23. Shen, Li, and Eng Chong Tan, Dimension reduction-based penalized logistic regression for cancer classification using microarray data, *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)*, **2**(2), 166-175 (2005)
24. Zhu, Ji, and Trevor Hastie, Classification of gene microarrays by penalized logistic regression, *Biostatistics*, **5**(3), 427-443 (2004)

25. Sartor, Maureen A., George D. Leikauf and Mario Medvedovic, LRpath: a logistic regression approach for identifying enriched biological groups in gene expression data, *Bioinformatics*, **25**(2), 211-217 (2009)
26. Nguyen, Danh V. and David M. Rocke, On partial least squares dimension reduction for microarray-based classification: a simulation study, *Computational statistics and data analysis*, **46**(3), 407-425 (2004)
27. Yeung, Ka Yee and Walter L. Ruzzo, Principal component analysis for clustering gene expression data, *Bioinformatics* **17**(9),763-774 (2001)
28. Huang, De-Shuang and Chun-Hou Zheng, Independent component analysis-based penalized discriminant method for tumor classification using gene expression data, *Bioinformatics*, **22**(15), 1855-1862 (2006)
29. Liu, Zhenqiu, Dechang Chen and Halima Bensmail, Gene expression data classification with kernel principal component analysis, *BioMed Research International*, **2**, 155-159 (2005)
30. Gao, Qingsong, et al, Gene-or region-based association study via kernel principal component analysis, *BMC genetics*, **12**(1), 75 (2011)
31. Bach Francis R. and Michael I. Jordan, Kernel independent component analysis, *The Journal of Machine Learning Research*, **3**, 1-48(2003)
32. Musa, Abdallah Bashir, Logistic Regression Classification for Uncertain Data, *Research Journal of Mathematical and Statistical Sciences -ISSN 2320: 6047* (2014)
33. Hyva`rinen A, Karhunen J, Oja E, Independent component analysis, Wiley, New York, (2001)
34. Hyvarinen A, Oja E, Independent component analysis: algorithms and applications, *Neural Network*, **13**, 411-430 (2000)
35. Musa, Abdallah Bashir, A comparison of  $\ell_1$ -regularization, PCA, KPCA and ICA for dimensionality reduction in logistic regression, *International Journal of Machine Learning and Cybernetics*, 1-13 (2013)
36. Hyvarinen A. and Oja E., A fast fixed-point algorithm for independent component analysis, *Neural Comput*, **9**(7), 483-1492 (1997)
37. Jin, Xin, et al, Kernel independent component analysis for gene expression data clustering, Independent Component Analysis and Blind Signal Separation, *Springer Berlin Heidelberg*, 454-461(2006)
38. van der Maaten L, Statistical pattern recognition toolbox for Matlab (stprtool) version 2.11, version 0.7.2b, (2010)
39. Gavert H., Hurri J., Sarela J. and Hyvarinen A., Fast ICA for Matlab 7.x and 6.x, Version 2.5, (2005)
40. Koh K., Kim S.J. and Boyd S.,  $\ell_1$ -logreg: A large-scale solver for  $\ell_1$ -regularized logistic regression problems. 0.8.2 Available at [http://wwwssw.stanford.edu/~boyd/l1\\_logreg/](http://wwwssw.stanford.edu/~boyd/l1_logreg/), (2009)