International Journal on Cybernetics & Informatics (IJCI) Vol. 3, No. 2, April 2014

# CANCER RECOGNITION FROM DNA MICROARRAY GENE EXPRESSION DATA USING AVERAGED ONE-DEPENDENCE ESTIMATORS

Shoon Lei Win, Zaw Zaw Htike, FaridahYusof, Ibrahim A. Noorbatcha

Faculty of Engineering, IIUM, Kuala Lumpur, Malaysia

## *ABSTRACT*

*Cancer is a major leading cause of death and responsible for around 13% of all deaths world-wide. Cancer incidence rate is growing at an alarming rate in the world. Despite the fact that cancer is preventable and curable in early stages, the vast majority of patients are diagnosed with cancer very late. Therefore, it is of paramount importance to prevent and detect cancer early. Nonetheless, conventional methods of detecting and diagnosing cancer rely solely on skilled physicians, with the help of medical imaging, to detect certain symptoms that usually appear in the late stages of cancer. The microarray gene expression technology is a promising technology that can detect cancerous cells in early stages of cancer by analyzing gene expression of tissue samples. The microarray technology allows researchers to examine the expression of thousands of genes simultaneously. This paper describes a state-of-the-art machine learning based approach called averaged one-dependence estimators with subsumption resolution to tackle the problem of recognizing cancer from DNA microarray gene expression data. To lower the computational complexity and to increase the generalization capability of the system, we employ an entropy-based geneselection approach to select relevant gene that are directly responsible for cancer discrimination. This proposed system has achieved an average accuracy of 98.94% in recognizing and classifyingcancer over 11 benchmark cancer datasets. The experimental results demonstrate the efficacy of our framework.*

## *KEYWORDS*

*Cancerrecognition; microarray gene expression; AODEsr*

## 1. INTRODUCTION

According to the World Health Organization (WHO), cancer is a leading cause of death and responsible for around 13% of all deaths world-wide [1]. Cancer incidence rate is growing at an alarming rate. Despite the fact that cancer is preventable and curable at an early stage, the vast majority of patients are diagnosed with cancer very late. Therefore, preventing and detecting cancer early is very important. Nonetheless, conventional methods of detecting and diagnosing cancer rely solely on skilled physicians, with the help of medical imaging, to detect certain symptoms that usually appear in the late stages of cancer. Therefore, an early cancer detection system is required to prevent people from dying as a consequence of this unfortunate disease.

Technically, cancer is a family of diseases that involve uncontrolled cell growth wherein cells divide and grow exponentially, generating malignant tumors and spreading to other parts of the body. The destructive power of the cancer is that it may not only spread to the neighboring tissues, but also to the whole body through the lymphatic system or bloodstream. There are a few hundreds of known cancers found in humans[2]. Because there are an astronomical number of

DOI: 10.5121/ijci.2014.3201

causes of cancer, researchers are still trying to understand the basis of cancer which still remain only partially understood. However, one thing that is apparent is that in order for a healthy cell to transmute into a cancer cell, the genes which regulate cell growth and differentiation must be modified[3]. It is known that cancers are caused are a chain of mutations to the genetic sequence. Figure 1[4] depicts the development of a cancer cell caused by a series of mutations which makes the cell proliferate more than its immediate neighbors by a process which transforms a normal healthy cell into a micro-invasive cell at the genetic level.

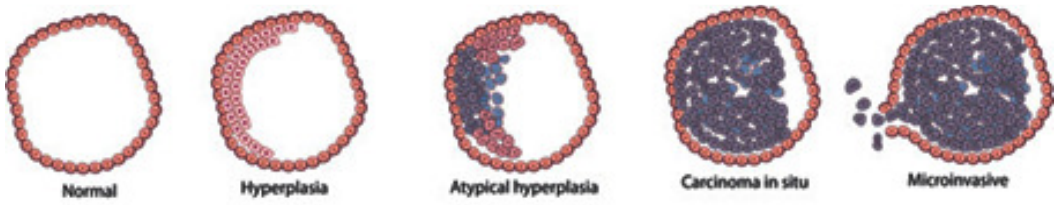


Figure 1. Formation of Cancer Cell[4].

The nucleus of a human cell contains 46 chromosomes, each of which comprises a single linear molecule of deoxyribonucleic acid (DNA), which is intimately complexed with proteins in the form of chromatin[5]. DNA is the building block of life, which contains encoded genetic instructions for living organisms. A DNA is transcribed to become a precursor mRNA, which is then spliced to become an mRNA, which is in turn translated to become a protein. Because all the cells (except some) in a human body contain an identical set of genes, the expression level of each gene must differ from cell to cell. If we can somehow measure the expression levels of individual genes in a cell, we can use machine learning techniques to predict whether a cell is cancerous and what type of cancer it is. Fortunately, the DNA microarray technology allows researchers to measure expression levels of genes in a cell. A DNA microarray, also known as DNA chip, gene chip, gene array or biochip, is a densely packed array of identified DNA sequences attached to a solid surface, such as glass, plastic or silicon chip[6]. On a microarray chip, DNA fragments are attached to a substrate and then probed with a known gene or fragment. DNA sequences representing tens of thousands of genes are spotted or in situ synthesized on a very small slide like the one in Figure 2 [6]. The microarray in Figure 2 [6] is comprised of more than 50,000 probe sets capable of evaluating expression level of over 40,000 transcripts, including 38,500 human genes [6].Microarray chips are scanned using an microarray scanner [7] and digitized on a computer. The scanner generates a 2D heat map, also known as, microarray image or microarray data. Therefore, DNA microarrays can be used to determine which genes are “turned on” (expressed) and which genes and “turn off” in a particular cell. They determine not only whether individual genes are expressed, but also the level at which these individual genes are expressed.


Figure2. DNA microarray chip[6].

In this paper, we tackle the problem of recognizing cancer from DNA microarray gene expression data. During the past few decades, applications of pattern recognition and machine learning techniques have emerged in many domains[8-17]. Pattern recognition and machine learning techniques have also recently become popular in the arena of microarray gene expression analysis.There have been some attempts to recognize and classify cancer using machine learning techniques. Zainuddin and Ong [18] proposed a novel approachto perform microarray gene expression data classification using wavelet neural networks (WNN). The types of activation functions used in the hidden layer of the WNN were varied. They also proposed an enhanced fuzzy c-means clustering algorithm—specifically, the modified point symmetry-based fuzzy c-means (MSFCM) algorithm—to select the locations of the translation vectors of the WNN. Similarly, Chen [19] employed strict ordinal regressions, including cumulative logit model in traditional statistics with dimensionality reduction, and distribution-free approaches of large margin rank boundaries implemented by the support vector machine, as well as an ensemble ranking scheme, to classify cancer stagesfrom gene expression microarray data. Sharma and Paliwal[20] proposed Gradient LDA technique which avoided the singularity problem associated with the within-class scatter matrix and their experimental results showed the usefulness of their cancer classification system. Their technique was applied on three gene expression datasets; namely, acute leukemia, small round blue-cell tumor (SRBCT) and lung adenocarcinoma. They alleged that their system achieved lower misclassification error as compared to several other previous techniques. Chakraborty[21] came up with a hierarchical Bayesian probit model for two-class cancer classification. Instead of assuming a linear structure for the function that relates the gene expressions with the cancer types, they assumed that the relationship was explained by an unknown function which belonged to an abstract functional space like the reproducing kernel Hilbert space. Their formulation automatically reduces the dimension of the problem from the large number of covariates or genes to a small sample size. Their model is highly flexible in terms of explaining the relationship between the cancer types and gene expression measurements and picking up the differentially expressed genes. Many other researchers [22-28] employ probabilistic approach to classify microarray gene expression data.This paper describes a state-of-the-art machine learning based approach called averaged one-dependence estimators with subsumption resolution to tackle the problem of recognizing cancer without any prior knowledge.

## 2. CANCER RECOGNITION FROM MICROARRAY GENE EXPRESSION DATA

The goal of cancer recognition is to predict, given a set of gene expression data, whether or not the genetic sequence comes from a cancerous cell and what type of cancer. We proposed athree-layered framework that consists of gene selection, numerosity reduction, and genetic data classification as shown in Figure 3. The complexity of any machine learning classifier depends upon the dimensionality of the input data [29]. There is also a phenomenon known as the ‘curse

of dimensionality' that arises with high dimensional input data [30]. In the case of genetic data classification, not all the genes in a genetic sequence might be responsible for discriminating cancer types. Therefore, we propose to employ a gene selection process to select relevant genes in an unsupervised manner and a numerosity reduction process to discretize the gene expression levels. Section 2.1 describes the process of gene selection and Section 2.2 describes the process of numerosity reduction. After dimensionality reduction, we propose to perform cancer classification using the averaged one-dependence estimators with subsumption resolution (AODEsr). Section 2.3 describes the process of classification

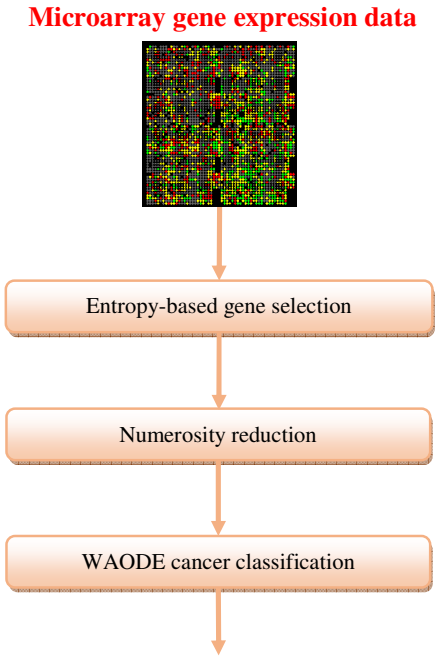


Figure3. High-level flow diagram of cancer classification framework.

## 2.1. Entropy-basedgene selection

The complexity of any machine learning classifier depends upon the dimensionality of the input data [29]. Generally, the lower the complexity of a classifier, the more robust it is. Moreover, classifiers with low complexity have less variance, which means that they vary less depending on the particulars of a sample, including noise, outliers, etc[29]. In the case of gene expression data classification, not all the genes in a genetic sequence might be responsible for discriminating cancer. Therefore, we need to have a gene selection method that chooses a subset of relevant genes that can discriminate cancer, while pruning the rest of the genes in the input genetic sequence[31, 32].

We are interested in finding the best subset of the set of genes that can sufficiently discriminate cancer. Ideally, we have to choose the best subset that contains the least number of genes that most contribute to the classification accuracy, while discarding the rest of the genes. There are $2^n$ possible subsets that can arise from an $n$-gene long genetic sequence. In essence, we have to choose the best subset out of $2^n$ possible subsets. Because performing an exhaustive sequential search over all possible subsets is computationally expensive, we need to employ heuristics to find a reasonably good subset that can sufficiently discriminate cancer. There are generally two common techniques: forward selection and backward selection [29]. In forward selection, we start

with an empty subset and add a gene (that increases the classification accuracy the most) in each iteration until any further addition of a gene does not increase the classification accuracy. In backward selection, we start with the full set of genes and remove a gene (that increases the classification accuracy the most) in each iteration until any further removal of a gene does not increase the classification accuracy. There are also other types of heuristics such as scatter search [33] and variable neighborhood search [34]. However, search-based gene selection techniques do not necessarily produce the best subset of the genes.

We employ a gene selection process based on an information-theoretic concept of entropy. Given a set of genes *X* and $p(x_i)$ which represents the probability of the $i^{th}$gene, then the entropy of genes, which measures the amount of 'uncertainty', is defined by:

$$H(X) = -\sum_{i=1}^{n} p(x_i)\, log\, p(x_i) \tag{1}$$

Entropy is a non-negative number. $H(X)$is 0 when X is absolutely certain to be predicted. The conditional entropy of class label *Y* given the genes is defined by:

$$H(Y \mid X) = \sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i,\, y_j)\, ln \frac{p(y_j)}{p(x_i,\, y_j)} \tag{2}$$

The information gain (IG) of the genes from the class label *Y* is defined to be:

$$IG(Y \mid X) = H(Y) - H(Y \mid X) \tag{3}$$

The gain ratio (GR) between the genes and the class label Y is defined to be:

$$GR(Y \mid X) = \frac{IG(Y \mid X)}{H(Y)} \tag{4}$$

The GR of a gene is a number between 0 and 1 which approximately represents the degree of 'significance' of the gene in discriminating cancer. A GR of 0 roughly indicates that the corresponding individual gene has no significance in cancer recognition while a GR of 1 roughly indicates that the gene is totally significant in cancer recognition. During the training phase, the GR for each gene is calculated according to (4). All the genes are then sorted by their GRs. Genes whose GRs are higher than a certain threshold value are selected as discriminating genes while the rest are discarded. Training needs to be carried out only once.

## 2.2. Numerosity reduction

Microarray gene expression heat map is essentially a matrix of gene expression levels. Each gene expression level is a continuous number. It has been demonstrated in a number of studies that many classification algorithms seem to work more effectively on discrete data or even morestrictly, on binary data[35]. Therefore, discretization is a desired step. Discretization is a process in which continuous gene expression levels are transformed into discrete representation which is comparable to linguistic expressions such as 'very low', low', 'high', and 'very high'. There are numerous discretization techniques in the literature [36]. However, we have adopted EMD(Entropy Minimization Discretization)[37] because of its reputation in discretization of

high-dimensional data. The training instances are first sorted in an ascending order. The EMD algorithm thenevaluates the midpoint betweeneach successive pair of the sorted values of an attribute as a potential cut point [38]. While evaluating each candidate cut point, the data arediscretized into two intervals and the resulting class informationentropy is calculated. A binary discretizationis determined by selecting the cut point for which theentropy is minimal amongst all candidates[35]. The binarydiscretization is applied recursively, always selecting thebest cut point. A minimum description length criterion(MDL) is applied to decide when to stop discretization[37]. The results of the discretization process are carried forward to the classification stage.

## 2.3. Classification

Naive Bayes (NB), which is fundamentally built on the strong independence assumption, is a very popular classifier in machine learning due to its simplicity, efficiency and efficacy [39-42]. There have been numerous applications of NB and variants thereof. The conventional NB algorithm uses the following formula for classification [43]:

$$Output = \underset{y}{\operatorname{argmax}}\big(P(y \mid x_1, \cdots, x_n)\big) \tag{1}$$

NB performs fairly accurate classification. The only limitation to its classification accuracy is the accuracy of the process of estimation of the base conditional probabilities. One clear drawback is its strong independence assumption which assumes that attributes are independent of each other in a dataset. In the field of genetic sequence classification, NB assumes that genes are independent of each other in a genetic sequence despite the fact that there are apparent dependencies among individual genes. Because of this fundamental limitation of NB, researchers have proposed various techniques such as one-dependence estimators (ODEs) [44] and super parent one-dependence estimators (SPODEs) [45] to ease the attribute independence assumption. In fact, these approaches alleviate the independence assumption at the expense of computational complexity and a new set of assumptions. Webb [39] proposed a semi-naive approach called averaged one-dependence estimators (AODEs) in order to weaken the attribute independence assumption by averaging all of a constrained class of classifiers without introduction of new assumptions. The AODE has been shown to outperform other Bayesian classifiers with substantially improved computational efficiency [39]. The AODE essentially achieves very high classification accuracy by averaging several semi-naive Bayes models that have slightly weaker independence assumptions than a pure NB. The AODE algorithm is effective, efficient and offers highly accurate classification. The AODE algorithm uses the following formula for classification [43]:

$$Output = \underset{y}{\operatorname{argmax}}\left(\sum_{i:\, 1 \le i \le n \,\wedge F(x_i) \ge m} P(y, x_i) \prod_{j=1}^{n} P\big(x_j \mid y,\, x_i\big)\right) \tag{2}$$

Semi-naive Bayesian classifiers attempt to preserve the numerous strengths of NB while reducing error by relaxing the attribute independence assumption [43]. Backwards sequential elimination (BSE) is a wrapper technique for attribute elimination that has proved to be effective at this task. Zheng et al. [43] proposed a new approach called *lazy estimation* (LE), which eliminated highly related attribute values at classification time without the computational overheads that are intrinsic in classic wrapper techniques. Their experimental results show that LE significantly reduces bias and error without excessive computational overheads. In the context of the AODE algorithm, LE has a significant advantage over BSE in both computational efficiency and error. This novel derivative of the AODE is called the averaged one-dependence estimators with

subsumption resolution (AODEsr). In essence, the AODEsr enhances the AODE with a subsumption resolution by detecting specializations among attribute values at classification time and by eliminating the generalization attribute value [43]. Because the AODEsr has a very weak independence assumption, it is very suitable for classification of gene expression data. Therefore, we employ an AODEsr classifier to recognize cancer from gene expression data.

## 3. EXPERIMENTS

We tested our proposed system using 11cancer datasets as listed in Table 1extracted from the biological literature[46]. Each dataset contains more than 60 samples with more than 2000 genes.

Table 1. 11 cancer datasets.

| Dataset | #Genes | #Samples |
|---|---|---|
| Colon Tumor | 2000 | 60 |
| Central Nervous System | 7129 | 60 |
| ALL-AML | 7129 | 72 |
| Breast Cancer | 24481 | 97 |
| Lung Cancer | 12533 | 181 |
| Ovarian Cancer | 15154 | 253 |
| ALL-AML-3 | 7129 | 72 |
| ALL-AML-4 | 7129 | 72 |
| Lymphoma | 4026 | 62 |
| MLL | 12582 | 72 |
| SRBCT | 2308 | 83 |

We carried out leave-one-outcross-validations (LOOCV) where an *N*-sized dataset was partitioned into *N* equal-sized sub-datasets. Out of the *N* sub-datasets, a single sub-dataset was retained as the validation data for testing the model, and the remaining *N*- 1 sub-datasets were used as training data. The whole cross-validation process was then repeated *N*- 1 more times such that each of the *N* sub-datasets got used exactly once as the validation data. The results were then averaged over all the *N* trials. We used a critical value of 1, frequency limit of 250, an M-estimate weight value of 0.03 for the AODEsr model for all the trails.

For each dataset, we performed one LOOCV experiment for varying number of selected genes ranging from 1 to 150. The genes for each trail were selected using the entropy-based technique outlined in Section 2.2. Figure 4 illustrates the results of our LOOCV experiments for each of the 11 datasets. The vertical axis represents the accuracy of the classifier in percentage while the horizontal axis represents the number of selected genes. Table 2 lists the same set of results in a tabular format for a certain number of selected genes. The results show that accuracy does increase with the number of selected genes, albeit without perfect monotonicity. Results also show that at certain instances (bolded and italicized in Table 2) accuracy decreases with an increase in the number of genes. This may not be because of the classifier because AODEsr, like other Bayesian classifiers, is not sensitive to irrelevant features. Therefore, adding an extra gene should not theoretically downgrade accuracy. The disruptions in monotonicity might be because of the intrinsic imperfection in the gene selection procedure. Because our proposed system is able to classify cancer accurately even with a very few genes, the results reinforce the clinical belief that cancers are initiated by glitchin a few genes. The maximum LOOCV accuracy of our cancer classifier is 100% for 7 out of 11 datasets. The average maximum LOOCV accuracy of our cancer classifier across all the 11 datasets is 98.94%. It is worth iterating the fact that we used the AODEsr classifier with the same set of parameters (critical value of 1, frequency limit of 250, an M-estimate weight value of 0.03) throughout all the experiments in order to avoid bias. To the best of our knowledge, the accuracy rate of the proposed cancer recognition system using the

AODEsr classifier with the entropy-based selection process seems to be higher than those of other cancer classification systems in the literature.

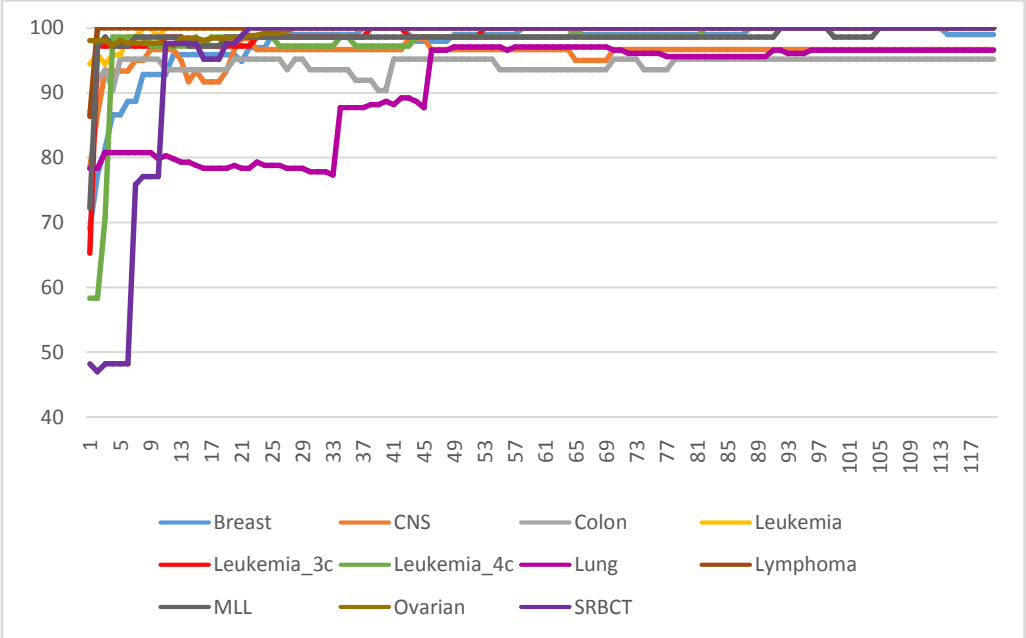


Figure 4. LOOCV accuracy (Y-axis) vs. number of genes (X-axis) [note: the plot maybe hard to read in monochrome print].

Table 2. LOOCV accuracy of the system on 11 datasets with varying number of selected genes.

| # of genes | Breast | CNS | Colon | Leuk. | Leuk_3c | Leuk_4c | Lung | Lymph | MLL | Ovarian |
|---|---|---|---|---|---|---|---|---|---|---|
| 5 | 86.6 | 93.3 | 95.2 | 95.8 | 97.2 | 98.6 | 80.8 | 100 | 97.2 | 98.0 |
| 10 | 92.8 | 96.7 | 95.2 | 100 | 97.2 | ***97.2*** | 80.8 | 100 | 98.6 | ***97.6*** |
| 25 | 99.0 | 96.7 | 95.2 | 100 | 100 | 98.6 | ***78.8*** | 100 | 98.6 | 99.2 |
| 50 | 99.0 | 96.7 | 95.2 | 100 | ***98.6*** | 98.6 | 97.0 | 100 | 98.6 | 100 |
| 75 | 99.0 | 96.7 | ***93.5*** | 100 | 100 | 98.6 | 96.1 | 100 | 98.6 | 100 |
| 110 | 100 | 96.7 | 95.2 | 100 | 100 | 100 | 96.6 | 100 | 100 | 100 |

## 4. CONCLUSION

Many people succumb to cancer every day. An early cancer detection and classification system is essential in order tosave countless lives.We have presented a machine learning based approach to recognize cancer from microarray gene expression data. We employ a state-of-the-art machine learning approach called the averaged-on dependence estimator with subsumption resolution (AODEsr) to tackle the problem of recognizing cancer. Given a set of gene expression data, the system predicts whether the gene expression data come from a cancerous cell or a non-cancerous cell. We have carried out experiments on 11 cancer datasets extracted from the biological literature. The proposed system has achieved an average maximum accuracy of 98.94%in recognizing cancer. The accuracy of the proposed system was found to be higher than those of other cancer classifiers in the literature. The experimental results demonstrate the efficacy of our framework. As future work, we would like to extend this framework to other applications such as cancer recurrence prediction and survivability prediction.

## ACKNOWLEDGEMENTS

This research was supported by Research Acculturation Collaborative Effort (RACE) Grant from the Ministry of Education, Malaysia.

## REFERENCES

[1] Worldwide cancer statistics. Retrieved from: http://www.cancerresearchuk.org/cancer-info/cancerstats/world/ Retrieved on: 15 November 2013.
[2] How many different types of cancer are there?Retrieved from: http://www.cancerresearchuk.org/cancer-help/about-cancer/cancer-questions/how-many-different-types-of-cancer-are-there Retrieved on: 15 November 2013.
[3] C. M. Croce, "Oncogenes and Cancer," New England Journal of Medicine, vol. 358, pp. 502-511, 2008.
[4] C. M. O'Connor and J. U. Adams, Essentials of Cell Biology. Cambridge: NPG Education, 2010.
[5] N. R. Colledge, et al., Davidson's Principles and Practice of Medicine, 21st ed.: Churchill Livingstone, 2010.
[6] M. M. R. Khondoker, "Statistical Methods for Preprocessing Microarray Gene Expression Data," Doctor of Philosophy, University of Edinburgh, 2006.
[7] ArrayIT. Retrieved from: http://shop.arrayit.com Retrieved on: 19 November 2013.
[8] Z. Z. Htike, "Multi-horizon ternary time series forecasting," in Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA), 2013, 2013, pp. 337-342.
[9] S. L. Win, et al., "Gene Expression Mining for Predicting Survivability of Patients in Early Stages of Lung Cancer," International Journal on Bioinformatics & Biosciences, vol. 4, 2014.
[10] Z. Z. Htike, "Can the future really be predicted?," in Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA), 2013, 2013, pp. 360-365.
[11] E.-E. M. Azhari, et al., "Brain Tumor Detection And Localization In Magnetic Resonance Imaging," International Journal of Information Technology Convergence and services, vol. 4, 2014.
[12] N. A. Mohamad, et al., "Bacteria Identification from Microscopic Morphology Using Naïve Bayes," International Journal of Computer Science, Engineering and Information Technology, vol. 4, 2014.
[13] E.-E. M. Azhari, et al., "Tumor Detection in Medical Imaging: A Survey," International journal of Advanced Information Technology, vol. 4, 2014.
[14] S. N. A. Hassan, et al., "Vision Based Entomology – How to Effectively Exploit Color and Shape Features," Computer Science & Engineering: An International Journal, vol. 4, 2014.
[15] N. A. Mohamad, et al., "Bacteria Identification from Microscopic Morphology: A Survey," International Journal on Soft Computing, Artificial Intelligence and Applications, vol. 3, 2014.
[16] S. N. A. Hassan, et al., "Vision Based Entomology: A Survey," International Journal of Computer science and engineering Survey, vol. 5, 2014.
[17] S. L. Win, et al., "Cancer Recurrence Prediction Using Machine Learning," International Journal of Computational Science and Information Technology, vol. 6, 2014.
[18] Z. Zainuddin and P. Ong, "Reliable multiclass cancer classification of microarray gene expression profiles using an improved wavelet neural network," Expert Systems with Applications, vol. 38, pp. 13711-13722, 2011.
[19] C.K. Chen, "The classification of cancer stage microarray data," Computer Methods and Programs in Biomedicine, vol. 108, pp. 1070-1077, 2012.
[20] A. Sharma and K. K. Paliwal, "Cancer classification by gradient LDA technique using microarray gene expression data," Data & Knowledge Engineering, vol. 66, pp. 338-347, 2008.
[21] S. Chakraborty, "Bayesian binary kernel probit model for microarray based cancer classification and gene selection," Computational Statistics & Data Analysis, vol. 53, pp. 4198-4209, 2009.
[22] J. M. Arevalillo and H. Navarro, "Exploring correlations in gene expression microarray data for maximum predictive–minimum redundancy biomarker selection and classification," Computers in Biology and Medicine, vol. 43, pp. 1437-1443, 2013.
[23] C. Bielza, et al., "Regularized logistic regression without a penalty term: An application to cancer classification with microarray data," Expert Systems with Applications, vol. 38, pp. 5110-5118, 2011.
[24] W. Engchuan and J. H. Chan, "Apriori Gene Set-based Microarray Analysis for Disease Classification Using Unlabeled Data," Procedia Computer Science, vol. 23, pp. 137-145, 2013.

[25] Z.J. Lee, "An integrated algorithm for gene selection and classification applied to microarray data of ovarian cancer," Artificial Intelligence in Medicine, vol. 42, pp. 81-93, 2008.
[26] R. Ruiz, et al., "Incremental wrapper-based gene selection from microarray data for cancer classification," Pattern Recognition, vol. 39, pp. 2383-2392, 2006.
[27] B. Sahu and D. Mishra, "A Novel Feature Selection Algorithm using Particle Swarm Optimization for Cancer Microarray Data," Procedia Engineering, vol. 38, pp. 27-31, 2012.
[28] H.S. Wong and H.Q. Wang, "Constructing the gene regulation-level representation of microarray data for cancer classification," Journal of Biomedical Informatics, vol. 41, pp. 95-105, 2008.
[29] E. Alpaydin, Introduction to Machine Learning, 2nd ed.: The MIT Press, 2010.
[30] C. M. Bishop, Pattern Recognition and Machine Learning: Springer, 2007.
[31] Z. Z. Htike and S. L. Win, "Recognition of Promoters in DNA Sequences Using Weightily Averaged One-dependence Estimators," Procedia Computer Science, vol. 23, pp. 60-67, 2013.
[32] Z. Z. Htike and S. L. Win, "Classification of Eukaryotic Splice-junction Genetic Sequences Using Averaged One-dependence Estimators with Subsumption Resolution," Procedia Computer Science, vol. 23, pp. 36-43, 2013.
[33] F. García López, et al., "Solving feature subset selection problem by a Parallel Scatter Search," European Journal of Operational Research, vol. 169, pp. 477-489, 2006.
[34] M. García-Torres, et al., "Solving Feature Subset Selection Problem by a Hybrid Metaheuristic," presented at the First International Workshop on Hybrid Metaheuristics, 2004.
[35] H. Liu and R. Setiono, "Feature selection via discretization," IEEE Transactions on knowledge and Data Engineering, vol. 9, pp. 642-645, 1997.
[36] V. Bolón-Canedo, et al., "A combination of discretization and filter methods for improving classification performance in KDD Cup 99 dataset," presented at the Proceedings of the 2009 international joint conference on Neural Networks, Atlanta, Georgia, USA, 2009.
[37] U. M. Fayyad and K. B. Irani, "Multi-Interval Discretization of Continuous-Valued Attributes for Classification Learning," in 13th International Joint Conference on Artificial Intelligence, Morgan Kaufmann, 1993, pp. pp. 1022-1029.
[38] S. Kotsiantis and D. Kanellopoulos, "Discretization techniques: A recent survey," GESTS International Transactions on Computer Science and Engineering, vol. 32, pp. 47-58, 2006.
[39] G. I. Webb, et al., "Not So Naive Bayes: Aggregating One-Dependence Estimators," Machine Learning, vol. 58, pp. 5-24, 2005.
[40] D. Hand and K. Yu, "Idiot's Bayes---Not So Stupid After All?," International Statistical Review, vol. 69, pp. 385-398, 2001.
[41] P. Domingos and M. Pazzani, "On the Optimality of the Simple Bayesian Classifier under Zero-One Loss," Mach. Learn., vol. 29, pp. 103-130, 1997.
[42] I. Rish, "An empirical study of the naive Bayes classifier," in IJCAI-01 workshop on "Empirical Methods in AI".
[43] F. Zheng and G. I. Webb, "Efficient lazy elimination for averaged one-dependence estimators," presented at the Proceedings of the 23rd international conference on Machine learning, Pittsburgh, Pennsylvania, 2006.
[44] M. Sahami, "Learning Limited Dependence Bayesian Classifiers," in Second International Conference on Knowledge Discovery and Data Mining, 1996, pp. 335-338.
[45] Y. Yang, K. Korb, K. Ting, and G. Webb, "Ensemble Selection for SuperParent-One-Dependence Estimators," in AI 2005: Advances in Artificial Intelligence. vol. 3809, S. Zhang and R. Jarvis, Eds., ed: Springer Berlin Heidelberg, 2005, pp. 102-112.
[46] Z. Zhu, et al., "Markov blanket-embedded genetic algorithm for gene selection," Pattern Recognition, vol. 40, pp. 3236-3248, 2007.