Classification of Cancer using DNA Microarray Data with Deep Belief Network Technique

Wannipa Saelim

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Science in Computer Science

Prince of Songkla University

2014

Thesis Title	Classification of Cancer using DNA Microarray Data with Deep Belief Network Technique
Author	Miss Wannipa Saelim
Major Program	Computer Science

Major Advisor :

……

(Asst. Prof. Dr. Pattara Aiyarak)

Co-advisor :

……

(Asst. Prof. Dr. Wiphada Wettayaprasit)

Examining Committee :

……..…….Chairperson

(Dr. Noppamas Pukkhem)

……...……

(Asst. Prof. Dr. Pattara Aiyarak)

……...……

(Asst. Prof. Dr. Wiphada Wettayaprasit)

……...……

(Asst. Prof. Dr. Ladda Preechaveerakul)

The Graduate School, Prince of Songkla University, has approved this thesis as partial fulfillment of the requirements for the Master of Science Degree in Computer Science

……..…..…..

(Assoc.Prof.Dr.Teerapol Srichana)

Dean of Graduate School

This is to certify that the work here submitted is the result of the candidate's own investigations. Due acknowledgement has been made of any assistance received.

……...…Signature

(Asst. Prof. Dr. Pattara Aiyarak)

Major Advisor

……..….Signature

(Miss Wannipa Saelim)

Candidate

I hereby certify that this work has not been accepted in substance for any degree, and is not being currently submitted in candidature for any degree.

……...…Signature

(Miss Wannipa Saelim)

Candidate

ชื่อวิทยานิพนธ์ การจำแนกประเภทข้อมูลโรคมะเร็ง จากข้อมูลไมโครอาร์เรย์ของดีเอ็นเอ ด้วยเทคนิค Deep Belief Network

ผู้เขียน นางสาววรรณิภา แซ่ลิ้ม

สาขาวิชา วิทยาการคอมพิวเตอร์

ปีการศึกษา 2556

บทคัดย่อ

เทคโนโลยีไมโครอาร์เรย์ เป็นหนึ่งในวิธีการศึกษาการแสดงออกของยีนที่สามารถศึกษาการแสดงออกของยีนแต่ละยีนที่มีจำนวนมากได้ในเวลาเดียวกัน ซึ่งในการวิเคราะห์การแสดงออกของยีน เพื่อการจำแนกประเภทข้อมูลของโรคมะเร็งสำหรับงานวิจัยนี้ ได้นำเทคนิคการคัดเลือกคุณลักษณะมาใช้ในการคัดเลือกยีนที่มีความเกี่ยวข้องกับเนื้อเยื่อมะเร็งแต่ละประเภท และทำการออกแบบและพัฒนาตัวจำแนกประเภทข้อมูลที่มีประสิทธิภาพ โดยการปรับปรุงเทคนิค Deep Belief Network ด้วยวิธีการ Quadratic Discriminant Analysis และใช้ชื่อว่า Quadratic Deep Belief Network จากนั้นทำการทดสอบประสิทธิภาพด้วยชุดข้อมูลโรคมะเร็ง 3 ชุด ดังนี้ ข้อมูลโรคมะเร็งเม็ดเลือดขาว โรคมะเร็งต่อมน้ำเหลือง และโรคมะเร็งต่อมลูกหมาก ซึ่งผลการทดลองแสดงให้เห็นว่า เทคนิคที่พัฒนาขึ้นมีประสิทธิภาพในการจำแนกประเภทโรคมะเร็งสูงกว่าเทคนิคอื่นที่นำมาเปรียบเทียบ ได้แก่ Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Naive Bayes (NB), Multilayer Perceptron (MLP) และ J48

Thesis Title	Classification of Cancer using DNA Microarray Data with Deep Belief Network Technique
Author	Miss Wannipa Saelim
Major Program	Computer Science
Academic Year	2013

ABSTRACT

Microarray technology is an advance tool that has been used in molecular biology and biomedicine. It allows researchers to measure the thousands of gene expression levels simultaneously in microarray experiment. Cancer microarray data normally contains a small number of samples that have a large number of features.

Efficient and reliable methods in cancer classification task that can find a small sample of informative genes amongst thousands are of great importance. In this field, much research is investigating the combination of advanced search strategies to find subset of features, and classification methods. This study designs and develops DNA microarray data analysis using Deep Belief Network along with Quadratic Discriminant Analysis called Quadratic Deep Belief Network (QDBN) for cancer classification. Three benchmark datasets were applied to evaluate the proposed method such as leukemia, lymphoma cancer and prostate cancer. The experimental results showed that the QDBN method outperforms the existing methods such as DBN, Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Naive Bayes (NB), Multilayer Perceptron (MLP) and J48 on three datasets.

TABLE OF CONTENTS

LIST OF TABLES

1. INTRODUCTION

Cancer classification task is one of the main researches in cancer diagnosis that requires an efficient tool for observing the expression of thousands of genes at the same time. Nowadays microarray technology is proposed to deal with the great number of gene expression simultaneously. Hence, several gene selection method and machine learning algorithms have been applied for selecting relevant genes and classifying tumor in normal and cancer cells, respectively. For example, the new cancer discovery and the class of leukemia cases of prediction were studied using Self Organizing Maps (SOM) (Golub *et. al.*, 1999). Hybrid method between Genetic Algorithm (GA) and K-Nearest Neighbor (KNN) were studied to used in the colon dataset (Li et. al., 2001). Moreover, deep architecture such as Deep Belief Network (DBN) have been used in many tasks and shown a good performance in various real-world applications. Recently, DBN was proposed in computational intelligence techniques in bioinformatics (Hassanien *et. al*., 2013). The step of DBN classification approach start with the direct training of the input data using Restricted Boltzmann Machine (RBM) to provide the visible layer with significant features for modeling in the hidden layers. Later, the hidden layers are trained until a defined number is reached using the mean activation function. This process is called learning features of features. Finally, fine-tuning parameter is performed using back-propagation technique.

The best option treatment for patients requires the high prognosis value together with diagnosis efficiency in cancer classification. Consequently, we developed an approach for cancer classification based on the use of DBN and improved the performance with Quadratic Discriminative Analysis (QDA), which will be called Quadratic Deep Belief Network (QDBN). The experimental results showed that the QDBN method obtained a good performance in cancer classification using DNA microarray data.

2. OBBJECTIVES OF THIS STUDY

The objectives of this study are as follows:

1. To propose the classification approach based on Deep Belief Network using cancer microarray data.
2. To improve Deep Belief Network model with Quadratic Discriminant Analysis
3. To compare the performance between the proposed method and the well-known classification techniques.

3. RESULT AND DISCUSSION

3.1 Testing DBN experiment.

We compared the classification performance of DBN with three classifiers, Support Vector Machine (SVM), K-nearest neighbor (KNN) and Naive Bayes (NB) on four microarray datasets.

The results of prediction accuracy are shown in Table 1. DBN reaches the best classification accuracy 98.61% on leukemia dataset, 99.45% on lung cancer dataset, 100% on lymphoma cancer dataset and 94.85% on prostate cancer dataset.

Table 1

The classification accuracy (%)

Method	DBN	SVM	KNN	NB
Leukemia	**98.61**	94.45	88.89	97.22
Lung Cancer	**99.45**	98.90	94.46	97.80
Lymphoma cancer	**100**	100	92.42	96.97
Prostate Cancer	**94.85**	91.18	81.62	55.15

3.2 Results of the improving DBN with QDA

After testing the DBN classifier on four microarray datasets, the traditional DBN with successful results were implemented on the uses of a full training set in evaluation process without feature selection methods. However, it is not surprising that the DBN results with full training provided the promising output because the training and testing datasets are identical. The k-fold cross validation was implemented to separate the training and testing datasets. We set k=10 in k-fold cross validation to evaluate classifier performance. Moreover, in order to reduce the noise genes, three feature selection method were used, namely Correlation-based feature selection (CFS), Information Gain (IG) and ReliefF. The classification accuracy decreased when compare with previous experiment. We solved this problem by using the capability of QDA to concentrate on the original DBN.

For an impartial comparison, we selected equal number of features as shown in Table 2.

Table 2

The number of selected genes

Feature Selection	**None of Feature selection**	**CFS**	**IG**	**ReliefF**
Leukemia	7129	81	81	81
Lymphoma Cancer	12600	193	193	193
Prostate Cancer	12600	75	75	75

Table 3, 4 and 5 illustrate different cancer type using three feature selection algorithms. The highest accuracy values with 10-fold cross validation are represented with numbers in bold.

Table 3

The classification accuracy of 10-fold cross validation on leukemia dataset (%)

Feature Selection	**QDBN**	**DBN**	**SVM**	**KNN**	**NB**	**MLP**	**J48**
CFS	**100**	98.61	98.61	98.61	**100**	97.22	84.72
IG	**100**	97.22	97.22	97.22	98.61	94.44	84.72
ReliefF	**98.61**	97.22	**98.61**	97.22	97.22	97.22	83.33

Table 4

The classification accuracy of 10-fold cross validation on lymphoma cancer dataset(%)

Feature Selection	**QDBN**	**DBN**	**SVM**	**KNN**	**NB**	**MLP**	**J48**
CFS	**100**	**100**	**100**	**100**	**100**	**100**	98.48
IG	**100**	98.48	**100**	98.48	98.48	98.48	95.45
ReliefF	**100**	96.96	98.48	98.48	98.48	98.48	93.94

Table 5

The classification accuracy of 10-fold cross validation on prostate cancer dataset (%)

Feature Selection	**QDBN**	**DBN**	**SVM**	**KNN**	**NB**	**MLP**	**J48**
CFS	**92.65**	86.76	86.76	88.97	61.76	91.18	88.24
IG	**93.38**	87.5	91.18	89.71	58.82	90.44	86.03
ReliefF	**93.38**	86.03	88.97	88.97	62.5	91.91	86.76

The prediction accuracy on leukemia dataset are shown. In Table 3, the QDBN with feature selection CFS, IG and ReliefF provide the highest accuracies of 100%, 100% and 98.61%, respectively. Next, Table 4 shows the QDBN accuracy on lymphoma dataset that is 100% all cases combined with feature selection method. Finally, The classification accuracy of the QDBN in Table 5 shows that it outperforms DBN, SVM, KNN, NB, MLP and J48 on prostate cancer dataset with 92.65%, 93.38% and 93.38% when selected by CFS, IG and ReliefF, respectively.

4. CONCLUSION

In this thesis, we proposed a comparative study of three feature selection methods and evaluate their performance using the DBN and the other methods. The experiments have been implemented on the three benchmark datasets, i.e. leukemia, lymphoma cancer and prostate cancer. Our study results demonstrated that the k-fold cross validation impacted the performance of DBN directly. In order to anticipate this problem, we used the advantages of QDA to enhance the original DBN. Hence, the classification performance is increased.

BIBLIOGRAPHY

Adjouadi, M., M. Ayala, M. Cabrerizo, N. Zong, G. Lizarraga and M. Rossman. 2010. Classification of leukemia blood samples using neural networks. *Annals of Biomedical Engineering*, *38*: 1473-1482.

Chen, E., X. Yang, H. Zha, R. Zhang and W. Zhang. 2008. Learning object classes from image thumbnails through deep neural networks. *Proceeding of IEEE International Conference on Acoustics Speech and Signal Processing*: 829-832.

Cho, S. B., and H. H. Won. 2003. Machine learning in DNA microarray analysis for cancer classification. *Proc*eeding of *First Asia-Pacific bioinformatics conference on Bioinformatics,* 19: 189-198.

Golub, T. R., D. K. Slonim, P. Tamayo, C. Huard, M. Gassenbeek, J. P. Mesirov, H. Coller, M. L. Loh, J. R. Downing, M. A. Caligiuri, C. D. Bloomfield and E. S. Lander. 1999. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science*, *286*: 531-537.

Hinton, G. 2002. Training products of experts by minimizing contrastive divergence. *Neural Computation*, *14*: 1771-1800.

Hinton, G., S. Osindero and Y. W. Teh. 2006. A fast learning algorithm for deep belief nets. *Neural Computation*, *18*: 1527-1554.

Hong, J.H. and S.B. Cho. 2006. The classification of cancer based on DNA microarray data that uses diverse ensemble genetic programming. *International Journal of Artificial Intelligence in Medicine*, *36*: 43-58.

Horster, E. and R. Lienhart. 2008. Deep networks for image retrieval on large-scale databases. *Proceeding* of *16th ACM International Conference on Multimedia*, 643-646.

Igor, K. 1994. Estimating attributes: analysis and extensions of RELIEF. Proceedings of the European conference on machine learning on Machine Learning, 171-182.

Jaitly, N. and G. Hinton. 2011. Learning a better representation of speech sound waves using restricted Boltzmann machines. *Proceeding of 11th IEEE International Conference on Acoustics Speech and Signal Processing*, 5884-5887.

Li, Jinyan. 2012. Kent Ridge Bio-medical Data Set Repository. http://levis.tongji.edu.cn/gzli/data/mirror-kentridge.html. (accessed June, 30 2012).

Larochelle, H., D. Erhan, A. Courville, J. Bergstra and Y. Bengio. 2007. An empirical evaluation of deep architectures on problems with many factors of variation. *Proc*eeding of *24th International Conf*erence *on Machine Learning*, 473-480.

Larochelle, H., M. Mandel, R. Pascanu and Y. Bengio. 2012. Learning algorithms for the classification restricted Boltzmann machine. *Journal of Machine Learning Research*, 13: 643-669.

Lee, H., R. Grosse, R. Ranganath and A. Y. Ng. 2011. Unsupervised Learning of Hierarchical Representations with Convolutional Deep Belief Networks. *Communications of the ACM*., 54: 95-103.

Li, M., B. Wang, Z. Momeni, and F. Valafar. 2002. Pattern recognition techniques in microarray data analysis. *Annals of the New York Academy of Sciences,* 980: 41-64.

Li, X., T. A. Darden, C.R. Weinberg, A. J. Levine and L. G. Pedersen. 2001. Gene assessment and sample classification for gene expression data using a genetic algorithm/k-nearest neighbor method. *Comb. Chem. High Throughput Screen*, 4: 727-739

Li, X., S. Peng, X. Zhan, J. Zhang and Y. Xu. 2011. Comparison of feature selection methods for multiclass cancer classification based on microarray data. *4th International Conference on Biomedical Engineering and Informatics*, 1692-1696.

Liu, Y., S. Zhou and Q. Chen. 2011. Discriminative deep belief networks for visual data classification. *Pattern Recognition*, *44*: 2287-2296.

Mark, A. H. 2000. Correlation-based Feature Selection for Discrete and Numeric Class Machine Learning. *Proceeding of 17th International Conference on Machine Learning,* 359-366.

Mobahi, H., R. Collobert and J. Weston. 2009. Deep learning from temporal coherence in video. *Proceeding of International Conference on Machine Learning*, 737-744.

Nguyen, D. V., and D. M. Rocke. 2002. Tumor classification by partial least squares using microarray gene expression data. *Bioinformatics*, 18: 39-50.

Ranzato, M. A., C. Poultney, S. Chopra and Y. Lecun. 2006. Efficient learning of sparse representations with an energy-based model. *Advances in Neural Information Processing Systems*, 19: 1137-1144.

Ranzato, M. A., J. Susskind, V. Mnih and G. Hinton. 2011. On deep generative models with applications to recognition. *Proceeding of 24th IEEE Conference on Computer Vision and Pattern Recognition*, 2857-2864.

Saelim, W., W. Wettayaprasit and P. Aiyarak. 2013. Leukemia classification using deep belief network. *Proceeding of the 12th IASTED International Conference Artificial Intelligence and Applications*, 59-63.

Saeys, Y., I. Inza and P. Larrañaga. 2007. A review of feature selection techniques in bioinformatics. *Bioinformatics*, 23: 2507-2517.

Salakhutdinov, R., and G. Hinton. 2008. Using deep belief nets to learn covariance kernels for Gaussian processes. *Advance in Neural Information Processing Systems*, 20: 1249-1256.

Salakhutdinov, R., and I. Murray. 2008. On the quantitative analysis of deep belief networks. *Proceeding of the 25th International Conf. on Machine Learning*, 872-879.

Salama, M. A., A. E. Hassanien and A. A. Fahmy. 2010. Deep belief network for clustering and classification of a continuous data. *Proceeding of the 10th IEEE International Symposium on Signal Processing and Information Technology*, 473-477.

Smolensky, P. 1986. Information Processing in dynamical systems: Foundations of harmony theory. *Parallel distributed processing: explorations in the microstructure of cognition,* 1: 194-281.

Tang, K. L., W. J. Yao, T. H. Li, Y. X. Li, and Z. W. Cao. 2010. Cancer classification from the gene expression profiles by discriminant kernel-PLS, *Journal of Bioinformatics and Computational Biology*, 8: 147-160.

Teck, D., C. Lai and J. M. Garibaldi. 2013. Improving Semi-supervised Fuzzy C-Means Classification of Breast Cancer Data using Feature Selection. *Proceeding of the 2013 IEEE International Conference on Fuzzy Systems*, 1-8.

Thomas, J. G., J. M. Olson, S. J. Tapscott and L. P. Zhao. 2001. An efficient and robust statistical modeling approach to discover differentially expressed genes using genomic expression profiles. *Genome Res*, 11: 1227-1236.

Zhou, X., and D. P. Tuck. 2007. MSVM-RFE: extensions of SVM-RFE for multiclass gene selection on DNA microarray data. *Bioinformatics*, 23: 1106-1114.

APPENDIX A

SUBMITTED MANUSCRIPT

Title	Quadratic Deep Belief Network of Cancer Classification using Feature Selection
Journal	Artificial Intelligence in Medicine

QUADRATIC DEEP BELIEF NETWORK OF CANCER CLASSIFICATION USING FEATURE SELECTION

Wannipa Saelim, Wiphada Wettayaprasit, Pattara Aiyarak*

Department of Computer Science, Faculty of Science, Prince of Songkla University, Hat Yai, Songkhla, 90110, Thailand

* Corresponding author, e-mail: pattara@aiyarak.net

ABSTRACT

In recent years, cancer classification based on microarray data is a well-known hard problem in the field of computational intelligence techniques. A challenging task of microarray analysis is to select a reasonable number of the most relevant gene. The objective of this paper is to investigate the combination of feature selection and classification methods. We propose a novel classifier based on Deep Belief Network (DBN) called Quadratic Deep Belief Network (QDBN). QDBN applies a novel deep architecture to combine the general capability of DBN with quadratic discriminant. The experimental results showed that the QDBN method outperforms the existing methods such as DBN, k-nearest neighbor (KNN), Naive Bayes (NB) and support vector machine (SVM), in terms of the prediction accuracy on leukemia dataset. QDBN, KNN and SVM reach the same highest accuracy on lymphoma dataset but KNN provides the best performance on prostate cancer dataset.

KEYWORDS: quadratic deep belief network, restricted Boltzmann machine, microarray, feature selection, cancer classification

1. Introduction

The DNA microarray technology has been used with valuable success in molecular biology and biomedicine for diagnosis and prognosis in order to treat or considerably prolong the life of patients and to ensure the best possible quality of life to cancer survivors. Microarray data for analysis contains the thousands of genes as features, many of which may

be irrelevant or insignificant to the cancer classification task in clinical diagnosis [1]. Hence, several machine learning algorithms have been applied for microarray data to select the important genes and classifying tumor using microarray data, as we have known in feature or gene selection methods and cancer classification techniques. Cancer classification is expected to obtain the diagnosis efficiency and the high prognosis value in patients for providing the best option treatment. For example, Golub *et. al.* [2] proposed self-organising maps (SOM) approach for discovering the new cancer classes and predicting the class of leukemia cases. Li *et. al.* [3] classified the colon dataset with a hybrid method, Genetic Algorithm (GA) and KNN were used to analyze the microarray data.

Feature selection is an important technology not only in terms of removing the noise genes from the great number of genes and reducing the computational time but also in terms of maintaining the reliable semantic of variables [4]. Moreover, the classification performance could be improved by relevant genes [5]. One simple method to select genes is *t*-test that presented by Thomas *et. al.* [6]. A Correlation-based Feature Selection (CFS) which is a filter approach has been shown in Hall [7] that compared with wrapper method and found the execution time faster. In addition, Teck *et.al.* [8] proposed the use of feature selection in combination with semi-supervised Fuzzy C-Means classification of breast cancer data.

Deep Belief Network (DBN) is a deep neural network that was first described in Smolensky (1986) [9]; at that time, this structure was called harmonium. The feature of the harmonium is similar to the classical Boltzmann Machine but in fact, the DBN architecture is constructed without connections among the visible or hidden units. Due to the difference cited, when the researchers explained in their works, Hinton et al. renamed the harmonium of DBN to a Restricted Boltzmann Machine (RBM) in 2006 [10]. In case of using class labels with feature vectors (supervised training), DBN could be used for classification. Whereas, when not using class labels and back-propagation in the DBN architecture (unsupervised training), DBN is exploited as a feature extraction in dimensional reduction task.

In our previous work [11], the traditional DBN with successful results on leukemia dataset were implemented on the same dataset of training set and test set. To avoid using the same sets for training and testing. the k-fold cross validation was implemented. In this study, 10-fold cross validation was performed in three microarray datasets and the classification accuracy decreased significantly when compared with previous work. To solve this problem, we used capability of QDA to concentrate on the original DBN.

The objective of this paper is to use the combination of feature selection and QDBN in the classification of continuous datasets like microarray data such as leukemia, lymphoma and

prostate cancers. We also compare QDBN to the well known classifiers of DBN, KNN, NB and SVM.

2. Feature Selection

2.1 Correlation based Feature Selection (CFS)

CFS [7] is an algorithm for selecting a subset of attributes that have a high relevant with the class and irrelevant with each other by using a heuristic search strategy. The following equation gives the merit of a feature subset $\boldsymbol{S}$ containing $\boldsymbol{k}$ features.

$$\boldsymbol{M_s = \frac{k\bar{r}_{cf}}{\sqrt{k + k(k-1)\bar{r}_{ff}}}}, \quad (\mathbf{1})$$

where $\bar{\boldsymbol{r}}_{\boldsymbol{cf}}$ is the average feature to class correlations and $\bar{\boldsymbol{r}}_{\boldsymbol{ff}}$ is the average feature to feature correlations.

2.2 Information Gain (IG)

The IG [13] is a measure based on entropy. The information gain value of each gene is computed by

$$\boldsymbol{InfoGain = H(Y) - H(Y|X)} \quad (\mathbf{2})$$

where $\boldsymbol{X}$ and $\boldsymbol{Y}$ are features, and

$$\boldsymbol{H(Y) = -\sum_{y \in Y} p(y) log_2(p(y))}, \quad (\mathbf{3})$$

$$\boldsymbol{H(Y|X) = -\sum_{x \in X} p(x) \sum_{y \in Y} p(y|x) log_2(p(y|x)) = -\sum_{y,x} p(y,x) log_2\left(\frac{p(y,x)}{p(x)}\right)}, \quad (\mathbf{4})$$

2.3 ReliefF

A nearest neighbor method is used to calculate relevant score for each feature. The ReliefF algorithm is defined the feature in the same class with the group of the same feature values and different classes with the different feature values [14].

3. Quadratic Deep Belief Network for Cancer Classification

3.1 Problem Formulation

We have reviewed many documents about DBN in classification tasks that were developed the visible and hidden layers using the units of binary data. In case of microarray data, they are the continuous datasets that obtained from image processing so the architecture of the proposed method is constructed by the pattern of continuous data. We applied data preprocessing to three datasets by replacing the missing values using mean along the axis. After that, we selected the significant genes with three feature selection methods as described in feature selection section. The performance of classification is evaluated by cross-validation and we improved the DBN with quadratic discriminant analysis (QDA) and we were found that classification accuracy increased using QDBN.

3.2 Deep Belief Network

DBN structure is constructed from the multiple stacks of Restricted Boltzmann Machine (RBM) that each layer consists of a set of binary or real-valued units [12]. The RBM per se is limited to what it represents. Its real power appears when RBMs are stacked to form a DBN - a generative model consisting of many layers.

The process of DBN classification approach is as follows; first, RBM is trained directly on the input data as its visible layer for modeling it in the hidden layer by capturing the significant features that are a representation of the input data. Then the mean activations of the trained features are used as input data for training the second RBM. This learning process called learning features of features which is continued until a defined number of hidden layers have been trained. Finally, all the parameters with back-propagation are fine-tuned.

3.3 Quadratic Deep Belief Network

Quadratic Discriminant Analysis (QDA) is a common method of classification task that similar to Linear Discriminant Analysis (LDA), The main difference between the QDA and LDA is their assumptions. LDA assumes the identification of the covariance of each class whereas the QDA assumes the normal distribution. QDA is more flexible for the covariance matrix, leads to fit the data better than LDA.

The likelihood ratio test is the best possible test for QDA. Suppose the $y \in \{0,1\}$, likelihood ratio is shown in equation (5)

$$\boldsymbol{g(x) = (x - \mu_{y=1})' \sum_{y=1}^{-1} (x - \mu_{y=1}) - (x - \mu_{y=0})' \sum_{y=0}^{-1} (x - \mu_{y=0}) - ln \left[\frac{|\sum_{y=0}|}{|\sum_{y=1}|} \right]} \quad \textbf{(5)}$$

where $\mu_{y=0}$, $\mu_{y=1}$, $\sum_{y=0}$, $\sum_{y=1}$. are the means of each class and the covariance, respectively. And

$$\boldsymbol{y_i = \begin{cases} 0 & if \quad g(x) > 0 \\ 1 & if \quad g(x) < 0 \end{cases}} \quad \textbf{(6)}$$

To train the QDBN, we have used the training procedure as described in Algorithm 1.

Algorithm 1. Algorithm of quadratic deep belief network

Algorithm 1. QDBN Classifier

Input: data x, y

- Training hidden layers with Greedy-layer wise using RBM
- Calculate

$$p(h_{j,s} = 1|v) = \text{sigm}\left(\sum_i W_{ij} v_{i,s} + c_j \right)$$

$$p(v_{i,s} = 1|h) = \text{sigm}\left(\sum_j W_{ij} h_{j,s} + b_i \right)$$

- Supervised learning the DBN with QDA, update the weight and bias

Output: Model of QDBN

4. Experiments

4.1 Experimental Setting

In preprocessing step, we used ReplaceMissingValues in WEKA for replacing the blank values by using mean along the axis. The QDBN algorithm was implemented in MATLAB while the feature selection methods and KNN, SVM and NB classifier were based on the WEKA platform. The k of KNN was set to be 1, polynomial kernel function with filterType was used in SVM, and NB was run with default settings. The 10-fold cross validation was used for measuring the performance of different feature selection methods and classifiers.

The three benchmark datasets were applied to evaluate the proposed method. Table 1. describe the characteristics of the three datasets.

Table 1 The Characteristics of the Three Datasets

Dataset	No. of probe	No. of samples	No. of classes
Leukemia	7129	72	2
Lymphoma Cancer	4026	66	3
Prostate Cancer	12600	136	2

Leukemia dataset: consists of 72 samples from two different types of leukemia in which 47 samples are acute lymphoblastic leukemia (ALL) and 25 samples are acute myeloid leukemia (AML). There are 38 bone marrow samples in training set (27 ALL and 11 AML). The testing set consists of 24 bone marrow and 10 peripheral blood samples (20 ALL and 14 AML). Each sample contains 7129 genes expression levels.

Lymphoma Cancer: consists of 66 samples with 4026 different gene expression levels and three different types of lymphoma. The 66 samples containing 46 B-cell lymphoma (DLBCL), 9 follicular lymphoma (FL) and 11 chronic lymphoma leukemia (CLL).

Prostate Cancer: consists of 136 samples with 12600 gene expression levels, the 136 samples containing 77 prostate tumor (PTS) and 59 normal prostate samples (NPS). There are 102 samples in training set (52 PTS and 50 NPS) and 34 samples in testing set (25 PTS and 9 NPS).

4.2 Experimental Results

Table 2 - 4 show the results of three datasets that evaluated classification performance with 10-fold cross validation and using three feature selection algorithms to selected genes. The bold numbers represent the highest accuracy of each feature selection method.

To summarize, the prediction accuracy on leukemia dataset are shown in Table 3. The QDBN reaches the highest accuracies of 100%, 100% and 98.61% when the features of data have been selected by CFS, IG and ReliefF, respectively. Next, Table 4. show the QDBN accuracy that is 100% all cases combined with feature selection method. Finally, The classification accuracy of the QDBN in Table 5. show that it outperforms DBN, SVM, KNN, NB, MLP and J48 on prostate cancer dataset with 92.65%, 93.38% and 93.38% when selected by CFS, IG and ReliefF, respectively.

Table 2 The classification accuracy of 10-fold cross validation on leukemia dataset (%)

Feature Selection	**QDBN**	**DBN**	**SVM**	**KNN**	**NB**	**MLP**	**J48**
CFS	**100**	98.61	98.61	98.61	**100**	97.22	84.72
IG	**100**	97.22	97.22	97.22	98.61	94.44	84.72
ReliefF	**98.61**	97.22	**98.61**	97.22	97.22	97.22	83.33

Table 3 The classification accuracy of 10-fold cross validation on lymphoma cancer dataset (%)

Feature Selection	**QDBN**	**DBN**	**SVM**	**KNN**	**NB**	**MLP**	**J48**
CFS	**100**	**100**	**100**	**100**	**100**	**100**	98.48
IG	**100**	98.48	**100**	98.48	98.48	98.48	95.45
ReliefF	**100**	96.96	98.48	98.48	98.48	98.48	93.94

Table 4 The classification accuracy of 10-fold cross validation on prostate cancer dataset (%)

Feature Selection	**QDBN**	**DBN**	**SVM**	**KNN**	**NB**	**MLP**	**J48**
CFS	**92.65**	86.76	86.76	88.97	61.76	91.18	88.24
IG	**93.38**	87.5	91.18	89.71	58.82	90.44	86.03
ReliefF	**93.38**	86.03	88.97	88.97	62.5	91.91	86.76

5. CONCLUSION

In this paper, we presented a novel classifier based on deep belief network, called Quadratic Deep Belief Network that combined with feature selection method such as CFS, IG and ReliefF in order to compare the performance between proposed method and the traditional neural networks. Leukemia, lymphoma cancer and prostate cancer dataset is used in this experimental. The 10-fold cross validation was used for evaluation all of three datasets.

Acknowledgements

The authors would like to thank the Research and Development Office (RDO) of Prince of Songkla University (PSU), Thailand for their grant support. The first author would also like to express her special appreciations to the Graduates School of PSU for the graduate student research grant. Special thanks go to Dr. Pitipol Meemak of the Department of Molecular Biotechnology and Bioinformatics, Faculty of Science, PSU for his initiation in this research.

REFERENCES

[1] X. Zhou & D. P. Tuck, MSVM-RFE: extensions of SVM-RFE for multiclass gene selection on DNA microarray data, *Bioinformatics*, *23*, 2007, 1106-1114.

[2] T. R. Golub, D. K. Slonim, P. Tamayo, C. Huard, M. Downing, M. A. Caligiuri, C. D. Bloomfield & E. S. Lander, Molecular classification of cancer: class discovery and class prediction by gene expression monitoring, *Science, 286(5439),* 1999, 531-537.

[3] L. Li, T. A. Darden, C.R. Weinberg, A. J. Levine & L. G. Pedersen, Gene assessment and sample classification for gene expression data using a genetic algorithm/k-nearest neighbor method, *Comb. Chem. High Throughput Screen, 4*, 2001, 727-739.

[4] X. Li, S. Peng, X. Zhan, J. Zhang & Y. Xu, Comparison of feature selection methods for multiclass cancer classification based on microarray data, *Proc. 4th International Conf. on Biomedical Engineering and Informatics*, Shanghai, China, 2011, 1692-1696.

[5] Y. Saeys, I. Inza & P. Larrañaga, A review of feature selection techniques in bioinformatics, *Bioinformatics, 23*, 2007, 2507-2517.

[6] J. G. Thomas, J. M. Olson, S. J. Tapscott & L. P. Zhao, An efficient and robust statistical modeling approach to discover differentially expressed genes using genomic expression profiles, *Genome Res.*, *11*, 2011, 1227-1236.

[7] A. H. Mark, Correlation-based Feature Selection for Discrete and Numeric Class Machine Learning, *Proc. 17th International Conf. on Machine Learning,* (Morgan Kaufmann Publishers Inc., 2000).

[8] D. Teck. C. Lai & J. M. Garibaldi, Improving Semi-supervised Fuzzy C-Means Classification of Breast Cancer Data using Feature Selection, *Proc. 2013 IEEE International Conference on Fuzzy Systems,* Hyderabad, India, 2013, 1-8.

[9] P. Smolensky, Information Processing in dynamical systems: Foundations of harmony theory, *Parallel distributed processing: explorations in the microstructure of cognition, 1*, 1986, 194-281.

[10] E. Hinton, S. Osindero & Y. W. Teh, A Fast Learning Algorithm for Deep Belief Nets, *Neural Computational*, *18,* 2006, 1527-1554.

[11] W. Sae-Lim, W. Wettayaprasit & P. Aiyarak, Leukemia classification using deep belief network, *Proc. 12th IASTED International Conf. Artificial Intelligence and Applications*, Innsbruck, Austria, 2013, 59-63.

[12] H. Lee, R. Grosse, R. Ranganath & A. Y. Ng, Unsupervised Learning of Hierarchical Representations with Convolutional Deep Belief Networks, *Communications of the ACM.*, *54*, 2011, 95-103.

[13] Y. Wang, I. V. Tetko, M. A. Hall, E. Frank, A. Facius, K. F. Mayer, et al., Gene selection from microarray data for cancer classification --a machine learning approach, *Comput Biol Chem.*, *29*, 2005, 37-46.

[14] K. Igor, Estimating attributes: analysis and extensions of RELIEF, Proc. European Conf. on Machine Learning, 1994, 171-182.

APPENDIX B

PUBLICATION OF FULL PROCEEDING

Title	Leukemia Classification using Deep Belief Network
Conference	The 12th IASTED International Conference on Artificial Intelligence and Applications (AIA 2013)
Place	Innsbruck, Austria
Date	11-13 February 2013

Proceedings of the IASTED International Conference
Artificial Intelligence and Applications (AIA 2013)
February 11 - 13, 2013 Innsbruck, Austria

LEUKEMIA CLASSIFICATION USING DEEP BELIEF NETWORK

Wannipa Sae-Lim, Wiphada Wettayaprasit, Pattara Aiyarak*
Artificial Intelligence Research Laboratory, Department of Computer Science, Faculty of Science
Prince of Songkla University
Songkhla, Thailand
giftzn@hotmail.com, wwettayaprasit@yahoo.com, pattara@aiyarak.net

ABSTRACT
This paper proposes a novel approach for leukemia classification based on the use of Deep Belief Network (DBN). DBN is a feedforward neural network with a deep architecture that consists of a stack of restricted Boltzmann machine (RBM). The study used the benchmark DNA microarray of leukemia data from Kent Ridge Bio-medical Data Set Repository. The classification performance was compared between the proposed method and the traditional neural networks. In conclusion, the DBN outperforms the state-of-the-art learning models such as support vector machine (SVM), k-nearest neighbor (KNN) and Naive Bayes (NB).

KEY WORDS
Deep belief network, restricted Boltzmann machine, neural networks, microarray data, leukemia classification

1. Introduction

One important method to deal with the gene expression microarray data is the classification of the type of tumor [1]. Cancer classification is a main research area in the medical field. Such classification is an essential step in prediction and diagnosis of diseases [2]. Acute leukemia is a type of cancer of the bone marrow characterized by an abnormal increase of immature white blood cells that cannot fight infection. When lymphoid cells are affected, the disease is called acute lymphoblastic leukemia (ALL); however, it is called acute myeloid leukemia (AML), when myeloid cells are affected [3]. The classification of leukemia is a way to differentiate cancer tissues such as ALL and AML for using in medical diagnosis and hence a suitable treatment. The classification accuracy is essentially very important point for cancer treatment. Many machine- learning techniques have been developed to reach highly accurate classification performance when classifying DNA microarray data.

DNA microarray technology is an advanced tool used in molecular biology and biomedicine for measuring the large-scale gene expression profiles under varying biologic tissues and for producing a great amount of genetic data. DNA microarray consists of an arrayed series of spots of DNA, called features [4]. The analysis and understanding of microarray data includes a search for genes that have similar or correlated patterns of expression. Among the various frameworks in which pattern recognition has been traditionally formulated, statistical approach, neural network techniques and methods imported from statistical learning theory are among those that have been applied in microarray data analysis [5]. This technology has been widely used in many fields such as drug screening, agriculture as well as the clinical diagnosis of human diseases. To address the classification of cancer problem, we use Deep Belief Network (DBN) that is one technique in machine learning.

Several machine learning techniques have been previously used in classifying gene expression data including k nearest neighbor, decision tree, multilayer perceptron, support vector machine, booting and self-organizing map [6]. Moreover, deep architecture such as DBN have used in many tasks and shown good performance in various real-world applications. For example, Larochelle *et. al.* presented the experiments which indicate that deep architecture can solve the learning problems with many factors of variation and outperform SVMs and single hidden layer feedforward neural networks [7]. In addition, Salakhutdinov and Hinton showed how to use unlabeled data and a DBN to learn a good covariance kernel for a Gaussian process [8]. Horster and Lienhart also investigated deep network for image retrieval on large-scale databases [9]. Mobahi *et. al.* proposed a learning method for deep architectures that took advantage of sequential data [10]. Furthermore, Salama *et. al.* applied the DBN with continuous data for clustering and classification [11]. Currently, DBN has been applied to different applications, such as audio classification [12], object recognition [13] and many visual data analysis tasks [14,15].

The discovery to efficient training procedure for deep learning came in 2006 with the algorithm for training deep belief network by Hinton *et. al*. [16] and stacked auto-encoders, which are based on a similar method, greedy layer-wise unsupervised pre-training followed by supervised fine-tuning by Ranzato *et. al*. in 2007 [17]. DBN is a probabilistic generative models that contain many layers of hidden layers including a greedy layer-wise learning algorithm. Each layer is pre-trained with an unsupervised learning algorithm, learning a nonlinear transformation of its input that captures the interesting feature as the input data for the next higher layer. The main building block of a DBN is a bipartite undirected graphical model called a restricted Boltzmann Machine (RBM) [18]. An RBM is undirected generative models that use a layer of hidden variables to model a distribution over visible variables. It has a single layer of hidden units, which are not connected to each other, and have undirected symmetrical connections to a layer of visible units. This RBM is called Harmonium RBM [19].

* To whom correspondence should be addressed.

DOI: 10.2316/P.2013.793-043

This study used the benchmark DNA microarray of leukemia data from Kent Ridge Bio-medical Data Set Repository [20] which was first adopted into the classification of leukemia by Golub *et. al*. in 1999 [21].

In this paper, we focus on a machine learning approach for identifying cancer types based on DBN and evaluated its performance on DNA microarray data of leukemia. We also compare DBN to the well known classifier of SVM, KNN and NB. The experimental results indicate that DBN provides better performance than SVM, KNN and NB in leukemia classification.

This paper is organized as follows. After introducing the concept in Section 1, the backgrounds of RBM and DBN are shown in Section 2. Section 3 describes the DBN approach for classification of leukemia with training data algorithm. Section 4 presents the experimental setting and experimental results along with the corresponding discussions. The conclusions and the direction for future study of this paper are addressed in Section 5.

2. Deep Belief Network for Leukemia Classification

2.1 Generative vs. Discriminative models

Two main types of probabilistic models are generative and discriminative models. The difference between the two models is the probability distribution. Basically the goal of training is to find the conditional distribution $p(y|x)$ to predict some output y given the value of an input x. Discriminative models (such as traditional feedforward neural networks) generate the probability of an output given an input that the result can be used to make predictions of y for a new value of x. On the other hand, the generative models (such as DBN) generate the joint probability distribution of the input and the output. The generative model estimate $p(x,y)$ that is possible to obtain either $p(y|x)$ or $p(x|y)$ using Bayes' theorem. Although the classification error rate of generative model is generally greater than discriminative model, the last and only supervised fine-tuning phase in DBN training is sufficient update the weights to minimize the appropriate loss function directly.

2.2 Restricted Boltzmann Machine

A Restricted Boltzmann Machine (RBM) is the core component of the DBN with no connection among the same layer. An RBM is an energy-based undirected generative model [13] that has a two-layer architecture in which the visible binary stochastic units $v \in \{0,1\}^D$ are connected to hidden binary stochastic units $h \in \{0,1\}^K$ using symmetrically weighted connections as shown in Fig. 1.

The energy of the state $\{v,h\}$ is defined as:

$$E(v,h;\,\theta) = -\sum_{i=1}^{D}\sum_{j=1}^{K} v_i W_{ij} h_j - \sum_{i=1}^{D} b_i v_i - \sum_{j=1}^{K} c_j h_j \qquad (1)$$

where $v \in V$ is the observation nodes, $h \in H$ is the hidden random variables and $\theta = \{W, b, c\}$ are the model parameters: W_{ij} is the symmetric interaction term between unit i in the visible layer and unit j in the hidden layer. b_i is the ith bias of visible layer and c_j is the jth bias of hidden layer. The probabilistic semantics for an RBM is defined by its energy function as follows:

$$p(v;\theta) = \frac{1}{Z(\theta)} \sum_{h \in H} \exp\big(-E(v,h;\,\theta)\big) \qquad (2)$$

$$Z(\theta) = \sum_{v \in V}\sum_{h \in H} \exp\big(-E(v,h;\,\theta)\big) \qquad (3)$$

where $Z(\theta)$ is the normalizing constant or partition function. The conditional distributions over hidden unit h and visible vector v are given:

$$p(h|v) = \prod_j p\big(h_j|v\big), \qquad p(v|h) = \prod_i p(v_i|h) \qquad (4)$$

the probability of turning on unit j is a logistic function of the states of v and W_{ij}:

$$p\big(h_j = 1\big|v\big) = \text{sigm}\left(\sum_i W_{ij} v_i + c_j\right) \qquad (5)$$

the probability of turning on unit i is a logistic function of the states of h and W_{ij}:

$$p(v_i = 1|h) = \text{sigm}\left(\sum_j W_{ij} h_j + b_i\right) \qquad (6)$$

where the logistic function is

$$\text{sigm}(x) = 1/\big(1 + \exp(-x)\big) \qquad (7)$$

The average of the log-likelihood with respect to the parameters W can be retrieved from the Contrastive Divergence (CD) method [22]

$$\frac{\partial \ln p(\mathbf{v})}{\partial W_{ij}} = \langle v_i h_j \rangle_{P_{data}} - \langle v_i h_j \rangle_{P_{Model}} \qquad (8)$$

where $\langle \cdot \rangle_{P_{data}}$ is an expectation with the data distribution and $\langle \cdot \rangle_{P_{Model}}$ is a distribution of samples from running the Gibbs sampler. In this case the term $\langle \cdot \rangle_1$ will be used such that it indicates an expectation with the distribution of samples from running the Gibbs sampler initialized at the data for one full step.

Then the parameter W can be adjusted by:

$$W_{ij} = \vartheta W_{ij} + \eta\left(\langle v_i h_j\rangle_{P_{data}} - \langle v_i h_j\rangle_{P_1}\right) \quad (9)$$

where ϑ is the momentum and η is the learning rate.

The above description is based on one sample datum. In our research, we use all labeled data by inputting them one by one from the first layer. The deep architecture is constructed layer by layer from bottom to top and the weight is trained by calculated data in the k-1th layer.

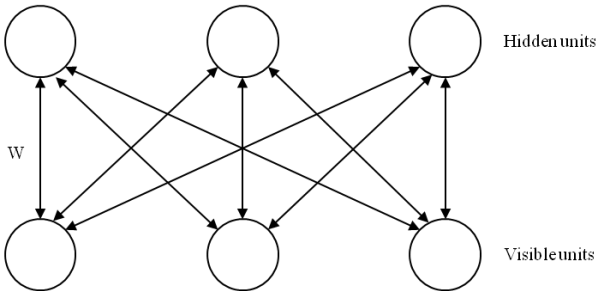

Figure 1. RBM Architecture

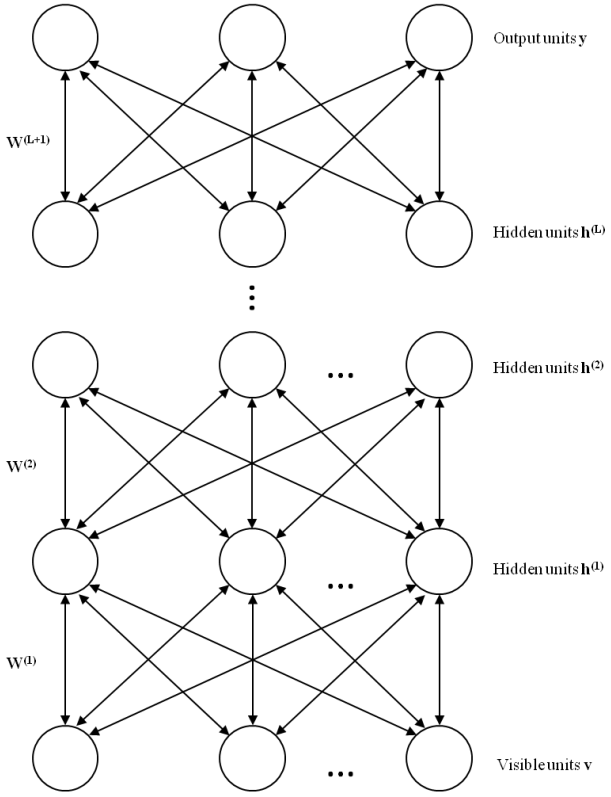

Figure 2. Deep Belief Network Architecture

2.3 Deep Belief Network

Deep Belief Network (DBN) is a generative model consisting of multiple stacked levels of neural networks that each layer contains a set of binary or real-valued units. The main building block networks for the DBN are restricted Boltzmann machine (RBM) layers and a backpropagation (BP) layer. An example of a DBN for classification is shown in Fig. 2. It consists of an input layer which contains the input units or called visible units, the hidden layers and finally an output layer that has one unit for each class. There is a full connections between two adjacent layers, but no two units in the same layer are connected.

To construct a DBN we train a stack of RBMs as many as the number of hidden layers in the DBN. First, the lowest layer is trained directly on the input data, so-called feature vector, then training each next higher layer by capturing the important feature of the hidden units of the previous layer as the input data for the next higher layer in order to get the weights in each RBM layer. RBM manages the feature vectors only excluding the label information. This procedure is continued until a number of hidden layers in the DBN have been trained unsupervisedly. Each RBM layer learns the parameters independently and makes the optimal parameters locally not for entire model. To address this optimization problem, there is a supervised BP layer on top of the model that fine-tunes the entire model in the learning process and generates the output in the inference process. Finally, the feature vector composes of some complicated features which reflect the structured information among the original features. When the stack of RBMs is trained, it can be used to initialize a multilayer neural network for classification tasks. The classification performance with the new feature vector is better than using the original feature vector.

3. Classification using DBN

The aim of this study is to use DBN in the classification of DNA microarray data like leukemia dataset. To train the DBN, we have used the training procedure as Algorithm 1.

Algorithm 1. Algorithm of deep belief network

Input: data x,y
number of hidden units N
number of layers *G*
number of epochs *Q*
number of labeled data *L*
hidden layer *h*
weight *W*
biases of hidden layer *b*
biases of visible layer *c*
momentum ϑ
learning rate η

Output: model of deep architecture
for $k = 1$; $k < G$ do
  for $r = 1$; $p \leq E$ do
    for $s = 1$; $s \leq L$ do

$$p(h_{j,s} = 1|\mathbf{v}) = \text{sigm}\left(\sum_i W_{ij} v_{i,s} + c_j\right)$$

$$p(h_{j,s} = 1|\mathbf{v}) = \text{sigm}\left(\sum_i W_{ij} v_{i,s} + c_j\right)$$

    update the weight and biases

$$W_{ij} = \vartheta W_{ij} + \eta\left(\langle v_{i,s} h_{j,s}\rangle_{P_{data}} - \langle v_{i,s} h_{j,s}\rangle_{P_1}\right)$$

$$b_i = b_i + \eta\left(\langle h_{j,s}\rangle_{P_{data}} - \langle h_{j,s}\rangle_{P_1}\right)$$

$$c_j = c_j + \eta(v_0 - v_1)$$

    end
  end
end

4. Experimental

4.1 Experimental setting

The proposed method has been evaluated by microarray datasets, which are lung cancer and prostate cancer in the first experiment before using leukemia data. The DBN successfully demonstrated the best test classification accuracies of 99.45% and 94.85% for lung cancer and prostate cancer dataset, respectively as shown in Table 1. Therefore, we adopted the DBN on the leukemia dataset. Leukemia data contains the expression levels of 7129 genes of 72 patients with two classes, acute lymphoblastic leukemia (ALL) and acute myeloid leukemia (AML). The training dataset consisted of 38 bone marrow samples with 27 ALL and 11 AML from adult patients. The testing dataset consisted of 24 bone marrow samples as well as 10 peripheral blood specimens from adults and children (20 ALL and 14 AML). Table 2 describe the specification of the leukemia dataset.

We compared the classification performance of DBN with the three representative classifiers, support vector machines (SVM), k-nearest neighbor (KNN) and naive bayes (NB) SVM and NB are the powerful classification methods and KNN is a typical nonlinear classifier that used as the standard for performance comparison.

Table 1
Performance comparison with different classification methods (%)

Method	DBN	SVM	KNN	NB
Lung Cancer	99.45	98.90	94.46	97.80
Prostate Cancer	94.85	91.18	81.62	55.15

Table 2
Specification of the leukemia dataset

Dataset	Gene expression levels	Training set	Testing set	Categories
Leukemia	7,129	38	34	2

4.2 Experimental results on leukemia data

The DBN classification has been applied on leukemia dataset. The performance of DBN was compared with SVM, KNN and NB by computing the accurate of the classification. We trained the weights of each layer with the number of epochs equal to 50 and the learning rate equal to 1.5 while the initial momentum is 0.5.

As shown in Table 3, the DBN classifier can obtain the classification accuracy of 98.61% for testing set of leukemia dataset. This shows that the DBN outperforms classifiers based SVM, KNN and NB algorithm.

Table 3
Performance comparison with different classification methods on the leukemia dataset (%)

Method	DBN	SVM	KNN	NB
Leukemia	98.61	94.45	88.89	97.22

5. Conclusion

DNA microarray experiment with a precise and reliable data analysis method can be used for cancer diagnosis. One of the methods that could help identifying the new pattern effectively is using machine learning approaches. In this paper we focused on the use of Deep Belief Network and applies it successfully to leukemia classification. We studied the details of DBN training and evaluated the performance of our approach on DNA microarray of leukemia data. Our result were compare with SVM, KNN and NB. The comparative results indicate that DBN has a better performance than the other classifiers for our data. The further work, we will consider gene selection such as Principle Component Analysis (PCA), Information Gain (IG) and Correlation-based Feature Selection (CFS) work with the DBN for improving the performance of leukemia classification and comparing with the other cancer datasets.

Acknowledgements

The authors would like to thank the Research and Development Office (RDO) of Prince of Songkla University (PSU), Thailand for their grant support. The first author would also like to express her special appreciations to the Graduates School of PSU for the graduate student research grant. Special thanks go to Mr.Pitipol Meemak of the Department of Molecular Biotechnology and Bioinformatics, Faculty of Science, PSU for his initiation in this research.

References

[1] D.V. Nguyen & D.M. Rocke, Tumor classification by partial least squares using microarray gene expression data, *Bioinformatics*, *18*(1), 2002, 39-50.

[2] K.L. Tang, W.J. Yao, T.H. Li, Y.X. Li, & Z.W. Cao, Cancer classification from the gene expression profiles by discriminant kernel-PLS, *Journal of Bioinformatics and Computational Biology*, *8*, 2010, 147-160.

[3] M. Adjouadi, M. Ayala, M. Cabrerizo, N. Zong, G. Lizarraga & M. Rossman, Classification of leukemia blood samples using neural networks, Annals *of Biomedical Engineering*, *38*, 2010, 1473-1482.

[4] J.H. Hong, & S.B. Cho, The classification of cancer based on DNA microarray data that uses diverse ensemble genetic programming, *International Journal of Artificial Intelligence in Medicine*, *36*, 2006, 43-58.

[5] M. Li, B. Wang, Z. Momeni, & F. Valafar, Pattern recognition techniques in microarray data analysis, *Annals of the New York Academy of Sciences*, 980, 2002, 41-64.

[6] S.B. Cho, & H.H. Won, Machine learning in DNA microarray analysis for cancer classification. *Proc. First Asia-Pacific bioinformatics conference on Bioinformatics*, *19*(APBC '03), Darlinghurst, AUS, 2003, 189-198.

[7] H. Larochelle, D. Erhan, A. Courville, J. Bergstra & Y. Bengio, An empirical evaluation of deep architectures on problems with many factors of variation. *Proc. 24th International Conf. on Machine Learning*, New York, USA, 2007, 473-480.

[8] R. Salakhutdinov & G. Hinton, Using deep belief nets to learn covariance kernels for Gaussian processes, *Advance in Neural Information Processing Systems*, *20*, 2008, 1249-1256.

[9] E. Horster & R. Lienhart, Deep networks for image retrieval on large-scale databases. *Proc. 16th ACM International Conf. on Multimedia*, Vancouver, British Columbia, Canada, 2008, 643-646.

[10] H. Mobahi, R. Collobert & J. Weston, Deep learning from temporal coherence in video. *Proc. International Conf. on Machine Learning*, Canada, 2009, 737-744.

[11] M.A. Salama, A.E. Hassanien & A.A. Fahmy, Deep belief network for clustering and classification of a continuous data. *Proc. 10th IEEE International Symposium on Signal Processing and Information Technology*, Washington, DC, USA, 2010, 473-477.

[12] N. Jaitly & G. Hinton, Learning a better representation of speech sound waves using restricted Boltzmann machines. *Proc. 11th IEEE International Conf. on Acoustics, Speech and Signal Processing*, Prague, Czech Republic, 2011, 5884-5887.

[13] M.A. Ranzato, J. Susskind, V. Mnih & G. Hinton, On deep generative models with applications to recognition. *Proc. 24th IEEE Conf. on Computer Vision and Pattern Recognition*, Colorado Springs, CO, USA, 2011, 2857-2864.

[14] E. Chen, X. Yang, H. Zha, R. Zhang & W. Zhang, Learning object classes from image thumbnails through deep neural networks. *Proc. IEEE International Conf. on Acoustics Speech and Signal Processing*, Las Vegas, USA 2008, 829-832.

[15] Y. Liu, S. Zhou & Q. Chen, Discriminative deep belief networks for visual data classification, *Pattern Recognition*, *44*(10-11), 2011, 2287-2296.

[16] G. Hinton, S. Osindero & Y.W. Teh, A fast learning algorithm for deep belief nets, *Neural Computation*, *18*(7), 2006, 1527-1554.

[17] M.A. Ranzato, C. Poultney, S. Chopra & Y. Lecun, Efficient learning of sparse representations with an energy-based model, *Advances in Neural Information Processing Systems*, *19*, 2006, 1137-1144.

[18] R. Salakhutdinov & I. Murray, On the quantitative analysis of deep belief networks. *Proc. 25th International Conf. on Machine Learning*, Helsinki, Finland, 2008, 872-879.

[19] H. Larochelle, M. Mandel, R. Pascanu & Y. Bengio, Learning algorithms for the classification restricted Boltzmann machine, *Journal of Machine Learning Research*, *13*, 2012, 643-669.

[20] Kent Ridge Bio-medical Data Set Repository, http://levis.tongji.edu.cn/gzli/data/mirror-kentridge.html

[21] T.R. Golub, D.K. Slonim, P. Tamayo, C. Huard, M. Gassenbeek, J.P. Mesirov, H. Coller, M.L. Loh, J.R. Downing, M.A. Caligiuri, C.D. Bloomfield & E.S. Lander, Molecular classification of cancer: class discovery and class prediction by gene expression monitoring, *Science*, *286*(5439), 1999, 531-537.

[22] G. Hinton, Training products of experts by minimizing contrastive divergence, *Neural Computation*, *14*, 2002, 1771-1800.