CLASSIFICATION OF BREAST CANCER MICROARRAY DATA USING RADIAL BASIS FUNCTION NETWORK

UMI HANIM BINTI MAZLAN

UNIVERSITI TEKNOLOGI MALAYSIA

CLASSIFICATION OF BREAST CANCER MICROARRAY DATA USING RADIAL BASIS FUNCTION NETWORK

UMI HANIM BINTI MAZLAN

A project report submitted in partial fulfillment of the
requirements for the award of the degree of
Master of Science (Computer Science)

Faculty of Computer Science and Information Systems
Universiti Teknologi Malaysia

October 2009

To my beloved dad, my lovely mum and my precious siblings

ACKNOWLEDGMENT

In the name of Allah, Most Gracious, Most Merciful

Praise to The Almighty, Allah S.W.T for giving me the strength and passion towards completing this research throughout the semester. By His willing and bless, I had managed to complete my work within the period given.

First of all, my thanks go to Dr. Puteh Bt Saad, who responsible as my Master Project's supervisor for her guidance, counsels, motivation, and most importantly for sharing her thoughts with me.

My deepest thanks go to my lovely family for their endless prayers, love and support. Their moral supports that always right by my side had been the inspiration along the journey and highly appreciated.

I would also like to extend my heartfelt appreciation to all my friends for their valuable assistance and friendship.

May Allah S.W.T repay all their sacrifices and deeds with His Mercy and Bless.

ABSTRACT

Breast cancer is the number one killer disease among women worldwide. Although this disease may affect women and men but the rate of incidence and the number of death is high among women compared to men. Early detection of breast cancer will help to increase the chance of survival since the early treatment can be decided for the patients who suffer this disease. The advent of the microarray technology has been applied to the medical area in term of classification of cancer and diseases. By using the microarray, thousands of genes expression can be determined simultaneously. However, this microarray suffers several drawbacks such as high dimensionality and contains irrelevant genes. Therefore, various techniques of feature selection have been developed in order to reduce the dimensionality of the microarray and also to select only the appropriate genes. For this study, the microarray breast cancer data, which is obtained from the Centre for Computational Intelligence will be used in the experiment. The Relief-F algorithm has been chosen as the method of the feature selection. As the comparison, another two methods of feature selection which are Information Gain and Chi-Square will also be used in the experiment. The Radial Basis Function, RBF network will be used as the classifier to distinguish between the cancerous and non-cancerous cells. The accuracy of the classification will be evaluated by using the chosen metric namely Receiver Operating Characteristic, ROC.

ABSTRAK

Barah payudara merupkan pembunuh nombor satu di kalangan wanita di seluruh dunia. Walaupun penyakit ini boleh menyerang wanita dan lelaki namun kadar kejadian dan kematian adalah lebih tinggi di kalangan wanita berbanding lelaki. Pengesanan awal barah payudara boleh membantu meningkatkan peluang untuk hidup memandangkan rawatan awal boleh disarankan kepada pesakit yang menghidapi penyakit ini. Kemunculan teknologi mikroarray telah diaplikasikan dalam bidang perubatan untuk pengkelasan bagi penyakit barah. Dengan menggunakan mikroarray, berjuta-juta pengekspresan gen boleh ditentukan secara serentak. Walaubagaimanapun, mikroarray ini berhadapan dengan beberapa kelemahan seperti mempunyai dimensi yang tinggi dan juga mengandungi gen-gen yang tidak relevan. Oleh sebab itu, pelbagai teknik pemilihan fitur telah dibangunkan bertujuan mengurangkan dimensi mikroarray dan hanya memilih gen-gen yang bersesuaian sahaja. Bagi kajian ini, data barah payudara mikroarray yang digunakan dalam eksperimen diperoleh dari *Centre for Computational Intelligence*. Algoritma *ReliefF* telah dipilih sebagai teknik untuk pemilihan fitur. Sebagai perbandingan, dua lagi teknik pemilihan fitur iaitu *Information Gain* dan *Chi-square* juga akan digunakan dalam eksperimen. Bagi membezakan di antara sel barah dan bukan barah, rangkaian *Radial Basis Function*, RBF, akan digunakan sebagai pengkelas. Ketepatan pengkelasan akan dinilai oleh metrik yang telah dipilih iaitu *Receiver Operating Characteristic*, ROC.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AUC	-	Area Under Curve
ARFF	-	Artificial Relation File Format
cDNA	-	Complementary Deoxyribonucleic Acid
cRNA	-	Complementary Ribonucleic Acid
DNA	-	Deoxyribonucleic Acid
EST	-	Expressed Sequence Tag
FPR	-	False Positive Rate
k-NN	-	*k*-nearest neighbor
MLP	-	Multilayer Perceptron
mRNA	-	Messenger Ribonucleic Acid
NB	-	Naive Bayes
RBF	-	Radial Basis Function
RNA	-	Ribonucleic Acid
ROC	-	Receiver Operating Characteristics
SOM	-	Self Organising Maps
SVM	-	Support Vector Machine
TPR	-	True Positive Rate

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Introduction

Cancer or known as malignant neoplasm in medical term is the number one killer diseases in most of the countries worldwide. The number of deaths and people suffering with cancer are increasing year by year. This number will continue to increase with an approximate 12 million deaths in 2030 (Farzana Kabir Ahmad, 2008). This killer disease may affect people at all ages even fetuses and also the animals but the risk increases with age. According to the National Cancer Institute, cancer is defined as a disease causes by uncontrolled division of abnormal cells. Cancer cells are capable to invade other tissues and can spread through the blood and lymph systems to other parts of the body.

There are various factors that can cause cancer namely mutation, viral or bacterial infection, hormonal imbalances and others. Mutation can be classified into two categories which are chemical carcinogens and ionizing radiation. Carcinogens are the mutagen, substances that cause DNA mutations, which cause cancer. Tobacco

smoking which is accounted for 90 percent of lung cancer (*Biesalski HK, Bueno de Mesquita B, Chesson A, et al., 1998)* and prolonged exposure asbestos are the examples of the substances (*O'Reilly KM, Mclaughlin AM, Beckett WS, Sime PJ, 2007)*.

Radon gas which can cause cancer while prolonged exposure to ultraviolet radiation from the sun that can lead to melanoma and other skin malignancies (*English DR, Armstrong BK, Kricker A, Fleming C,1997)* are the sources of the ionizing radiation. Based on the experiment and epidemiological data, it shows that viruses become the second most important risk factor of the cancer development in human (*zur Hausen H, 1991)*. Hormonal Imbalances can lead to cancer because some hormones are able to behave similarly like the non-mutagenic carcinogens. This condition may stimulate the uncontrolled cell growth.

Cancer can be classified into various types with various names which are mostly named based on the part of the body where the cancer originates (American Cancer Society). Lung cancer, breast cancer, ovary cancer, colon cancer are several examples of the common cancer types. Breast cancer is the disease which fears every woman since it is the number one killer cancer for woman. Breast cancer remains a major health burden since it is a major cause of cancer-related morbidity and mortality among female worldwide (Bray, D., and Parkin, 2004).

It has been reported that, each year, more than one million new cases of female breast cancer are diagnosed, worldwide (Ferlay J, Bray F, Pisani P, Parkin DM, 2001). In United States of America, an estimated 192,370 new cases of invasive breast cancer and an estimated 40,610 breast cancer death are expected in 2009 (American Cancer Society, 2009). The following figure shows the percentage of ten most cancers in female in Malaysia from 2003 to 2005. From the figure, significant shows that breast cancer hold the highest percentage among all the cancers in female.

As discussed in this part, it is proven that breast cancer is one of the main health problems. A lot of researches done to study the best way to prevent, detect and treat the breast cancer in order to reduce the number of deaths. This research also study about detection of breast cancer using the microarray data. The microarray data of the breast cancer will be classified to determine whether it is benign (non-cancerous) or malignant (cancerous) using Radial Basis Function, RBF, after doing the feature selection.

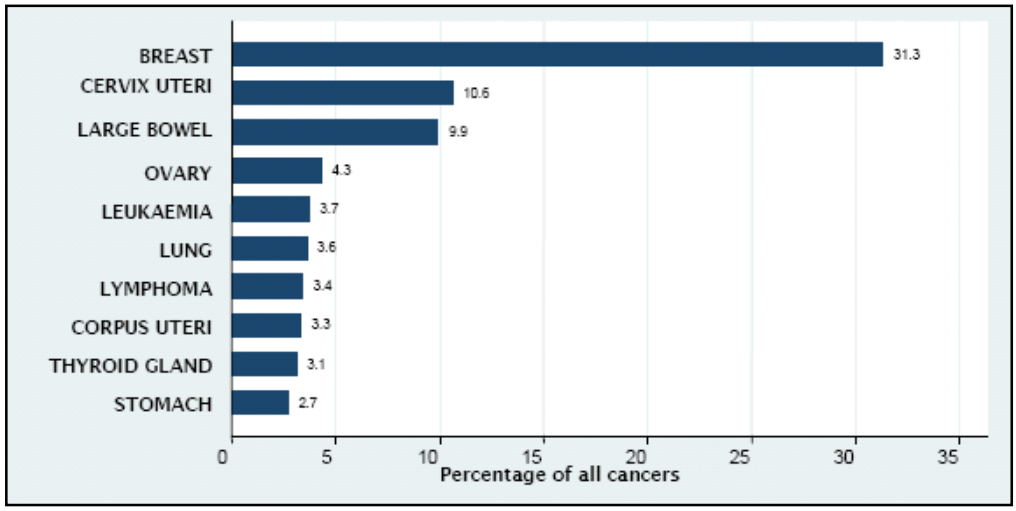

Figure 1.1: Ten most frequent cancer in females, Peninsular Malaysia 2003-2005 (Lim, G.C.C., Rampal, S., and Yahaya, H., 2008)

1.2 Problem Background

This project will use the microarray data of breast cancer. According to Tinker et al., 2006, microarray has become a standard tool in many genomic research laboratories because it has revolutionized the approach of biology research. Now, scientist can study thousand of genes at once instead of working on a single gene

basis. However, microarray data are often overwhelmed, over fitting and confused by the complexity of data analysis. Using this overwhelmed data during classification will lead to the increase of the dimensionality of the classification problem presents computational difficulties and introduces unnecessary noise to the process.

Other than that, due to improper scanning, it also contains multiple missing gene expression values. In addition, mislabeled data or questioned tissues result by experts also the drawback of microarray that decreases the accuracy of the results (Furey *et al*., 2000). However, feature selection can be utilized to filter the unnecessary genes before the classification stage is implemented. Based on Farzana Kabir Ahmad, 2008, (cited from Ben-Dor *et al*., 2000; Guyon *et al*., 2002; Li, Y. *et al*., 2002; Tago & Hanai, 2003) in many large scale of gene expression data analysis, feature selection has become a prior step and a pre-requisite. The high reliance of this gene selection technique is due to its important purpose that requires small sets of genes in order to informatively sufficient to differentiate between cancerous and non-cancerous cells.

1.3 Problem Statement

The main issue in the classification of microarray data is the presence of noise and irrelevant genes. According to Liu and Iba, 2002, many genes are not relevant to differentiate between classes and caused noise in the process of classification. Hence, this will lead to the drowning out of the relevant ones Shen et al., 2007. Therefore, the prior step before doing the classification which is feature selection is important in order to reduce the huge dimension of microarray data.

1.4 Objectives of the Projects

The aim of the project is to classify a selected significant genes from microarray breast cancer data. In order to achieve the aim, the objectives are as follows:

i. To select significant features from microarray data using a suitable feature selection technique.
ii. To classify the selected features into non-cancerous and cancerous using Radial Basis Function network.
iii. To evaluate the performance of the classifier using Receiver Operating Characteristic, ROC.

1.5 Scopes of the Project

This project will use a breast cancer microarray data which is obtained from Centre for Computational Intelligence. This data originates from the Zhu, Z., Ong, Y.S., and Dash, M., (2007). The dataset contains of 97 instances with 24481 genes and 2 classes.

1.6 Importance of the Study

Early detection of breast cancer can help to avoid the cancer from getting worse and eventually leads to the decreasing of the rate of death. Mammogram is one of the tools that can detect early stage of breast cancer. However, biopsy is needed in order to confirm the present of the malignant (cancerous) tissue that cannot be performed by the mammogram. From this biopsy process the microarray data will be obtained.

As discussed in the problem background there are several drawbacks of microarray data. However, it can be solved by implementing the feature selection before it can proceed to the classification. In this study, several feature selection will be studied before choosing the most suitable method. Feature selection is very important since it will determine the accuracy of the result after the classification.

The accuracy of the result is also important because it will classify between the cancerous and non-cancerous. Since the purpose of biopsy is to confirm the presence of the cancer cells after the early detection, the confirmation must not take a long time. The best feature selection method must be used before classifying the microarray data in order to reduce the time and produce the accurate result.

Through this project, the best feature selection method will be selected and implemented to microarray data before it will be classified. The performance of the classifiers will be evaluated based on the suitable metric. The result of this study will be discussed in order to give the information for the future research in breast cancer area.

1.7 Summary

As a conclusion, this chapter discusses the overview of the project. This project will classify the breast cancer microarray data using the Radial Basis Function to two classes namely benign (non-cancerous) and malignant (cancerous). Feature selections will be studied and the best method will be chosen to implement the microarray data before the classification.

REFERENCES

Abu-Khalaf, M.M., Harris, L.N., and Chung, G.C. (2007). Chapter 10 DNA and Tissue Microarrays. In Patel, H.R.H., Arya, M., and Shergill, I.S. *Basic Science Techniques in Clinical Practice.* London: Springer

Adomas, A., Heller, G., Olson, A., Osborne, J., *et al.* (2008). Comparative analysis of transcript abundance in Pinus sylverstric after challenge with a saprotrophic, pathogenic or mutualistic fungus. *Tree Physiology.* 28, 885-897

Ahmad, A., and Dey, L. (2005). A feature selection technique for classificatory analysis. *Pattern Recognition Letters.* 26(1), 43-56

Alizadeh A.A., Eisen M.M., Davis, R.E., Ma, C., Lossos, I.S., *et al.,* (2000). Diffuse types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature.* 403, 503-511

Arauzo-Azofra, A., Benitez, J.M., Castro, J.L. (2004). A feature set measure based on Relief. *The 5th International Conference on Recent Advances in Soft Computing (RASC2004).* Nottinhham Trant Univ. 104-109

Ben-Bassat, M. (1982) Pattern Recognition and reduction of dimensionality. In Krishnaiah, P. and Kanal, L., (eds) *Handbook of Statistics II,* Vol I. North-Holland, Amsterdam; 773-791

Ben-Dor, A., Bruhn, L., Friedman, N., Nachman, I., Schummer, M. E., and Yakhini, Z. (2000). Tissue classification with gene expression profiles. *Journal of Computational Biology.* 7; 559-584

Biesalski HK, Bueno de Musqeito B, Chesson A, *et al*., (1998). European Consensus Statement on Lung Cancer: Risk Factors and Prevention. *CA : a cancer journal for clinicians.* 48(3),167-176

Boldrick, J.C., Alizadeh, A.A., Diehn, M., Dudoit, S., Liu, C.L., *et al.* (2002). Stereotyped and specific gene expression programs in human innate immune responses to bacteria. *Proceedings of the National Academy of Sciences of the USA.* 99, 972-977

Bors, A. G., Gabbouj, G., (1994). Minimal topology for radial basis function neural network for pattern classification. *Digital Signal Processing: a review journal.* 3, 302-309

Bors, A.G., Pitas, I., (1996). Median radial basis functions neural network. *IEEE Trans. On Neural Networks.* 7(6), 1351-1364

Bray, F., D., P. M., and Parkin, M (2004). The changing global patterns of female breast cancer incidence and mortality. *Breast Cancer Research.* 6(6)

Burchard, J., unpublished.

Chang, C. Y., and Chung, P. C. (1999). A Contextual-Constraint Based Hopfield Neural Cube for Medical Image Segmentation. *Proceedings of The IEEE Region 10 Conference on TENCON.* 2, 1170-1173

Cheung, VG., Morley, M., Aguilar, F., Massimi, A., Kucherlapati, R., and Childs, G. (1999). Making and reading microarrays. *Nature Genetics.* 21, 15-19

Christian Harwanegg and Reinhard Hiller. (2005). Protein microarrays for the diagnosis of allergic diseases: State-of-the-art and future development. *Clinical Chemistry and Laboratory Medicine.* 29 (4), 272-277. Walter de Gruyter.

Chu, S., DeRisi, J., Eisen, M., Mulholland, J., *et al.* (1998). The transcriptional program of sporulation in budding yeast. *Science.* 282, 699-705

Compton, R., Lioumi, M., and Falciani, F. (2008). Microarray Technology. *Encyclopedia of Molecular Pharmacology.* Berlin Heidelberg, New York: Springer-Verlag

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

Dasarathy, B.(1991). *Nearest Neighbor Norms: NN Pattern Classification tehniques.* IEEE Computer Society Press, Los Alamitos, CA,USA

Dash, M., and Liu, H. (1997). Feature selection for classification. *Intelligent Data Analysis.* 1, 131-156

Downey, T.J., Jr., Meyer, D. J., Price, R.K., and Spitznagel, E.L. (1999). Using the receiver operating characteristic to asses the performance of neural classifier. *International Joint Conference on Neural Networks (IJCNN'99)*. 10-16 July. Washigton, DC, USA. 5, 3642-3646

Driouch, K., Landemaine, T., Sin, S., Wang, SX., and Lidereau, R. (2007). Gene arrays for diagnosis, prognosis and treatment of breast cancer metastasis. *Clinical & Experimental Metastasis.* 24(8), 575-585

Duda, P., *et al*. (2001). *Pattern Classification*. Wiley, New York.

English, DR., Armstrong, BK., Kricker, A., Fleming, C. (1997). Sunlight and cancer. *Cancer Causes and Contol.* 8(3); 271-83, 283

Farzana Kabir Ahmad (2008). *An Integrated Clinical and Gene Expression Profiles for Breast Cancer Prognosis Model.* Master, Universiti Utara Malaysia

Fawcett, T. (2005). An introduction to ROC analysis. *Pattern Recognition Letters.* Palo Alto, USA. 27(2006), 861-874

Fayyad, U., and Irani, K.(1993). Multi-interval discretization of continuous-values attributes for classification learning. *Proceedings of the 13th International Joint Conference on Artificial Intelligence.* Portland, OR. 1022-1027

Ferlay J., Bray F., Pisani P., and Parkin D. (2001). *GLOBOCAN 2000: Cancer incidence, mortality and prevalence worldwide*

Ferran, E.A., Pflugfedder, B., and Ferrarap. (1994). Self Organized neural maps of human protein sequences. *Protein Science.* 3,507-521

Ferri, F., et al. (1994). *Pattern Recognition in Practice IV, Multiple Paradigms, Comparative Studies and Hybrid Systems.* Elsevier, Amsterdam. 403-413

Fodor, S.P., *et al.* (1991). Ligth-directed, spatially addressable parallel chemical synthesis. *Science.* 251, 767-773

Fodor, S.P., *et al.* (1993). Multiplexed biochemical assays with biological chips. *Nature.* 364, 555-556

Furey, T.S., Cristianni, N., Duffy N., Bednarski, D. W., Schummer, M. and Haussler, D. (2000). Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics.* 16(10), 906-914

Gianluca Bontempi and Benjamin Haibe-Kains. 2008. *Feature selection methods for mining bioinformatics data*. Bruxelles, Belgium: ULB Machine Learning Group

Green, D. M., and Swets, J. A. (1996). *Signal detection theory and psychophysics.* New York: John Wiley and Sons, Inc

Gruvberger-Saal, S.K., Cunliffe, H.E., and Carr, M.K. 2006. Microarrays in breast cancer research and clinical practice- the future lies ahead. *Endocrine-Related Cancer.* 13, 1017-1031

Guyon, I., Weston, J., Stephen Barnhill, M. D., and Vapnik, V. (2002). Gene Selection for Cancer Classification using Support Vector Machines. *Machine Learning.* 46; 389-422

Guyon , I., and Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research.* 3, 1157-1182

Hall, M. (1999). *Correlation-based feature selection for machine learning*. Doctoral Philosophy, Waikato University, New Zealand.

Haykin, S. (1994). *Neural Networks: A Comprehensive Foundation.* New York: Macmillan College Publishing Co. Inc

Holland, J. (1975). *Adaptation in Natural and Artificial Systems.* University if Michigan Press, Ann Arbor.

Horng, M-H. (2008). Texture Classification of the Ultrasonic Images of Rotator Cuff Disease based on Radial Basis Function Network. *International Joint Conference on Neural Networks (IJCNN 2008).* Hong Kong. 91-97

Inza, I., *et al*. (2000). Feature subset selection by Bayesian networks based optimization. *Artificial Intelligence.* 123, 157-184

Jin, X., Xu, A., Bie, R., and Guo, P. (2006). Machine Learning and Chi-Square Feature Selection for Cancer Classification Using SAGE Gene Expression Profiles. In Li, J. *et al.* (Eds) *Data Mining for Biomedical Applications.* (pp: 106-115). Berlin Heidelberg: Springer-Verlag

Kasabov, N.K. (2002). *Evolving Connectionist Systems, Methods and Applications in Bioinformatics, Brain Study and Intelligent Machines*. Verlag: Springer

Kerekes, J. (2008). Receiver Operating Characteristic Curve Confidence Intervals and Regions. *IEEE Geoscience and Remote Sensing Letters.* 5(2), 251-255

Kim, YS., Street W.N., Menczer, F. (2003). Feature selection in data mining. In John Wang (Ed.) *Data Mining: opportunities and challenges.* (80-105) Hershey, PA, USA : IGI Publishing

Kira, K., and Rendell, L.A. (1992). A practical approach to feature selection. *Proceedings of the ninth international workshop on Machine learning.* Morgan Kaufmann Publishers Inc. 249-256

Kittler, J. (1978). *Pattern Recognition and Signal Processing, Chapter Feature Set Search Algorithms* Sijthoff and Noordhoff, Alphen aan den Rijn, Netherlands; 41-60

Kohavi, R., and John, G.H. (1997). Wrappers for feature subset selection. *Artificial Intelligence.* 97, 273-324

Koller, D. and Sahami, M. (1996). Toward optimal feature selection. *Proceedings of the Thirteenth International Conference on Machine Learning*, Bari, Italy, 284-292

Kononenko, I. (1994). Estimating attributes: Analysis and extensions of RELIEF. *European Conference on Machine Learning.* 171-182

Konvacevic, D., and Loncaric, S. (1997). Radial Basis Function Based Image Segmentation Using A Receptive Field. *Proceedings of Tenth IEEE Symposium on Computer-Based Medical Systems.* 126-130

Kuan, M. M., Lim, C. P., Ismail, O., Yuvaraj, R. M., and Singh, I. (2002). Application of Artificial Neural Networks to The Diagnosis of Acute Coronary Syndrome. *Proceedings of International Conference on Artificial Intelligence in Engineering and Technology.* Kota Kinabalu, Malaysia. 470-475

Kégl, B., Krzyzak, A., and Niemann, H. (1998). Radial Basis Function Network in Nonparametric Classification and Function Learning. *Proceedings of the Fourteenth International Conference on Pattern Recognition.* 16-20 August. 1, 565-570

Leonard, J.A., and Kramer, M.A. (1991). Radial basis function network for classifying process faults. *Control Systems Magazine, IEEE.* 11(3), 31-38

Li, Y., Campbell, C., and Tipping, M. (2002). Bayesian automatic relevance determination algorithms for classifying gene expression data. *Bioinformatics.* 18(10); 1332-1339

Li, X., Zhou, G., and Zhou, L. (2006). Novel Data Classification Method on Radial Basis Function Networks. *Proceedings of the Sixth International Conference on Intelligent Systems Design and Applications (ISDA'06)*

Lim, G.C.C., Rampal, S., and Yahaya, H. (Eds.) (2008). *Cancer Incidence in Peninsular Malaysia, 2003-2005*. Kuala Lumpur: National Cancer Registry

Lin, C.T., and Lee, C. S. G. (1996). *Neural Fuzzy Systems - A Neuro-Fuzzy Synergism to Intelligent Systems.* New Jersey: Prentice Hall.

Liu, H., and Setiono, R. (1995). Chi2: feature selection and discretization of numeric attributes. *Proceedings of Seventh International Conference on Tools with Artificial Intelligence.* 11 May-11 August 1995. 388-391

Liu, H., and Motoda, H. (1998). *Feature Selection for Knowledge Discovery and Data Mining.* Boston: Kluwer Academic Publishers.

Liu. J. and Iba, H. (2002). Selecting Informative Genes Using a Multiobjectives Evolutionary Algorithm. *Proceedings of the 2002 congress.* 297-302

Lockhart, DJ., Dong, H., Byrne, MC., *et al*. (1996). Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nature Biotechnology*. 14, 1675-1680

Markov, R.S., and Igor, K. (2003). Theoretical and empirical analysis of relief and rreliefF. *Machine Learning Journal.* 53,23-69

Mashor, M. Y., Mat Isa, N. A., and Othman, N. H. (2002). Automatic Pap Smear Screening Using HMLP Network. *Proceedings of International Conference on Artificial Intelligence in Engineering and Technology.* Kuala Lumpur, Malaysia. 453-457

Mat Isa, N.A., Mashor, M. Y., and Othman, N.H. (2002). Diagnosis of Cervical Cancer Using Hierachical Radial Basis Function (HRBF) Network. *Proceedings of International Conference on Artificial Intelligence in Engineering and Technology.* Kota Kinabalu, Malaysia. 458-463

Matej, S., Lewitt, R.M., (1996). Practical considerations for 3-D image reconstruction using spherically symmetric volume elements. *IEEE Trans. On Medical Imaging.*15(1), 68-78

Mat Isa, N.A., Mat Sakim, H.A., Zamli, K.Z., Haji A Hamid, N., and Mashor, M.Y. (2004). Intelligent Classification System for Cancer Data Based on Artificial Neural Network. *Proceedings of the 2004 IEEE Conference on Cybernatics and Intelligent Systems.* 1-3 December. Singapore.

Miki, R., Kadota, K., Bono, H., Mizuno, Y., Tomaru, Y., *et al.* (2001). Delineating developmental and metabolic pathways in vivo by expression profiling using the RIKEN set of 18,816 full-length enriched mouse cDNA arrays. *Proc Natl Acad Sci USA.* 98(2), 199-204

Miller, L.D and Liu, E.T. 2007. Expression genomics in breast cancer research: microarrays at the crossroads of biology and medicine. *Breast Cancer Research.* 9, 206

Mitchell, T.M. (1997). *Machine Learning.* New York : McGraw-Hill

Momin, B. A., Mitra, S., and Gupta, R. N. (2006). Reduct Generation and Classification of Gene Expression Data. *International Conference on Hybrid Information Technology (ICHIT'06).* 6-11 November. 1, 699-708

Moody, J., (1989). Fast learning in networks of locally-tuned processing units. *Neural Computation.* 1, 281-294

National Cancer Institute. (2005). *What You Need To Know About Breast Cancer.* National Institutes of Health, U.S Department of Health and Human Services

Ng, E. Y. K., Peh, Y. C., Fok, S. C., Ng, F. C., and Sim, L. S. J. (2002). Early Diagnosis of Breast Cancer Using Artificial Neural Network with Thermal Images. *Proceedings of Kuala Lumpur International Conference on Biomedical Engineering.* Kuala Lumpur , Malaysia. 458-463

Orr, M. J. L. (1996). *Radial Basis Function Networks*. Edinburgh, Scotland

Oyang, Y.J., Hwang S.C., Ou, Y.Y., et al. (2005). Data Classification with radial basis function networks based on a novel kernel density estimation algorithm. *IEEE Transaction on Neural networks.* 16(1), 225-236

O'reilly KM, MClaughin AM, Backett WS, Sime PJ (2007). Asbestos-Related Lung Disease. *American Academy of Family Physician.* 75(5); 683-8,690

Palacios, G., Quan, P-L., Jabado, O.J., Conlan, S., Hirschberg, D.L., *et al.* (2007). Panmicrobial Oligonucleotide Array for Diagnosis of Infectious Diseases. *Emerging Infectious Diseases.* 13, 73-81

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

Park, J., and Sandberg, I.W. (1993). Approximation and Radial-Basis-Function Networks. *Neural Computation.* 5, 305-316

Park, S. H., Goo, J. G., and Jo, C-H. (2004). Receiver Operating Characteristic (ROC) Curve: Practical Review for Radiologists. *Korean Journal of Radiology.* 5, 11-18

Pease, A.C. *et al.* (1994). Light-generated oligonucleotide arrays for rapid DNA sequence analysis. *Proceedings of National Academy Science of USA.* 91, 5022-5026

Pedrycz, W., editor. (1998). *Computational Intelligence : An Introduction.* NY: CRC Press, Boca Raton

Poggio, T., Girosi, F., (1990). Networks for approximation and learning. *Proceedings IEEE.* 78(9), 1481-1497

Powell, M.J.D. (1977). Restart Procedures for the Conjugate Gradient Method. *Mathematical Programming.* 12, 241-254

Ramaswamy, S. and Golub, TR. (2002). DNA microarrays in clinical oncology. *Journal Clinical Oncology.* 20, 1932-1941

Reisert, M., and Burkhardt, H. (2006). Feature Selection for Retrieval Purposes. In Lectures Notes in Computer Science. *Image Analysis and Recognition.* 4141, 661-672. Heidelberg: Springer-Verlag

Sankupellay, M., and Selvanathan, N. (2002). Fuzzy Neural and Neural Fuzzy Segmentation of Magnetic Resonance Images. *Proceedings of Kuala Lumpur International Conference on Biomedical Engineering.* Kuala Lumpur, Malaysia. 47-51

Sauter, G., and Simon, R. (2002). Predictive molecular pathology. *New England Journal Medicine.* 347, 1995-1996

Sanner, R.M., Slotine, J.-J. E., (1994). Gaussian networks for direct adaptive control. *IEEE Trans. On Neural Networks.* 3(6), 837-863

Saeys, Y. *et al*. (2007). Gene Selection, A review of feature selection techniques in bioinformatics. *Bioinformatics.* 23(19), 2507-2517. Oxford University Press.

Schena, M., Shalon, D., Davis, R., and Brown, P.O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science.* 270, 467-470

Siedelecky, W. and Sklansky, J. (1998). On automatic feature selection. *International Journal Pattern Recognition.* 2; 197-220

Sikonja, M.R., and Kononenko, I. (1997). An adaptation of relief for attribute estimation in regression. In Morgan Kaufmann, editor. *Machine Learning: Proceedings of the Fourteenth International Conference.* 296-304

Sikonja, M.R., and Kononenko, I. (2003). Theroretical and Empirical Analysis of ReliefF and RReliefF. *Journal of Machine Learning.* 53(1-2), 23-69

Skalak, D. (1994). Prototype and feature selection by sampling and random mutation hill climbing algorithms. *Proceedings of the Eleventh International Conference on Machine Learning,* 293-301

Sorlie, T., Perou, C.M., Tibshirani, R., *et al.* (2001). Gene expressiom patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proceedings of National Academy Science of USA.* 98, 10869-10874

Sun, J., Shen, R.M., Han, P. (2003). An original RBF network learning algorithm. *Chinese journal of computers.* 26(11), 1562-1567

Tago, C., and Hanai, T., (2003). Prognosis prediction by microarray gene expression using support vector machine. *Genome Informatics.* 14; 324-325

Tinker, A.V., Boussioutas, A., and Bowtell, D.D.L. (2006). The Challenges of gene expression microarrays for the study of human cancer. *Cancer Cell.* 9(5), 333-339

Tokan, F., Turker, N., and Tulay, Y., (2006). ROC Analysis as a Useful Tool for Performance Evaluation of Artificial Neural Networks. In Kollias, S., *et al*. (Eds). *ICANN 2006, Part II, LNCS 4132*. 923-931. Springer-Verlag.

Tou, J.T., and Gonzalez, R.C. (1974). *Pattern Recognition Principles.* London: Addison-Wesley.

Tzouvelekis, A., Patlakas, G., and Bouros, D. 2004. Application of microarray technology in pulmonary diseases. *Respiratory Research.* 5(26)

van de Vijver, M.J., He, Y.D., van't Veer L.J., Dai, H., Hart, A.A., *et al.* A gene-expression signatures as a predictor of survival in breast cancer. *New England Journal of Medicine.* 98, 2199-2204

Vapnik, V.N. (1998). *Statistical Learning Theory*. Wiley-Interscience, New York, USA

Wang, D.H., Lee, N.K., Dillon, S.T., and Hoogenraad, N.J. (2002) Protein Sequences Classification using Radial Basis Function (RBF) Neural Networks. *Proceedings of the 9th International Conference on Neural information Processing (ICONIP'02).* 2,

Wang, H.C., Daparzo J., De La Fraga, L.G., Zhu, Y.P., Carazo, J.M.(1998). Self-Organizing tree-growing network for the classification of protein sequences. *Protein Science,* 2613-2622

Wang, Y., Makedon, F.S., Ford, J.C., and Pearlman, J. (2005). HykGene: a hybrid approach for selecting marker genes for phenotype classification using microarray gene expression data. *Bioinformatics.* 21(5), 1530-1537

Westin, L. K. (2001). *Receiver operating characteristic (ROC) analysis: Evaluating discriminance effects among decision support systems*. Umea, Sweden

Weston, J., Mukherjee, S., Chapelle, O., Pontil, M., Poggio, T., and Vapnik, V. (2000). Feature selection for SVMs. *Advances in Neural Information Processing Systems.* 13

Whitfield, M.L., Sherlock, G., Saldanha, A.J., *et al.* (2002). Identification of genes periodically expressed in the human cell cycle and their expression in tumors. *Mol Biol Cell.* 13, 1997-2000

Witten, I.H., and Frank, E. (1999). *Data mining: practical machine learning tools and techniques with Java implementations.* Morgan Kaufmann, San Francisco, CA

Wu, C.H., Berry, M., Shivakumar, S., and McLarty, J.(1995). Neural networks for full-scale protein sequence classification: sequence encoding with singular value decomposition. *Machine Learning.* 21, 177-193

Wu, C.H. Artificial neural networks for molecular sequence analysis. (1997). *Computers Chemistry.* 21, 237-256

Yu, L. and Liu, H. (2004). Efficient feature selection via analysis of relevance and redundancy. *Journal of Machine Learning Respiratory.* 5; 1205-1224

zur Hausen H (1991). Viruses in human cancers. *Science*. 254(5035); 1167-1173

Zhu, Z., Ong, Y-S., and Dash, M. (2007). Markov blanket-embedded genetic algorithm for gene selection. *Pattern Recognition.* 40, 3236-3248