Classification of Laser Induced Fluorescence Spectra from Normal and Malignant Tissues using Learning Vector Quantization Neural Network in Bladder Cancer Diagnosis

Gopal Karemore, Kim Komal Mascarenhas, Choudhary K.S., Ajeethkumar Patil, Unnikrishnan V.K., Vijendra Prabhu, Arunkumar Chowla, Mads Nielsen and Santhosh C

Abstract: **In the present work we discuss the potential of recently developed classification algorithm, Learning Vector Quantization (LVQ), for the analysis of Laser Induced Fluorescence (LIF) Spectra, recorded from normal and malignant bladder tissue samples. The algorithm is prototype based and inherently regularizing, which is desirable, for the LIF spectra because of its high dimensionality and features being settled at widely spaced intervals (sparseness). We discuss the effect of different parameters influencing the performance of LVQ in LIF data classification. Further, we compare and cross validate the classification accuracy of LVQ with other classifiers (eg. SVM and Multi Layer Perceptron) for the same data set. Good agreement has been obtained between LVQ based classification of spectroscopy data and histopathology results which demonstrate the use of LVQ classifier in bladder cancer diagnosis.**

I. INTRODUCTION

Bladder cancer refers to any form of cancers that affect the urinary bladder[1]. The incidence of bladder epithelial tumors has been steadily increasing in the past years. Each year, this cancer is diagnosed in approximately 275000 people worldwide, and about 108000 die from this disease[1]. In fact, it is the fifth most common cancer both in the western world and in developing countries [2,4]. About 95% of bladder tumors are of epithelial origin, the remainder being mesenchymal tumors[5]. Bladder cancer is the fourth most common type of cancer in men and the ninth most common cancer in women[6]. The American Cancer Society estimates that in 2008 there will be about 68,810 new cases of bladder cancer diagnosed in the United States. In which about 51,230 are men and 17,580 are women respectively [8]. One reason for its higher incidence in men is that the androgen receptor, which is much more active in men than in women, plays a major role in the development of cancer [9].

Manuscript received July 5, 2008. This work was supported partially by the Department of Science and Technology, Government of India. Project No:SR/S2/LOP/05/2003.

Gopal Karemore and Mads Nielsen is with the Department of Computer Science, University of Copenhagen, and affiliated to Nordic Bioscience A/S,Denmark-2730.(phone: +45-44547771; fax:45-35321401; e-mail: gp@nordicbioscience.com). Kim Komal Mascarenhas, Choudhary. K..S., Ajeethkumar Patil, Unnikrishnan. V.K., Vijendra Prabhu and Santhosh C are with, Centre for Atomic and Molecular Physics and Arunkumar Chowla is with KMC, Manipal University, Manipal, India -576104

The possible symptoms for bladder cancer are blood in the urine [5] pain during urination and frequent urination or feeling the need to urinate without results [5]. Exposure to carcinogens, habit of smoking increases the risk for bladder cancer [2]. Occupational risk factors include recurrent and early exposure to hair dye, and exposure to dye containing aniline, a chemical used in medical and industrial dyes [2]. Incidence of bladder cancer increases with age [1,10,11]. People over the age of 70 develop the disease 2 to 3 times more often than those aged 55–69 and 15 to 20 times more often than those aged 30–54[10,11]. Developments of methods for the early diagnosis of cancers by detecting tumor makers are getting momentum these days [12].

Several spectroscopic techniques are coming up as useful methods for the diagnosis of malignancies. Raman spectroscopy and Fluorescence spectroscopy are getting importance as promising techniques in cancer diagnosis [13-16]. These techniques provide information about the biochemical changes that occur during disease conditions. Laser induced fluorescence technique is one of the matured methods to discriminate between normal and malignant conditions in oral, cervical and column cancers [17-20]. All the above methods generate large quantity of spectroscopy data and they need reliable classification algorithms for clinical applications.

Exploration and analysis of data in the field of clinical proteomics have become one of the key problems in computational proteomics [21]. From mathematical point of view, the data space to be explored is sparsely filled. Thereby, the spectra may be overlaid by noise such that the contained signal is difficult to extract. Another problem expected to arise for the data analysis methods is due to high dimensionality: The data can always be separated by using more or less independent separation criteria [22]. Thus any method is confronted with the problem of detection of the underlying regularities. Therefore, advanced methods are required to deal with high dimensional, sparse, noisy data analysis.

In the present work we will give insights to recently developed prototype based classifier LVQ, which fulfill the above mentioned requirements for classifying LIF data. However application of LVQ in mass spectroscopy data has been studied previously in [21]. In order to test the

application of LVQ we have applied the LVQ method for Laser Induced Fluorescence data recorded from normal and malignant bladder tissues. However, this work is not meant for the early diagnosis of the diseases, where as, it may be a useful tool for surgeon for the demarcation of the normal site with the diseased part of the bladder. In corporation of reliable classification algorithms along with the LIF technique will be useful for the objective diagnosis of the disease.

II. Materials and methods

The essential parts of the Laser Induced Fluorescence (LIF) system (Figure 1) consist of a HeCd laser (Kimmon, Japan) used for the 325 nm excitation of the tissue, Spectrograph (PI-Acton, USA) disperses the collected radiation in wavelength components which is recorded using a Charge Coupled Device (CCD) (Andor, Ireland). A seven fibers bundle probe was used to deliver laser line on the tissue surface as well as to collect the resulting fluorescence signal.

A. Sample collection and storage

Bladder tissue samples were collected from patients, with informed consent, who had under gone surgery at Kasturba Medical College, Manipal. Ethical clearance has obtained from the Manipal University Ethical Committee for the present study. Clinically normal samples were also collected during the surgical procedure. Samples obtained were immediately kept in normal saline and taken for the spectroscopy experiments, otherwise stored at -85^{o} C deep freezer. Fluorescence was recorded at 2-12 different sites on both upper and lower surfaces of the tissue. We have used 30 spectra each from normal and bladder cancer patients for the present study.

B. Data Processing

1) Preprocessing and feature selection: The LIF spectra as shown in figure 2, have to be preprocessed before the analysis. Preprocessing aims to correct intensity and wavelength values in order to: (i) remove background (ii) reduce noise and (ii) make spectra comparable (normalization) [23]. This includes Baseline correction, Normalization and Peak alignment Baseline subtraction uses an iterative algorithm to attempt to remove the baseline slope and offset from spectrum by iteratively calculating the best fit straight line through a set of estimated baseline points. Normalization enables the comparison of different samples since absolute peak values of different fraction of spectrum could be slightly shifted [24, 25].

Spectroscopy data classification problem is typical in the sense that the number of features is much larger than the number of observations (1024 data points each for 60 spectra in total), but in which no single feature achieves a correct classification, therefore we need to find a classifier which appropriately learns how to weight multiple features and at the same time produce a generalized mapping which is not over-fitted. A simple approach for finding significant features is to assume that each wavelength value is independent and compute a two-way t-test and wilcoxon test [26,27]. The effect of feature selection on the basis of these testes on accuracy of LVQ classification is reported in discussion part of this paper. After extracting significant features, next step is to build a classifier.

2) Feature classification: The automation of classification through the use of Artificial Neural Network is a common practice today, giving remarkable benefits [27]. Classification is inherently a discrimination problem. Recent research [21,24,29,30] shows that for problems where discrimination is the main concern, attacking discrimination problems by density estimation may be inferior to more direct approaches. Still, it may be desirable to formulate the models in terms of generative, probabilistic models, while the learning procedure aims at being able to discriminate well.

3) LVQ classifier: In this paper, we used LVQ, or Learning Vector Quantization, which is a prototype-based supervised classification algorithm. LVQ is a special case of an artificial neural network, more precisely, it applies a winner-take-all Hebbian learning-based approach [31]. Learning Vector Quantization is a precursor of the well-known self-organizing maps (also called Kohonen feature maps) and like them it can be seen as a special kind of artificial neural network [32]. Both types of networks represent a set of reference vectors, the positions of which are optimized w.r.t. a given dataset. LVQ is a learning codebook-based classifiers, where the codebook is expressed in terms of probabilistic models, but where the training procedure is discriminative in nature [32,33,34]. The training algorithm is derived from the LVQ algorithm that uses vector quantization [35,36,37], where we assume a codebook which is defined by a set of M prototype vectors. (M is chosen by the user and the initial prototype vectors are chosen arbitrarily). An input belongs to cluster i if i is the index of the closest prototype (closest in the sense of the normal Euclidean distance). This has the effect of dividing up the input space into a Dirichlet tessellation which is a special kind of decomposition of a metric space determined by distances to a specified discrete set of objects in the space [38,39].

A neural network for Learning Vector Quantization consists of two layers: an input layer and an output layer. It represents a set of reference vectors, the coordinates of which are the weights of the connections leading from the input neurons to an output neuron. Hence, one may also say that each output neuron corresponds to one reference vector. LVQ is also defined by learning function, the activation function of neurons that allows non-linearity to be introduced into LVQ training and determines the elasticity of weight changes. It therefore can improve convergence in training, in present study we have used LVQ version 1 [31] and 2 [40,41]. LVQ is also characterized by introducing a hidden layer in training a feed-forward neural network that allows for a multitude of functions to be learned and represented. Performance of LVQ is also governed by number of hidden neurons in hidden layer [36].

The learning method of learning vector quantization is often called competition learning, because it works as

follows: For each training pattern the reference vector that is closest to it is determined. The corresponding output neuron is also called the winner neuron. The weights of the connections to this neuron - and this neuron only: the winner takes all - are then adapted. The direction of the adaptation depends on whether the class of the training pattern and the class assigned to the reference vector coincide or not. If they coincide, the reference vector is moved closer to the training pattern, otherwise it is moved farther away. This movement of the reference vector is controlled by a parameter called the learning rate [41,42]. It states as a fraction of the distance to the training pattern how far the reference vector is moved. Usually the learning rate is decreased in the course of time, so that initial changes are larger than changes made in later epochs of the training process. Learning may be terminated when the positions of the reference vectors do hardly change anymore as shown in LVQ algorithm below. In learning vector quantization classes are predefined ie a set of labeled data act as a target. Here the goal is to determine a set of prototypes the best represent each class label. The significant features identified will act as the inputs to the LVQ. Figure 3 shows a typical LVQ based classifier.

4) The generalized learning vector quantization algorithm:

Step 1: Choose the number of clusters M

Step 2: Initialize the prototypes w_{*1},... w_{*m} (one simple method for doing this is to randomly choose M vectors from the input data)

Step 3: Repeat until stopping criterion is satisfied.

Step 4: Randomly pick an input x.

Step 5: Determine the "winning" node k by finding the prototype vector that satisfies following

$$|w_{*k}-x| <= | w_{*i} - x | \text{ (for all i)} \quad (1)$$

note: if the prototypes are normalized, this is equivalent to maximizing $w_{*i}x$

Step 6: Update *only the winning prototype weights* according to

$$w_{*k}(new) = w_{*k}(old) + \mu (x - w_{*k}(old)) \quad (2)$$

Where μ is learning rate.

III. Validation and Testing

The dataset is divided into training, cross-validation and testing. Cross-validation dataset is used to measure the training performance of LVQ during training and stop training if necessary. The testing dataset is not used in any way during training and hence provides an independent measure of training performance. Divided up into training, validation and test sets. We have analyze the LVQ performance with different combination of validation and test sets containing each 10, 20 and 25 % of total data samples (60 spectra) , leaving 80,60 and 50% of total data samples (60 spectra) for training respectively.

Classification rate is calculated by building a classification matrix from number of detected samples(Cancerous classified as cancerous), false positives samples (cancer samples classified as normal), true positives samples (normal samples classified as normal), false alarms samples (normal samples classified as cancerous).

IV. Result and Discussion

A. Parameters influencing the performance of LVQ

1) Criteria for feature selection: Figure 4 shows the variation of classification rate of LVQ with respect to the feature selection criteria on the basis of ranking by ttest and Wilcoxon test as explained earlier. From figure It is worth nothing that Wilcoxon based feature selection gives more significant increase in classification rate ($P<0.05$) as compare to ttest. This is also justified by the fact that ttest assumes normal distribution while wilcoxon is non parametric test.

2) Learning rate: A range of learning rates was looked at for different combination of validation and testing input samples as explained earlier. This is done in order to give a good idea of what learning rates are most suitable for training a LVQ on the LIF dataset. This is a fairly broad investigation though and so although it was the intention to find good leaning rates, show other characteristics related resulting from training. Figure 5 shows the variation of LVQ classification rate with respect to learning rate 0.01 and 0.1 for 150 cycles. It is observed that network is relatively stable at around 0.01, and provide better classification rate ($P<0.05$). However this in contrary to the fact that, the error tends to converge more quickly as the learning rate is increased, although this increases the deviation of results obtained at around 0.1 learning rate. The results could be made more significant by looking at more runs of the network for each learning rate, and by examining more learning rates within certain ranges.

3) Learning function: Figure 6 shows the effect of learning function version 1 and 2 on LVQ classification rate. The results shows that on average, the LVQ version 2 function performs better than the other function ie LVQ version 1 since LVQ 2($P<0.01$) has more significant increase in classification rate than LVQ 1($P<0.05$).

4) Number of hidden neurons: Figure 7 shows the variation of classification rate with number of hidden neurons 2 ($P<0.05$) and 5 ($P<0.01$). It shows that rate of classification increases as number of hidden neurons increases. Using only one or two hidden neurons, the network is incapable of learning and the high error rates on

both testing and training sets indicate this. Statistically there is no significant difference between the results where the number of hidden neurons is greater than 5.

Table I shows the optimal parameters of LVQ chosen for LIF data classification. It is also worth nothing that, all figures shows either increase in classification rate or no change, as we move along x axis so it is also concluded that reduction in LIF data sample improves LVQ performance.

B. Comparison with other classifier: After selecting an optimal parameters of LVQ, we compare the performance of LVQ with comprehensive classification algorithms, support vector machines (SVM) [43]. Sequential Minimal Optimization (or SMO) [44] is applied to train Support Vector Machines (SVMs). Training an SVM requires the solution of a very large quadratic programming (QP) optimization problem. SMO breaks this large QP problem into a series of smallest possible QP problems. Both LVQ and SVM belong to the class of prototype based classifier. The basic difference between them is that SVM takes data points on the class borders as prototypes, whereas LVQ prototypes are local averages of data points, which are nearby or on the class borders. We have trained SVM with same number of rank features as used for LVQ using Wilcoxon test. We have generated classification result using SVM for LIF data set using different kernels which includes linear, quadratic, and polynomial. Kernel is one of the important concepts in SVM and plays very important role. Kernel function computes the inner product of two vectors in the feature space and thus implicitly defines the mapping function [43]. We also compared LVQ with Multi Layer Perceptron (MLP) [45] which estimates the discriminating functions from training data with fairly simple algorithms such as backpropagation that is based on gradient descent. TABLE II shows the classification accuracy for different SVM kernels. We found that In all cases LVQ has better classification accuracy than SVM on LIF data set. However, the total numbers of available samples in our LIF dataset are small at least from a neural network perspective it is limited and fragmentary. To get better understanding and comparison with SVM result, sensitivity of LVQ should be evaluated in larger LIF dataset.

V. Conclusion

In this article recently developed prototype based method, Learning Vector Quantization is reviewed in the application to build a classifier that can distinguish between cancer and control patients from the Laser Induced Fluorescence data. It is worth noting that, LVQ is an adaptive machine learning approach and inherently regularizing by non-vanishing neighborhood cooperativeness, such that they are able to handle sparse, high dimensional and noisy data generated by LIF. However, the performance of LVQ is dependant on its various tuning parameter and feature selection criteria as seen from results. The generated classification LVQ model show good performance compared to other machine learning methods thereby demonstrate the ability of Laser Induced Fluorescence technique for bladder cancer diagnosis.

VI. Figures and Tables

It should be noted that In figures 3 to 6, **X** axis represent the variation in number of selected features from original 1024 to reduced 200 data points selected from each spectrum and **Y** axis represents variation in the classification rate of LVQ in %.

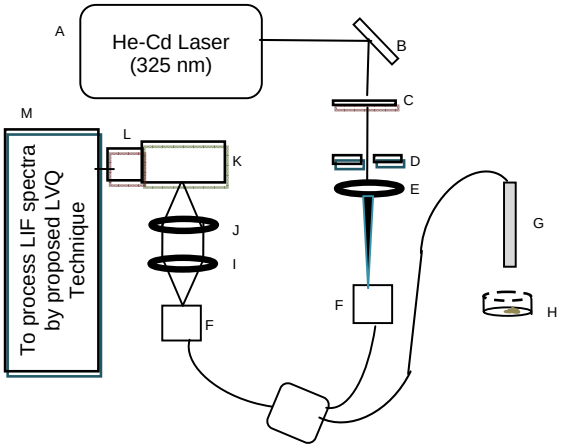

Fig. 1. A-He-Cd laser, B-Dichroic Mirror, C-Wavelength Filter, D-Iris Diaphragm, E-Focusing Lens,F-Optical Coupler, G-Probe, H-Sample ,I-Collimating lens, J-Focusing lens, K-Spectrograph, L-CCD, M-Computer

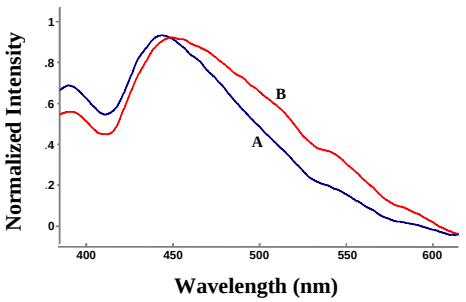

Fig. 2. Typical LIF spectra obtained from Bladder tissue of A: Healthy volunteer, B: Bladder Cancer patient

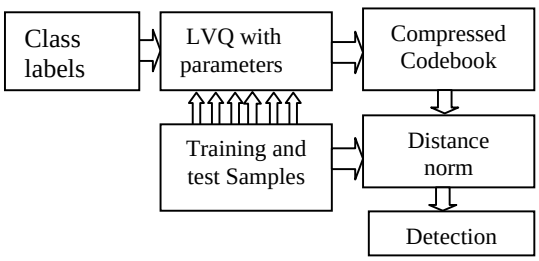

Fig. 3. Block diagram of typical LVQ classifier

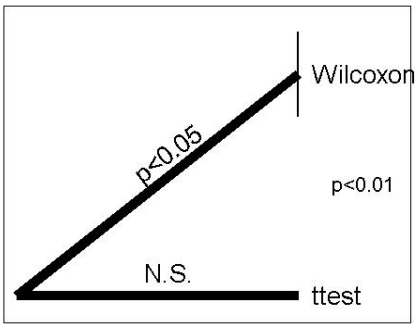

Fig. 4. Influence of feature selection criteria on the classification rate of LVQ.

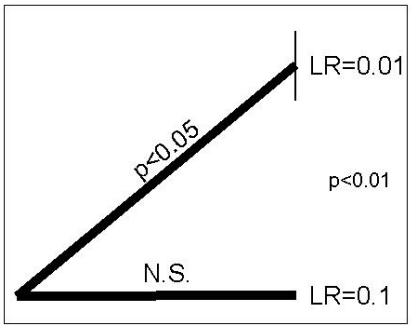

Fig. 5. Influence of learning rate on the classification rate of LVQ.

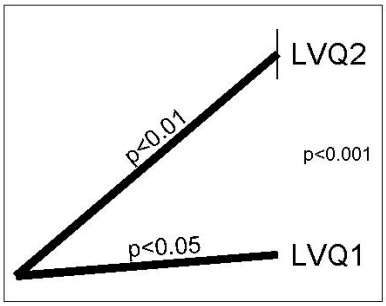

Fig. 6. Influence of selection of learning function during training on the classification rate of LVQ.

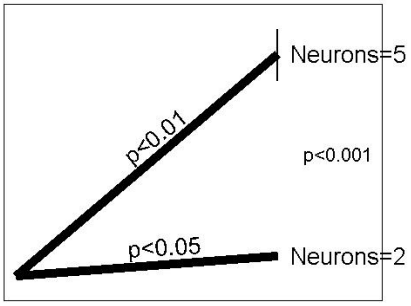

Fig. 7. Influence of Number of hidden neurons on the classification rate of LVQ.

TABLE I
OPTIMAL LVQ PARAMETERS FOR LIF DATA CLASSIFICATION

FEATURE SELECTION	LEARNING RATE	LEARNING FUNCTION	NUMBER OF HIDDEN NEURONS
WILCOXON	0.01	LVQ2	5

TABLE II
COMPARISON OF CLASSIFICATION ACCURACY OF LVQ WITH OTHER CLASSIFIERS

SVM_1	SVM_2	SVM_3	MLP	LVQ
57.7%	57.9%	59.17%	58.9%	86.11%

Where, SVM_1 uses linear kernel, SVM_2 uses quadratic kernel, SVM_3 uses polynomial kernel. MLP is multilayer percepron.

REFERENCES

[1] Robbins, *Pathologic Basis of disease*, 7th Ed, ch 21,(Elsevier)

[2] Adi F. Gazdar, "Filling the Void: Urinary Markers for Bladder Cancer Risk and Diagnosis", *J. National Cancer Institute* (2001) 93, pp. 413-414.

[3] Paweena Kreunin, "Bladder Cancer Associated Glycoprotein Signatures Revealed by Urinary Proteomic Profiling", *Journal of Proteome Research* (2007), 6, pp. 2631-2639.

[4] Antonia Vlahou," Development of a Novel Proteomic Approach for the Detection of Transitional Cell Carcinoma of the Bladder in Urine", *American Journal of Pathology* (2001)158, pp. 1491 – 1502.

[5] Herbert A. Fritsche, National Academy of Clinical Biochemistry Guidelines for the Use of Tumor Markers in Bladder Cancer, (2005)

[6] Monica McGrath, "Hormonal and Reproductive Factors and the Risk of Bladder Cancer in Women", *American Journal of Epidemiology* (2006) 163(3),pp. 236-244

[7] Morten Ostergaard, Proteome Profiling of Bladder Squamous Cell Carcinomas: Identification of Markers That Define Their Degree of Differentiation, *Cancer Research* (1997) 57,pp. 4111-4117.

[8] American Cancer Society. *Cancer Facts & Figures 2008*. Atlanta: American Cancer Society-2008

[9] Schiffer E, Challenges of using mass spectrometry as a bladder cancer biomarker discovery platform; *World J. Urol.*. (2008)26(1),pp. 67-74.

[10] Hiroshi Miyamoto, Promotion of Bladder Cancer Development and Progression by Androgen Receptor Signals, *Journal of the National Cancer Institute* (2007**)** 99(7), pp. 558 - 568.

[11] Jian Gu, Roles of tumor suppressor and telomere maintenance genes in cancer and aging—an epidemiological study; *Carcinogenesis* (May 2005) 26(10),pp. 1741-1747.

[12] Carmen J.Marsit, Promoter hypermethylation is associated with current smoking, age, gender and survival in bladder cancer; *Carcinogenesis* (2007) 28(8), pp. 1745-1751

[13] D Hleem J. Issaq,Detection of Bladder Cancer in Human Urine by Metabolomic Profiling Using High Performance Liquid Chromatography/Mass Spectrometry, *J. of. Urology* (2007), 179, pp. 2422-2426.

[14] Ramachandra R. Dasari ,Diagnosing breast cancer by using Raman spectroscopy, *PNAS* (2005), 102,pp. 12371-12376

[15] Frank Koenig, Laser Induced Autofluorescence Diagnosis Of Bladdercancer , *J. of Urology*,(1996) 156, 1597-1601.

[16] Diana C.G. de Veld, Autofluorescence characteristics of healthy Oral Mucosa at different Anatomical Sites, *Lasers in Surgery and Medicine* (2003) , 32, pp.367-376.

[17] Kevin T. Schomacker, Novel Optical Detection system for in vivo identification and localization of cervical intraepithelial Neoplasia, *J. Biomed. Optics*, (2006) 11(3),pp. 034009-1

[18] Kevin T. Schomacker , Ultraviolet Laser –Induced fluorescence of colonic tissue: Basic Biology and Diagnostic potential, *Lasers in Surgery and Medicine*, (1992)12, pp. 63-78.

[19] Mahadevan-Jansen A, Raman Spectroscopy for the Detection of Cancers and Precancers. *J. of Biomed. Optics*, (1996)1(1),pp. 40-79.

[20] B. K. Manjunath, "Autofluorescence of oral tissue for optical pathology in oral malignancy" , *Journal of Photochem. and Photobiol. B: Biology*,(2004) 73,pp. 49-58.

[21] Thomas Villmann, "Classification of mass spectroscopic data in clinical proteomics using learning vector quantization methods", *Briefings in Bioinformatics*,(2008), 9(2), pp. 129-143.

[22] Verleysen M, "The curse of dimensionality in data mining and time series prediction", In *Computational Intelligence and Bioinspired system*, IWANN(2005), pp. 758-70

[23] Wanger M, "Protocols for disease classification from mass spectroscopy data", in *Proteomics*, 3 (9), pp. 1692-8

[24] Wu B, "Comparison of statistical methods for classification of ovarian cancer using mass spectroscopy data", *Bioinformatics* 2003, 1(19), pp. 1636-43

[25] Yasui Y, "An automated peak identification/ Calibration procedure for high-dimensional protein measures from mass spectrometers ", *J. Biomed Biotechnol* 2003,1 (4),pp. 242-48

[26] T.P. Conrads, et al., "High-resolution serum proteomic features for ovarian detection", *Endocrine-Related Cancer*, 2004, 11, pp. 163-178.

[27] E.F. Petricoin, et al., "Use of proteomic patterns in serum to identify ovarian cancer", *Lancet*, 2002, 359(9306), pp. 572-577.

[28] Pierangelo Veltri, "Algorithms and tool for analysis and management of mass spectroscopy data", *Briefings in Bioinformatics*, (2008), 9 (2), pp. 144-145

[29] Annalisa Barla, "Machine learning methods for predictive proteomics", *Briefings in Bioinformatics*, (2008), 9 (3), pp. 119-128

[30] Chambers J, "Graphical Methods for data analysis", New york : Chapman and Hall, 1983.

[31] Teuvo Kohonen, " Self-Organizing Maps", Springer-Verlag, 1995.

[32] Vladimir Cherkassky, *Learning from data: Concepts, Theory and Method*, John Wiley & Sons, 1998.

[33] Richard O. Duda and Peter E. Hart. *Pattern Recognition and Scene Analysis*, John Wiley & Sons, 1973.

[34] Tommi Jaakkola, " Exploiting generative models in discriminative classifiers" In M. Kearns, S. Solla, and D.A. Cone, editors, *Advances in Neural Information Processing Systems*:(1998) NIPS, pages 487–493.

[35] Hollmen Jaakko ,"A learning vector quantization algorithm for probabilistic models". EUSIPCO 2000- X European Signal Processing Conference, 2, pp. 721-724 .

[36] S N Sivanandam, Introduction to Neural Networks using Matlab 6.0,McGraw-Hill, 2005.

[37] Neuro-dynamic programming an overview, Dimitri Bertsekas, Dept. of Electrical Engineering and computer science, MIT.

[38] Byers, J.A. "Dirichlet tessellation of bark beetle spatial attack points". Journal of Animal Ecology, (1992) , 61, pp. 759-768

[39] Halls P. J, "Dirichlet neighbours: revisiting Dirichlet tessellation for neighbourhood analysis", *Computers, environment and urban systems*, GISRUK Conference, York, 25, pp. 105-117.

[40] Poechmueller, W , "Is LVQ really good for classification?-an interesting alternative", *Neural Networks*, (1993) , 3, pp. 1207 – 1212

[41] Teuvo Kohonen, "LVQ PAK: The Learning Vector Quantization Program Package" Technical Report A30, Otaniemi (1996)

[42] *M. Ho,"*Improving Neural Network Generalization Ability Using Outlier Analysis and Voronoi Tessellation", Session 139, *AAS 207th Meeting*, 8-12 January 2006.

[43] Cristianini, N., " An Introduction to Support Vector Machines and Other Kernel-based Learning Methods", (2000).First Edition (Cambridge: Cambridge University Press)

[44] Platt, J.C.. "Sequential Minimal Optimization:A Fast Algorithm for Training Support Vector Machines". In Advances in Kernel Methods - Support Vector Learning, B. Scholkopf, J.C. Burges, and A.J. Smola, eds. (1999) (Cambridge MA: MIT Press), pp. 185–208.

[45] Gilhan Kim, " Feature selection using Multi-Layer Perceptron in HIV-1 protease cleavage data" International Conference on BioMedical Engineering and Informatics, (2008)