Anal. Chem. **2007,** *79,* 6670−6674

Application of Linear Discriminant Analysis in Prostate Cancer Research by Synchrotron Radiation-Induced X-Ray Emission

Krzysztof Banaś,*,† Andrzej Jasiński,†,‡ Agnieszka M. Banaś,† Mariusz Gajda,§ Grzegorz Dyduch,‖ Bohdan Pawlicki,⊥ and Wojciech M. Kwiatek†

Institute of Nuclear Physics PAN, Radzikowskiego 152, 31-342 Kraków, Poland, Pedagogical University, Podchorążych 2 30-084 Kraków, Poland, Department of Histology, Collegium Medicum, Jagiellonian University, Kopernika 9, 31-034 Kraków, Poland, Department of Pathomorphology, Collegium Medicum, Jagiellonian University, Grzegórzecka 16, 31-034 Kraków, Poland, and Gabriel Narutowicz Hospital, Prądnicka 37, 31-202 Kraków, Poland

The ability to visualize an object of interest is one of the cornerstones of advancement in science. For this reason, synchrotron radiation-induced X-ray emission (μ-SRIXE) holds special promise as a imaging technique in structural biology, biochemistry, and medicine. It gives the possibility to image concentration of most of the elements in a sample at high space resolution. Statistical analysis of data obtained for samples of prostate tissues in an experiment at L-beam line HASYLAB (Hamburg, Germany) is presented in this paper. The regions for the measurements were selected according to the histological view of the sample. By histological examination, samples were divided into five groups (from healthy to Gleason4, most advanced stage of cancerogenesis). Data obtained in μ-SRIXE experiments on prostate cancer samples provide information about concentrations of certain elements in these groups. The rising problem is to find out concentrations of which elements allow the researcher to discriminate between different (early mentioned) groups. Linear discriminant analysis, a basic technique for feature extraction, was used in statistical analysis of the data. Our results indicate that the use of synchrotron radiation and discriminant analysis in the study of prostate cancer tissues provide information that can be key to better understanding of biomolecular functions.

Adenocarcinoma of the prostate is now the most frequently diagnosed male malignancy with 1 in every 11 men developing the disease.[1] It is the second most common cause of cancer deaths in men.[2] Grading prostatic adenocarcinomas remains an important problem. Various systems of grading exist, but none of them seems able to reliably predict either the lethal potential of a tumor in an individual patient or the responsivness of an individual tumor to various forms of therapy.

The most frequently used grading system as proposed by Gleason[3,4] is essentially based on the description of tumor growth pattern. Gleason did not take cytological abnormalities into account in his system because he considered that these abnormalities largely correlate with gland architecture modifications. The histological slides of all cases are reviewed by the Gleason classification for which the Gleason primary pattern, i.e., the most prevalent pattern within a tumor is kept as the final grade. The grading scale runs from 1 to 5, with 1 being the least aggressive form of cancer. The lower the number, the slower the cancer growth. High-grade tumors (4 and 5) grow more quickly and are more likely to spread to other parts of the body than low-grade tumors (1 and 2). Proper assessing cancerous stage depends on the ability, experience, and attention of the histopathologist, but obviously sometimes misclassification appears. The histopathologist's analysis discriminates structures only by their microscopic appearance but not by any biochemical processes taking place in them. In order to help with this difficulty of classification, one can try to solve the discrimination problem by using a completely different set of parameters as predicators.

Synchrotron radiation is used for a number of applications to biology. It has long been recognized that the ability to visualize an object of interest is one of the cornerstones of advancement in science. For this reason, X-ray imaging holds special promise as a technique in structural biology. Synchrotron radiation-induced X-ray emission (μ-SRIXE) gives the ability to image the concentration of most of the elements in a sample at high resolution.

Linear discriminant analysis (LDA) is a well-known technique for dimension reduction and feature extraction. It was used here to find an optimal transformation that maps the data into a lower-dimensional space (while preserving the class structure) that minimizes the within-class distance and simultaneously maximizes the between-class distance, thus achieving maximum discrimination. The original thought of LDA came from Fisher's classical

* To whom correspondence should be addressed. Tel: +48-12-6628256. Fax: +48-12-6628089. E-mail: Krzysztof.Banas@ifj.edu.pl.

† Institute of Nuclear Physics PAN.

‡ Pedagogical University.

§ Department of Histology, Jagiellonian University.

‖ Department of Pathomorphology, Jagiellonian University.

⊥ Gabriel Narutowicz Hospital.

(1) Hayward, S.; Cunha, G. *Radiol. Clin. North Am.* **2000**, *38* (1), 1−15.

(2) Landis, S. H.; Murray, T.; Bolden, S.; Wingo, P. A. *Cancer J. Clin.* **1999**, *49*, 8−31.

(3) Gleason, D. F. *Cancer Chemother. Rep.* **1966**, *50*, 125−128.

(4) Gleason, D. F. In *Urologic Pathology: The Prostate*; Tannenbaum, M., Ed.; Lea and Febiger: Philadelphia, PA, 1977: pp 171−197.

10.1021/ac070931u CCC: $37.00
Published on Web 08/03/2007

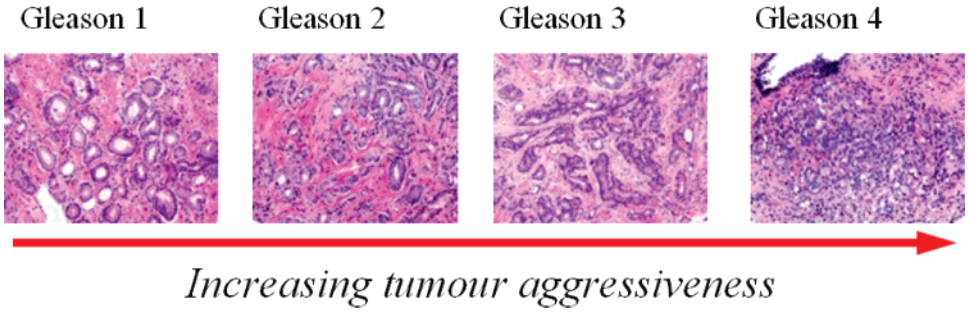

Figure 1. Examples of the prostate tissue sections with different Gleason scores (size about 700 by 900 μm^2).

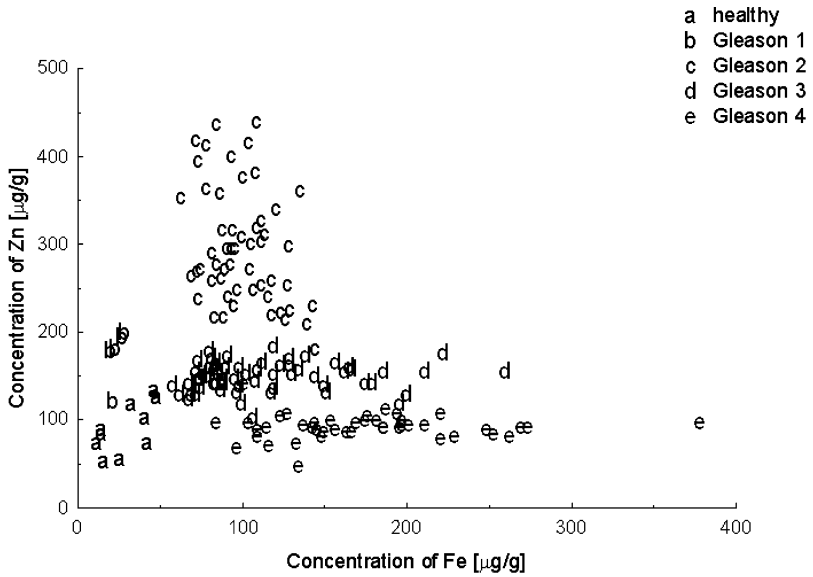

Figure 2. Concentrations of iron and zinc in different Gleason score-related groups (μg/g).

paper.[5] The basic idea underlying discriminant function analysis is to establish whether groups differ with regard to the mean of a variable and then to use that variable to predict group membership (e.g., of new cases).

Multivariate methods in statistics and LDA among them are a perfect tool for multidimensional data analysis.[6,7] These statistical techniques make it possible to investigate the dependences and connections among variables, helping to reduce and to simplify data, and also, what is more important, to classify objects into groups. It has been used widely in many applications such as face recognition, text classification, microarray data classification, etc.[8−13]

Data obtained in a μ-SRIXE experiment on prostate cancer samples provide information about concentrations of certain elements in these groups. We were focused on the issue: if is possible to grade prostate tissue with results obtained during SRIXE experiments?

EXPERIMENTAL SECTION

Samples. The samples of prostate tissue were obtained from six patients operated due to *Adenocarcinoma prostatae*. The specimens were frozen in liquid nitrogen and cut into 14-μm-thick slices with the use of a cryomicrotome. The slice designed for elemental analysis was immediately placed onto mylar foil. The commercially available mylar foil is transparent for X-rays, thin (2.5 μm), and pure (contains no trace metals).

Adjacent sections were put on a microscopic glass for later histological examination.[14] All samples were classified based on patomorphological examination into different Gleason grades (examples of prostate stained tissue sections are presented in Figure 1). In order to obtain concentrations of the elements, two multielemental external standards[15] were prepared.

Measurements. The measurements were carried out on the L-beam line at the HASYLAB, DESY (Hamburg, Germany), which is characterized by the following parameters: electron energy 4.45 GeV, critical energy 16.04 keV, and maximum ring current ~120 mA.

(5) Fisher, R. A. *Ann. Eug.* **1936**, *7*, 178−188.
(6) Fukunaga, K. *Introduction to Statistical Pattern Recognition*; Academic Press: Boston, 1990.
(7) Hastie, T.; Tibshirani, R.; Friedman, J. H. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*; Springer: New York, 2001.
(8) Belhumeour, P. N.; Hespanha, J. P.; Kriegman, D. J. *IEEE Trans. Pattern Anal. Machine Intelligence* **1997**, *19* (7), 711−720.
(9) Dudoit, S.; Fridlyand, J.; Speed, T. P. *J. Am. Stat. Assoc.* **2002**, *97*, 77−87.
(10) Gardner, S.; Roux, N. J. *J. Classification* **2005**, *22* (1), 59−86.
(11) Ke, W.-M.; Ye, Y.-N.; Huang, S. *J. Gastroenterology* **2003**, *38* (9), 861−864.
(12) Airoldi, J.-P.; Flury, B. D.; Salvioni, M. *J. Theor. Biol.* **1995**, *177* (3), 247−262.
(13) Kwak, N. K.; Kim, S. H.; Lee, Ch. W.; Choi, T. S. *J. Med. Syst.* **2002**, *26* (5), 427−438.
(14) Kwiatek, W. M.; Banas, A.; Gajda, M.; Galka, M.; Pawlicki, B.; Falkenberg, G.; Cichocki, T. *J. Alloys Compd.* **2005**, *401*, 173−177.
(15) Kwiatek, W. M.; Hanson, A. L.; Paluszkiewicz, C.; Galka, M.; Gajda, M.; Cichocki, T. *J. Alloys Compd.* **2004**, *362*, 83−87.

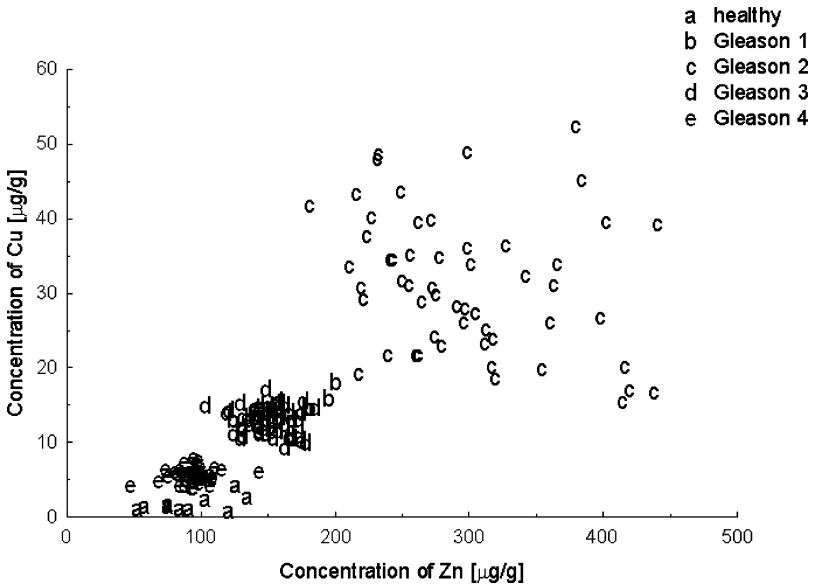

Figure 3. Concentrations of zinc and copper in different Gleason score-related groups (μg/g).

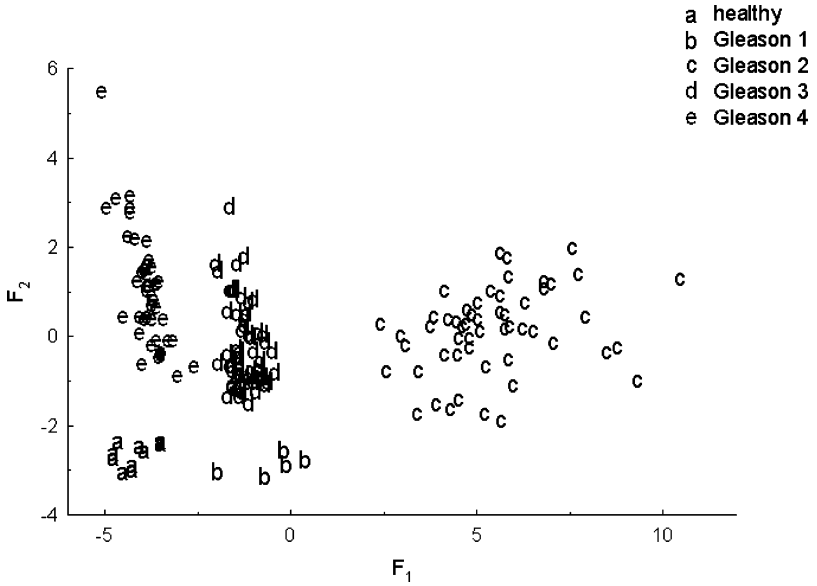

Figure 4. Results of the linear discrimination analysis with concentrations of Fe, Zn, Mn, and Cu as predicators, (F_1 and F_2 are two discriminant functions defined by eq 1.)

The capillary was used to obtain spectra from single points (beam size of 14 μm in diameter was achieved).[16,17] X-ray fluorescence was detected by use of a conventional HPGe X-ray detector positioned at 90° with respect to the incident beam. The energy resolution of the detector was 140 eV for 5.9 keV. All experiments were performed in air. At this beam line with setup described above, it is possible to achive the minimum detection limit for Fe, Zn, Cu, and Mn ~0.1 μg/g, using the K lines as analytical signal.[16,17]

The position of the sample, with respect to the X-ray beam, was changed by means of a computer-controlled stepper motor allowing micrometric movements. This equipment allows us to analyze different structures of prostate tissues under the same conditions.

The acquisition time of the single spectrum was equal to 300 s. The collected spectra were deconvoluted with the use of the AXIL program working under Linux OS. All the computed X-ray line intensities were normalized to the value of the incident photon flux.

RESULTS AND DISCUSSION

After the experiments, all spectra were carefully analyzed in order to obtain reliable concentration levels of certain elements present in cancerous tissues with different Gleason grades and healthy tissues. Special attention was paid to calculate the concentrations of Fe, Zn, Cu, and Mn.

Fe, Cu, and Mn are suspected to take part in cancerogenesis processes as a catalyzer of the Fenton reaction.[18] On the other hand Cu, Mn, and Zn are the prosthetic groups of several

(16) Haller, M.; Knöchel, A. *J. Trace and Microprobe Techniques* **1996**, *14* (3), 461.

(17) Lechtenberg, F.; et al. *J. Trace Microprobe Tech.* **1996**, *14* (3), 561.

(18) Ozen, O. A.; Yaman, M.; Sarsilmaz, M. *J. Trace Elem. Med. Biol.* **2003**, *16*, 119−122.

Table 1. Canonical Discriminant Coefficients (Eigenvectors)

	F_1	F_2
a^{Fe}	−0.006	0.022
a^{Zn}	0.017	−0.002
a^{Mn}	0.275	0.260
a^{Cu}	0.140	−0.026
a^0	−6.257	−3.393

Table 2. Centers of the Groups and Squared Mahalanobis Distance between the Groups

centers of groups			squared Mahalanobis distance			
	F_1	F_2	Healthy	Gleason1	Gleason2	Gleason3
Healthy	−4.241	−2.618				
Gleason1	−0.530	−2.860	13.830			
Gleason2	5.440	0.158	**101.428**	44.749		
Gleason3	−1.261	−0.255	14.464	7.320	45.074	
Gleason4	−3.875	1.085	13.846	26.752	87.629	8.629

metalloenzymes including superoxide dismutase, which is an important antioxidant enzyme in the cellular protection from reactive oxygen species.[19]

Figures 2 and 3 present calculated concentrations of (Fe, Zn) and (Zn, Cu) obtained from the same place in a tissue sample. Different letters and different gray levels show spots belonging to different groups. As one can see when taking into account only concentrations of two elements, problems with misclassification appear (Figures 2 and 3). The natural solution to this problem is to use statistical analysis to improve performance in discrimination between various stages of prostate carcinoma by taking into account concentrations of more than two elements.

Linear Discriminant Analysis. The basic idea underlying discriminant function analysis is to establish whether groups differ with regard to the mean of a variable and then to use that variable to predict group membership (e.g., of new cases). Computationally, discriminant function analysis is very similar to analysis of variance.[20,21]

In this paper, we show how discriminant analysis methodology can be combined with synchrotron radiation-based experiments leading to highly informative visual displays of the respective class separations. The same relevant quantities were measured for all samples in the set, and there is some distinct outcome or occurrence that determines group membership for each unit. The objective of LDA is to find a set of factor weights that best separates the groups, given a set of units for which group membership is already known. The resulting set of weights may then be used to predict group membership for new units.

Classification of the different groups was based on the concentration of Fe, Zn, Cu, and Mn. The discriminant functions were calculated based on 175 classified observations (Healthy, 10; Gleason1, 5; Gleason2, 54; Gleason3, 62; Gleason4, 44).

Table 3. Confusion Matrices for Discrimination

		Fe Zn						Zn Cu						Fe Zn Cu Mn					
		predicted						predicted						predicted					
actual	N	Healthy	Gleason1	Gleason2	Gleason3	Gleason4	corr discrim (%)	Healthy	Gleason1	Gleason2	Gleason3	Gleason4	corr discrim (%)	Healthy	Gleason1	Gleason2	Gleason3	Gleason4	corr discrim (%)
Healthy	10	9	**1**	0	0	0	90.0	8	0	0	0	**2**	80.0	10	0	0	0	0	100.0
Gleason1	5	**1**	4	0	0	0	80.0	0	4	0	**1**	0	80.0	0	5	0	0	0	100.0
Gleason2	54	0	**1**	47	**6**	0	87.0	0	**2**	52	0	0	96.3	0	0	54	0	0	100.0
Gleason3	62	**2**	**1**	0	51	**8**	82.3	0	**10**	0	52	0	83.9	0	0	0	60	**2**	96.8
Gleason4	44	**2**	0	0	**5**	37	84.1	**1**	0	0	0	43	97.7	0	0	0	**2**	42	95.5
Apparent Error Rate (%)		**15.43**						**9.14**						**2.29**					

(19) Halliwell, B.; Gutteridge, J. M. C. *Free Radicals in Biology and Medicine,* 3rd ed.; Oxford University Press: New York, 1999.

(20) Sanathanan, L. Discriminant analysis. In *Introductory multivariate analysis*; Amick, D.; Walberg H., Eds.; McCutchan Publishing Corp.: Berkeley, CA, 1975; pp 236−256.

(21) Brown, M. T.; Wicker, L. R. Discriminant analysis. In *Handbook of Applied Multivariate Statistics and Mathematical Modeling,* 1st ed.; Tinsley, H. E. A., Brown, S. D., Eds.; Elsevier Inc.: New York, 2000; pp 209−235

Table 4. Judgment of the New Cases Based on Concentration of the Elements and Discriminant Analysis Results

	distance					input				
judgment	Healthy	Gleason1	Gleason2	Gleason3	Gleason4	Fe	Zn	Mn	Cu	minimum score
Gleason3	39.115	22.475	29.108	**14.679**	34.545	**50**	**200**	**10**	**10**	14.679
Healthy	**0.300**	12.770	92.158	11.014	11.402	**40**	**100**	**2**	**3**	0.300

Two discriminant functions representing 93.7 and 6.3% of the variance were separated. The original features of the objects (concentrations of the elements in the tissue) were replaced by coordinates of the discriminant functions in the scatterplot of observations (Figure 4). Results presented in Figure 4 show five evidently separable classes corresponding to five groups: Healthy, Gleason1, Gleason2, Gleason3, and Gleason4.

In the Table 1, canonical discriminant coefficients (eigenvectors) are presented for the functions F_1 and F_2.

Discriminant functions are defined by the following equation:

$$F_j = a_j^0 + a_j^{Fe}C_j^{Fe} + a_j^{Zn}C_j^{Zn} + a_j^{Mn}C_j^{Mn} + a_j^{Cu}C_j^{Cu} \qquad (1)$$

where $j = 1, 2$; a_j^{Fe}, a_j^{Zn}, a_j^{Mn}, and a_j^{Cu} are the canonical discriminant coefficients for function F_j; and C_j^{Fe}, C_j^{Zn}, C_j^{Mn}, and C_j^{Cu} are the concentrations of Fe, Zn, Mn, and Cu, respectively.

The measure of the differentiation between groups in discriminant analysis is the squared Mahalanobis distance. In Table 2, coordinates of the centers of the groups and squared Mahalanobis distances between these centers are presented. The largest Mahalonobis distance was noticed between the group denoted as Healthy and Gleason2.

Table 3 shows the confusion matrices of the classification when the discriminant analysis is carried out. Description Actual used in these tables means genuine assessment of the structure of prostate tissues done by histopathologist, whereas Predicted means the judgment done by taking into account parameters calculated by discriminant function analysis.

As can be seen, the apparent error rate is 2.3% (Table 3) in the case of LDA with four concentrations of the elements (Fe, Zn, Cu, Mn) and is much lower than results obtained in analysis with only two concentrations as variables (Fe−Zn, apparent error rate 15.4%; Zn−Cu, apparent error rate 9.1%).

Using the results of linear discriminant analysis (canonical discriminant coefficients and coordinates of the centers of the groups), one can classify new cases into one of the categories based on minimum distance criteria. An example of this kind of calculations with two different input concentrations data sets is shown in Table 4.

CONCLUSIONS

The histological analysis of tissue samples by focusing on the bulk morphological features lacks quantitative accuracy and gives no account for the biochemical background of the changes that occur during cancerogenesis.

The results show that concentrations of only two elements are weak discriminators for the various grades of the Gleason classification system and, particularly, show misclassification between the high (Gleason4) and the other Gleason grades, i.e., the low (Gleason1 and 2) and intermediate (Gleason3) ones. In the case of LDA with concentrations of four elements (Fe, Zn, Cu, Mn), the percentage of correct classifications is 97.7% (only 4 misclassification cases appear in a group of 175 observations).

In the discriminant analysis, the squared Mahalanobis distance was used as a measure of the differences between examined populations.

Discriminant analysis, like any predictive method, focuses on the relationship between a data matrix of predictor variables and a response variable. Primarily, it aims to optimally separate different classes of objects to provide a classification rule for allocating entities of unknown origin to one of a known set of classes and in addition serves as an aid in interpreting the role of the different predictor variables in class separability.

The presented statistical investigations indicated that multivariate methods such as discriminant analysis is a very useful technique that can be successfully used for the analysis. The use of this method permits classification of the biological ensemble representing different stages of prostate cancer cases in separate groups and also application of these results to unknown cases.

Our results show that the use of synchrotron radiation in order to obtain concentrations of the elements on prostate cancer tissues and statistical processing of the data by the discriminant function analysis provide information that can be key to full understanding of biomolecular functions. Both these methods can be regarded as complementary techniques generating a synergetic outcome in the form of valuable and accurate results.

ACKNOWLEDGMENT

This work was supported by the European Community-Research Infrastructure Action under the FP6 Structuring the European Research Area Programme (through the Integrated Infrastructure Initiative Integrating Activity on Synchrotron and Free Electron Laser Science) Contract RII3-CT-2004-506008. HASYLAB Project II-04-79EC.

Received for review May 8, 2007. Accepted July 2, 2007.

AC070931U