

Computer-Based Malignancy Grading of Astrocytomas Employing a Support Vector Machine Classifier, the WHO Grading System and the Regular Hematoxylin-Eosin Diagnostic Staining Procedure

Dimitris Glotsos, M.Sc., Panagiota Spyridonos, Ph.D., Panagiotis Petalas, M.Sc., Dionisis Cavouras, Ph.D., Panagiota Ravazoula, M.D., Petroula-Arampatoni Dadioti, M.D., Ioanna Lekka, M.D., and George Nikiforidis, Ph.D.

OBJECTIVE: To investigate and develop an automated technique for astrocytoma malignancy grading compatible with the clinical routine.

STUDY DESIGN: One hundred forty biopsies of astrocytomas were collected from 2 hospitals. The degree of tumor malignancy was defined as low or high according to the World Health Organization grading system. From each biopsy, images were digitized and segmented to isolate nuclei from background tissue. Morphologic and textural nuclear features were quantified to encode tumor malignancy. Each case was represented by a 40-dimensional feature vector. An exhaustive search procedure

in feature space was utilized to determine the best feature combination that resulted in the smallest classification error. Low and high grade tumors were discriminated using support vector machines (SVMs). To evaluate the system performance, all available data were split randomly into training and test sets.

RESULTS: The best vector combination consisted of 3 textural and 2 morphologic features. Low and high grade cases were discriminated with an accuracy of 90.7% and 88.9%, respectively, using an SVM classifier with polynomial kernel of degree 2.

CONCLUSION: The proposed methodology was based

From the Computer Laboratory, School of Medicine, University of Patras, and Department of Pathology, University Hospital, Rio, Patras; Department of Medical Instrumentation Technology, Technological Education Institution of Athens, Athens; and Department of Pathology, METAXA General Anticancer Hospital, Piraeus, Greece.

Dr. Glotsos is Doctoral Candidate, Computer Laboratory, School of Medicine, University of Patras.

Dr. Spyridonos is Postdoctoral Researcher, Computer Laboratory, School of Medicine, University of Patras.

Dr. Petalas is Master of Science Candidate, Computer Laboratory, School of Medicine, University of Patras.

Dr. Cavouras is Professor, Department of Medical Instrumentation Technology, Technological Education Institution of Athens.

Dr. Ravazoula is Physician, Department of Pathology, University Hospital.

Drs. Dadioti and Lekka are Physicians, Department of Pathology, METAXA General Anticancer Hospital.

Dr. Nikiforidis is Professor, Computer Laboratory, School of Medicine, University of Patras.

Funded by the Greek Ministry of Education and the European Union under research fund HRAKLEITOS—Development of New Machine Learning Methods for Medical Image Processing and Analysis Applications.

Address reprint requests to: George Nikiforidis, Ph.D., Computer Laboratory, School of Medicine, University of Patras, 265 00 Rio, Patras, Greece (gnikif@med.upatras.gr).

Financial Disclosure: The authors have no connection to any companies or products mentioned in this article.

on standards that are common in daily clinical practice and might be used in parallel with conventional grading as a second-opinion tool to reduce subjectivity in the classification of astrocytomas. (*Analyt Quant Cytol Histol* 2004;26:77-83)

Keywords: astrocytoma, computer-assisted diagnosis, support vector machine.

Astrocytomas (ASTs) constitute the majority of primary brain tumors and are considered one of the most lethal and difficult-to-treat forms of cancer.¹ In the diagnosis of astrocytomas, grade characterization is of major clinical importance since it provides an index of disease severity and influences treatment.² Grading is performed by visual observation and recognition of structural tissue characteristics during microscopic inspection. According to the World Health Organization (WHO) grading system,³ the existence or absence of cellular pleomorphism, mitoses, endothelial proliferation and necrosis classifies astrocytic tumors on the basis of their aggressiveness in 3 grades of malignancy: ASTs, anaplastic astrocytomas (ANAs) and glioblastoma multiforme (GBM). ASTs are the least malignant low grade tumors and generally have a good prognosis.⁴ High grade tumors (ANAs and GBMs) are the most aggressive, characterized by rapid growth and a tendency to invade nearby healthy tissue.⁵ It is essential for clinicians to define disease severity because tumor grade is strongly related to survival and treatment selection.^{1,2}

Pathologists' subjective grade assessment has been shown to influence diagnostic accuracy.⁶ Factors that may provoke diagnostic errors include biopsy size, tumor heterogeneity, subjectivity and previous experience.⁷ Problems with tumor grade assignment are most apparent in the discrimination of low from high grade tumors, mostly because AST malignancy develops along a biological continuum.⁸ Indeed, as emphasized by Decaestecker et al,⁹ there is no specific biological criterion to establish boundaries between low and high grade tumors. Misdiagnosis may lead to inadequate therapy of high grade or to aggressive therapy for low grade tumors.⁷ It is thus necessary to objectify tumor classification.

Automated tumor grading has been extensively examined¹⁰⁻¹⁶ and remains an active research area.¹⁷ Most previous studies¹⁰⁻¹⁴ utilized modifications of the WHO grading system (SAMS, RTOG, HOM, Burger and Vogel) and specialized staining

procedures (Feulgen, Ki-67). Decaestecker et al¹⁵ analyzed cytometry-related variables extracted from Feulgen-stained biopsies. Their system is based on the nearest neighbor classifier and discriminates different grades of astrocytic tumor with accuracy of 55%. As emphasized by the authors, their system had difficulty discriminating grade 2 from grade 3 tumors. Sallinen et al¹⁰ developed a decision-tree model evaluating histologic features closely related to the WHO grading criteria. The authors concluded that the combined analysis of Feulgen-stained slides for the generation of nuclear features, Ki-67-stained slides for mitosis counting and hematoxylin-eosin (HE)-stained slides for counting tumor necrosis provides a means to improving the prognostic efficiency of the WHO scheme. Belacel et al¹⁶ presented a fuzzy logic-based system analyzing quantitative nuclear features extracted from Feulgen-stained slides for automated grading of ASTs. The method is compatible with the WHO scheme and discriminates different-grade tumors with overall accuracy of 66%. In a recent study, Reinhold et al¹⁷ investigated the differences in the nuclear placement pattern in low (grade 2) and high (grade 3) grade astrocytomas. The authors evaluated quantitative topometric nuclear features by analyzing Ki-67-stained images as possible descriptors of cell nuclear malignancy. Their system discriminates grade 2 from grade 3 tumors with accuracy of 88%.

Little effort has been made¹⁸ to construct an automated grading approach that may be incorporated into clinical practice. It is essential to investigate computer-aided grading techniques based on HE staining because this protocol is the one used in clinical practice⁹ but is not as accurate as the specialized nuclear staining methods used in other studies.¹⁰⁻¹⁷

The aim of this study was to investigate the possibility of performing automated grading based on the WHO classification scheme and HE staining. These protocols are the most widely accepted by clinicians for everyday clinical use.¹⁸ Computer-assisted grading was performed by the analysis of quantitative morphologic and textural nuclear features. Low and high grade tumors were discriminated utilizing support vector machines (SVMs),¹⁹ state-of-the-art learning systems originating with statistical learning theory. The support vector classifier is designed to find the optimal separating hyperplane among the existing categories based on geometric criteria. Data are initially mapped into a

higher-dimensional feature space. In this feature space, the SVM constructs a separating hyperplane that has the maximum distance from the closest training data. This geometrically defined hyperplane enables the machine to create a classifier with a low probability of making errors; that means that it will perform well on unseen data. SVM does not risk trapping in local minima, needs no iterative procedures and does not assume a priori knowledge of the underlying data probability functions. Its setting is much easier when compared to those of conventional classifiers (e.g., multilayer neural networks).¹⁹ The possibility that SVM may be generalized, ensuring good performance even with limited training samples,²⁰ made the selection of this approach most attractive. To the best of our knowledge, SVM have never been used before for the classification of histopathological images. In contrast to those in previous studies,¹⁰⁻¹⁷ the proposed methodology was validated for clinical material collected from 2 hospitals.

Materials and Methods

Archival material from formalin-fixed, paraffin-embedded tissue samples was obtained from 140 patients who had undergone surgery at the University Hospital, Patras (92 cases), and METAXA General Anticancer Hospital, Piraeus, Greece (48 cases) between 1993 and 2002. Patients' ages ranged from 11 to 70 years for the AST group, from 14 to 77 for the ANA group and from 13 to 83 for the GBM group. All patients were treated with partial or total

tumor resection. Most patients with high grade tumors were treated postoperatively with radiation and/or chemotherapy.

On average, 5 HE-stained sections were generated from the same block for each patient. Sections were placed on slides for microscopic examination. Slides were previewed to ensure that crucial histologic characteristics (e.g., presence or absence of necrosis) were similar for all sections for each case.⁶ Tumor grade was defined as low or high according to the WHO grading system by 2 independent pathologists. In cases of discrepant diagnoses, pathologists reviewed the slides on a multiheaded microscope to make a unanimous decision concerning tumor grade. Of the 140 astrocytomas, 61 were classified as low grade (ASTs) and 79 as high grade (30 ANAs and 49 GBMs). For each slide, a pathologist specified the most representative region. From this region, images were digitized (Figure 1) (768×576×8 bit) at a magnification of 400× using a light microscopy imaging system consisting of a Zeiss Axiostar plus microscope (Welwyn Garden City, U.K.) and an Ikegami color video camera (Maywood, New Jersey, U.S.A.).

The automatic classification methodology consisted of 5 distinct stages: image segmentation, feature generation, feature selection, classification and system performance evaluation.

A segmentation algorithm²¹ was applied (Figure 2) to separate nuclei from surrounding tissue in order to quantify diagnostic features from cell nuclei. Two kinds of features were generated: 18 morphologic features related to the size and shape of cell nuclei and 22 textural features (first-order, co-occurrence, run-length based) that encoded chromatin distribution and nuclear DNA content (Table I).²² After feature generation, each case was represented by a 40-dimension feature vector, with each vector element formed by the mean feature value of all examined nuclei.

To safeguard the classifier generalization, it is essential to reduce feature dimensionality and select the optimum feature subset with the highest-class discrimination information. The exhaustive search algorithm guaranteed finding the optimal subset by combining features in all possible ways (combinations of 2, 3, 4, 5 features, etc.).²³ The best feature combination was the 1 that led to the smallest classification error according to SVM classifier self-consistency.²⁴ The self-consistency method estimates the reliability of the classifier using all available data, first for training and then for testing.

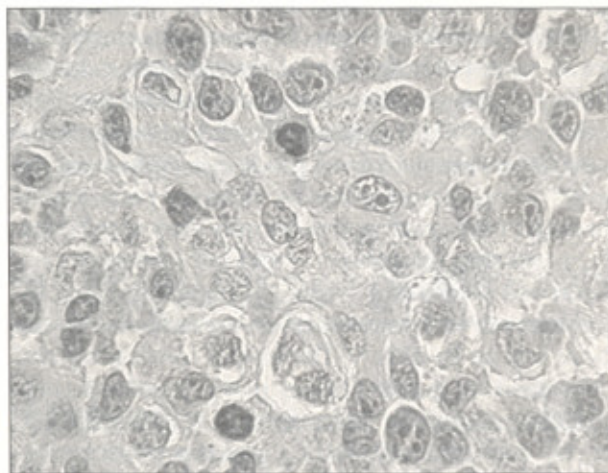


Figure 1 Typical HE-stained image of brain AST (×400).

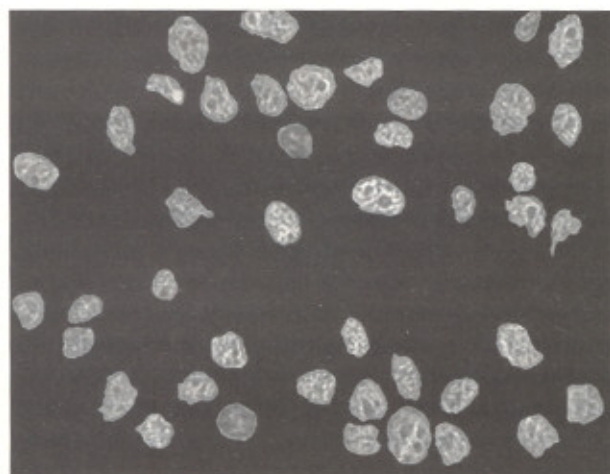


Figure 2 Segmented image.

To evaluate system performance, data from two-thirds cases were randomly selected from each class and used to train the SVM classifier. The remaining one-third was used for testing the classifier. This procedure was repeated 100 times. The average performance on the 100 test sets was calculated.

The basic idea behind applying SVM to classification problems is 2-fold: the machine initially maps the input space in a higher-dimensional feature space through a nonlinear transformation function (kernel) and consequently "seeks" the optimal separating hyperplane with the maximum distance from the closest training data.²⁰ The discriminant function of the SVM classifier for binary classification problems is:

$$g(x) = \text{sign} \left(\sum_{i=1}^N a_i y_i K(x, x_i) + b \right), \quad (1)$$

where x_i is the training data belonging to either class $y_i \in \{+1, -1\}$, N is the number of training samples, and a_i , b weight coefficients and $K(x, x_i)$ are the kernel function.¹⁹

Construction of the SVM classifier required the determination of 2 parameters: the kernel function and the cost factor C , which specifies the importance of misclassifications.²⁵ The support vector classifier was constructed with polynomial kernels of degree 1, 2, 3 and 4 and the radial basis function (RBF) kernel with values of $\gamma = 1/(2\sigma^2)$, ranging from 0.005 to 6. For either kernel function, C was experimentally set equal to 100:

$$K_{RBF}(x, x_i) = \exp \left(\frac{-\|x - x_i\|^2}{2\sigma^2} \right), \quad \sigma = \text{spread}, \text{ and} \quad (2)$$

$$K_{POLYNOMIAL}(x, x_i) = ((x^T x_i) + 1)^d, \quad d = \text{degree}. \quad (3)$$

The optimization problem²⁰ was solved by using the routine quadprog provided with the MATLAB optimization toolbox (MathWorks Inc., Natick, Massachusetts, U.S.A.).

Results

The best vector combination consisted of 3 textural features— inertia, inverse different moment and correlation—and 2 morphologic ones, roundness and kurtosis of roundness. Using all available data for training as well as testing (self-consistency), low and high grade cases were discriminated with an accuracy of 100%. Optimum kernel configuration was polynomial with degree 2.

Retaining the SVM model configuration and using the selected feature vector, system performance was evaluated. On average, for the 100 simulations, the success rates of correct prediction of low and high grade cases were 90.7% and 88.9%, respectively. The performance of alternative SVM models with different kernel functions was also tested (Table II).

Discussion

The classification of ASTs into low or high grade defines groups of patients who are significantly different in regard to duration of symptoms and length of survival.¹ Thus, accurate distinction between low and high grade tumors is fundamentally important. Furthermore, as emphasized by Prayson et al,⁶ the grade of a tumor determines treatment planning and clinical management. For example, many high grade tumors are frequently treated with radiation therapy, whereas many low grade ones are not.⁷ Although the WHO grading scheme provides accurate definitions for tumor grade determination, the relative importance given by different pathologists to each of the grading criteria may vary significantly, promoting interobserver and intraobserver variation and decreasing diagnostic reproducibility. Interobserver and intraobserver reproducibility can be as low as 36–51%.^{6,7} The lack of consistency and consensus in grading raises many problems complicating diagnostic interpretation. It is necessary to revise conventional grading practice by introducing more objective techniques.

Toward this goal, automated tumor grading ap-

Table I Morphologic and Textural Features of the Cell Nucleus

Features encoding DNA content and chromatin distribution			
1st Order statistical		2nd Order statistical	
Based on DNA histogram	Co-occurrence matrix*	Run length matrix	Features describing shape and size
Density	Energy	Short run emphasis	Area**
Standard deviation	Inertia	Long run emphasis	Roundness**
Kurtosis	Entropy	Gray level nonuniformity	Concavity**
Skewness	Correlation	Run length nonuniformity	
	Cluster shade		
	Cluster prominence		
	Contrast		

*Each feature was computed with interpixel distance $d=1$ and $d=3$.

**For each feature its maximum value, standard deviation, skewness and kurtosis were computed.

proaches have been developed. Most previous studies utilized modifications of the WHO grading scheme¹¹⁻¹³ and evaluated quantitative nuclear features generated by Feulgen- and Ki-67-stained sections.¹⁰⁻¹⁶ The Feulgen staining procedure selectively stains DNA, but it must be analyzed the same day or under controlled conditions after several days.²⁶ Ki-67 staining is used specifically for mitosis counting,¹⁰ which accounts for only 1 of the 4 grading criteria in the WHO scheme.³

To support the regular diagnostic procedure followed by experts, it would be essential to investigate and develop automated techniques that are compatible with clinical routine. The present work extended the research on automated grading and formulated a methodology that relies on the WHO classification scheme and the analysis of HE stained nuclei images. These are the standards widely accepted by clinicians for use in daily practice.

Automated grading was performed using the SVM principle and extracting diagnostic information from nuclear appearance. The feature vector combination that carried the greatest discriminatory information between low and high grade tumors was composed of 2 morphologic features (roundness, kurtosis of roundness) and 3 textural ones (in-

ertia, inverse different moment, correlation). Roundness is an index of the irregularity of nuclear boundaries. Kurtosis of roundness is a measure of uniformity of nuclear shape distribution. Inertia describes the inhomogeneity of chromatin texture of the nucleus. In high grade nuclei, chromatin clusters form clumps and are more irregular, resulting in lower values of inertia in comparing low grade nuclei. Concerning inverse different moment, it accepts high values for low-contrast nuclei and describes the existence of dense and sparser areas within the cell nucleus. Finally, correlation encodes information on chromatin texture and is higher for higher grade tumors.²²

The SVM classifier self-consistency for the best feature vector was 100%. This result is an optimistically biased estimate of the classifier performance since the same dataset was used for both training and testing.²³ However, it is an important step for assessing the reliability of the classifier in correctly classifying the data that we used for its design. Thus, the best feature combination determined was employed in evaluating the classifier's performance. The evaluation was estimated by a more objective method, splitting all available data into random training and test sets.²⁴

Table II Performance Evaluation of the SVM Classifier with Different Kernel Functions by Splitting All Available Data into Random Training and Test Sets

Classification task	SVM with RBF kernel (best $\gamma=1$)	SVM with polynomial kernel of degree d (%)				
		$d=1$	$d=2$	$d=3$	$d=4$	
Low versus high grade tumors	Low grade	88.4	85.1	90.7	94.3	84.8
	High grade	87.6	85.4	88.9	86.7	94.6
	Overall accuracy	88.0	85.8	89.7	88.7	89.7

SVM configuration with a polynomial kernel of degree 2 optimized classification performance, resulting in 89.7% overall accuracy in discriminating low from high grade tumors (Table II). Different kernel configurations resulted in reduced performance due to overfitting.

The results were satisfactory considering that low and high grade tumors form a biological continuum. Indeed, there is no specific criterion to establish distinct boundaries between low and high grade tumors, promoting sources of diagnostic misinterpretation.⁹ Reduced performance in the discrimination of high grade tumors may be due to the large heterogeneity identified in the cytologic characteristics of tumors belonging to this category.¹⁵ Although it is believed that the misclassification of high grade cases is more severe, recent studies have indicated that diagnostic errors in identifying low grade tumors are equally important.⁴ As stated by Coons, incorrect identification of a low grade tumor as high grade may lead to aggressive therapy, with adverse effects on survival.⁷ From this point of view, although the polynomial kernel of degree 4 more accurately discriminates high grade tumors (Table II), it is not selected as optimum due to its reduced performance in low grade tumor identification. The polynomial kernel of degree 2 combines relatively high classification rates for both low (90.7%) and high grade (88.9%) cases.

The results obtained with the SVM classifier are comparable to those obtained by cross-validation discriminant analysis¹⁷ and better than those obtained by the nearest-neighbor approach¹⁵ and fuzzy logic models¹⁶ proposed for automated grading of ASTs. However, the staining method employed in the present study is the one used in daily clinical practice and is not as accurate in staining nuclei as the specialized methods used in other studies. Additionally, in contrast to previous studies,¹⁰⁻¹⁷ our methodology was validated for clinical material collected from 2 hospitals.

The method described above is accurate and compatible with standards that are common in clinical practice and could be used in parallel with conventional grading as a second-opinion tool to reduce subjectivity in the classification of ASTs.

References

- DeAngelis LM: Brain tumors. *N Engl J Med* 2001;344:114-123
- Shapiro W, Shapiro J: Biology and treatment of malignant gliomas. *Oncology* 1998;12:233-240
- World Health Organization: International Histological Classification of Tumors: Histological Typing of Tumors of the Central Nervous System. Second edition. Berlin, Springer-Verlag, 1993
- Kaye A, Walker D: Low grade astrocytomas: Controversies in management. *J Clin Neurosci* 2000;7:475-483
- Herfarth K, Gutwein S, Debus J: Postoperative radiotherapy of astrocytomas. *Semin Surg Oncol* 2001;20:13-23
- Prayson RA, Agamanolis DP, Cohen ML, Estes ML: Interobserver reproducibility among neuropathologists and surgical pathologists in fibrillary astrocytoma grading. *J Neurol Sci* 2000;175:33-39
- Coons W, Johnson P, Sceithauer B, Yates A, Pearl D: Improving diagnostic accuracy and interobserver concordance in the classification and grading of primary gliomas. *Cancer* 1997;79:1381-1393
- Kleihues P, Cavenee WK (editors): Pathology and Genetics of Tumors of the Nervous System. Lyon, International Agency for Research on Cancer, 2000
- Decaestecker C, Camby I, Nagy N, Brotchi J, Salmon I: Improving morphology-based malignancy grading schemes in astrocytic tumors by means of computer-assisted techniques. *Brain Pathol* 1998;8:29-38
- Sallinen P, Sallinen S, Helen T, Rantala I, Rautiainen E, Helin H, Kalimo H, Haapsalo H: Grading of diffusely infiltrating astrocytomas by quantitative histopathology, cell proliferation and image cytometric DNA analysis. *Neuropathol Appl Neurobiol* 2000;26:319-331
- Scarpelli M, Bartels P, Montironi R, Galluzzi C, Thompson D: Morphometrically assisted grading of astrocytomas. *Analyt Quant Cytol Histol* 1994;16:351-356
- McKeown MJ, Ramsay D: Classification of astrocytomas and malignant astrocytomas by principal component analysis and a neural net. *J Neuropathol Exp Neurol* 1996;55:1238-1245
- Kolles H, van Wangenheim A, Rahmel J, Niedermayer I, Feiden W: Data-driven approaches for decision making in automated tumour grading: An example in astrocytoma grading. *Analyt Quant Cytol Histol* 1996;18:298-304
- Scarpelli M, Montironi R, Thompson D, Bartels P: Computer-assisted discrimination of glioblastomas. *Analyt Quant Cytol Histol* 1997;19:369-375
- Decaestecker C, Salmon I, Dewitte O, Camby I, Van Ham P, Pasteels J, Brotchi J, Kiss R: Nearest-neighbor classification for identification of aggressive versus nonaggressive astrocytic tumors by means of image cytometry-generated variables. *J Neurosurg* 1997;86:532-537
- Belacel N, Boulassel M: Multicriteria fuzzy assignment method: A useful tool to assist medical diagnosis. *Artif Intell Med* 2001;21:201-207
- Reinhold N, Schlote W: Topometric analysis of diffuse astrocytomas. *Analyt Quant Cytol Histol* 2003;25:12-18
- Spyridonos P, Zolota V, Cavouras D, Zenebissis G, Glotsos D, Nikiforidis G: A computer-based image analysis system for classification of astrocytomas according to the WHO grading system. Proc of International Conference on Signal Processing, Pattern Recognition and Applications, Crete, Greece, June 25-28, 2002, pp 371-374

19. Burges CJC: A tutorial on support vector machines for pattern recognition. *Data Mining Knowl Discov* 1998;2:121-167
20. Kechman V: Support vector machines. *In Learning and Soft Computing*. Cambridge, Massachusetts, Massachusetts Institute of Technology, 2001, pp 121-184
21. Spyridonos P, Cavouras D, Ravazoula P, Nikiforidis G: Neural network-based segmentation and classification system for the automated grading of histologic sections of urinary bladder carcinoma. *Analyt Quant Cytol Histol* 2002;26:317-324
22. Theodoridis S, Koutroubas K: Feature generation: II. *In Pattern Recognition*. San Diego, Academic Press, 1999, pp 233-270
23. Fellow J, Duin R, Mao J: Statistical pattern recognition: A review. *IEEE Trans Pattern Anal Machine Intell* 2000;22:4-37
24. Fukunaga K, Hayes R: Estimation of classifier performance. *IEEE Trans Pattern Anal Machine Intell* 1989;11:1087-1101
25. Christianini N, Shawe-Taylor J: Support vector machines. *In An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge, Cambridge University Press, 2000, pp 93-122
26. Bancroft JD, Stevens A: *Theory and Practice of Histological Techniques*. Edinburgh, Churchill Livingstone, 1990