journal homepage: www.intl.elsevierhealth.com/journals/cmpb

Improving accuracy in astrocytomas grading by integrating a robust least squares mapping driven support vector machine classifier into a two level grade classification scheme

Dimitris Glotsos[a,*], *Ioannis Kalatzis*[a], *Panagiota Spyridonos*[b], *Spiros Kostopoulos*[b], *Antonis Daskalakis*[b], *Emmanouil Athanasiadis*[b], *Panagiota Ravazoula*[c], *George Nikiforidis*[b], *Dionisis Cavouras*[a]

[a] *Department of Medical Instruments Technology, Technological Educational Institution of Athens, Ag. Spyridonos Street, Aigaleo, Athens 122 10, Greece*
[b] *Medical Image Processing and Analysis Laboratory, Medical Physics, School of Medicine, University of Patras, Rio, Patras 265 00, Greece*
[c] *Department of Pathology, University Hospital, Rio, Patras 265 00, Greece*

ARTICLE INFO

Article history:
Received 11 July 2007
Received in revised form
16 January 2008
Accepted 16 January 2008

Keywords:
Astrocytomas
Support vector machines
Least squares mapping
Computer-assisted microscopy

ABSTRACT

Grading of astrocytomas is an important task for treatment planning; however, it suffers from significantly great inter-observer variability. Computer-assisted diagnosis systems have been propose to assist towards minimizing subjectivity, however, these systems present either moderate accuracy or utilize specialized staining protocols and grading systems that are difficult to apply in daily clinical practice. The present study proposes a robust mathematical formulation by integrating state-of-art technologies (support vector machines and least squares mapping) in a cascade classification scheme for separating low from high and grade III from grade IV astrocytic tumours. Results have indicated that low from high-grade tumours can be correctly separated with a certainty as high as 97.3%, whereas grade III from grade IV tumours with 97.8%. The overall performance was 95.2%. These high rates have been a result of applying the least squares mapping technique to features prior to classification. A significant byproduct of least squares mapping is that the number of support vectors of the SVM classifiers dropped dramatically from about 80% when no mapping was used to less than 5% when mapping was used. The latter is a clear indication that the SVM classifier has a greater potential to generalize well to new data. In this way, digital image analysis systems for automated grading of astrocytomas are brought closer to clinical practice.

* *Corresponding author at*: Medical Image & Signal Processing Lab (MEDISP), Department of Medical Instruments Technology, Technological Educational Institute of Athens, Ag. Spyridonos Street, Aigaleo, Athens 122 10, Greece. Tel.: +30 210 5385375.
E-mail address: dimglo@teiath.gr (D. Glotsos).
0169-2607/$ – see front matter © 2008 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.cmpb.2008.01.006

1. Introduction

Malignancy grading of astrocytomas is fundamentally important since it affects accurate treatment planning and patient management [1]. Pathologists decide on the aggressiveness of astrocytic tumours by visually examining tissue section slides (biopsies) with the microscope [2]. According to guidelines published by the World Health Organization (WHO) [3], three grades are established on the basis of histological criteria: grade II and grade III and grade IV. Grade II (low-grade) astrocytomas are the least malignant tumours and have generally good prognosis with survival up to 5 years. Astrocytomas of grade III and IV (high grade) are the most aggressive tumours, characterized by a rapid growth pattern and a tendency to invade nearby healthy tissue; survival time for high-grade tumours ranges on average from 6 months to 1 year [2].

However, grade differentiation, has been shown susceptible to great inter-observer variability [4]. Even though the WHO grading scheme retains its popularity among existing grading schemes (such as Daumas-Dupport [5], the Kernohan [6], the St. Anne/Mayo [7], and the HOM [8]), it has been questioned by many experts regarding the accuracy of its descriptions used to define each grade has been questioned by many experts [4,9]. The latter has generated a lack of consensus among experts regarding to the selection of a single best grading scheme. As a consequence the exchange of histological data among different laboratories, which could improve standardization and reproducibility, remains the most problematic issue in grading of astrocytomas.

Another source of complication is the utilization of different staining protocols from different laboratories. Again there is a lack of consensus regarding the selection of a single, or even a combination of staining protocols able to unravel and highlight the histological evidence on tumour slides, which would clarify the grade of a tumour [10–13]. The routine (fast, cost-effective, simple) staining protocol is the Haematoxylin–Eosin (H&E) [1]. However, H&E stained images present the highest degree of complexity regarding image-processing tasks due to the diversity of the structures stained and the severe variations in staining intensity as compared to Feulgen [14] and to the ki-67 [15] stained images. Finally, laboratories use different kinds of measuring equipment, from simple oculars to expensive microscopy imaging systems, which, on one hand are fundamentally important in adding value to ongoing research, however, on the other hand are rarely used in clinical practice.

In this study we continue our investigation and we propose a mathematical formulation able, on one hand, to significantly boost up the accuracy in the crucial separation of both low from high grade and grade III from grade IV tumours, while ensuring on the other hand good generalization to new data. The proposed method differs from others in two key issues. (a) Accuracy; we will show that the combination of support vector machines [16] and least squares mapping techniques, which has been evaluated by our group in the field of ultrasonic image analysis [17], results in the highest classification rates presented in literature in astrocytomas grading based on analysis of H&E-stained images. The latter is essential since although H&E staining produces complex to process images, H&E remains the standard choice in daily clinical practice. Thus, automated grading based on such images brings digital image analysis systems closer to real clinical practice. (b) Robustness; we will prove that the proposed method ensures good generalization to unseen data. The support vector machine is among the few algorithms that allow a mathematical assessment of its ability to generalize to unseen data, an issue that has not been comprehensively investigated in the field of astrocytomas grading by previous studies.

2. Methods

2.1. *Material*

The clinical material comprised 150 biopsies of astrocytomas collected from the Department of Pathology of the University Hospital of Patras, Greece. Five H&E stained sections were generated from the same block (patient). Of the 150 astrocytomas, 61 were classified as low grade (grade II) and 89 as high grade (40 grade III and 49 grade IV) according the WHO grading system. Table 1 illustrates the variants of different astrocytic tumours utilized in this study. For each slide, a histopathologist (P.R.) marked the most representative region. From this region, images (Fig. 1) were digitized ($768 \times 576 \times 8$ bit) using a light Zeiss Axiostar plus microscope (Zeiss; Germany) connected to an Leica DC 300 F (Leica; Germany) color video camera.

2.2. *Segmentation of H&E-stained images of astrocytomas*

Subsequently, images were segmented using a pixel-based pattern recognition methodology designed to identify textural differences among regions of nuclei and surrounding tissue. Segmentation is an essential process for this application, since from segmented nuclei, which are the most important structures in our images, features were subsequently extracted to encode distinct tumour grade characteristics. The algorithm for nuclei segmentation has been presented by our group elsewhere [18].

To facilitate evaluation of the segmentation procedure, for each patient the original and the segmented images were displayed on the screen. Segmented nuclei were automatically labeled on the original and segmented images by processing the binary-segmented image using connecting component analysis as described in [19]. Following, the labeled binary image was superimposed to the original image (see Fig. 1a–d). The histopathologist (P.R.) verified the correctness of the labeled segmented nuclei against those of the original image and also observed the non-labeled nuclei. The number of wrongly identified and/or missed nuclei was noted against the number of correctly identified ones. Wrongly identified nuclei were considered by the pathologist cases of overlapped, ‘missed’ and corrupted nuclei. The number of wrongly identified over correctly identified was calculated. The accuracy of the segmentation algorithm for each patient’s set of images was calculated as the ratio of the number of correctly over the number of all identified nuclei.

Table 1 – Variants of different astrocytic tumours utilized in this study

	Variants	Number of cases	Total number of cases
Astrocytoma (WHO grade II)	Gemistocytic	8	61
	Fibrillary	19	
	Mixed	34	
Astrocytoma (WHO grade III)	Anaplastic	40	40
Glioblastoma multiforme (WHO grade IV)	Giant cell	13	49
	Gliosarcoma	1	
	Pleomorphic	35	
			150

2.3. *Automatic grade classification*

2.3.1. *Design of the classification scheme*

Grade classification scheme was built as a two-level hierarchical cascade scheme (Fig. 2). In the first level low (grade II) from high-grade (grade III–IV) tumours were discriminated. In the second level, correctly classified high-grade cases were further categorized as grade III or IV tumours. At each level, a set of morphological and textural features were extracted to encode tumour malignancy. Features we transformed using the least squares mapping [20] technique. Based on the transformed features, classification was performed using an SVM classifier.

2.3.2. *Feature extraction*

After segmentation, a set of 40 morphological and textural features was extracted from correctly segmented nuclei describing each individual case (patient). Morphological features describe the shape and size of nuclei (18 features) and comprised area, roundness and concavity. For each one of these features the mean value, standard deviation, range, skewness, kurtosis, and maximum value was calculated [10,21]. Textural features (first-order, co-occurrence, run-length based) were used to encode chromatin distribution and nuclear DNA content (22 features). These features have been proven to encode significant information concerning malignancy status [10,21].

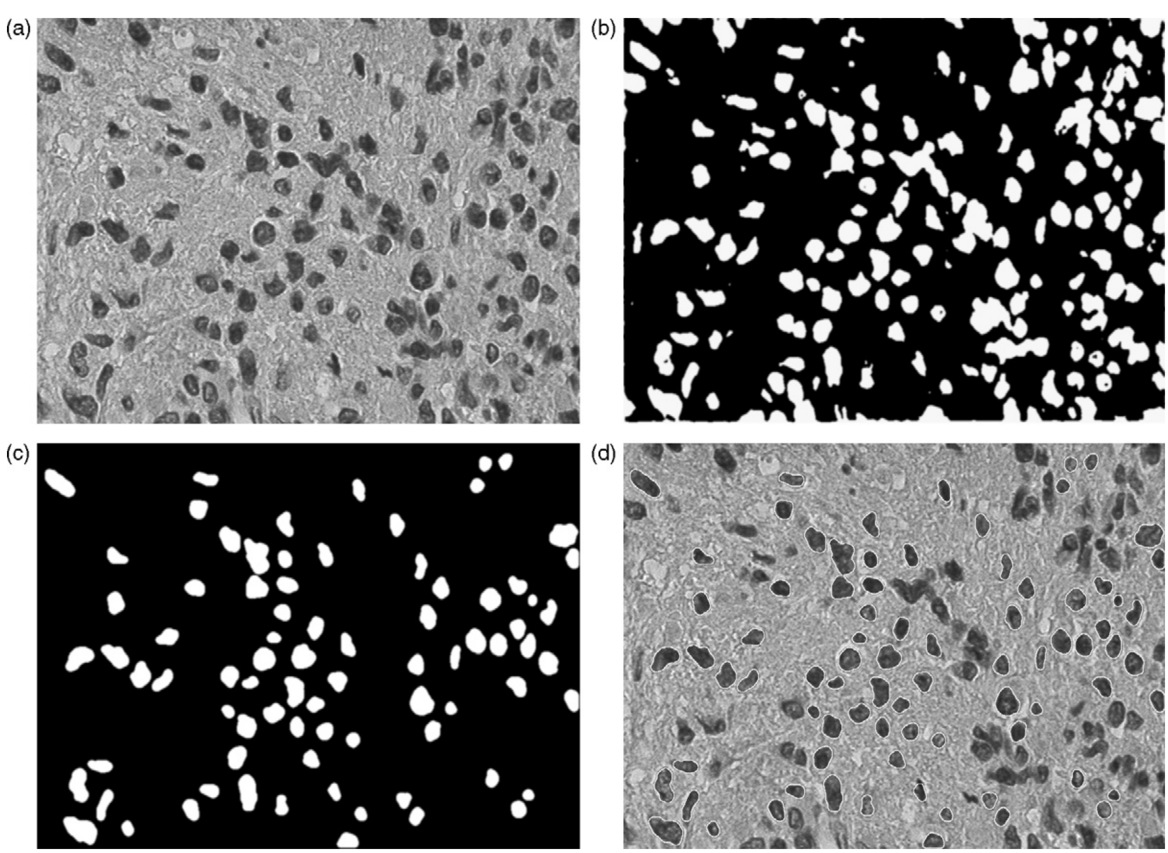

Fig. 1 – (a) H&E stained image of astrocytomas (magnification: 400×). (b) Resulted binary image after applying the pixel-based classification algorithm. (c) Small and noisy regions elimination by size and morphological filtering. Overlapped nuclei were eliminated. (d) Final segmented nuclei image.

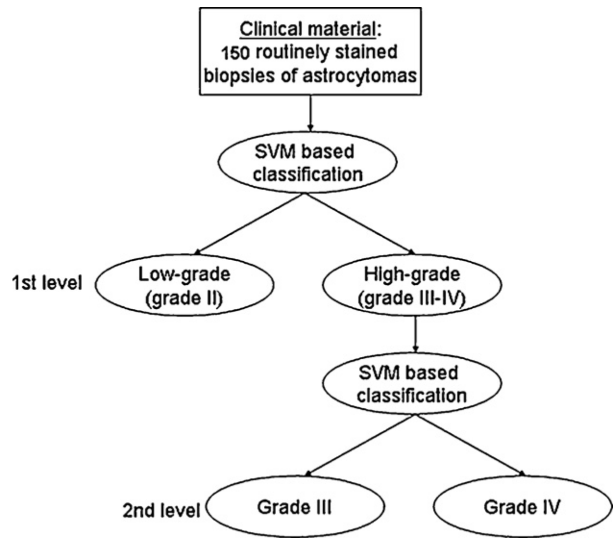

Fig. 2 – SVM-based cascade classification tree methodology.

2.3.3. *Feature selection and classification*

Feature dimensionality was initially reduced to 14 features using a ranking feature selection method based on a Wilcoxon class separability criterion [22]. Subsequently, the 14 pre-selected features were split into 13 subsets (first subset comprised the two most important features as described by the ranking method, second subset the three most important features, . . ., last subset all 14 features). Each subset was transformed according to the least squares mapping technique described in the following Section 2.3.4 using three different mapping degrees, first, second and third degree.

Each feature subset, thus formed, was next used in the design of each classifier with purpose of determining the feature subset that provided highest classification accuracy following a k-fold cross validation procedure [22]. Accordingly, data were randomly split into $k = 10$ subsets of approximately same size. One of the k subsets was held out for testing (test data), whereas the remaining $k - 1$ subsets were used to train the SVM classifier (training data). This process was repeated k times, such as each subset was treated only once as test data. At each repetition, we made sure that the training data comprised at the first level 43 low-grade and 52 high-grade cases and at the second level 26 grade III and 32 grade IV cases. Additionally, we made sure that the left-out dataset (test data) comprised at the first level 18 low-grade and 37 high-grade cases and at the second level 14 grade III and 17 grade IV cases.

Finally, the probabilistic neural network (PNN) [23], and the k-nearest neighbor classifiers [24] were implemented as alternatives to the SVM classifier, and the sequential floating forward selection (SFFS) [25] as alternative to the Wilcoxon-based ranking feature selection method for comparative purposes.

2.3.4. *Least squares mapping technique*

The least squares mapping technique targets to increase class separability and it consist of the transformation of pattern vectors around arbitrary pre-selected points in the $\mathbb{R}^C$ space (where C is the number of classes) called the *decision space*, in such a way that the least squares transformation error is minimized [20].

Prior to the transformation, pattern vectors are expanded so they contain all polynomial terms of 2nd, 3rd, . . ., kth-degree combinations of their feature elements. As an example, in case of $k = 3$, the pattern vector $\mathbf{x} = [x_1\ x_2]$ will expand to $\hat{\mathbf{x}} = [x_1\ x_2\ x_1^2\ x_2^2\ x_1x_2\ x_1^3\ x_2^3\ x_1^2x_2\ x_1x_2^2]$. Given that the dimensionality of a pattern vector is d, then the expanded pattern vector dimensionality is equal to [24]:

$$\hat{d} = \frac{(d+k)!}{k!d!} - 1 \tag{1}$$

In the present study, the polynomial expansion has limited to second and third degree terms, due to the increased computational demands that higher dimensionality pattern vectors create [16,24], especially when employing classifiers with polynomial kernels; moreover, higher dimensionality decision boundaries often lead to over-fitting results.

After the polynomial expansion, the least squares transformation of the expanded feature vectors follows. Let $\hat{\mathbf{x}}_k$ an expanded pattern vector of class k in $\mathbb{R}^{\hat{d}}$ space (where $k = 1,2,\ldots,C$ and C the number of classes) and $\mathbf{p}_k = [p_{1k}\ p_{2k} \ldots p_{Ck}]$ an arbitrary selected vector point in the $\mathbb{R}^C$ space. Let the transformation:

$$\hat{\mathbf{x}}_k \in \mathbb{R}^{\hat{d}} \mapsto \mathbf{z}_k = \mathbf{T} \cdot \hat{\mathbf{x}}_k \in \mathbb{R}^C \tag{2}$$

where $\mathbf{T}$ the transformation matrix, which is define by means of minimizing the least squares error (e_{LS}) between vectors $\mathbf{z}_k$ and $\mathbf{p}_k$ for all k:

$$e_{\text{LS}} = \sum_{k=1}^{C} \left(\frac{1}{C} \sum_{i=1}^{N_k} (\mathbf{T}\hat{\mathbf{x}}_{ki} - \mathbf{p}_k)^T (\mathbf{T}\hat{\mathbf{x}}_{ki} - \mathbf{p}_k) \right) \tag{3}$$

where N_k is the number of patterns of class k.

The e_{LS} minimization is performed by solving the following equation over $\mathbf{T}$:

$$\nabla_{\text{T}} e_{\text{LS}} = 0 \tag{4}$$

which, in conjunction to Eq. (3), leads to:

$$\mathbf{T} = \left[\sum_{k=1}^{C} \left(\frac{1}{N_k} \sum_{i=1}^{N_K} \mathbf{p}_k \hat{\mathbf{x}}_{ki}^T \right) \right] \left[\sum_{k=1}^{C} \left(\frac{1}{N_k} \sum_{i=1}^{N_K} \hat{\mathbf{x}}_{ki} \hat{\mathbf{x}}_{ki}^T \right) \right]^{-1} \tag{5}$$

2.3.5. *SVM classifier configuration*

The basic idea behind the SVM approach for binary classification problems is to (a) map the input space into a higher dimensional feature space through a linear or non-linear transformation function (kernel) and (b) in that feature space, compute a separating hyper plane that effectively splits data into the two classes of interest. This hyper plane is optimal in the sense that it has maximum distance from the closest-to the hyper plane-training data (the so-called support vectors). The discriminant function of the SVM classifier has the following form [16,26,27]:

$$g(x) = \left(\sum_{i=1}^{N} \alpha_i y_i K(\mathbf{x}, \mathbf{x}_i) + b \right) \tag{6}$$

where x_i represent each of the $i = 1, \ldots, N = 40$ three-dimensional training input feature vectors, $y_i \in \{+1, -1\}$, α_i are the Langrage multiplies, b is a weighting coefficient, and K is the kernel function. To find the optimum performance structure, the SVM classifier was alternatively constructed with the radial basis function (RBF) and polynomial kernels of degree $d = 1$ and 2 (see Eqs. (7) and (8)). Parameter σ was experimentally determined after examining values from $\sigma = 0.1$ to 4. The optimization problem of finding the Langrage multiplies was solved by using the routine *quadprog* provided with the *MATLAB* optimization toolbox [28].

$$K_{\mathrm{RBF}}(x, x_i) = \exp\left(\frac{-\|x - x_i\|^2}{2\sigma^2} \right) \tag{7}$$

$$K_{\mathrm{polynomial}}(x, x_i) = ((x^T x_i) + 1)^d \tag{8}$$

3. Results

3.1. *H&E image segmentation*

According to the evaluation procedure described in Section 2.2, the algorithm correctly recognized and segmented 88% of all nuclei. About 12% of nuclei were erroneously or inadequately circumscribed. Fig. 1a illustrates an H&E image of astrocytomas. The binary image in Fig. 1b consists of nuclei (white) and surrounding tissue (black) as recognized by the pixel-based segmentation algorithm. 'Damaged' nuclei located at the image boundaries, small (less than 150 pixels) and noisy regions were omitted by size filtering and morphological operations (Fig. 1c). In Fig. 1d the final segmentation result is illustrated by superimposing the original image (Fig. 1a) with the final corrected binary image (Fig. 1c).

3.2. *Grade classification*

Tables 2 and 3 demonstrate the error estimates for all classifiers at both levels of the cascade classification scheme using different degrees of least squares mapping. The SVM with polynomial kernel of degree 2 using seven features and least squares mapping of third degree was selected as the best classifier configuration for the first level of the cascade scheme, since it gave the highest classification accuracy (97.3% overall accuracy), and the minimum number of support vectors (≈2% of all training data). The best feature combination comprised five textural features (from the co-occurrence matrix with inter-pixel distance $d = 1$ and $d = 3$ *homogeneity*$^{d=1}$, *correlation*$^{d=1}$, *energy*$^{d=1}$, *homogeneity*$^{d=3}$, *correlation*$^{d=3}$, *energy*$^{d=3}$ and from the run length matrix *gray level non-uniformity*) and one morphological (*maximum of roundness*).

Considering the second level of the cascade scheme, the highest classification accuracy (97.8% overall accuracy) was achieved using an SVM classifier with RBF kernel, 10 features and least squares mapping of third degree. Best features com-

Table 2 – First level of the cascade classification[a] scheme: average classification rates of SVM, PNN, and *k*-nn classifiers over 10 random splits of available data into a training set (95 cases consisting of 43 low-grade and 52 high-grade cases) and a test set (55 cases, comprising 18 low-grade and 37 high-grade cases)

Classifier	No mapping				Mapping LS^1				Mapping LS^2				Mapping LS^3			
	Features	Low	High	Overall	Features	Low	High	Overall	Features	Low	High	Overall	Features	Low	High	Overall
SVM RBF	13	91.8	98.9	96.0	12	63.9	87.6	78.0	13	93.4	96.6	95.3	12	90.2	100.0	96.0
SVM $d = 1$	12	60.7	86.5	76.0	14	63.9	86.5	77.3	13	95.1	95.5	95.3	8	96.7	96.6	96.7
SVM $d = 2$	12	75.4	91.0	84.7	12	62.3	87.6	77.3	13	95.1	95.5	95.3	7	98.4	96.6	97.3
PNN	12	75.4	91.0	84.7	14	21.3	100.0	68.0	14	95.1	95.5	95.3	8	96.7	96.6	96.7
KNN	13	65.6	87.6	78.7	4	73.8	91.0	84.0	13	96.7	95.5	96.0	8	97.5	95.9	96.6

[a] Low vs. high-grade tumours.

Table 3 – Second level of the cascade classification[a] scheme: average error rates of SVM, PNN, and *k*-nn classifiers over 10 random splits of available data into a training set (58 cases consisting of 26 grade III and 32 grade IV cases) and a test set (48 cases, comprising 14 grade III and 17 grade IV cases)

Classifier	No mapping				Mapping LS^1				Mapping LS^2				Mapping LS^3			
	Features	GIII	GIV	Overall	Features	GIII	GIV	Overall	Features	GIII	GIV	Overall	Features	GIII	GIV	Overall
SVM RBF	12	90.0	100.0	95.5	9	75	75.5	75.3	9	92.5	100	96.6	10	95.0	100.0	97.8
SVM $d = 1$	9	80.0	71.4	75.3	9	90	63.3	75.3	14	95.0	98.0	96.6	8	97.5	95.9	96.6
SVM $d = 2$	12	92.5	93.9	93.3	10	70	77.6	74.2	13	97.5	98.0	97.8	8	92.5	100.0	96.6
PNN	12	97.5	79.6	87.6	9	57.5	89.8	75.3	14	95.0	98.0	96.6	8	97.5	95.9	96.6
KNN	10	85.0	77.5	80.9	8	82.5	81.6	82.0	14	95.0	98.0	96.6	8	97.5	95.9	96.6

[a] Grade III vs. Grade IV tumours.

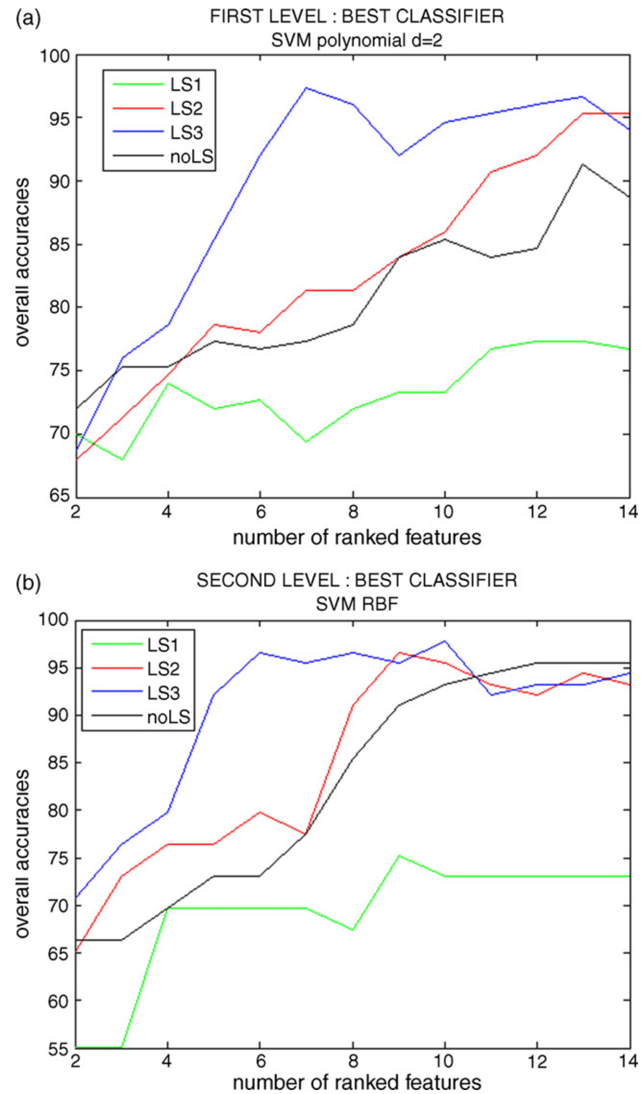

Fig. 3 – (a) Results for the best classifier (SVM polynomial of second degree) when no mapping and least squares mapping of different degrees were used for the first level of the cascade scheme. (b) Results for the best classifier (SVM RBF) when no mapping and least squares mapping of different degrees were used for the second level of the cascade scheme.

prised the same seven features resulted in the first level plus *range of roundness*, *maximum of roundness* and *maximum of area*. The number of support vectors for this configuration was as low as 5% of training data used to build the classifier for each run of the cross validation procedure. Table 4 presents classification results for the variants of different astrocytic tumours utilized in this study.

The overall performance of system was computed as the product of best overall performances at each node of the cascade classification tree (95.2%) [29]. Fig. 3 indicates the performance of best classifiers for the first and second level of the cascade scheme when designed without mapping, mapping of first, second and third degree. Fig. 4 illustrates the number of support vectors needed to construct the SVM classifier for

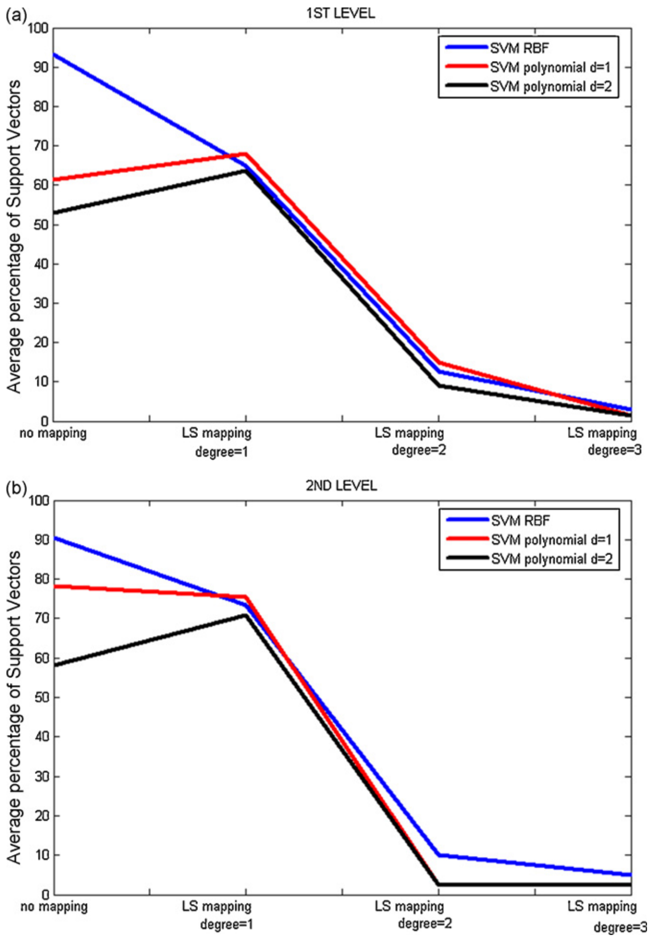

Fig. 4 – (a) Number of support vectors of SVM classifiers when no mapping and least squares mapping of different degrees were used for the first level of the cascade scheme. (b) Number of support vectors of SVM classifiers when no mapping and least squares mapping of different degrees were used for the second level of the cascade scheme.

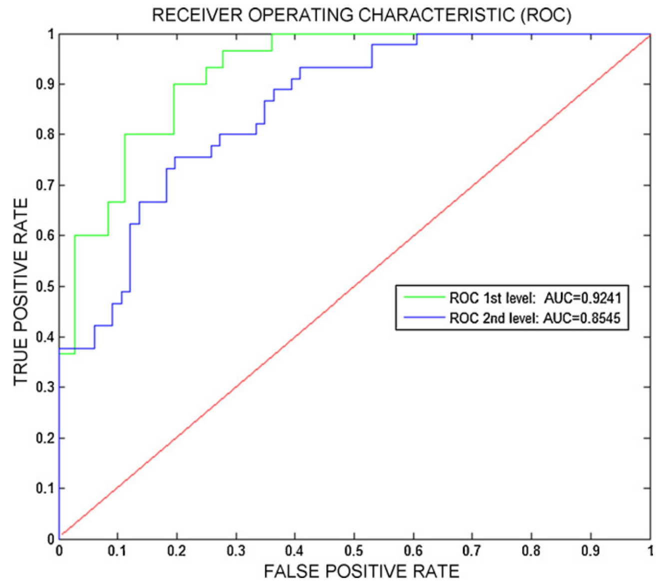

Fig. 5 – ROC curves and corresponding area under the ROC curve (AUC) values for the first and second level of the cascade classification scheme.

different degrees of least squares mapping and for both level of the cascade scheme. Fig. 5 demonstrates the system's performance in terms of receiver operating characteristic (ROC) curves. ROC curves were designed as described in [30].

4. Discussion

Grading of astrocytomas is an important task for treatment planning, however, it suffers from significantly great inter-observer variability [4,9]. Computer-assisted diagnosis systems [13,31-36] have been proposed to assist towards minimizing subjectivity, however, these systems present either moderate accuracy or utilize specialized staining protocols and grading systems that are difficult to apply in daily clinical practice. Schad et al. [13] have used Feulgen-staining, semi quantitative nuclear features and quadratic discriminant analysis to establish a system able to classify tumours according to kernohan grades with 94% accuracy. Decaestecker et al. [32] proposed a nearest neighbor classification technique with 55% success rate based on the WHO guidelines, Feulgen staining and quantitative nuclear features. Belacel and Boulassel [31] presented a fuzzy-logic system analysing nuclear features extracted from H&E stained images with 66% discrimination accuracy concerning the WHO grades. Kolles et al. [33] suggested that digital image analysis systems performance improves when using the HOM (>90% accuracy) system rather than when using the WHO system (about 60% accuracy) analysing nuclear features derived from quantitative measurement from both Feulgen and ki-67 stained biopsies. Nafe and Schlote [34] used cross-validated discriminant analysis, ki-67 and the WHO system for discriminating only low (grade II) from high-grade (grade III) tumours with 88% accuracy. In a most recent study by the same author, classification accuracy was boosted up to 97.4% [37]. Although this study gives significantly high accuracy rates, it examines only discrimination of low from high-grade tumours; it does not examine the separation of high-grade tumours into subgroups of grade III and grade IV tumours. The latter has become an important task following the results from investigations proving that grade III tumours are more chemosensitive than grade IV tumours in respect to certain agents. Thus, their accurate separation would contribute to more effective treatment planning. In our previous studies, we have also proposed digital image analysis systems to accurate grading of astrocytomas using, in contrast to previous studies, the routine combination of clinical protocols adopted by most laboratories: simple measuring equipment, the WHO grading, H&E staining, and quantitative nuclear features formulated to reflect visual observations of experts when examining astrocytic tumour slides. Additionally we have examined not only the separation of low from high-grade tumours (with 89.7% accuracy) but also the critical discrimination of grade III from grade IV tumours (with 83.8% accuracy) [8].

What is missing here is a comprehensive investigation of ways not only to improve the accuracy of digital image analy-

Table 4 – Classification results for the best classifiers' configuration regarding tumours variants

Variants	Number of cases	Correct classification rate (%)[a]
Gemistocytic	8	≈83
Fibrillary	19	≈95
Mixed	34	≈98
Anaplastic	40	≈95
Giant cell	13	≈97
Gliosarcoma	1	≈0
Pleomorphic	35	≈95

[a] The classification rates are presented as average accuracies, since these are derived from the cross-validation process.

sis systems, but also to bring them closer to clinical practice by (a) building such systems using routine clinical protocols and (b) proving that such systems generalize well to data that have not been used for its construction. The latter is most significant and has not been investigated by previous studies to the best of our knowledge. The present study aims to comprehensively investigate these issues and proposes a mathematical formulation towards this direction by integrating state-of-art technologies (support vector machines and least squares mapping) in a cascade classification scheme for separating low from high and grade III from grade IV tumours.

The cascade classification scheme was preferred to using a 3-class classifier because it resembles the diagnostic procedure followed by histopathologists in clinical practice. Initially, tumours were separated into subgroups of low and high grade and subsequently high-grade cases were further categorized into grade III and grade IV. This scheme led to significantly high classification rates (see Tables 2 and 3), which are on one hand comparable to those presented in literature [13,31–36], on the other hand outperform similar studies that have utilized the WHO grading and H&E staining [10,31].

Additionally, it has to be stressed that in this study we have used routinely H&E stained biopsies. H&E produces high quality images for routine histopathological evaluation. For nuclear staining, hematoxilin is more reliable when chromatin is markedly condensed in the nuclei. Current applications of the Feulgen reaction [14] are mainly concerned with DNA quantification. Specific demonstration of DNA in cell structures at the light microscopic level is very little used nowadays. The Ki-67 antigen [15] is expressed during all phases of the cell cycle, consisting in a non-histone nuclear protein of unknown function. It was pointed to be the best marker of cellular proliferation, and its expression may predict the grade of astrocytic tumours. However, H&E stained images present the highest degree of complexity regarding image-processing tasks due to the diversity of the structures stained and the severe variations in staining intensity as compared to Feulgen and to the ki-67 stained images. It has to be mentioned that in this study high classification rates were accomplished using routinely H&E stained biopsies. This adds an additional value to the results.

More specifically, best classifier rates were as high as 97.3% for the first level and 97.8% for the second level (Table 3). The mapping process proved essential in boosting up all classifiers' performance at both levels. The effect of mapping was more pronounced for the PNN and k-nn classifiers (>10% improvement) boosting up their performance higher than 95% for both levels. The improvement for the SVM classifiers in terms of performance was in a lesser extent compared to the PNN and k-nn, however, in terms of support vectors the enhancement was drastic. For the best classifier for the first level accuracy was increased for 84.7% with no mapping to 97.3% with third degree mapping and the number of support vectors reduced from 52.9% of all training data to 1.5%. The latter is most essential since the number of support vectors is an indication of the generalization capability of SVM to unseen data [16,38]. The lesser the number of support vectors the better the generalization expected. The high performance of SVM classifiers when no mapping is applied (especially for the SVM RBF: first level, no mapping 96%, mapping^3 96%; second level, no mapping 95.5%, mapping^3 97.8%) is misleading, since high rates are results of over fitting/overtraining of the classifier. This can be easily observed by noting the number of support vectors. For the first level, for the SVM classifier to achieve 96% accuracy, about 93% of all data operated as support vectors, whereas for the second level about 90% of all data were needed. Thus, results are factitious, since the classifier gives high rates due to over fitting, meaning that its generalization capability is questionable. On the other hand, mapping not only retains the performance of the SVM RBF classifier to same high levels (96% for the first level and 97.8% for the second level), but most importantly reduces the number of support vectors from 93% to 3% for the first level and from 90% to 5% for the second level. The latter proves that the classifier becomes robust, canceling in this way over fitting. Thus, mapping is an important process and should be a part of digital image analysis systems, even in cases where the performance improvement is not vociferous since it affects generalization.

Additionally, we have tried the SFFS feature selection method and we got similar results in terms of features (for the first level seven features as compared to eight features using the Wilcoxon criterion, whereas for the second level 10 features as compared to 10 features using the Wilcoxon criterion) and slightly inferior results in terms of accuracies (for the first level 96.1% accuracy as compared to 97.3% using the Wilcoxon criterion, whereas for the second level 94.6% accuracy as compared to 97.8% using the Wilcoxon criterion). The 7 features derived using the SFFS process were the same as those derived with the Wilcoxon criterion for the first level of the cascade scheme, whereas for the second level we had slightly different results, over the 10 features 7 were the same.

Regarding segmentation, results were promising considering that the H&E staining protocol is not as accurate in staining nuclei as other specialized protocols used in previous studies [13,31–36]. It has to be stressed that the aim of the segmentation procedure was to extract a representative number of accurately segmented nuclei from every set of images describing each case (patient), in order to compute nuclear features. Under this perspective and considering that segmented nuclei for each case ranged from 275 to 415, the misclassification error of 12% may be regarded as of limited significance. The latter can be furthermore supported by the fact that it has been shown that even 200 correctly segmented nuclei per case are adequate for extracting nuclear features [35].

An important comment that can be made is that relevant features appeared at both levels of the cascade scheme, during the feature selection process. These were descriptors of nuclear shape (*roundness*) and chromatin cluster patterns (*homogeneity*, *energy* and *correlation* from the co-occurrence matrix). The latter enforces the belief that certain nuclear features carry significant diagnostic information. *Roundness* describes the circularity of nuclei. Irregular shaped nuclei are an indication of higher malignancy tumours [1,5], whereas in lower grade tumours nuclei shape is more circular with almost constant size. As far as the textural differences, the existence of coarser nuclei (which are described by the *correlation* and *energy*) is an important clue revealing abnormal DNA dissociation within nuclei, and is frequently found into higher-grade tumours. Additionally, an elucidating symptom of tumours' aggressiveness is the presence of chromatic clumps in nuclei, which are encoded by the *homogeneity* [1,5]. It would be a very interesting extension for the studies investigating the differences in chemosensitivity in different grade astrocytic tumours [39,40], to correlate chemosensitivity with the textural differences that seem to prevail between grade III and grade IV tumours. These features are mathematically described in Appendix A.

This study investigates features encoding the following WHO nuclear criteria: presence of giant cells; nuclear morphology, nuclear chromatin texture, pleomorphism, and multinucleated cells. Non-nuclear criteria such as necrosis, endothelial hyperplasia and mitotic activity are of crucial importance for grading of astrocytomas and are not dispensable criteria. The proposed system's effectiveness can be explained by the following: (a) nuclear features are automatically computed within a region of interest, carefully selected and designated on the histological image by the physician interactively. Thus, the physician decides which region from all available slides is the most representative regarding nuclei appearance with respect to grade differentiation. (b) Apart from calculating features that cover the definitions of WHO for nuclei appearance, we calculate nuclear features that cannot by resolved by the human eye [19]. (c) In most cases, expert physicians recognize nuclei attributes of different grade astrocytic tumours almost instantly and unconsciously. However, they do not know how precisely these features have to be taken into account in the decision process since exact definitions do not exist within the WHO guidelines. The proposed system can be used to automatically summarize this knowledge by quantifying nuclear features, and identify which of these features can be used for more effective grade differentiation. (d) The presence of necrosis has been defined by the WHO as an excellent criterion for discriminating anaplastic from glioblastoma multiforme [3]. However, even this critical criterion is not always visible on histological sections especially in cases where tumour's liquid is sucked up using cavitrons by neurosurgeons [41]. If necrosis is not visible, the physician has to perform diagnosis based on the remaining criteria. Our study, in accordance to other similar studies [31–34], indicates the potential of nuclear features to effectively discriminate tumours even when necrosis is not visible on histological sections.

The proposed methodology gives a new insight to building digital image analysis in grading of astrocytomas by combining state-of-art technologies (SVM and least squares mapping) with clinical routine protocols (WHO scheme, H&E staining). This study presented a method that leads to high classification rates for the crucial separation of low from high grade and grade III from grade IV astrocytic tumours while ensuring that high rates are most likely to generalize to unseen data, since the number of support vectors that the SVM classifier needs for its construction is significantly low. The latter is important since it brings digital image analysis systems closer to clinical practice.

Conflicts of interest

None.

Appendix A

Roundness characterizes the circularity of nuclei and takes low values for circular nuclei and high for irregular boundaries. Roundness is calculated according to Eq. (9):

$$\text{Roundness} = \frac{\text{perimeter}^2}{4\pi \times \text{area}} \tag{9}$$

Homogeneity describes image smoothness and takes minimum values for smooth textures nuclei. *Energy* increases for high contrast nuclei, i.e. malignant nuclei within which chromatin clumps are prominent. Finally, *correlation* encodes the gray-tone dependencies revealing irregular textural patterns. Features *homogeneity*, *energy* and *correlation* are calculated from the co-occurrence matrix as follows:

$$\text{Homogeneity (HOM)} = \sum_{i=0}^{N_g-1}\sum_{j=0}^{N_g-1}(p(i,j))^2 \tag{10}$$

$$\text{Energy (EN)} = \sum_{n=0}^{N_g-1} n^2 \left\{ \sum_{i=0}^{N_g-1}\sum_{j=0}^{N_g-1}(p(i,j))^2 \right\}, \quad |i-j| = n \tag{11}$$

$$\text{Correlation (COR)} = \frac{\sum_{i=0}^{N_g-1}\sum_{j=0}^{N_g-1}(ij)p(i,j) - m_x m_y}{\sigma_x \sigma_y} \tag{12}$$

where N_g is the number of gray levels in the image, $i,j = 1, \ldots, N_g$, $p(i,j)$ is the co-occurrence matrix, and m_x, m_y, σ_x and σ_y are the respective mean values and standard deviations of p_x and p_y, which are described in Eqs. (13) and (14).

$$p_x(i) = \sum_{j=1}^{N_{\text{rows}}} p(i,j) \tag{13}$$

$$p_y(j) = \sum_{j=1}^{N_{\text{columns}}} p(i,j) \tag{14}$$

REFERENCES

[1] L. DeAngelis, Medical progress: brain tumors, New Engl. J. Med. 344 (2001) 114–123.

[2] A. Hutter, K.E. Schwetye, A.J. Bierhals, R.C. McKinstry, Brain neoplasms: epidemiology, diagnosis, and prospects for cost-effective imaging, Neuroimag. Clin. North Am. 13 (2003) 237–250, x–xi.
[3] P. Kleihues, PC. Burger, B.W. Scheithauer, The new WHO classification of brain tumours, Brain Pathol. 3 (1993) 255–268.
[4] W. Coons, P. Jhonson, B. Sceithauer, A. Yates, D. Pearl, Improving diagnostic accuracy and interobserver concordance in the classification and grading of primary gliomas, Cancer 79 (1997) 1381–1393.
[5] C. Daumas-Duport, B. Scheitauer, J. O'Fallon, P. Kelly, Grading of astrocytomas. A simple and reproducible method, Cancer 62 (1988) 2152–2165.
[6] W. Kernohan, R.F. Mabon, H.J. Svien, A simplified classification of gliomas, Mayo Clin. Proc. (1949).
[7] H. Kolles, I. Niedermayer, W. Feiden, Grading of astrocytomas and oligodendrogliomas, Pathologe 19 (1998) 259–268.
[8] H. Kolles, H. Ludt, G.H. Vince, W. Feiden, Application of minimal spanning trees in glioma grading—a CLIPPER program for the calculation and construction of minimal spanning trees, Comput. Meth. Prog. Biomed. 42 (1994) 201–206.
[9] R. Prayson, D. Agamanolis, M. Cohen, M. Estes, B. Kleinschmidt-DeMasters, F. Abdul-Karim, S. McClure, B. Sebek, R. Vinay, Interobserver reproducibility among neuropathologists and surgical pathologists in fibrillary astrocytoma grading, J. Neurol. Sci. 175 (2000) 33–39.
[10] D. Glotsos, P. Spyridonos, P. Petalas, D. Cavouras, P. Ravazoula, P.A. Dadioti, I. Lekka, G. Nikiforidis, Computer-based malignancy grading of astrocytomas employing a support vector machine classifier, the WHO grading system and the regular hematoxylin-eosin diagnostic staining procedure, Anal. Quant. Cytol. Histol. 26 (2004) 77–83.
[11] C. Decaestecker, M. Petein, R. van Velthoven, T. Janssen, G. Raviv, J.L. Pasteels, C. Schulman, P. Van Ham, R. Kiss, The computer-assisted microscope analysis of Feulgen-stained nuclei linked to a supervised learning algorithm as an aid to prognosis assessment in invasive transitional bladder cell carcinomas, Anal. Cell Pathol. 10 (1996) 263–280.
[12] M. Scarpelli, P. Bartels, R. Montironi, C. Galluzzi, D. Thompson, Morphometrically assisted grading of astrocytomas, Anal. Quant. Cytol. Histol. 16 (1994) 351–356.
[13] L.R. Schad, H.P. Schmitt, C. Oberwittler, W.J. Lorenz, Numerical grading of astrocytomas, Med. Inform. 12 (1987) 11–22.
[14] A.M. Gurley, D.F. Hidvegi, J.W. Bacus, S.S. Bacus, Comparison of the Papanicolaou and Feulgen staining methods for DNA quantification by image analysis, Cytometry 11 (1990) 468–474.
[15] R. Hofmann-Wellenhof, J. Smolle, H. Kerl, The influence of staining procedures on the assessment of cell proliferation as defined by the monoclonal antibody Ki-67, Am. J. Dermatopathol. 12 (1990) 458–461.
[16] V. Kechman, Support Vector Machines, in Learning and Soft Computing, MIT, 2001, pp. 121–184.
[17] N. Piliouras, I. Kalatzis, N. Dimitropoulos, D. Cavouras, Development of the cubic least squares mapping linear-kernel support vector machine classifier for improving the characterization of breast lesions on ultrasound, Comp. Med. Imag. Graph. 28 (2004) 247–255.
[18] D. Glotsos, P. Spyridonos, D. Cavouras, P. Ravazoula, P.A. Dadioti, G. Nikiforidis, Automated segmentation of routinely hematoxylin-eosin-stained microscopic images by combining support vector machine clustering and active contour models, Anal. Quant. Cytol. Histol. 26 (2004) 331–340.
[19] R. Haralick, L. Shapiro, Computer and Robot Vision, Addison-Wesley, 1992.
[20] N. Ahmed, R. Rao, Orthogonal Transforms for Digital Signal Processing, Springer-Verlag, Berlin, 1975.
[21] N. Street, Cancer diagnosis and prognosis via linear programming based machine learning, PhD Thesis, University of Wisconsin, Madison, 1994.
[22] J. Fellow, R. Duin, J. Mao, Statistical pattern recognition: a review, IEEE Trans. Pattern Anal. Machine Intel. 22 (2000) 4–37.
[23] D. Specht, Probabilistic neural networks, Neural Netw. 3 (1990) 109–118.
[24] S. Theodoridis, K. Koutroubas, Pattern Recognition, Academic Press, 1999.
[25] J.N.P. Pudil, J. Kittler, Floating search methods in feature selection, Pattern Recogn. Lett. 15 (1994) 1119–1125.
[26] I. Kalatzis, D. Pappas, N. Piliouras, D. Cavouras, Support vector machines based analysis of brain SPECT images for determining cerebral abnormalities in asymptomatic diabetic patients, Med. Inform. Internet Med. 28 (2003) 221–230.
[27] I. Kalatzis, N. Piliouras, E. Ventouras, C.C. Papageorgiou, A.D. Rabavilas, D. Cavouras, Design and implementation of an SVM-based computer classification system for discriminating depressive patients from healthy controls using the P600 component of ERP signals, Comput. Meth. Prog. Biomed. 75 (2004) 11–22.
[28] MATLAB Software, The MathWorks Inc., Optimization Toolbox, 2005.
[29] S. Theodoridis, K. Koutroubas, Pattern Recognition, Academic Press, 1999, pp. 144.
[30] A.P. Bradley, The use of the area under the ROC curve in the evaluation of machine learning algorithms, Pattern Recogn. 30 (1997) 1145–1159.
[31] N. Belacel, M. Boulassel, Multicriteria fuzzy assignment method: a useful tool to assist medical diagnosis, Artif. Intel. Med. 21 (2001) 201–207.
[32] C. Decaestecker, I. Salmon, O. Dewitte, I. Camby, P. Van Ham, J.L. Pasteels, J. Brotchi, R. Kiss, Nearest-neighbor classification for identification of aggressive versus nonaggressive low-grade astrocytic tumors by means of image cytometry-generated variables, J. Neurosurg. 86 (1997) 532–537.
[33] H. Kolles, A. von Wangenheim, J. Rahmel, I. Niedermayer, W. Feiden, Data-driven approaches to decision making in automated tumor grading. An example of astrocytoma grading, Anal. Quant. Cytol. Histol. 18 (1996) 298–304.
[34] R. Nafe, W. Schlote, Topometric analysis of diffuse astrocytomas, Anal. Quant. Cytol. Histol. 25 (2003) 12–18.
[35] P.K. Sallinen, S.L. Sallinen, P.T. Helen, I.S. Rantala, E. Rautiainen, H.J. Helin, H. Kalimo, H.K. Haapasalo, Grading of diffusely infiltrating astrocytomas by quantitative histopathology, cell proliferation and image cytometric DNA analysis. Comparison of 133 tumours in the context of the WHO 1979 and WHO 1993 grading schemes, Neuropathol. Appl. Neurobiol. 26 (2000) 319–331.
[36] M. Scarpelli, R. Montironi, D. Thompson, P. Bartels, Computer-assisted discrimination of glioblastomas, Anal. Quant. Cytol. Histopathol. 19 (1997) 369–375.
[37] R. Nafe, W. Schlote, B. Schneider, Histomorphometry of tumour cell nuclei in astrocytomas using shape analysis, densitometry and topometric analysis, Neuropathol. Appl. Neurobiol. 31 (2005) 34–44.
[38] C. Burges, A tutorial on support vector machines for pattern recognition, Data Mining Knowl. Discov. 2 (1998) 121–167.
[39] Y. Iwadate, S. Fujimoto, A. Yamaura, Differential chemosensitivity in human intracerebral gliomas measured

by flow cytometric DNA analysis, Int. J. Mol. Med. 10 (2002) 187–192.
[40] M.H. Cohen, J.R. Johnson, R. Pazdur, Food and drug administration drug approval summary: temozolomide plus radiation therapy for the treatment of newly diagnosed glioblastoma multiforme, Clin. Cancer Res. 11 (2005) 6767–6771.
[41] C. Decaestecker, I. Camby, N. Nagy, J. Brotchi, R. Kiss, I. Salmon, Improving morphology-based malignancy grading schemes in astrocytic tumors by means of computer-assisted techniques, Brain Pathol. 8 (1998) 29–38.