Anal. Chem. **2003,** *75,* 5352−5361

A Chemometric Approach for Brain Tumor Classification Using Magnetic Resonance Imaging and Spectroscopy

Arjan W. Simonetti,[†] Willem J. Melssen,[†] Marinette van der Graaf,[‡] Geert J. Postma,[†] Arend Heerschap,[‡] and Lutgarde M. C. Buydens*,[†]

Laboratory for Analytical Chemistry, University of Nijmegen, Toernooiveld 1 6525 ED Nijmegen, The Netherlands, and Department of Radiology, University Medical Center Nijmegen, P.O. Box 9101, 6500 HB Nijmegen, The Netherlands

A new classification approach was developed to improve the noninvasive diagnosis of brain tumors. Within this approach, information is extracted from magnetic resonance imaging and spectroscopy data, from which the relative location and distribution of selected tumor classes in feature space can be calculated. This relative location and distribution is used to select the best information extraction procedure, to identify overlapping tumor classes, and to calculate probabilities of class membership. These probabilities are very important, since they provide information about the reliability of classification and might provide information about the heterogeneity of the tissue. Classification boundaries were calculated by setting thresholds for each investigated tumor class, which enabled the classification of new objects. Results on histopathologically determined tumors are excellent, demonstrated by spatial maps showing a high probability for the correctly identified tumor class and, moreover, low probabilities for other tumor classes.

Magnetic resonance imaging (MRI) has become one of the most important noninvasive aids in clinical diagnosis of brain tumors, because it enables the radiologist to assess the anatomical location and size of tumor tissue. Normally, multiple complementary images are recorded, as the acquisition of only one image may not show the desired information. Another use of clinical MR is to perform spatially resolved magnetic resonance spectroscopy (MRS). In contrast to MRI, which has a high spatial resolution, this method is performed at low spatial resolution, but it involves the acquisition of metabolic information. The method allows the use of several MR-sensitive nuclei, but often the proton nucleus is used as it is the most sensitive one. Proton MRS at multiple spatially resolved locations is referred to as proton magnetic resonance spectroscopic imaging (^{1}H-MRSI).[1]

When ^{1}H-MRSI is applied to the human brain, the MR spectra obtained on a clinical 1.5-T MR system commonly arise from volumes (called voxels) of approximately 2−4 cm^3. The spectra contain information from important brain metabolites such as glutamate, myoinositol, choline, creatine, *N*-acetyl aspartate (NAA), lactate, and fatty acids.[2] NAA is considered to be a neuronal marker of which the levels are decreased in brain tumors. The choline resonance arises from a group of compounds that are involved in membrane synthesis and degradation, which is often increased in tumors. The creatine signal reflects compounds involved in energy metabolism and may be decreased in brain tumors. Lactate is normally below detection limit, but it is often increased in brain tumors. The function of myoinositol is largely unknown, but its signal appears to be altered in some pathologies. The signals from fatty acids are known to be increased in several brain tumors, especially high grades, and are an indicator for necrosis (cell death). Glutamate is an important neurotransmitter in brain and elevated in oligodendrogliomas.[3]

The MRSI spectra could be used collectively with the anatomical MR images to spatially determine the type of tissue with a high reliability. However, in a clinical setting, the increased amount of information hampers the diagnosis, since radiologists have to investigate multiple MR images and spectra. This is time-consuming and often clinicians are not familiar with spectral information. To deal with this problem, a project called INTERPRET (http://carbon.uab.es/INTERPRET) was funded by the European commission. This project had the goal to develop new methodologies to automatically type tumors in the human brain with the use of MRS and MRI data. The resulting decision support system enables clinicians to utilize MRS information for the diagnosis of brain tumors. In this paper, a new approach is presented to explore the information content of the combined MRI and MRSI data, collected from patients with a brain tumor. Imagelike maps can be generated that show the classification of the tissue and the probability of classification in order to assist the clinician in the interpretation of the data.

The development of statistical models (especially multivariate ones) that are focused on the classification of brain tumors from MRSI data is rather new. Some attempts have been made,[4−7] but

* Corresponding author. Telephone: ++ 31 (0)24 36 53180. Fax: ++ 31 (0)-24 36 52653. E-mail: L.Buydens@sci.kun.nl.

[†] University of Nijmegen.

[‡] University Medical Center Nijmegen.

(1) Noworolski, S. M.; Nelson, S. J.; Henry, R. G.; Day, M. R.; Wald, L. L.; Star-Lack, J.; Vigneron, D. B. *Magn. Reson. Med.* **1999**, *41*, 21−29.

(2) Govindaraju, V.; Young, K.; Maudsley, A. A. *NMR Biomed.* **2000**, *13*, 129−153.

(3) Rijpkema, M.; Schuuring, J.; van der Meulen, Y.; van der Graaf, M.; Bernsen, H.; Boerman, R.; van der Kogel, A.; Heerschap, A. *NMR Biomed.* **2003**, *16*, 12−18.

(4) Preul, M. C.; Caramanos, Z.; Collins, D. L.; Villemure, J.; Leblanc, R.; Olivier, A.; Pokrupa, R.; Arnold, D. L. *Nat. Med.* **1996**, *2*, 323−325.

10.1021/ac034541t CCC: $25.00
Published on Web 09/16/2003

a dedicated multivariate determination of the location and distribution of classes in feature space has not been performed yet. Both Preul[4] and Szabo de Edelenyi[6] focused on the quantitation of metabolite regions from the spectroscopic data, followed by linear discriminant analysis. Others have used pattern recognition[8,9] on single voxel MR spectra to classify[10−13] or cluster[14,15] multiple tumor classes. The emphasis of data analysis is often put on preprocessing,[16,17] data reduction, and the search for appropriate classification algorithms. However, a very important part of a classification approach has not received much attention yet. This deals with the assumption of multivariate normality, the identification of class boundaries for each tumor class, the recognition of overlap between classes, and the probability of classification. Multivariate normality is important to consider if the classification method assumes it, especially when probabilities of class membership of a certain class are also to be given. The identification of class boundaries is important since there are many different types of brain tumors. The usually limited number of investigated tumor classes in a classification procedure should therefore have strict class boundaries. This leaves room for the noninvestigated classes in feature space, which will (hopefully) be classified as "unknown" by the classifier. The probability of classification is important, because it provides information on the reliability of classification.

Few papers deal with the combination of MRI and MRSI data (Szabo de Edelenyi[6] used information from the T_2-weighted image, combined with MRSI). If a patient with a brain tumor is investigated, MR images with different contrasts are always acquired. If a MRSI measurement is also performed, it would be a waste to leave out the MRI information in the statistical models, since the MR images are already available. It is possible to identify morphologically different regions in the brain with MRI information alone.[18] If this information is combined with MRSI data, the pathology of the regions might also be identified. Therefore, we combined MRI and MRSI data in all our experiments.

The goal of this paper was threefold. First, to find an appropriate feature selection technique for the MRSI data, while preserving the distances and distribution of the classes in feature space. We compared two techniques, simple quantitation of spectral regions and principal component analysis. The second goal was to find the isolated and overlapping classes of tissue types in feature space and to create plots that showed their distribution. These distribution plots can be used to generate spatial probability maps, which show for each voxel the probability that it belongs to a certain tissue class. The third goal was to develop a classification model that uses thresholds that are set with use of the distribution plots. These thresholds define class boundaries for each class, enabling the classification of voxels to a specific tissue type or as "unknown".

To address the above research issues, a training set was set up using data from volunteers and patients with a brain tumor. The construction of the distribution plots was performed with a cross-validation method. The evaluation of the classifier was performed with the leave-one-out method on the training set. Furthermore, the resulting classification maps of the patients were visually investigated. Most of the voxels within each classification map of a patient were not selected within the training set. If the classification map of a patient contained homogeneous regions of healthy tissue, cerebrospinal fluid, and tumor tissue, then this demonstrated robustness of the system to a large extent. Next to this, two patients from which no data were selected for the training set were validated.

EXPERIMENTAL SECTION

Data. Data from 18 patients with a brain tumor and 4 volunteers selected from the INTERPRET[19] database were used. Each patient case had passed strict quality control and validation procedures, including consensus histopathological determination of the tumor according to the rules of the World Health Organization. For 2 of the 18 patients (I-1172, I-1281), a consensus was not reached. Therefore, data from the tumor region of these patients were not used in the training set.

For each subject, stacked MR images of cross sections of the whole brain at four contrasts were acquired: T_1- and T_2-weighted images, a proton density image, and a gadolinium-enhanced T_1-weighted image (256 × 256, FOV = 200 mm, slice thickness 5 mm). In the T_1- and T_2-weighted images, the contrast reflects differences in T_1- and T_2-relaxation times of the water protons in the brain, respectively. In the proton density image, the distribution of protons produces the contrast. The contrast-enhanced image (GD image) is obtained after intravenous administration of Gd-DTPA solution. The paramagnetic Gd compound influences the relaxation times of water protons and therefore regions are highlighted that contain large blood vessels or where the blood−brain barrier is disrupted due to proliferation of malignant tissue. To ensure that image pixels from subsequent images originate from the same spatial location, the images have to be coaligned. The coalignment was performed by calculating the maximum correlation between two images while shifting them left−right and up−down. Coalignment is difficult but does not need to be perfect in this method, because the high-resolution images are combined with the low-resolution MRSI eventually.

(5) Mader, I.; Roser, W.; Hagberg, G.; Schneider, M.; Sauter, R.; Seelig, J.; Radue, E. W.; Steinbrich, W. *Magn. Reson. Mater. Phys., Biol., Med.* **1996**, *4*, 139−150.

(6) Szabo De Edelenyi, F.; Rubin, C.; Estève, F.; Grand, S.; Décorps, M.; Lefournier, V.; Le Bas, J.; Rémy, C. *Nat. Med.* **2000**, *6*, 1287−1289.

(7) Nelson, S. J.; Huhn, S.; Vigneron, D. B.; Day, M. R.; Wald, L. L.; Prados, M.; Chang, S.; Gutin, P. H.; Sneed, P. K.; Verhey, L.; Hawkins, R. A.; Dillon, W. P. *J. Magn. Reson. Imaging* **1997**, *7*, 1146−1152.

(8) Hagberg, G. *NMR Biomed.* **1998**, *11*, 148−156.

(9) El-Deredy, W. *NMR Biomed.* **1997**, *10*, 99−124.

(10) Somorjai, R. L.; Nikulin, A. E.; Pizzi, N.; Jackson, D.; Scarth, G.; Dolenko, B.; Gordon, H.; Russell, P.; Lean, C. L.; Delbridge, L.; Mountford, C. E.; Smith, I. C. P. *Magn. Reson. Med.* **1995**, *33*, 257−263.

(11) Meyerand, M. E.; Pipas, J. M.; Mamourian, A.; Tosteson, T. D.; Dunn, J. F. *Am. J. Neuroradiol.* **1999**, *20*, 117−123.

(12) Minguillón, J.; Tate, A. R.; Arús, C.; Griffiths, J. R. *Lect. Notes Comput. Sci.* **2002**, *2364*, 282−292.

(13) Tate, A. R.; Griffiths, J. R.; Martinez-Pérez, I.; Moreno, A.; Barba, I.; Cabañas, M. E.; Watson, D.; Alonso, J.; Bartumeus, F.; Isamat, F.; Ferrer, I.; Vila, F.; Ferrer, E.; Capdevila, A.; Arús, C. *NMR Biomed.* **1998**, *11*, 177−191.

(14) Hagberg, G.; Burlina, A. P.; Mader, I.; Roser, W.; Radue, E. W.; Seelig, J. *Magn. Reson. Med.* **1995**, *34*, 242−252.

(15) Howells, S. L.; Maxwell, R. J.; Peet, A. C.; Griffiths, J. R. *Magn. Reson. Med.* **1992**, *28*, 214−236.

(16) Soher, B. J.; Young, K.; Govindaraju, V.; Maudsley, A. A. *Magn. Reson. Med.* **1998**, *40*, 822−831.

(17) Bartha, R.; Drost, D. J.; Williamson, P. C. *NMR Biomed.* **1999**, *12*, 205−216.

(18) Clarke, L. P.; Velthuizen, R. P.; Camacho, M. A.; Heine, J. J.; Vaidyanathan, M.; Hall, L. O.; Thatcher, R. W.; Silbiger, M. L. *Magn. Reson. Imaging* **1995**, *13*, 343−368.

(19) International network for pattern recognition of tumours using magnetic resonance. http://carbon.uab.es/INTERPRET/interpret_home.shtml.

Besides the acquisition of images, ^{1}H-MRSI data were also acquired from each subject, both with and without water signal suppression using a 2D STEAM ^{1}H-MRSI sequence (16 × 16 × 1024 matrix size, Tr = 2000 or 2500 ms, TE = 20 ms, slice thickness 12.5 or 15 mm, FOV = 200 mm, spectral width 1000 Hz, NS = 2 (for the MRSI without water suppression, NS = 1)). When STEAM[20] is applied, a box is selected from which the signal is acquired by the application of three slice-selective radio frequency pulses in the presence of *x*-, *y*-, and *z*-gradients. Only the MRSI voxels within the STEAM box will contain signal, while signals in voxels from outside the box are suppressed. The STEAM box was positioned totally in the brain; therefore, disturbing signals arising from the fat tissue surrounding the skull were avoided. The location of the ^{1}H-MRSI slice was selected and aligned with the Gd image that showed the largest enhanced tumor area. Thus, the MRSI slice (of at most 15-mm thickness) was centered around a MRI slice of 5 mm. Since the MRI slices above and below the selected MRI slice have a gap of 1.5 mm with the selected slice, they are partially outside the view of the MRSI slice. Therefore, only MR images from one slice location have been used.

Proper preprocessing of the MRSI data is essential for successful classification. All data were processed in the same automatic way. In short: *k*-space data (16 × 16) was filtered by a hanning filter of 50% using the LUISE software package (Siemens, Erlangen, Germany), followed by zero filling to 32 × 32 and 2D Fourier transformation to obtain time domain signals for each voxel. Then, each voxel within the preselected STEAM box was corrected for eddy current effects in the spectra using a method that prevents the occasional occurrence of eddy current correction-induced artifacts.[21] This was followed by HLSVD[22] filtering to remove the residual water signal between 4.3 and 5.5 ppm. Next, a simple but efficient baseline correction was applied, which removed broad resonances. This was performed by exponential filtering of the time domain signal of each voxel with a filter width of 5 ms, followed by the subtraction of the residual from the original time domain signal. To correct for first-order phase effects (in principle the zero-order phase is removed by the eddy current correction), the mean spectrum, which was calculated from all spectra within the STEAM box of each patient's MRSI data set, was manually optimized. This correction value was then applied to each separate time domain signal of that data set. Finally, each time domain signal was Fourier transformed and only the real values of the complex spectrum was retained between 0.5 and 4.0 ppm (230 points). To normalize between patients, all spectra of one data set were divided by the patients' mean water signal of all voxels within the STEAM box, as calculated by singular value decomposition[23] from the unsuppressed MRSI data.

Next, the spatial resolution of the MR images and the MRSI grid had to be readjusted, so that the two were in agreement. Therefore, the resolution of the MR images was lowered to the resolution of the MRSI grid by averaging the image pixels within each spectroscopy voxel. The values within each low-resolution image were scaled to the same range as the spectral data. After preprocessing, each voxel within the grid was represented by a spectrum of 230 variables (the region between 0.5 and 4.0 ppm) and 1 variable from each MR image. This is schematically represented in Figure 1a−c.

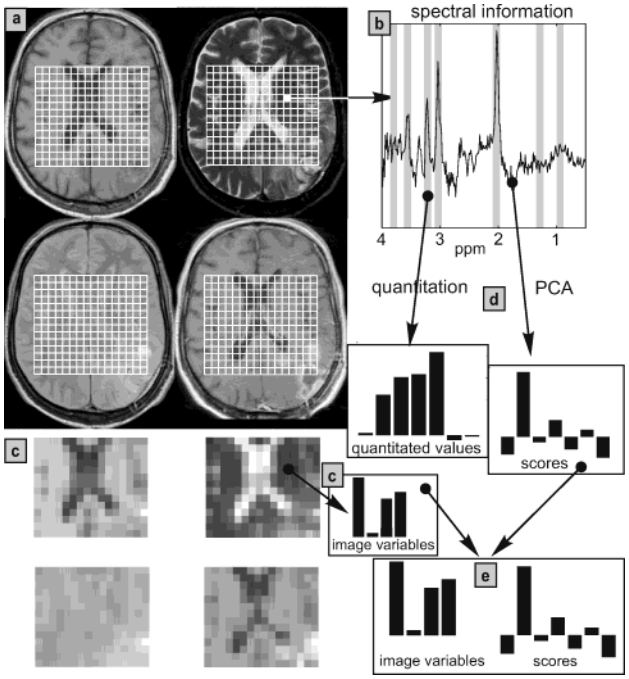

Figure 1. Representation of preprocessing and data reduction. (a) The images with the four contrasts (top, T_1- and T_2-weighted image; bottom, proton density and Gd image) are aligned (the lower right image has been shifted up) and the spectral grid within the STEAM box has been superimposed on the images; (b) a spectrum after preprocessing, obtained from the indicated voxel; (c) the resolution of the images is reset to the MRSI resolution. Each voxel is represented by four image variables now; (d) data reduction is performed either with quantitation of important regions in the spectrum (indicated with gray bands in Figure 1b) or with PCA of the spectra; (e) the image variables from each voxel are combined with the variables obtained from either data reduction technique. In this example, the image variables are combined with the variables (scores) obtained by PCA.

Setup of a Training Set. Based on the 16 patients for which consensus about the histopathology was reached, four different tumor classes were selected. The first three classes contained glial tumors with different grades, grade II (6 patients), grade III (4 patients), and grade IV (5 patients). A fourth class consisted of a patient (I-1283) with a meningioma (MNG). No classifier for this class was calculated, because the size was too small, but it was used to validate results of the classifier.

To create the training set, a selection of voxels from patients who were in the same tumor class was made using the MR spectra and the four MR images. The voxels were selected from regions that clearly consisted of tissue belonging to the correct class. Boundary voxels were not selected. Also, classes for healthy tissue and cerebrospinal fluid (CSF) were created. The voxels for healthy tissue were selected from the contralateral region of several patients and from healthy volunteers. Data for the CSF were selected from CSF voxels that were not in close contact with the tumor region. The total number of voxels selected from a subject was never more than 30%.

(20) Frahm, J.; Merboldt, K. D.; Hanicke, W. *J. Magn. Reson.* **1987**, 72, 502−8.
(21) Simonetti, A. W.; Melssen, W. J.; van der Graaf, M.; Heerschap, A.; Buydens, L. M. C. *J. Magn. Reson.* **2002**, *159*, 151−157.
(22) Pijnappel, W. W. F.; van den Boogaart, A.; de Beer, R.; van Ormondt, D. *J. Magn. Reson.* **1992**, *97*, 122−134.
(23) Vanhamme, L.; Fierro, R. D.; Van Huffel, S.; de Beer, R. *J. Magn. Reson.* **1998**, *132*, 197−203.

Table 1. Selected Patients, Volunteers, and Number of Voxels for Each Class in the Training Set

tumor class	patient number (I-)[a]	no. of voxels selected	total
grade II	1227 1233 1260 1275 1278 1318	07 20 28 12 14 09	90
grade III	1212 1229 1230 1232	17 06 26 4	53
grade IV	1231 1234 1285 1286 1322	12 06 16 17 08	59
MNG	1283	15	15
healthy (volunteer)	1236 1237 1238 1239	29 31 25 37	122
healthy (patient)	1172 1212 1231 1285 1318	13 32 21 39 17	122
CSF	1172 1230 1231 1234 1275 1318	15 09 26 03 09 16	78
no consensus	1281	used as independent test case	
total			539

[a] The patient numbers correspond with the INTERPRET [19] database.

Patient I-1172, for which no consensus was reached, was used for the selection of training set data for healthy tissue and cerebrospinal fluid. Data from patient I-1281 were not used for the training set at all but were used as an independent test case for evaluation, just like patient I-1283 (the MNG case). An overview of the training set is given in Table 1.

Data Reduction Methods. After selection of the training set, two different data reduction techniques were used to create two separate training sets with reduced dimensionality, which were validated for their robustness. The following techniques were used:

(a) Quantitation of Important Spectral Regions. Quantitation was performed in the following way: from the preprocessed spectra, intensities of seven spectral ranges encompassing the brain metabolites mentioned in the introduction were determined by integrating the intensities within a window of 0.13 ppm at 3.75 (glutamate), 3.56 (myoinositol), 3.20 (choline), 3.02 (creatine), 2.02 (NAA), 1.33 (lactate), and 0.90 ppm (fatty acids). These seven windows have been indicated with gray bands in Figure 1b. The quantitation procedure is simple, robust, and easily automated. It should be noted that the levels found with our procedure are only an approximate indication of the true levels: only part of the peak width is taken into account and negative values are acceptable in case of a negative baseline. However, for classification purposes, this appears to be adequate. The 7 spectroscopic variables (Figure 1d) and the 4 image variables (Figure 1c) were combined to create a training set that contained 11 variables for each voxel.

(b) Principal Component Analysis of the Spectral Data. The second data reduction technique used was principal component analysis (PCA).[24] PCA was performed on the spectra of the training set, which consisted of 524 (MNG excluded) spectra with 230 variables each. PCA finds directions in the data that explain most of the variance. These new axes are linear combinations of the original ones and are orthogonal and ordered with respect to the amount of explained variance. After PCA, the first seven principal components, which explained 95.3% of the variance, were retained. The number of principal components to use is normally not straightforward. We used seven to make a comparison with the seven variables obtained by the quantitation of important spectral ranges as described above. By chance, the seven principal components retained more than 95% of the variance, which is also a limit used to choose the number of principal components. After PCA, each voxel was described by seven variables (called scores, Figure 1d) obtained from the MRSI modality and four from MRI (Figure 1e).

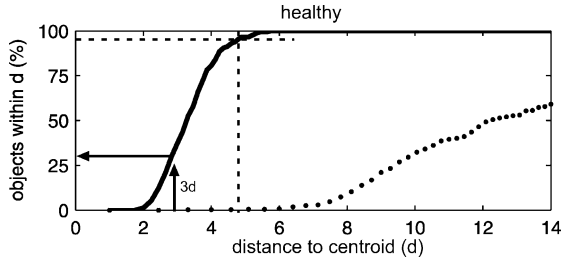

Figure 2. Distribution plot showing the distribution of "healthy" objects (solid) and objects from other classes (dotted) with respect to the centroid of the healthy class.

Exploration of the Classes in Feature Space. To develop a classifier that is able to classify a voxel as belonging to one of the investigated classes or as "unknown" or as "undecided" between some tissue classes, we explored the distribution of each class in feature space and compared it with the distribution of all other classes. This was done for each class in both training sets created by the two data reduction methods. First, the data representing the voxels (objects) of the class under investigation were split, separating at random 10% (A) from the rest (B). Then the covariance matrix of B was calculated and used to calculate the Mahalanobis distances (d) of all objects in A with respect to the centroid of B. Also the Mahalanobis distances of all objects in the other classes with respect to the centroid of B were calculated. This procedure was repeated successively leaving the next 10% of the class under investigation out, until all objects were left out once. This procedure was repeated for all classes. Eventually, for each tissue class, a plot was created that showed the percentage of objects against the Mahalanobis distance from the classes centroid for members of the class itself and for nonmembers. An example of such a distribution plot is shown in Figure 2. The solid line represents the percentage of healthy objects within a certain Mahalanobis distance d from the centroid of the healthy class. One hundred percent of the healthy objects in the training set (created by the first data reduction method) are within a distance of $6d$ from the healthy centroid. The dotted line represents the percentage of all other objects in the training set (not belonging to healthy) within a certain distance to the centroid of the healthy class. Only a small percentage of nonhealthy objects are within a

(24) Vandeginste, B. G. M.; Massart, D. L.; Buydens, L. M. C.; De Jong, S.; Lewi, P. J.; Smeyers-Verbeke, J. *Handbook of Chemometrics and Qualimetrics: Part B*; Elsevier: New York, 1998; Chapter 33.

distance of 6*d* from the healthy centroid, and the healthy class is therefore relatively isolated from the other classes.

Calculation of Probabilities of Class Membership. The probability of class membership of a new object for a class was calculated based on the true distribution of the class, which is reflected by the distribution plot. The density at point x_i of a multivariate normal distribution (with p dimensions), characterized by the mean μ and covariance matrix Σ is given by

$$\phi(x_i|\mu, \Sigma) = (2\pi)^{-p/2}|\Sigma|^{-1/2}e^{-(\mathbf{x}_i-\mu)^T\Sigma^{-1}(\mathbf{x}_i-\mu)/2} \quad (1)$$

In a nonnormalized way, the density is a direct estimator of the probability that an object at point x_i belongs to this multivariate normal distribution. If the Mahalanobis distance of a new object to a class centroid is known, the density and thus probability of class membership can be calculated using this equation. In the proposed approach, a slightly different method was used, based on the "true" distribution of each class shown in the distribution plots.

To calculate the probability of class membership of a new object to the healthy class (for example), first its Mahalanobis distance to the class centroid was calculated. Using the experimentally determined distribution plot in Figure 2, the percentage of objects of the training set within this distance could be red. For example, when an object has a distance of 3*d* from the centroid of the healthy class, then 28% of the objects in the training set was within this distance (arrow at 3*d* in Figure 2). Last, from a theoretical determined curve, which showed the number of objects a certain distance from the centroid of a multivariate normal distribution against the density, the density could be red. If the distribution plot in Figure 2 was precisely following the theoretical distribution of a multivariate normal distribution, then the density calculated by our approach would give the same result as the density calculated by eq 1.

Classification of Voxels. To use the distribution plots as a classifier, thresholds have to be set for each class. The value of the threshold is dependent on the errors one accepts to take for a specific tissue class. If one wants to classify 95% of the healthy objects correctly (the α error is 5%), then the threshold has to be set at the position shown in Figure 2 (dashed vertical line). However, this means that ~2% of the objects will be classified incorrectly (β error) as healthy.

Classification of a new voxel is performed by projecting it in the feature space. Then, its distance to the centroid of each class is calculated. The voxel is only assigned to the closest class when its distance to the centroid does not exceed the threshold. If the distance exceeds the threshold, then it is classified as "unknown". If overlapping classes are found, rules can be set up to classify voxels as "belonging to either one of these classes", as described in the Discussion. If a voxel is classified as such, it may be reclassified with another classifier, like a support vector machine, especially constructed for the classification between two classes.[25]

Evaluation of the System. Evaluation of the system was performed at four levels. First, the best data reduction method was selected by comparison of the distribution plots constructed with both methods. Second, the classifier was tested by a leave-one-out procedure on the training set (n = 524). Leave-one-out testing is an accepted validation procedure when the number of samples is low. The leave-one-out procedure was executed in the following way. From the training set, the first sample was left out. The remainder of the training set was used to calculate new (independent to the left out sample) distribution plots. Then the left-out sample was classified with use of the calculated distribution plots. The classification thresholds were kept the same during the leave-one-out procedure. This was repeated until all samples were left out of the training set once.

Third, the classifier was evaluated using all the data from all patients in the training set. This was performed by calculation of the classification maps and probability of classification maps. Most of the voxels of the patients were not used in the training set and therefore could be used for validation. Since no "true" tissue type for these voxels were available, all maps were visually inspected for obvious misclassifications. Regions that could be assigned visually as healthy tissue, CSF, or the tumor by means of the MR images had to be homogeneously classified as such in the classification map. Last, the classification results of patient I-1283 (the meningioma) and patient I-1281 (for which no consensus was reached) were also used for validation.

RESULTS AND DISCUSSION

Exploration of Classes in Feature Space. The distribution plots can be used to compare data reduction techniques as described above or the influence of other processing steps, e.g., normalization or scaling. In this study, only a comparison of data reduction techniques was made. Each subplot in Figure 3 shows the distribution plot of a tissue class. The solid black and dashed gray visualize the percentage of objects of the class itself within a certain Mahalanobis distance from its centroid. Theoretically, the squared Mahalanobis distance should follow the Hotelling T^2 statistics,[26] which is plotted as the gray thick line for each class. For all tissue classes, the plots follow T^2 statistics, suggesting that the multivariate data are normally distributed in space and that using the Mahalanobis distance is justified for both approaches. This is important, because the probabilities can only be calculated if the data follow a normal distribution. The approach using quantitated values seems to be better than the approach using PCA, since objects not belonging to a class are always further away from the class centroid for this method. This is especially visible for the grade III and grade IV class (Figure 3d and e) in which the black dotted line is lower than the gray dashed−dotted. This suggests that the classes are more separated in the approach using quantitated values and therefore this data set is used in the remainder of the paper.

To investigate the distribution of all tissue classes that surround a particular class, we plotted the Mahalanobis distance of all objects (including MNG) within a class to the centroid of the particular class under investigation (Figure 4). The surrounding of the healthy class is relatively empty. At a Mahalanobis distance of 6*d* from its centroid, objects from the cerebrospinal fluid class arise. Objects from other classes, however, are almost absent until

(25) Lukas, L.; Devos, A.; Suykens, J. A. K.; Vanhamme, L.; Van Huffel, S.; Tate, A. R.; Majós, C.; Arús, C. *ESANN'2002 Proceedings: European Symposium on Artificial Neural Networks*, 2002; pp 131−136.

(26) Candolfi, A.; De Maesschalck, R.; Jouan-Rimbaud, D.; Hailey, P. A.; Massart, D. L. *J. Pharm. Biomed. Anal.* 1999, *21*, 115−132.

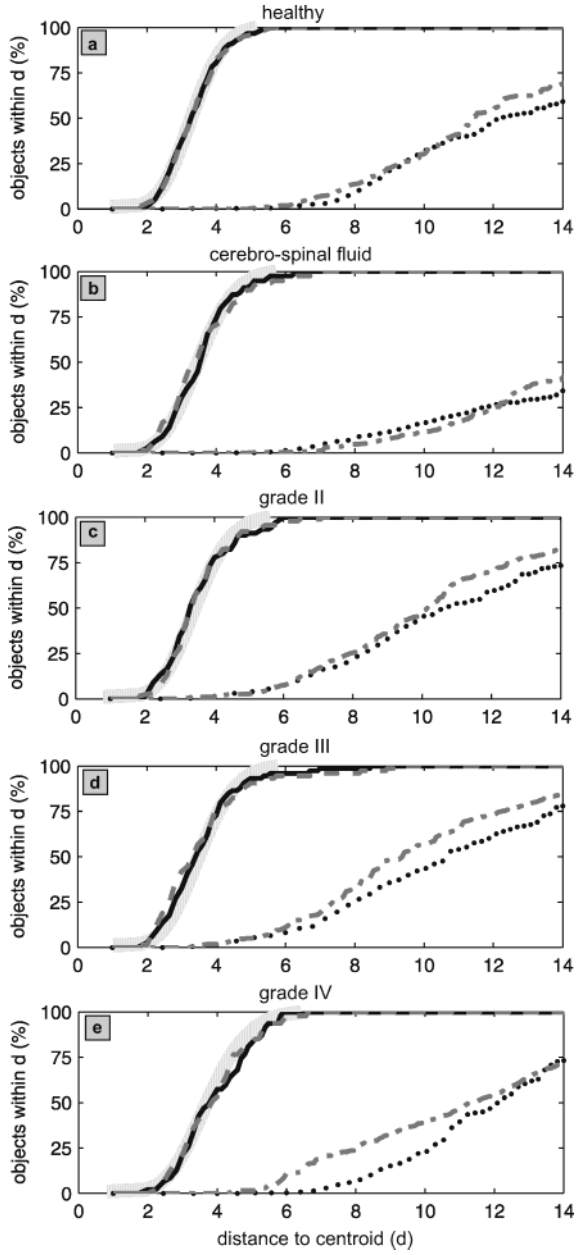

Figure 3. Distribution plots for each class using both data reduction approaches. (a) The distribution of the healthy class is plotted in solid black for the quantitation of spectral ranges method and in dashed gray for the PCA method. Both distributions follow Hotelling T^2 statistics, which is plotted as a thick gray band. Also, the percentage of nonhealthy training samples within a certain Mahalanobis distance from the centroid of the healthy class is plotted for each data reduction approach. The dotted black corresponds with the quantitation of spectral ranges method and the dashed−dotted gray for the PCA method. (b−e) The same as (a), but now for the CSF, grade II, grade III, and grade IV class, respectively.

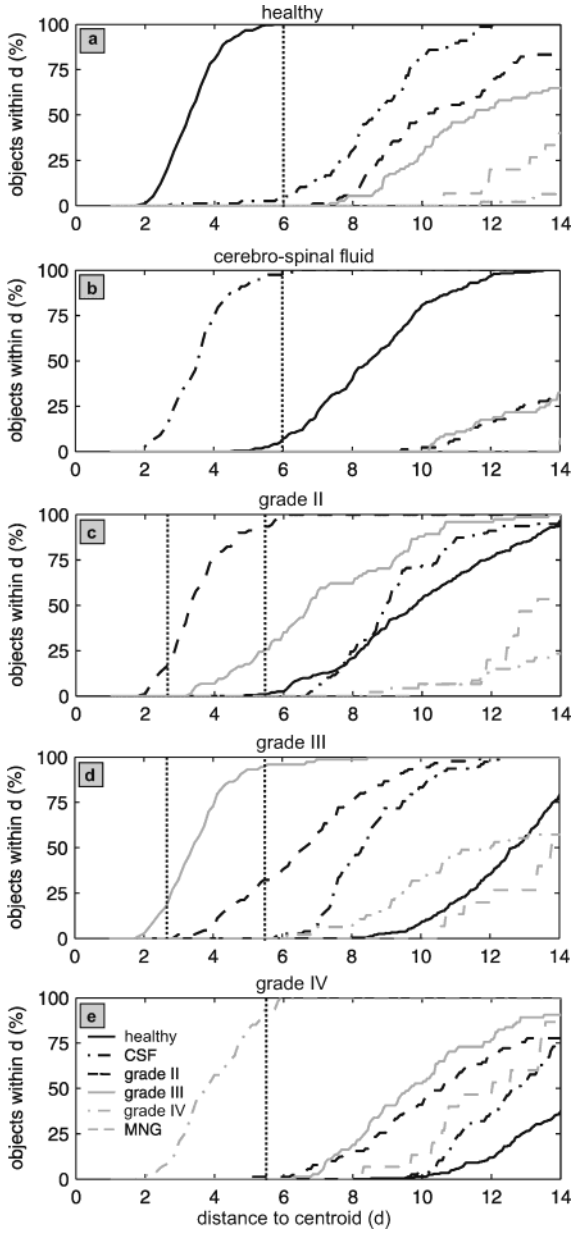

Figure 4. Distribution plots of each class and its surrounding classes using the quantitation of spectral ranges method for data reduction. In each subplot, the solid black line represent the percentage of healthy objects within a certain distance to the centroid of the class under investigation. The dashed−dotted black line represents CSF, the dashed-black line, grade II. The solid gray, dashed−dotted gray, and dashed gray respectively represent grade III, grade IV, and MNG. The vertical dotted lines indicate the thresholds used for the classification.

a distance of 8*d*. The CSF class has some overlap with objects from the healthy class at 6*d*, but no overlap with other classes at all (Figure. 4b). This suggests that healthy tissue and CSF are dense clouds in feature space, which have some overlap with each other but not with malignant tissue. Therefore, one may expect some error between these classes but low error between healthy/CSF and any of the other classes. This is in agreement with findings of other researchers,[6,12,13] who found that healthy tissue can be classified with a high reliability. This can be explained by the stringent concentrations of metabolites in the human brain for healthy tissue. The overlap of CSF with healthy is likely due to partial volume effects.

The grade II class overlaps with the grade III class, but not with grade IV or MNG (Figure 4c). It is therefore very well possible to separate grade II from grade IV and MNG objects, but separation of grade II from grade III will be hard. The same can be concluded from Figure 4d, which describes the grade III distribution. The grade IV class is separated from all other classes, suggesting that this class can be classified with a high reliability. As pointed out, we included the MNG class in Figure 4a−d,

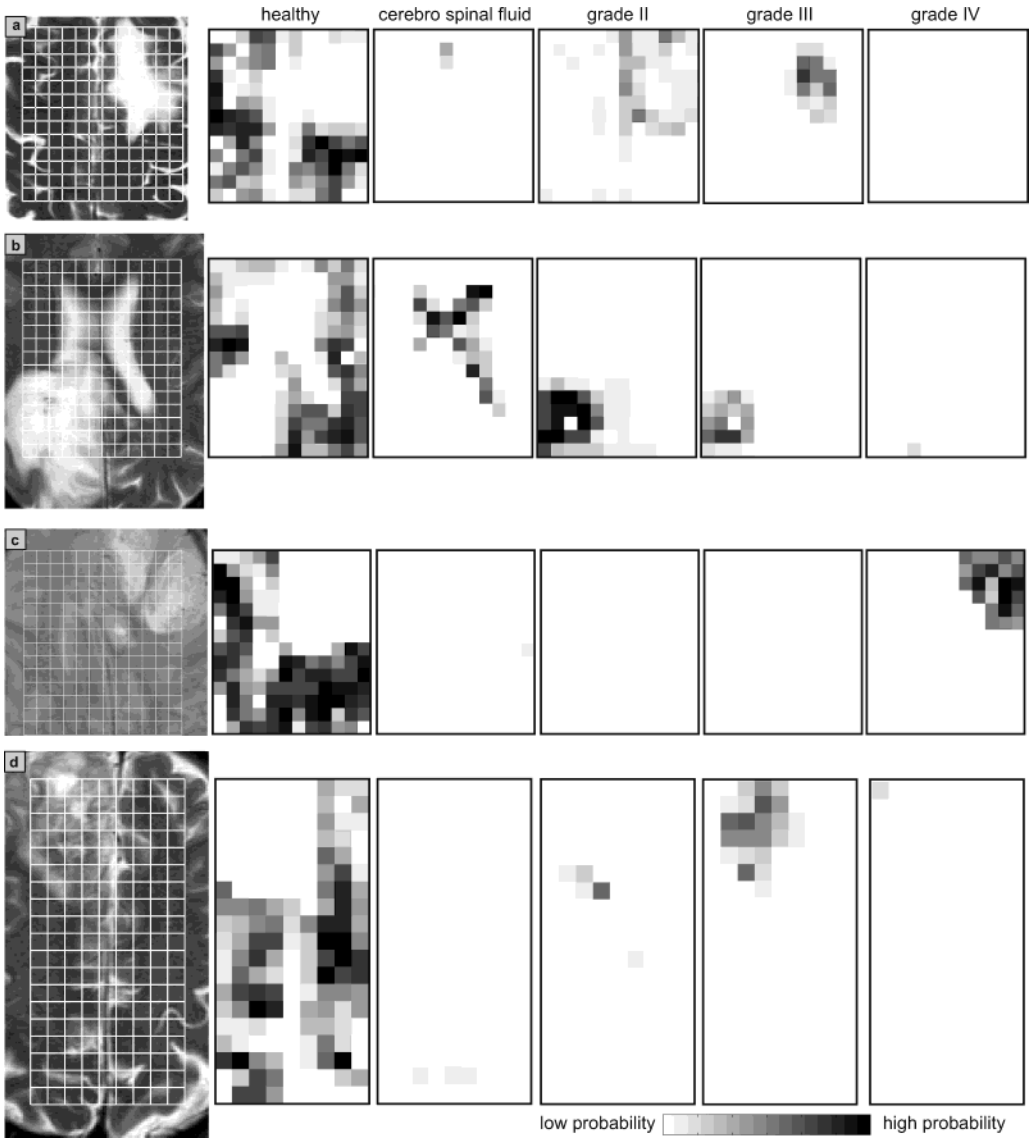

Figure 5. Overview of probability maps of four patients. For each patient, a MR image is shown on the left with a contrast that clearly identifies the tumorous region. The MR images are zoomed, and only the MRSI grid within the STEAM box is depicted. Next to the MR image, the probability maps for the five classes are shown. (a) Patient I-1229 (6/156), which has a grade III tumor. After the patient label, the number of voxels used in the training set and the total number of voxels for this patient are depicted between brackets. The MR image is T_2-weighted. (b) Patient I-1275 (21/180), which has a grade II tumor. The MR image is T_2-weighted. (c) Patient I-1285 (55/168), which has a grade IV tumor. The MR image is proton density-weighted. (d) Patient I-1212 (49/171), which has a grade III tumor. The MR image is T_2-weighted.

because it gives relevant information about the location of MNG in feature space.

The distribution plots are very import for the interpretation of the classes in feature space. If a new class is added, one can directly see whether this new class has a significant overlap with any of the other classes. Assume that a new class A is overlapping with class B. If a new object is classified to class A, but the probabilities for both classes A and B are high, then the reliability of classification is low. The object might also be a member of class B. If classes A and B are from totally different tumor types, which normally do not occur together in the brain, then the classification is A, with a low reliability. On basis of other information, the clinician might decide that the "true" classification is A. It can also be that the classes A and B are from the same tumor type, but with a different grade. This is, for example, the case for the grade II and grade III classes of glial tumors described in this paper. A mixture of these two tumor types can occur in the brain. If the probability for classes A and B is high now, and the two classes are from different grades, then the tissue could also be a mixture of both classes and thus heterogeneous. The expert user has to keep this in mind during use of the classifier.

As explained in the Experimental Section, probability maps for a new patient can be constructed using the distribution plots of Figure 4. Maps are presented from patients with a histologically proven grade III (Figure 5a and d), a grade II (Figure 5b), and a grade IV (Figure 5c) tumor. The darker a voxel in a map, the higher the probability that the underlying tissue belongs to the class represented by the map. The images give a very consistent indication of the different tissues in the brain, indicating robustness of the system. Remember that only several voxels of each patient have been used in the training set. The number of voxels used in the training set and the number of total voxels is specified

in the figure caption for each patient. For the interpretation of the probability maps, it is important to compare them region wise, because the probabilities are relative. When a region in the probability map of grade II is dark, while it is bright on the other maps, then the estimation of class membership is high. When the region is dark in more than one map, then the estimation of class membership is lower. As explained, the probabilities of classification might also be important for the assessment of tumor heterogeneity. When the probability in a region is high for only one tissue type, then the tissue will be, in general, homogeneous. If the probability in a region is high for two tissue types, then the region might be a mixture of these tissues and, therefore, heterogeneous. However, this can only be the case if the mixture of these two tissues is possible from a clinical point of view. Otherwise, on can only say that the reliability of classification is low. In addition to other classification procedures that try to assess tissue heterogeneity with a spatially resolved classifications algorithm[4,6] or the construction of metabolic maps,[27,28] the use of statistically calculated probabilities could be an important asset to the mapping of tumor heterogeneity.

Classification of Voxels. For the classification, we created classification rules based upon the results presented in Figure 4. The healthy and the CSF classes are isolated in feature space, having some overlap with each other. To minimize the α error, we have set the threshold to 6*d* for these classes as shown by the vertical dotted line in Figure 4a and b. The same can be concluded for the grade IV class, with a threshold of 5.5*d*. This leaves some α error for grade IV but minimizes the chance that a healthy or CSF object will be falsely classified as grade IV. Because of the overlap of the grade II and grade III classes, we defined different rules for these classes. When an object is very close ($\leq$2.5*d*) to the centroid of one of these classes it is appointed to it (threshold at 2.5*d*), but when the distance is between 2.5 and 5.5*d*, other rules were applied. (a) If the object was closest to the grade II class and the distance of the object to the grade III class was 1.3 times larger than to the grade II class, then it was assigned to the grade II class. This rule can be applied because all other classes are far away from the grade II class. (b) If the distance of the object to the grade III class is smaller than 1.3 times the distance to the grade II class, then it is a grade II or a grade III member and thus "undecided" between *these* two classes. We applied identical rules when the object was closest to the grade III class. The value of 1.3 was found by repeating the classification for different values and assessing the number of objects that were classified as "undecided".

Classification using leave one out is excellent for the healthy (100% correct), CSF (97% correct, the 3% being misclassified as healthy), and grade IV (100% correct) classes. Classification results for GII and GIII were promising. A total of 83% of the GII objects were classified correctly and 88% of the GIII class. However, the largest part of the misclassified objects in the GII and GIII classes were classified as "undecided" (12% for the GII and 4% for the GIII class), which is to be expected since these classes are overlapping. A total of 2% of the GII objects were misclassified as GIII and 3% of the GIII as GII, respectively. Last, some objects of the GII and GIII classes were classified as "unknown" (3 and 5%). All MNG objects were classified as "unknown", which is correct as we did not build the classifier to be able to take MNG into account.

To compare the method with a commonly used algorithm,[4,6] we performed linear discriminant analysis (LDA)[24] with use of leave one out. In general, the number of correctly classified objects for each class using LDA was in correspondence with the presented method, but when LDA was used, the misclassified objects were often assigned to classes not expected. Objects from the healthy and CSF classes were wrongly assigned to grade II (respectively, 0.5 and 1%). A total of 5% of the GII was misclassified as GIII, and 11% of the GIII as GII; 1.5% of the GIII was misclassified as GIV.

The results of LDA may be caused by the fact that LDA can only be used if the classes under investigation have equal covariance in size and shape. In our data set, this was not the case. The healthy and CSF classes are much more condensed than the tumor classes. This will probably always be the case, because the highly structured brain will give rise to spectra with low variance, while heterogeneous tumor tissue with a broader range of morphologic and metabolic abnormalities will give rise to MR images and spectra with high variance.

In Figure 6, the classification results of the patients presented in Figure 5 are shown, together with two other patients. The patient in Figure 6e (I-1318) has a grade II tumor; the pathology of the tumor of the patient in Figure 6f (I-1281) is unclear, since the pathologists who validated the tumor disagreed between grade II and grade III. Therefore, this patient was not used during training of the classifier but was used as a test case. Also, all other patients have been evaluated, showing results similar to those of the patients presented.

Although only a small part of the voxels from patients a−e are used for training the classifier, the results look very good. As an example we have indicated the voxels used in the training set for the patient in Figure 6a with white circles. Only 6 of the 156 voxels of this patient were used for training! Important considerations for the evaluation are as follows: (a) voxels that are identically classified tend to be spatially grouped, (b) the spatial distribution of classification is in agreement with morphological information from the MR images, (c) in most patients, only one tumor type is assigned, and this assignment is always in agreement with the pathology of the patient, (d) most grade III patients also show grade II voxels, which can be expected, because the pathologist types the patient with respect to the most malignant tissue present. However, grade II patients do not show grade III classified voxels. This is important since otherwise a clinician could assume grade III tumor based on one or two falsely classified grade II voxels. In Figure 6d and e, voxels have been misclassified as grade IV. In the case of (d), the misclassification could have occurred because it is a corner voxel. These voxels normally have a low signal, since the boundary setup by the STEAM box is never perfect.

The classification result of the patient in Figure 6f is fairly good for the healthy and CSF tissue. There are misclassified voxels in the border region of healthy and CSF that might be caused by partial volume. In the tumor area, none of the voxels have been classified as grade III, suggesting this is a grade II tumor, but

(27) Segebarth, C.; Balériaux, D.; Luyten, P. R.; den Hollander, J. A. *Magn. Reson. Med.* **1990**, *13*, 62−76.
(28) Furuya, S.; Naruse, S.; Ide, M.; Morishita, H.; Kizu, O.; Ueda, S.; Maeda, T. *NMR. Biomed.* **1997**, *10*, 25−30.

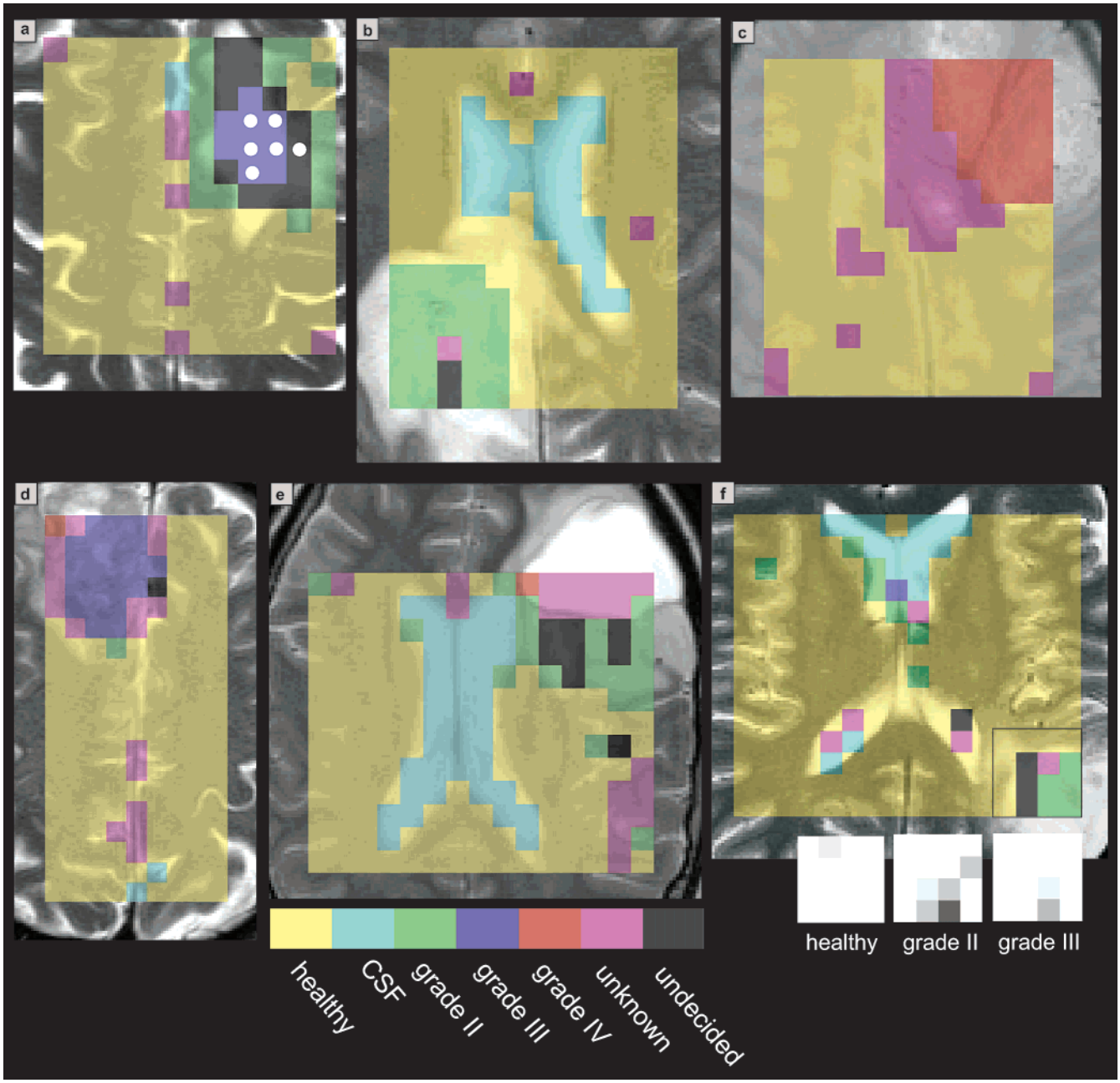

Figure 6. Classification result of six patients. The classification is projected over a MR image with a contrast that clearly identifies the tumorous region. (a) Patient I-1229, pathology: grade III, T_2-weighted image. (b) Patient I-1275, pathology: grade II, T_2-weighted image. (c) Patient I-1285, pathology: grade IV, proton density-weighted image. (d) Patient I-1212, pathology: grade III, T_2-weighted image. (e) Patient I-1318 (42/195), pathology: grade II, T_2-weighted image. (f) Patient I-1281, no consensus of the patient's pathology was reached. The radiologists differed between grades II and III. None of the voxels of this patient have been used for training. The inset shows a section (indicated by the black square on the classification map) of three probability maps of healthy, grade II, and grade III. Note that a voxel classified as "undecided" can be either a grade II or a grade III.

the probability of grade II shown on the probability maps (inset) is not very high for most voxels. The disagreement between the two radiologists of this patients pathology is therefore reflected in the results of the classifier and probability maps.

Last, the patient with the meningioma tumor was evaluated (classification result not shown). Again, no voxels from this patient were selected for the training set. The classification map shows normal tissue in regions outside the tumor. The area of the tumor is classified as "unknown", which is correct as the classifier was not built to be able to classify meningiomas.

CONCLUSIONS

A new classification approach was developed that enables the classification and probability of classification of brain tumors on the basis of magnetic resonance spectroscopy and imaging data. The model can be used to determine the best preprocessing or data reduction technique in feature space and to recognize overlap of classes. The results of classification, including probabilities, are presented as images to facilitate the interpretation in a clinical environment.

Comparison of PCA and quantification of spectral ranges to reduce the number of variables in feature space gave similar results. The last method was favored, because it was slightly better and because the variables retain a meaningful property. However, PCA is much easier to perform, is not subjective to quantification error, and is therefore a good competitor.

Exploration of the classes in feature space as presented is a powerful method to compare class distances, distributions, and class boundaries. It can be used to identify tumor classes that are not separable, because they overlap in feature space. The overlap between the grade II and III tumor classes can be understood if the method to identify the pathology is considered. When only a small part of the tumorous cells are found to be grade III, the pathology of the patient will be grade III, whereas the spectral features may arise from a bulk of grade II tissue. It is, therefore, never possible to separate these two classes perfectly. The classification thresholds of the model thus have to be set to minimize the number of voxels falsely classified as grade III in patients with a grade II tumor.

The calculation of probabilities of classification from MRI and MRS data is relatively new. The probability maps are very important for interpretation of the classification image. It provides information of the reliability of classification for each voxel, shows whether voxels have a high probability for a class they were not classified to, and can be used as a direct identifier of tissue heterogeneity in certain cases.

Unfortunately, as in any classifier, the classification approach is limited to classification of the classes in the training set. There are many different types of tumor, so an addition of classes to the training set in future would be highly desirable. Then it would also appear whether totally different tumor classes show overlap in feature space. However, the presented approach recognizes "unknown" tumor classes, so at the moment it is a valuable method to differentiate between glial tumors and other tumors. Furthermore, it provides the grade and probability of class membership for glial tumors, which are important parameters in tumor diagnosis.

ACKNOWLEDGMENT

The authors acknowledge Pieter Wesseling (Ph.D., M.D.) and Janneke Schuuring (M.D.) for histopathological identification and selection of patients. The EU funded INTERPRET project (IST-1999-10310; http://carbon.uab.es/INTERPRET/) is gratefully acknowledged.

Received for review May 22, 2003. Accepted August 1, 2003.

AC034541T