MRI-BASED CLASSIFICATION OF BRAIN TUMOR TYPE AND GRADE USING SVM-RFE

Evangelia I. Zacharaki, Sumei Wang, Sanjeev Chawla, Dong Soo Yoo, Ronald Wolf, Elias R. Melhem, Christos Davatzikos

Section of Biomedical Image Analysis, Department of Radiology, University of Pennsylvania

ABSTRACT

The objective of this study is to investigate the use of pattern classification methods for distinguishing different types of brain tumors, such as primary gliomas from metastases, and also for grading of gliomas. A computer-assisted classification method combining conventional magnetic resonance imaging (MRI) and perfusion MRI is developed and used for differential diagnosis. The proposed scheme consists of several steps including ROI definition, feature extraction, feature selection and classification. The extracted features include tumor shape and intensity characteristics as well as rotation invariant texture features. Features subset selection is performed using Support Vector machines (SVMs) with recursive feature elimination. The binary SVM classification accuracy, sensitivity, and specificity, assessed by leave-one-out cross-validation on 102 brain tumors, are respectively 87%, 89%, and 79% for discrimination of metastases from gliomas, and 87%, 83%, and 96% for discrimination of high grade from low grade neoplasms. Multi-class classification is also performed via a one-versus-all voting scheme.

Index Terms— classification, brain tumor, MRI, SVM, feature selection, texture, tumor grade

1. INTRODUCTION

The objective of this study is to provide an automated tool that may assist in the imaging evaluation of brain neoplasms by determining the glioma grade and differentiating between different tissue types, such as primary neoplasms (gliomas) from secondary neoplasms (metastases). These issues are of critical clinical importance in making decisions regarding initial and evolving treatment strategies, and traditional MR imaging is often not adequate in providing answers. Automated tools, if proven accurate, can ultimately be applied to (i) provide more reliable differentiation, especially when the neoplasm is heterogeneous and therefore cannot be adequately sampled by localized needle biopsy, (ii) avoid invasive procedures such as biopsy, especially in cases where the risks outweigh the benefits (iii) expedite or anticipate the diagnosis (histological examination is usually time consuming).

Toward a similar goal, researchers used pattern classification techniques for differentiating brain neoplasms based on Linear Discriminant Analysis (LDA) [1][2], or independent component analysis (ICA) [3] on spectral intensities. Others [4] applied variable selection and classification using Bayesian least squares Support Vector Machines (SVMs) and Relevance Vector Machines on microarray or spectroscopy data. Textural features from T1 post-contrast images were employed in [5] to discriminate between metastatic and primary brain tumors using a probabilistic neural network with a non-linear least squares features transformation method. These studies used a single MR sequence and didn't investigate the contribution of multiple imaging parameters. Multi-parametric features were explored by non-linear classification techniques in [6][7]. Li *et al.* [6] classify gliomas according to their clinical grade using linear SVMs trained on a maximum of 15 descriptive features (such as amount of mass effect or blood supply), which are estimated quantitatively by domain experts. The definition of such features is based on expert knowledge and therefore is not completely automated and reproducible. Devos *et al.* [7] combine standard MR intensities with spectroscopy imaging to improve classification performance using 3 classification techniques.

In this study, we explore the heterogeneous regions of brain tumors by combining imaging features from several sequences, extract morphological and textural characteristics, such as rotation invariant texture features based on Gabor filtering, and assess the significance of each feature in classification. This approach incorporates imaging data which are (or can be) acquired in a routine clinical protocol, including multi-parametric conventional MRI and perfusion. We apply the method for binary classification, but also investigate the multi-class classification problem for differentiating between the most common brain tumors: metastasis, meningioma (usually grade I) and glioma (grade II, II and IV) histopathologically diagnosed and graded according to the World Health Organization (WHO) system.

The proposed scheme consists of four parts: ROI definition, feature extraction, feature selection and classification based on SVMs. Leave-one-out cross-validation is used to test the robustness and accuracy of the proposed classification scheme.

2. METHODS

We propose a multi-parametric framework for brain tumor

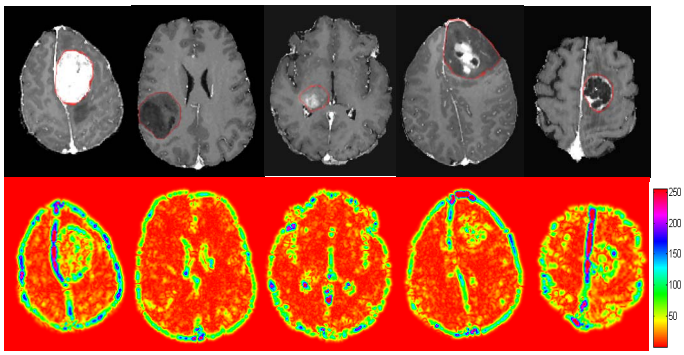

Fig. 1. Illustration of different of brain tumor types and the corresponding texture images. From left to right: meningioma, glioma grade II, III and IV, and metastasis. 1st row: T1ce image with the tumoral region of interest. 2nd row: One of the texture features ($\lambda = 4$, $c = 3$), which proved to be significant in discrimination between high grade gliomas (grade III versus grade IV).

classification and prediction of degree of malignancy by integrating intensity and texture based features into a pattern classification method. The features are first normalized to have zero mean and unit variance. A feature selection method is then used to select a small set of effective features for classification in order to improve the generalization ability and the performance of the classifier. The data and the methodology are described next with more details.

2.1. Data description and preprocessing

We examined 98 patients (52 women, 46 men; age 17-83 years) with a diagnosis of brain tumor who had not been treated at the time of MRI. Four patients had multiple (2), not related to each other, tumors which were regarded as independent masses. All patients underwent biopsy or surgical resection of the tumor with histopathological diagnosis. The total of 102 brain masses graded based on WHO criteria as metastasis (24), meningiomas grade I (4), gliomas grade II (22), which included ependymomas (2) and gliomatosis cerebri (2), gliomas grade III (17), and glioblastomas grade IV (35). Examples of different tumor types are shown in Fig. 1 (1st row).

The following sequences were used: axial 3D T1-weighted (T1), sagittal 3D T2-weighted (T2), Fluid Attenuated Inversion Recovery (FLAIR), axial 3D contrast enhanced T1-weighted (T1ce) images and relative cerebral blood volume (rCBV) maps generated off-line based on T2*-weighted dynamic susceptibility perfusion MRI.

The images were preprocessed following a number of steps including noise reduction, inhomogeneity correction, and rigid intra-subject registration using the public software package FSL [13]. The co-registration of all sequences is required in order to extract features and perform ROI analysis. T1ce was used as reference space (matrix size 192×256×192, voxel size 0.98×0.98×1 mm). The intensity levels were made comparable across subjects by extracting the brain region using BET [14] and applying histogram matching using a linear transformation.

2.2. Definition of ROIs and features

For feature extraction the following ROIs were manually selected:

– ROI1 (neoplastic, enhancing), ROI2 (neoplastic, non-enhancing): includes all non-necrotic enhancing neoplastic tissue, or, if the lesion did not show enhancement, the whole non-necrotic T1-hypointense neoplastic tissue avoiding peritumoral edema by tracing the FLAIR image.

– ROI3 (necrotic): this ROI was delineated only in cases including necrotic tumor tissue.

– ROI4 (edematous): FLAIR and T2 images were used to depict the peritumoral edema (possibly including neoplastic infiltration), drawing the ROI surrounding the high signal intensity seen on these sequences.

We chose a large number of features (141) for investigation which included age, tumor shape characteristics, image intensity characteristics within several regions of interest and Gabor texture features, as explained next.

1) Shape and statistical characteristics of tumor: (evaluated in $\text{ROI1} \cup \text{ROI2} \cup \text{ROI3}$): tumor circularity, irregularity, rectangularity, the entropy of radial length distribution of the boundary voxels, surface-to-volume ratio, percentage of tumor volume being enhancing and necrotic, and the ratio of edema versus tumor volume.

2) Image intensity characteristics: The mean and variance of image intensities are calculated in the central and marginal area of several ROIs. All intensity related features sum up to 52 features in total.

3) Gabor texture: For each radial frequency and orientation, the Gabor features are calculated by convoluting the Gabor filter with an MR sequence. In order to make the Gabor features rotation invariant, for each frequency the magnitudes of Fourier coefficients are calculated [15]. We calculate the texture only from the T1ce in $\text{ROI1} \cup \text{ROI2} \cup \text{ROI3}$ and use $N_\theta = 8$ orientations in $[0, \pi]$ and 5 radial frequencies with corresponding wavelengths $\lambda \in \left\{2\sqrt{2}, 4, 4\sqrt{2}, 8, 8\sqrt{2}\right\}$, and bandwidth $b = 1$. Due to the Fourier transformation, only the first $N_C = N_\theta/2 + 1 = 5$ Fourier coefficients for each frequency are unique. Therefore we finally obtain in total 25 rotation-invariant texture features for tumor classification. One of the texture images ($\lambda = 4$, $c = 3$, where $c = 1, \ldots, N_C$, is the index on the Fourier coefficients) is shown as example in Fig. 1 (2nd row). For each texture image the mean value inside the ROI is used as texture feature.

2.3. Feature selection

First the number of features is reduced by eliminating the less relevant features using a forward selection method based on a ranking criterion and then backward feature elimination is applied using a feature subset selection

method, as explained next.

1) Ranking-based criterion: We use a simple ranking-based feature selection criterion, a two-tailed *t-test*, which measures the significance of a difference of means between two distributions [10], and therefore evaluates the discriminative power of each individual feature in separating two classes. The features are assumed to come from normal distributions with unknown, but equal, variances. Since the correlation among features has been completely ignored in this feature ranking method, redundant features can be inevitably selected, which ultimately affects the classification results. Therefore, we use this feature ranking method to select the more discriminative features, e.g. by applying a cut-off ratio (*p-value* < 0.1), and then apply a feature subset selection method on the reduced feature space, as detailed next.

2) Features subset selection method based on SVM: The support vector machine recursive feature elimination (SVM-RFE) algorithm [11] is applied to find a subset of features that optimizes the performance of the classifier. This algorithm determines the ranking of the features based on a backward sequential selection method that removes one feature at a time. At each time, the removed feature makes the variation of SVM-based leave-one-out error bound smallest, compared to removing other features.

Classification is performed by following a leave-one-out strategy on the training samples. For each leave-one-out experiment, feature ranking is performed using data only from the training samples. The feature selection method is implemented in each training subset in order to correct for the selection bias [12]. It is important that cross-validation is external to the feature selection process in order to more accurately estimate the prediction error. Evidently, there is no guarantee that the same subset of features will be selected at each leave-one-out experiment. We combine the rankings of all leave-one-out experiments and report the total rank of features (in Table 1) according to the frequency of a feature appearing in a specific rank.

2.4. Classification

Classification is performed by starting with the more discriminative features and gradually adding less discriminative features, until classification performance no longer improves. Non-linear Support Vector Machines [8] with Gaussian kernel is used as classifier. Since the data are highly unbalanced and the sample size is rather small to produce balanced classes by subsampling the largest class, we used a weighted SVM [9] to apply larger penalty to the class with the smaller number of samples. If the penalty parameter is not weighted (equal C for both classes), there is an undesirable bias towards the class with the large training size; and thus we set the ratio of penalties for different classes to the inverse ratio of the training class sizes [9].

Moreover, the parameter (γ) that controls the size of the Gaussian radial basis function is chosen as $\gamma = 1/N_F$, where N_F is the number of features. Therefore the maximum kernel size is 1.0 and decreases with increasing number of features during feature selection.

The multi-class problem is solved by constructing and combining several binary SVM classifiers into a voting scheme. We apply majority voting from all one-versus-all binary classification problems. The predictive ability of the classification scheme is assessed by leave-one-out cross validation.

3. RESULTS

The results of leave-one-out cross validation using non-linear SVM are shown in Table 1. The first column shows the tumor type to be classified. The second column shows the number of selected features (N_F) giving the highest classification accuracy. The subsequent columns show the sensitivity, specificity, accuracy (percentage of correctly classified samples) and the area under the receiver operating characteristic curve (AUC), respectively.

The feature selection and ranking showed that parameters extracted from perfusion imaging performed well for most classification tasks. Also the enhancing portion in T1ce was selected as a single feature to classify metastasis from low grade glioma and for distinguishing between high grade gliomas (grade III and IV). A combination of multi-parametric features was selected for classifying primary versus secondary gliomas and low versus high grade gliomas. Gabor texture also seemed to be a relevant pattern.

Fig. 2 shows the classification accuracy with increasing number of retained features for two main classification problems: metastases versus primary and low versus high grade gliomas. The plots illustrate that the fluctuations of accuracy are small and the method is not very sensitive to the number of selected features. The feature selection method can eliminate the redundant features, reduce the noise and build groupings that are both robust and accurate.

For the multi-class problem, one-versus-all classification with features selection and majority voting is applied. The results illustrate that the highest classification accuracy is achieved for low grade gliomas (90.9%) whereas the classification accuracy for glioblastomas is slightly reduced (17.1% are classified as grade III and 14.3% as metastases). The lowest classification rate in the multi-class problem is for the grade III gliomas where the largest portion (52.9%) is classified as grade II and most of the rest as glioblastomas. The prediction of glioma grade is inherently difficult since brain neoplasms are often heterogeneous, meaning that different histopathologic features can be present throughout an individual neoplasm. The failure of the method to classify grade III gliomas possibly indicates that the extracted features do not form a separate cluster, but are rather similar to the features of the nearby classes (grade II and grade IV).

Table 1. Binary classification results for metastasis (MET), meningioma (MEN), and primary gliomas (GL) of grade II, III and IV (GL2, GL3 and GL4, respectively) obtained by leave-one-out cross validation using SVM-RFE.

Binary classification	N_F	Sensitivity (%)	Specificity (%)	Accuracy (%)	AUC
MET-MEN	2	75.0	95.8	92.8	88.5
MET-GL2	1	95.5	100	97.8	95.5
MET-GL3	16	82.4	100	92.7	90.0
MET-GL4	11	82.9	83.3	83.1	87.1
MEN-GL2	6	100	75.0	96.2	100
MEN-GL3	21	94.1	50.0	85.7	86.8
MEN-GL4	2	100	75.0	97.4	97.9
GL2-GL3	3	58.8	68.2	64.1	64.2
GL2-GL4	18	100	95.5	98.2	99.2
GL3-GL4	22	91.4	64.7	82.7	81.7
MET-GL	14	89.2	79.2	86.7	88.2
LOW-HIGH GRADE	19	82.7	95.5	86.5	89.9

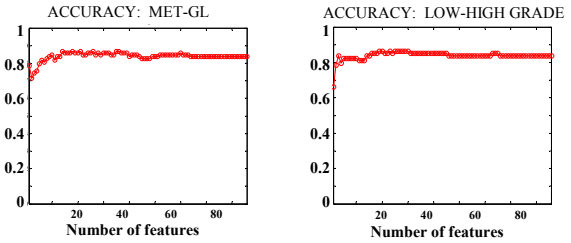

Fig. 2. Classification accuracy versus number of retained features for metastases versus primary gliomas (left) and low versus high grade gliomas (grade II versus III and IV) (right).

4. CONCLUSIONS

In this study we developed a computer-assisted classification method by combining conventional and perfusion MRI for differential tumor diagnosis. We exploited the potential of features extracted automatically from images and investigated the diagnostic value of each feature by applying the support vector machine recursive feature elimination algorithm. The proposed classification scheme achieved high accuracy for most classification problems.

5. ACKNOWLEDGEMENTS

This work was supported in part by NIH grant R01 NS042645.

6. REFERENCES

[1] A.R. Tate, et al., "Automated classification of short echo time in *in vivo* 1H brain tumor spectra: a multicenter study," *Magn. Reson. Med.*, vol. 49, no. 1, pp. 29–36, 2003.

[2] C. Majos, et al., "Brain tumor classification by proton MR spectroscopy: comparison of diagnostic accuracy at short and long TE," *Am. J. Neuroradiol.*, vol. 25, no. 10, pp. 1696–1704, 2004.

[3] Y. Huang, P.J.G. Lisboa, W. El-Deredy, "Tumour grading from magnetic resonance spectroscopy: a comparison of feature extraction with variable selection," *Statist. Med.*, vol. 22, no. 1, pp. 147–164, 2003.

[4] C. Lu, A. Devos, J.A.K. Suykens, C. Arus, S. Van Huffel, "Bagging linear sparse Bayesian learning models for variable selection in cancer diagnosis," *IEEE Trans. on Information Technology in Medicine*, vol. 11, no. 3, pp. 338–346, 2007.

[5] P. Georgiadis, et. al, "Improving brain tumor characterization on MRI by probabilistic neural networks and non-linear transformation of textural features," *Computer Methods and Programs in Biomedicine*, vol. 89, no. 1, pp. 24–32, 2008.

[6] G. Li, J. Yang, C. Ye, D. Geng, "Degree prediction of malignancy in brain glioma using support vector machines," *Computers in Biology and Medicine*, vol. 36, no. 3, pp. 313–325, 2006.

[7] A. Devos, et al., "The use of multivariate MR imaging intensities versus metabolic data from MR spectroscopic imaging for brain tumour classification," *J. Magn. Reson.* , vol. 173, no. 2, pp. 218–228, 2005.

[8] C.-C. Chang, C.-J. Lin, "*LIBSVM: a library for support vector machines*," 2001. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm

[9] Y.-M. Huang, S.-X. Du, "Weighted support vector machine for classification with uneven training class sizes," *Proc. of 2005 Int. Conf. on Machine Learning and Cybernetics*, vol. 7, pp. 4365–4369, 2005.

[10] W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing*, Cambridge University Press, pp. 616, 1992.

[11] I. Guyon, J. Weston, S. Barnhill, V. Vapnik, "Gene selection for cancer classification using support vector machines," *Mach. Learn.*, vol. 46, pp. 389–422, 2002.

[12] C. Ambroise, G.J. McLachlan, "Selection bias in gene extraction on the basis of microarray gene-expression data," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 99, no. 10, pp. 6562–6566, 2002.

[13] S.M. Smith, M. Jenkinson, et al., "Advances in functional and structural MR image analysis and implementation as FSL," *Neuroimage*, vol. 23(S1), pp. 208–219, 2004.

[14] S.M. Smith, "BET: Brain Extraction Tool," *FMRIB* technical report TR00SMS26.

[15] T.N. Tan, "Rotation invariant texture features and their use in automatic script identification," *IEEE Trans. PAMI*, vol. 20, pp. 751–756, 1998.