

Decision Forests for Tissue-specific Segmentation of High-grade Gliomas in Multi-channel MR

D. Zikic¹, B. Glocker¹, E. Konukoglu¹, A. Criminisi¹, C. Demiralp², J. Shotton¹,
O. M. Thomas^{3,4}, T. Das³, R. Jena³, S. J. Price^{3,5}

¹Microsoft Research Cambridge, UK

²Brown University, Providence, RI, USA

³Cambridge University Hospitals, Cambridge, UK

⁴Department of Radiology, Cambridge University, UK

⁵Department of Clinical Neurosciences, Cambridge University, UK

Abstract. We present a method for automatic segmentation of high-grade gliomas and their subregions from multi-channel MR images. Besides segmenting the gross tumor, we also differentiate between active cells, necrotic core, and edema. Our discriminative approach is based on decision forests using context-aware spatial features, and integrates a generative model of tissue appearance, by using the probabilities obtained by tissue-specific Gaussian mixture models as additional input for the forest. Our method classifies the individual tissue types simultaneously, which has the potential to simplify the classification task. The approach is computationally efficient and of low model complexity. The validation is performed on a labeled database of 40 multi-channel MR images, including DTI. We assess the effects of using DTI, and varying the amount of training data. Our segmentation results are highly accurate, and compare favorably to the state of the art.

1 Introduction

In this paper, we present our work on tissue-specific segmentation of high-grade gliomas in multi-channel MR images, with focus on grade IV glioblastoma tumors. Such high-grade gliomas (HGG) grow rapidly, infiltrate the brain in an irregular way, and often create extensive vasculature networks. HGGs contain a necrotic core (NC), surrounded by a varyingly thick layer of active cells (AC). Together, necrotic core and active cells form the gross tumor (GT). Usually, the tumor itself is surrounded by a varying amount of edema (E). In consequence, HGGs have extremely heterogeneous shape, appearance and location (cf. Figs. 1,2), which makes their automatic analysis challenging.

Our goal is to segment high-grade gliomas as well as the individual tissue components automatically and reliably. This would 1) speed-up accurate delineation of the tissue components, which is crucial for radiotherapy and surgery planning and is currently performed manually in a labor intensive fashion, and 2) allow direct volume measurements. Volume measurements are critical for the evaluation of treatment [17], however seldom performed since manual tumor segmentation

is often impractical in a routine clinical setting. Instead more straightforward but less accurate measures are used, such as a pair of perpendicular tumor diameters [17]. Distinguishing between volumes of individual tissue types, especially active cells and necrotic core is an important step for assessment of treatment response. For example, an effective drug might not change the gross tumor volume, while transforming active into necrotic cells. To detect this change, the volumes of both these tissues must be monitored.

This paper proposes an efficient method for automatic segmentation of glioblastoma in multi-channel MR images. While most of the previous research focuses on segmentation of gross tumor, or tumor and edema, we perform a tissue-specific segmentation of three relevant tissues types: active cells (AC), necrotic core (NC), and edema (E). Our method is based on decision forests (DF) [3], a discriminative model which we combine with a generative model of tissue appearance. We achieve this by using the probability estimates based on Gaussian mixture models (GMM) as additional input for the forest. An important advantage of DFs is that they are inherently multi-label classifiers, which allows us to classify the different tissues simultaneously. Such simultaneous separation of classes has the potential to simplify the modeling of the distributions of the individual classes. Through the use of context-sensitive features in the forest, our approach yields a natural smoothness of the segmentation results without explicit regularization. Our method has a low model complexity and reduces the necessity for a large number of pre- and post-processing steps.

The accuracy of our method is evaluated quantitatively on a database of 40 high-grade glioma patients – to our knowledge, the largest annotated database of this kind so far – and compares favorably to the results in the literature.

1.1 Related Work

In recent years several approaches for segmentation of brain tumors have been proposed. The settings differ from one another in many respects, such as the type of tumor being handled (e.g. low-grade [6], high-grade [4, 13], pediatric tumors [16]), the type of anomalous tissues being detected (e.g. GT [7, 10, 16]; GT and E [4, 6, 12, 15]; AC, NC, E [2, 15]), input data, and the evaluation procedure.

A popular group of methods is based on the registration of patient images to a probabilistic brain atlas [6, 9, 10, 12]. The main idea is that – given an accurate atlas and registration – the tumor can be detected as deviation of patient data from the atlas. Since the presence of the tumor makes the registration challenging, some approaches use manual interaction [9], while others integrate tumor growth models [6].

A large group of discriminative methods applies learning techniques to the problem [2, 4, 7, 13, 15, 16]. Our method belongs to this group. Mostly, a learning method is combined with a regularization step, e.g. by modeling the boundary [8, 11], or by applying a variant of a random field spatial prior (MRF/CRF) [4, 7, 16]. Works which classify multiple labels [2, 15] often use SVMs, which are inherently *binary* classifiers. In order to classify different tissues, they are applied

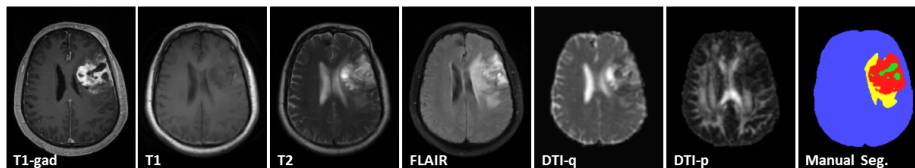


Fig. 1: Example of one of 40 patients in our high-grade glioma database, with tissues labeled as active cells (red), necrotic core (green), and edema (yellow).

hierarchically [2], or in the one-versus-all manner [15]. For these approaches, several classes have to be grouped together, a step which can make the distribution inside the aggregate group more complex than the distribution of each individual class. For example, the intensity distribution of a tumor consisting of AC and NC tissues, which have very different representations in the multi-channel data, is likely to be more complex than the distributions of the single classes. In our technique, we circumvent this potential problem by classifying all tissues simultaneously, which allows us to only handle distributions of individual classes.

Finally, discriminative methods are sometimes seen as requiring heavy data processing and mandatory spatial regularization [10]. In our discriminative approach, despite using only minimal amounts of pre-processing, we achieve high accuracy results without post-hoc regularization.

2 The Labeled High-grade Glioma Database

We acquired a set of multi-channel 3D MR data for 40 patients suffering from high-grade gliomas, with 38 cases of grade IV tumors (glioblastomas) and 2 grade III tumors (anaplastic oligodendrogliomas). The data is acquired prior to treatment. For each patient we have the following 6 channels: T1 post gadolinium (T1-gad), T1, T2 turbo spin echo (T2-tse), and FLAIR, and 2 channels from diffusion tensor imaging (DTI-p and DTI-q). Fig. 1 gives an example for one patient. All acquisitions were performed on a 3T Siemens TrioTim. We will refer to the multi-channel data as I_{MR} . For all 40 patients, a manual segmentation of the three classes of AC, NC, and E is obtained in 3D (see Figs. 1,2).

We try to keep the amount of data pre-processing at a minimum. We perform skull stripping of MR channels [14], and for each patient we perform an affine intra-patient registration of all channels to the T1-gad image. No inter-patient registration is required. We also avoid a full bias-field correction, and only align the mean intensities of the images within each channel by a global multiplicative factor. All these steps are fully automatic.

3 Method: Decision Forests with Initial Probabilities

In our approach we use decision forests (DF)[3,5] as a discriminative multi-class classification method, and combine them with a generative model of tissue

appearance. This is achieved by using initial tissue probability estimates based on trained GMMs as additional input channels for the forest, along with the MR data I_{MR} . We classify four classes AC, NC, and E, and background (B), and gross tumor remains defined as GT=AC \cup NC.

As the first step of our approach, we estimate the initial class probabilities for a given patient as posterior probabilities based on the likelihoods obtained by training a set of GMMs on the training data. For each class c , we train a single GMM, which captures the likelihood $p(I_{\text{MR}}|c)$ of the multi-dimensional intensity for this class. For a given patient data set I_{MR} , the GMM-based posterior probability p_c^{GMM} for the class c is estimated for each point $x \in \mathbb{R}^3$ by $p^{\text{GMM}}(c|x) = p(I_{\text{MR}}(x)|c)p_c / \sum_{c_j} p(I_{\text{MR}}(x)|c_j)p_{c_j}$, with p_c denoting the prior probability for the class c , based on its relative frequency. We can now use the probabilities $p_c^{\text{GMM}}(x)=p^{\text{GMM}}(c|x)$ directly as input for the decision forests, in addition to the multi-channel MR data. So now, our data for one patient is a set of n_c channels $I=(\text{T1-gad}, \text{T1}, \text{T2}, \text{FLAIR}, \text{DTI-q}, \text{DTI-p}, p_{\text{AC}}^{\text{GMM}}, p_{\text{NC}}^{\text{GMM}}, p_{\text{E}}^{\text{GMM}}, p_{\text{B}}^{\text{GMM}})$.

For simplicity, we denote single channels by I_j , and the data for a patient k by $I^{(k)}$. Please note that we can use the GMM-based probabilities for maximum a posteriori classification by $\hat{c} = \arg \max_c p^{\text{GMM}}(c|x)$. We will use this for a base line comparison in Sec. 4.

3.1 Decision Forests

We employ decision forests (DF) to determine a class $c \in C$ for a given spatial input point $x \in \Omega$, based on the representation of x by a feature vector $f(x, I)$. DFs are ensembles of (binary) decision trees, indexed by $t \in [1, n]$. As a supervised method, DFs operate in two stages: training and testing.

During training, each tree t learns a weak class predictor $p_t(c|f(x, I))$ for a n_f -dimensional feature representation $f(x, I) \in \mathbb{R}^{n_f}$ of a spatial point x from the data set I . The input training data set is $\{(f(x, I^{(k)}), c^{(k)}(x)) : x \in \Omega^{(k)}\}$, that is, the feature representations of all spatial points $x \in \Omega^{(k)}$, in all training patient data sets k , and the corresponding manual labels $c^{(k)}(x)$. We refer to all spatial points in all training data sets by $X = \bigcup_k \Omega^{(k)}$. We will use $x \in X$ to identify single training examples in most part, thus writing e.g. $p_t(c|x)$ for $p_t(c|f(x, I))$.

In a decision tree, each node i contains a set of training examples X_i , and a class predictor $p_i^c(c|x)$, which is the probability corresponding to the fraction of points with class c in X_i . Starting with X at the root, the training is performed by successively splitting the training examples at every node based on their feature representation, and assigning the partitions X_L and X_R to the left and right child node. At each node, a number of splits along randomly chosen dimensions of the feature space is considered, and the one maximizing the Information Gain is applied. Tree growing is stopped at a certain tree depth d .

At testing, a point x to be classified is pushed through each tree t , by applying the learned split functions. Upon arriving at a leaf node l , the leaf probability is used as the tree probability, i.e. $p_t(c|x)=p_t^l(c|x)$. The overall probability is computed as the average of tree probabilities, i.e. $p(c|x)=\frac{1}{n} \sum_{t=1}^n p_t(c|x)$. The actual class estimate \hat{c} is chosen as the most probable class, i.e. $\hat{c} = \arg \max_c p(c|x)$.

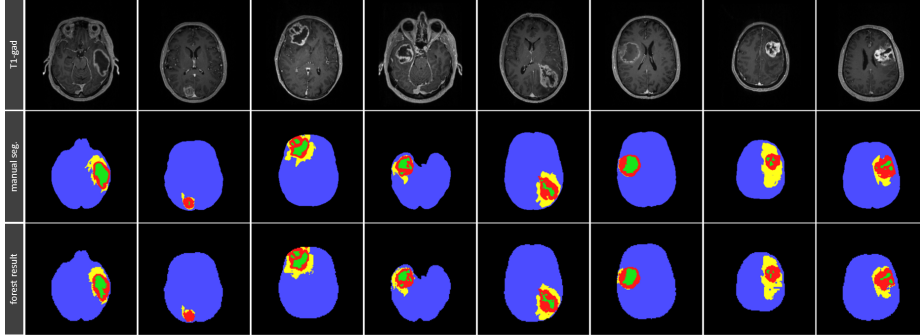


Fig. 2: Examples of results on 8 patients. Obtained by a forest with GMM, MR, and DTI input, with training on 30 patients. The high accuracy of our results is quantitatively confirmed in Figs. 3,4 (AC=red, NC=green, E=yellow).

3.2 Context-aware Feature Types

We employ three spatial and context-aware features types, which are intensity-based and parametrized. Two of these feature types are generic [5], while the third one is designed with the intuition of detecting structure changes. Every instantiated feature with its unique parameters corresponds to one dimension of the feature space used by decision trees.

We use the following notation: Again, $x \in \Omega$ is a spatial point, to be assigned a class, and I_j is an input channel. $N_j^s(x)$ denotes an x -centered and axis aligned 3D cuboid in I_j with edge lengths $s=(s_x, s_y, s_z)$, and $v \in \mathbb{R}^3$ is an offset vector.

Feature Type 1: Intensity difference between x in a channel I_{j_1} and an offset point $x+v$ in a channel I_{j_2} (note that $I_{j_1}=I_{j_2}$ is allowed)

$$f^1(x, I)_{j_1, j_2, v} = I_{j_1}(x) - I_{j_2}(x+v) . \quad (1)$$

Feature Type 2: Difference between intensity means of a cuboid around x in I_{j_1} , and around an offset point $x+v$ in I_{j_2}

$$f^2(x, I)_{j_1, j_2, s_1, s_2, v} = \mu(N_{j_1}^{s_1}(x)) - \mu(N_{j_2}^{s_2}(x+v)) . \quad (2)$$

Feature Type 3: Intensity range along a 3D line between x and $x+v$ in one channel. This type is designed with the intuition that structure changes can yield a large intensity change, e.g. NC being dark and AC bright in T1-gad.

$$f^3(x, I)_{j, v} = \max_{\lambda} (I_j(x + \lambda v)) - \min_{\lambda} (I_j(x + \lambda v)) \quad \text{with } \lambda \in [0, 1] . \quad (3)$$

4 Evaluation

We perform an extensive series of cross-validation experiments to evaluate our method. For this, the 40 patients are randomly split into non-overlapping training and testing data sets. To investigate the influence of the size of the training

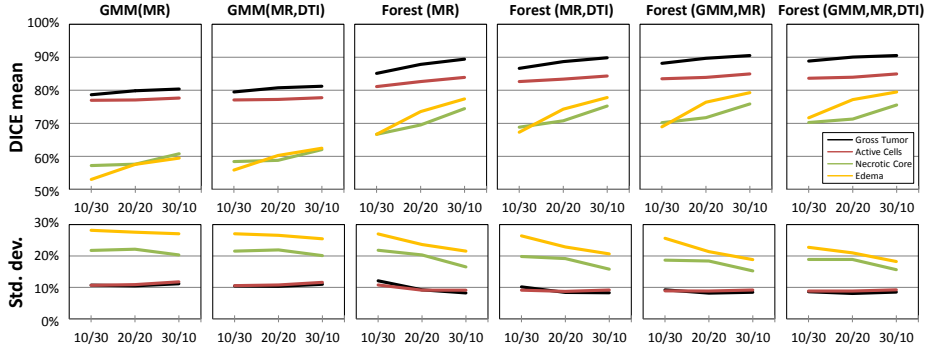


Fig. 3: Average mean and standard deviations of DICE scores, for experiments on 10 random folds, with the training/testing data set sizes of 10/30, 20/20, and 30/10. From left to right, the approaches yield higher mean scores, with lower std. devs. Our approach (rightmost) shows increased robustness to amount of training data, resulting in more horizontal lines, indicating better generalization.

set and generalization properties of our method, we perform experiments with following training/testing sizes: 10/30, 20/20, 30/10. For each of the three ratios, we perform 10 tests, by randomly generating 10 different training/testing splits.

To demonstrate the influence of the single components of the method, we also perform tests on Forests without GMMs, and compare to the results of GMM only. Finally, we investigate the influence of using DTI, by performing all experiments also with MR input only. Overall, this results in 30 random training sets, and 600 tests for each of the 6 approaches. The evaluation is performed with all images sampled to isotropic spatial resolution of 2mm, and forests with $n=40$ trees of depth $d=20$. With these settings, the training of one tree takes between 10-25 min, and testing 2-3 min, depending on the size of training set and the number of channels. The algorithm and feature design were done on a single independent 20/20-fold, which was not used for evaluation.

Fig. 2 shows a visual example of the results, while the quantitative evaluation and more details are given in Figs. 3,4. We observe an improvement of the segmentation accuracy by the proposed method (Forest(GMM,MR,DTI)) compared to the other tested configurations. The amount of training data influences NC and E more than AC and GT. The effect of using DTI seems to be most visible for smaller training data sets.

Comparison to Quantitative Results of Other Approaches is difficult for a number of reasons, most prominently the different input data. To provide some indicative context, we cite results of a recent work from [2]. There, the mean and standard deviation for a leave-one-out cross-validation on 10 glioma patients, based on multichannel MR are as follows: GT: 77 ± 9 , AC: 64 ± 13 , NC: 45 ± 23 , E: 60 ± 16 . Our results compare favorably. For our 30/10-tests we get: GT: 90 ± 9 , AC: 85 ± 9 , NC: 75 ± 16 , E: 80 ± 18 , and for the more challenging 10/30-tests (less training data), we get GT: 89 ± 9 , AC: 84 ± 9 , NC: 70 ± 19 , E: 72 ± 23 .

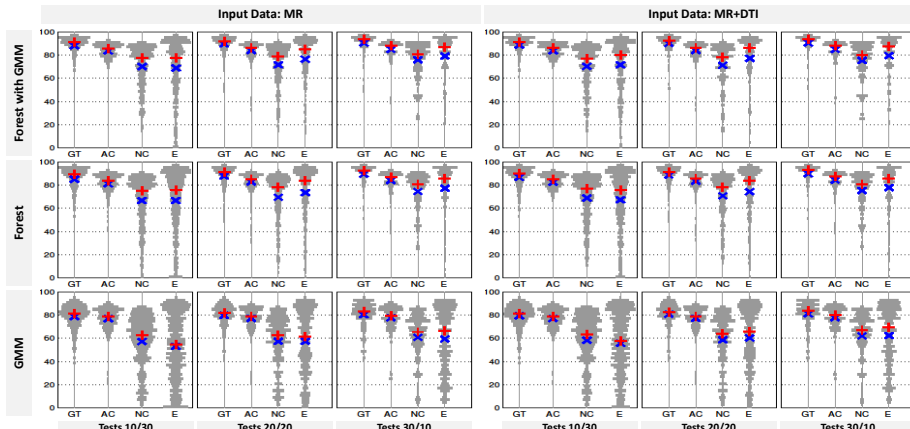


Fig. 4: Evaluation of 10 random fold experiments with varying amount of training data. Distribution of DICE scores is indicated along y-axis, by plotting the histogram of the scores, grouped per test and tissue, with medians (+) and means (\times). Our forest method clearly reduces the number of lower score outliers compared to GMM. The outliers occur mostly for NC and E, and visual inspection confirms that the misclassification of NC and E is the most significant error of our method. Increasing the amount of training data reduces the number of outliers. DTI has most effect for less training data, and the GMM estimates.

Sensitivity to Variation of Parameters is tested by varying $n \in [15, 40]$ and $d \in [12, 20]$, for the ten 30/10-tests. We observe robustness to the selection of these values, especially n . Details are given in the supplementary material [1].

5 Summary and Conclusion

We propose a method for automatic and tissue-specific segmentation of high-grade gliomas. Our discriminative approach is based on decision forests using context-aware features, integrates a generative model of tissue appearance, and classifies different tissues simultaneously. Our method requires comparably little pre-processing, and no explicit regularization, thus resulting in a low model complexity. The approach is computationally efficient, reasonably robust to parameter settings, and achieves highly accurate segmentation results. The automatic results are suitable for volume measurements, and can be used as high-quality initial estimates for interactive treatment planning.

Acknowledgments S. J. Price is funded by a Clinician Scientist Award from the National Institute for Health Research (NIHR). O. M. Thomas is a Clinical Lecturer supported by the NIHR Cambridge Biomedical Research Centre.

References

1. Sup. material: <http://research.microsoft.com/apps/pubs/default.aspx?id=164382>.
2. S. Bauer, L.-P. Nolte, and M. Reyes. Fully automatic segmentation of brain tumor images using support vector machine classification in combination with hierarchical conditional random field regularization. In *MICCAI*, 2011.
3. L. Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.
4. J. J. Corso, E. Sharon, S. Dube, S. El-saden, U. Sinha, and A. Yuille. Efficient multilevel brain tumor segmentation with integrated bayesian model classification. *IEEE Trans. medical Imaging*, 27(5), 2008.
5. A. Criminisi, J. Shotton, and E. Konukoglu. Decision forests: A unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning. *FnT Computer Graphics and Vision*, 2012.
6. A. Gooya, K.M. Pohl, M. Bilello, G. Biros, and C. Davatzikos. Joint segmentation and deformable registration of brain scans guided by a tumor growth model. In *MICCAI*, 2011.
7. L. Gorfiltz, B. H. Menze, M.-A. Weber, B. M. Kelm, and F. A. Hamprecht. Semi-supervised tumor detection in magnetic resonance spectroscopic images using discriminative random fields. In *Proc. of DAGM*, 2007.
8. S. Ho, E. Bullitt, and G. Gerig. Level-set evolution with region competition: automatic 3-d segmentation of brain tumors. In *ICPR*, 2002.
9. M.R. Kaus, S.K. Warfield, A. Nabavi, P.M. Black, F.A. Jolesz, and R. Kikinis. Automated segmentation of brain tumors. *Radiology*, 218, 2001.
10. B. H. Menze, K. V. Leemput, D. Lashkari, M.-A. Weber, N. Ayache, and P. Golland. A generative model for brain tumor segmentation in multi-modal images. In *MICCAI*, 2010.
11. K. Popuri, D. Cobzas, A. Murtha, and M. Jägersand. 3d variational brain tumor segmentation using dirichlet priors on a clustered feature set. *Int J CARS*, 2011.
12. M. Prastawa, E. Bullitt, S. Ho, and G. Gerig. A brain tumor segmentation framework based on outlier detection. *Medical Image Analysis*, 2004.
13. M. Schmidt, I. Levner, R. Greiner, A. Murtha, and A. Bistriz. Segmenting brain tumors using alignment-based features. In *Proc. of ICMLA*, 2005.
14. S. M. Smith. Fast robust automated brain extraction. *Hum. Br. Map.*, 2002.
15. R. Verma, E. I. Zacharaki, Y. Ou, H. Cai, S. Chawla, A.-K. Lee, E.R. Melhem, R. Wolf, and C. Davatzikos. Multi-parametric tissue characterisation of brain neoplasm and their recurrence using pattern classification of MR images. *Acad Radiol.*, 15(8), 2008.
16. M. Wels, G. Carneiro, A. Aplas, M. Huber, D. Comaniciu, and J. Hornegger. A discriminative model-constrained graph-cuts approach to fully automated pediatric brain tumor segmentation in 3D MRI. In *MICCAI*, 2008.
17. P.Y. Wen, D.R. Macdonald, D.A. Reardon, T.F. Cloughesy, A.G. Sorensen, E. Galanis, J. Degroot, W. Wick, M.R. Gilbert, A.B. Lassman, C. Tsien, T. Mikkelsen, E.T. Wong, M.C. Chamberlain, R. Stupp, K.R. Lamborn, M.A. Vogelbaum, M.J. van den Bent, and S.M. Chang. Updated response assessment criteria for high-grade gliomas: response assessment in neuro-oncology working group. *Am. J. Neuroradiol.*, 2010.