Comparison of unsupervised classification methods for brain tumor segmentation using multi-parametric MRI

N. Sauwen[1,2], M. Acou[3], S. Van Cauter[4,5], D.M. Sima[1,2], J. Veraart[6], F. Maes[7], U. Himmelreich[8], E. Achten[3] and S. Van Huffel[1,2]

[1] KU Leuven, Department of Electrical Engineering (ESAT), STADIUS Centre for Dynamical Systems, Signal Processing and Data Analytics, Leuven, Belgium
[2] iMinds, Department of Medical Information Technologies, Belgium
[3] Ghent University Hospital, Department of Radiology, Ghent, Belgium
[4] University Hospitals of Leuven, Department of Radiology, Leuven, Belgium
[5] Ziekenhuizen Oost-Limburg, Department of Radiology, Leuven, Belgium
[6] University of Antwerp, iMinds Vision Lab, Department of Physics, Antwerp, Belgium
[7] KU Leuven, Department of Electrical Engineering (ESAT), PSI Centre for Processing Speech and Images, Leuven, Belgium
[8] KU Leuven, Biomedical MRI/MoSAIC, Department of Imaging and Pathology, Leuven, Belgium

Abstract
Tumor segmentation is a particularly challenging task in high-grade gliomas (HGGs), as they are amongst the most heterogeneous tumors in oncology. An accurate delineation of the lesion and its main subcomponents contributes to optimal treatment planning, prognosis and follow-up. Conventional MRI (cMRI) is the imaging modality of choice for manual segmentation, and is also considered in the vast majority of automated segmentation studies. Advanced MRI modalities such as perfusion-weighted imaging (PWI), diffusion-weighted imaging (DWI) and magnetic resonance spectroscopic imaging (MRSI) have already shown their added value in tumor tissue characterization, hence there have been recent suggestions of combining different MRI modalities into a multi-parametric MRI (MP-MRI) approach for brain tumor segmentation. In this paper, we compare the performance of several unsupervised classification methods for HGG segmentation based on MP-MRI data including cMRI, DWI, MRSI and PWI. Two independent MP-MRI datasets with a different acquisition protocol were available from different hospitals. We demonstrate that a hierarchical non-negative matrix factorization variant which was previously introduced for MP-MRI tumor segmentation gives the best performance in terms of mean Dice-scores for the pathologic tissue classes on both datasets.

Keywords: Segmentation, glioma, multi-parametric MRI, unsupervised classification, non-negative matrix factorization, clustering

Abbreviations: *1H-MRSI: proton magnetic resonance spectroscopic imaging; ADC: apparent diffusion coefficient; Cho: total choline; cMRI: conventional magnetic resonance imaging; Cre: total creatine; DKI: diffusion kurtosis imaging; DSC-MRI: dynamic susceptibility-weighted contrast-enhanced magnetic resonance imaging; DTI: diffusion tensor imaging; DWI: diffusion-weighted imaging; FA: fractional anisotropy; FCM: fuzzy C-means clustering; FLAIR: fluid-attenuated inversion recovery; GBM: glioblastoma multiforme; Glx: glutamine+glutamate; Gly: glycine; GMM: Gaussian mixture modelling; HALS: hierarchical alternating least squares; HGG: high-grade glioma; hNMF: hierarchical non-negative matrix factorization; Lac: lactate; LGG: low-grade glioma; Lip: lipids; MD: mean diffusivity; mI: myo-inositol; MK: mean kurtosis; MP-MRI: multi-parametric magnetic resonance imaging; NAA: N-acetyl-aspartate; NMF: non-negative matrix factorization; NNLS: non-negative linear least-squares; PWI: perfusion-weighted imaging; rCBV: relative cerebral blood volume; ROI:*

region of interest; SC: spectral clustering; SPA: successive projection algorithm; T1c: contrast-enhanced T1; UZ Gent: University hospital of Ghent; UZ Leuven: University hospitals of Leuven

1 Introduction

High-grade gliomas (HGGs) are the most common type of malignant primary brain tumors [1]. Despite considerable advances in understanding their biological behavior, patient prognosis remains poor due to their rapid infiltrative growth into the surrounding healthy tissue. Anaplastic astrocytomas (WHO grade III) have a 5-year survival rate of approximately 30%, whereas glioblastoma multiforme (GBM, WHO grade IV), the most malignant type, has a 5-year survival rate of only 5% [2]. HGGs are amongst the most heterogeneous tumors in oncology. They are diffuse, exhibiting unclear and irregular boundaries, preferentially invading the surrounding tissue along white matter tracts [3]. Different stages of the disease can occur within the same lesion [4], with varying degrees of tumor enhancement, mitotic activity and necrosis. Furthermore, tumor structures vary considerably in terms of size, shape and location. This makes tumor segmentation, i.e. the imaging-based delineation of the tumor and its subcomponents, particularly challenging in HGGs.

Tumor segmentation is a crucial task in the treatment planning and follow-up of HGG patients. Manual segmentation by a neuro-radiologist, usually based on conventional MRI, is currently still the gold standard. It is a tedious and time-consuming task, and it is susceptible to subjective interpretation. Intra- and inter-observer variabilities of 20% and higher have been reported for manual tumor delineation [5, 6]. To alleviate these limitations and due to the high clinical relevance, there has been an increasing interest for automated brain tumor segmentation. Classification methods, which rely primarily on voxel intensity-based features for differentiating tissue types, have been receiving the most attention [6]. They have the advantage of being able to deal with multi-parametric imaging datasets, combining voxel-wise information from different MRI modalities. Classification methods can be further categorized into supervised and unsupervised methods. Supervised classification methods naturally incorporate large amounts of prior knowledge in the form of a training dataset with known tissue labels. From this training dataset, the algorithm learns decision boundaries between tissue classes in high-dimensional feature space, which can then be applied to unlabeled test data. Supervised classification methods require extensive training datasets, to account for the wide heterogeneity in tumor appearance among glioma grades and occasional labelling errors. Unsupervised classification methods reveal structure in a dataset by modelling the similarity within the data itself. These methods can directly be applied to any imaging dataset, irrespective of the acquisition protocol and without the need for any training data. Due to the lack of manually annotated ground truth, unsupervised methods depend more strongly on the incorporation of additional prior knowledge, e.g. by imposing spatial coherence [7] or feature-specific knowledge [8], to achieve a valid tissue segmentation.

Fuzzy C-means clustering (FCM) is one of the most popular unsupervised algorithms for brain tumor segmentation [9]. FCM takes into account the MRI feature overlap between tissue classes, assigning fuzzy membership values to different tissue types. FCM was introduced for brain tumor segmentation by Philips et al. [10] and was later combined with knowledge-based techniques for improved performance [8, 11]. Several studies have applied finite mixture modelling for unsupervised tumor segmentation. Mostly, each tissue type is modelled by a multi-variate Gaussian distribution, resulting in a Gaussian mixture model (GMM) [12, 13]. Recently, GMM has been shown to be competitive with state-of-the-art supervised classification algorithms for GBM segmentation based on cMRI data [14]. Over the last years graph-cut based methods have become popular as well. A semi-supervised procedure based on graph-cut has been presented to differentiate tumor tissue

from healthy brain using MRSI data [15]. Padole et al. have initiated a normalized-cut method using the mean-shift algorithm for increased computational efficiency [16].

Nearly all of the proposed segmentation algorithms have so far mainly been applied to cMRI data [10-14]. However, along with the recent emergence of advanced MRI modalities such as diffusion-weighted MRI (DWI), perfusion-weighed MRI (PWI) and magnetic resonance spectroscopic imaging (MRSI), numerous studies have suggested that tumor characterization might benefit from the additional structural, biological and biochemical information provided by these advanced MRI modalities [17-19].

DWI probes diffusion of water molecules and its interaction with a local microstructure. Minimal apparent diffusion coefficient (ADC) values have been shown to inversely correlate with tumor cellularity [20]. Diffusion tensor imaging (DTI) models the diffusion process three-dimensionally, providing insight in diffusion directionality and tissue structure through mean diffusivity (MD) and fractional anisotropy (FA) [21]. DTI has been shown to better delineate tumor margins in gliomas than cMRI [3]. Jones et al. presented a DTI based segmentation to delineate tumor volumes of interest using isotropic and anisotropic components of the diffusion tensor [22]. Diffusion kurtosis imaging (DKI) is a recent extension of DTI, quantifying the non-Gaussian component of diffusion by the mean kurtosis (MK) [23]. The additional information from DKI is thought to indicate the complexity of the microstructural environment [24].

PWI is widely used for studying tumor angiogenesis, mainly through the quantification of relative cerebral blood volume (rCBV) [25]. HGGs are known to promote vascular ingrowth under hypoxic conditions [26]. Strong correlations have been reported between rCBV and cell density, and between rCBV and microvessel density in HGGs [27]. PWI has been reported to detect tumor recurrence at an earlier stage than cMRI [28] and has been shown to be useful in differentiating active glioma from radiation necrosis [29].

Proton MRSI (1H-MRSI) provides biochemical information about the spatial distribution of metabolites in a localized region of the brain. The principal metabolites seen in brain tumors include lactate (Lac), lipids (Lip), N-acetyl-aspartate (NAA), glutamine+glutamate (Glx), total creatine (Cre), total choline (Cho), glycine (Gly) and myo-inositol (mI). NAA levels progressively decrease with increasing glioma grade, as it is a marker of neuronal cell density [30]. Cho indicates cell membrane density and integrity, showing elevated levels in HGGs with increased cell density. Elevated levels of Lip are mostly seen in GBM, as it is a hallmark of necrosis [31]. 1H-MRSI has been reported to detect metabolically active tumor beyond the radiological boundaries defined by cMRI [32].

Thus far, only few studies have combined different MRI modalities for glioma segmentation. Di Costanzo et al. used DWI, PWI and 1H-MRSI to classify a set of manually delineated regions of interest (ROIs) in the non-necrotic part of the tumor and the peritumoral region [33]. Step-wise linear discriminant analysis was applied to the ROI averaged MP-MRI parameters to differentiate between regions of gross tumor, edema, tumor/edema, tumor infiltration and normal tissue. Highest classification accuracy was reported when all MRI modalities were included. Verma et al. applied support vector machines to a combined set of cMRI and DTI parameters to segment intensity-based tissue profiles on a voxel-wise basis [34]. They provide tissue probability maps as well as hard segmentation of enhancing tumor, non-enhancing tumor and edema in the pathologic region. Zikic et al. applied a decision forest classifier on a set of voxel-wise intensities and context-aware features from cMRI and DTI [35]. They report an improvement of the segmentation accuracy when DTI is included compared to using only cMRI data. As extensive MP-MRI datasets might be more prone to variations in the acquisition protocol, unsupervised methods benefit from not requiring a uniform

training dataset. Kazerooni et al. are among the first to explore unsupervised classification methods for MP-MRI based glioma segmentation [8]. They employed spatial fuzzy C-means clustering to a combination of cMRI, DWI and PWI data, followed by a region growing step. Unsupervised classification based on non-negative matrix factorization (NMF) has been proposed for differentiating brain tumor from healthy tissue in MRSI data [36, 37]. In a previous study, we applied hierarchical NMF (hNMF) for tumor segmentation to an MP-MRI dataset combining cMRI, DWI, PWI and MRSI data [38]. Segmentation results were shown to be significantly better when using the MP-MRI data compared to using cMRI data only.

In this paper, we present an hNMF variant with improved computational efficiency, making it approximately 10 times faster while maintaining segmentation performance compared to [38]. We compare its performance to 5 other state-of-the-art unsupervised classification algorithms for brain tumor segmentation using MP-MRI data, namely 2 single-level NMF methods (hierarchical alternating least-squares NMF and convex NMF) and 3 clustering methods (fuzzy C-means, Gaussian mixture modelling and spectral clustering). In order to make our conclusions more general, all methods are applied to 2 independent MP-MRI datasets, acquired at different hospitals and using a different acquisition protocol. To assess the added value of the advanced MRI modalities, all methods were also applied when considering only cMRI data.

2 MP-MRI datasets

Two independent MP-MRI datasets were acquired at the university hospitals of Ghent (UZ Gent) and Leuven (UZ Leuven). Both datasets included cMRI, DWI, PWI and 1H-MRSI, but the acquisition protocols and the resulting sets of MRI features were different. Table 1 gives a concise overview of the acquisition protocols and MP-MRI features from both datasets.

2.1 UZ Gent

21 HGG patients (12 GBMs, 1 grade III astrocytoma, 1 grade III oligodendroglioma, 7 grade III oligoastrocytomas) were included in the study. The lesions were classified according to grade using the 2007 WHO classification [39]. Retrospective analysis of the data was approved by the local ethics committee. MR examinations were performed on a 3T scanner (Siemens Trio Tim, Erlangen, Germany), using a standard 12-channel phased array head coil. cMRI included T1 with and without contrast (T1c and T1, voxel size 0.9×0.85×0.85mm^3) and fluid-attenuated inversion recovery (FLAIR, voxel size 1×0.98×0.98mm^3) imaging. Axial DWI images were acquired using a fast single-shot gradient echo echo-planar imaging sequence (voxel size 2×2×3mm^3). The raw DWI data were averaged over 3 orthogonal directions and ADC maps were calculated using weighted linear least-squares fitting [40]. The T2-weighted b0 images were also added to the feature set. PWI was performed by using a T2*-weighted dynamic susceptibility-weighted contrast-enhanced MRI (DSC-MRI) sequence (voxel size 1.8×1.8×5mm^3). rCBV maps were quantified from the dynamic signal intensity curves using the method proposed by Boxerman et al., which accounts for leakage correction [41]. A 3D 1H-MRSI protocol with long echo-time was included (voxel size 10×10×15mm^3). In the two-slice MRSI examination, a volume of interest was positioned manually to include tumor, perilesional edema and normal brain tissue. Metabolite quantification was performed using AQSES-MRSI [42], resulting in Lac, Lip, NAA, Glx, Cre and Cho maps. A detailed description of the UZ Gent MP-MRI acquisition protocol and post-processing of the MRI features can be found in [43].

2.2 UZ Leuven

14 HGG patients were enrolled in the study (11 GBMs, 2 grade III astrocytomas and 1 grade III oligoastrocytoma). Written informed consent was obtained from every patient before participation.

MRI acquisition was performed on a 3T MR system (Philips Achieva, Best, The Netherlands), using a body coil for transmission and an 8-channel head coil for signal reception. cMRI consisted of T2-weighted imaging (voxel size 0.45×0.45×4mm^3), T1-weighted imaging after contrast administration (0.98×0.98×1mm^3) and FLAIR (0.45×0.45×4mm^3). An extensive DWI echo-planar imaging sequence was used to acquire DKI data (voxel size 2.5×2.5×2.5mm^3). Diffusion and kurtosis tensors were estimated in each voxel using a constrained weighted linear least-squares algorithm [40]. MD, FA and MK maps were calculated according to [44, 45]. PWI consisted of a DSC-MRI sequence (voxel size 1.56×1.56×3mm^3). rCBV maps were derived from the dynamic signal intensity curves using the method by Boxerman et al [41]. Short echo-time 1H-MRSI data were acquired for a 2-dimensional ROI positioned in the center of the tumor (voxel size 10×10×10mm^3). AQSES-MRSI was used to quantify Lac, Lip, NAA, Glx, Cre, Cho, Gly and mI. For a detailed description of the UZ Leuven acquisition protocol and post-processing of the MRI features, the reader is referred to [38].

	UZ Gent dataset (21 patients)	**UZ Leuven dataset** (14 patients)
cMRI	**T1, T1c, FLAIR**	**T2, T1c, FLAIR**
DWI	*DWI (3 b-values, 1+3+3 directions)* **ADC, b0**	*DKI (4 b-values, 10+25+40+75 directions)* **MD, MK, FA**
1H-MRSI	*3D long echo 1H-MRSI* **Lac, Lip, NAA, Glx, Cre, Cho**	*2D short echo 1H-MRSI* **Lac, Lip, NAA, Glx, Cre, Cho, Gly, mI**
PWI	*DSC-MRI* **rCBV**	*DSC-MRI* **rCBV**

Table 1: Schematic overview of the MP-MRI protocols and derived parameters of the 2 datasets acquired at UZ Gent and UZ Leuven.

2.3 Image coregistration and voxel selection

To allow for voxel-wise segmentation, all MP-MRI parameters were coregistered and resampled to the same spatial resolution of 1×1×3mm^3. cMRI data were skull-stripped and T1c served as a reference for rigid coregistration in SPM8 (Wellcome Trust Centre for Neuroimaging, University College London), using the normalized mutual information criterion [46] and cubic B-spline interpolation for reslicing. The MRSI data were spatially aligned and resliced to the T1c reference without coregistration, due to their low spatial resolution and limited spatial extent. Only voxels within the MRSI ROI have a full set of MP-MRI features, so only those voxels were included in the segmentation analyses. For the UZ Gent dataset, voxels from 10 slices located within the 3D MRSI ROI were included. For the UZ Leuven dataset, voxels in a central slice intersecting the 2D MRSI ROI were considered. Additional intensity-based features were added to the feature set to include localized spatial information. An in-plane local neighbourhood of 3×3 and 5×5 voxels was used to calculate average intensity values that were assigned to the central voxel. These spatially averaged intensity values were added for all MP-MRI features except for the MRSI data, because of the low spatial MRSI resolution. Finally, each feature's full range was rescaled linearly to [0-1].

3 Methods

In this paper, several state-of-the-art unsupervised MP-MRI classification methods were applied and evaluated on the UZ Gent and UZ Leuven datasets for brain tumor segmentation, 3 of which are based on NMF and 3 are based on clustering.

3.1 Non-negative matrix factorization

Given a non-negative input matrix *X*, NMF will provide a low-rank (rank *r*) approximation of *X* as the product of 2 non-negative factor matrices *W* and *H*:

$$X \approx WH \;\; with \;\; X \in R_+^{m \times n}, W \in R_+^{m \times r} \; and \; H \in R_+^{r \times n} \quad (1)$$

NMF aims at finding meaningful basic components which are present in a dataset, using an additive parts-based representation to model the data. As we are dealing with image intensities, the non-negativity constraint applies naturally. Each column of *X* corresponds to one data point, i.e. one voxel's MP-MRI feature vector. The columns of *W*, the so-called sources, will represent tissue-specific MP-MRI signatures. Each row of *H* will then contain the relative weights (the so-called abundances) of the corresponding source in *W* for all voxels. As such, each data point is modelled as a weighted sum of the tissue-specific signatures. The abundances in *H* tell us which tissue types are most prominent in each voxel. To convert the multi-linear NMF result into a hard segmentation, k-means clustering is applied to the abundance values in *H*. As we assume the NMF sources to correspond to the tissue classes, we initialize each cluster centroid with a different abundance value set to 1, and all other abundance values set to 0. The most commonly used cost function to assess the factor approximation is the Frobenius norm, which is based on the squared Euclidian distance as a similarity metric:

$$\min_{W,H} f(W,H) = \frac{1}{2}\|X - WH\|_F^2 \, , such \; that \;\; \forall i,j : W_{i,j} \geq 0, H_{i,j} \geq 0 \;\; (2)$$

Many algorithms have been developed to solve this non-convex optimization problem. In the next sections, we will propose 2 commonly used NMF algorithms for brain tumor segmentation, along with the hNMF algorithm [38].

3.1.1 HALS NMF

Hierarchical alternating least-squares (HALS) NMF belongs to the family of alternating least-squares algorithms (ALS) [47]. These methods rely on the observation that optimizing *W* when *H* is fixed, and optimizing *H* when *W* is fixed are convex problems, as opposed to the original non-convex NMF problem as defined in Equation 2. ALS NMF methods iteratively update *W* and *H* by solving the normal equations:

$$H^{(k)}H^{(k)^T}W^{(k+1)^T} = H^{(k)}X^T \quad for \; updating \; W \quad (3)$$

$$W^{(k)^T}W^{(k)}H^{k+1} = W^{(k)^T}X \quad for \; updating \; H \quad (4)$$

HALS NMF successively updates the individual columns of *W* and the individual rows of *H*, as one obtains a computationally efficient closed-form solution when decoupling the variables in this way. In the current study we apply an accelerated HALS NMF algorithm because of its fast convergence compared to other ALS based methods [48].

3.1.2 Convex NMF

Convex NMF imposes the source vectors (columns) of *W* to lie within the column space of *X* [49]. Moreover, each source must be a weighted sum of some of the data points. An auxiliary non-negative matrix *A* is introduced to enforce this additional constraint:

$$W = XA \quad such \; that \quad X \approx XAH \quad (5)$$

Multiplicative update rules are defined to alternatingly revise *A* and *H* towards convergence. Convex NMF naturally leads to a sparse abundance matrix *H*.

3.1.3 Hierarchical NMF

Hierarchical NMF variants have been previously used for document clustering [50], for unmixing hyperspectral images [51] and for tumor tissue differentiation based on MRSI data [33]. Our hNMF method was introduced in a previous paper for MP-MRI based glioma segmentation [38]. We have currently adapted the original hNMF algorithm, making it computationally more efficient while maintaining similar performance. The hNMF method consists of 2 levels of NMF, using HALS NMF at both levels. First, rank-2 NMF is applied to the input matrix X, resulting in 2 sources (W_1 and W_2) and corresponding abundance maps (H_1 and H_2). It is assumed that each tissue class is mainly represented by one source, such that we can assign each tissue type to either W_1 or W_2. The voxels are then divided over the 2 sources based on their abundance values H_1 and H_2. Whereas hNMF originally used iterative thresholding for dividing the voxels, the algorithm has been adapted such that k-means clustering is applied to the H_1 and H_2 values for efficient voxel assignment. The clusters are initialized at [1 0] and [0 1], i.e. we expect to find one cluster (X_1) with high H_1 and low H_2 values, and another cluster (X_2) with low H_1 and high H_2 values. By doing so, we were able to make the hNMF computation approximately 10 times faster while maintaining segmentation performance. A second level of NMF is then applied to X_1 and X_2 separately, with the rank set to the number of tissue types represented by each cluster, k_1 and k_2. k_1 and k_2 are determined based on visual inspection of the abundance maps H_1 and H_2 in combination with T1c: each tissue class is assigned to the source for which it has the highest abundance values. The sources found at this second level are the actual tissue signatures. Finally, we recombine all the voxels and use non-negative least-squares fitting (NNLS) with the tissue signatures to obtain abundance maps for each tissue type over the full ROI. A schematic overview of the hNMF procedure is shown in Figure 1.

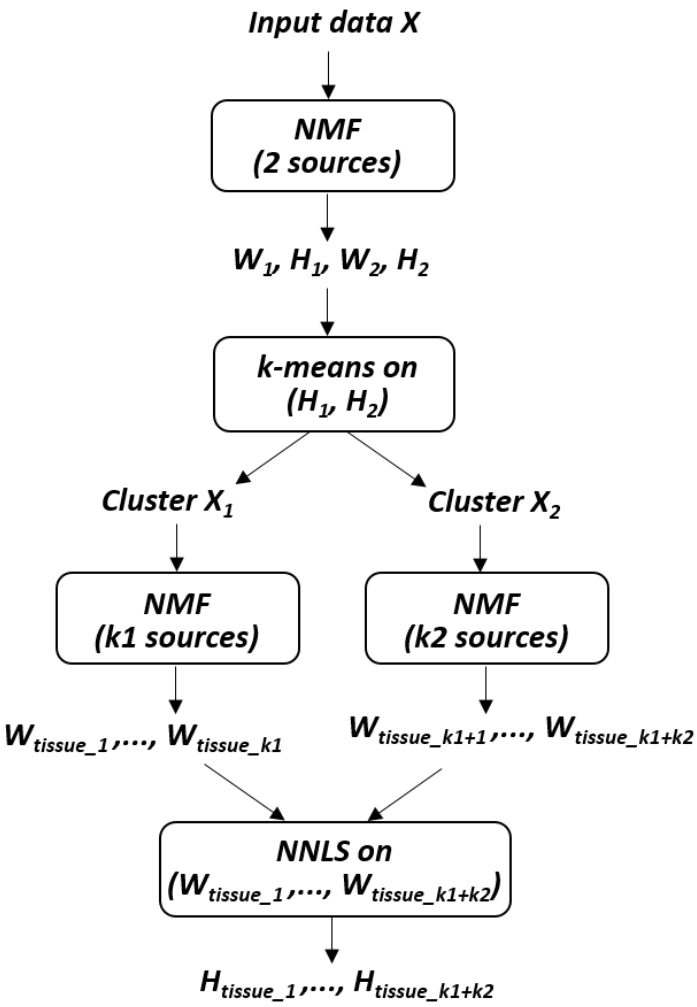


Figure 1: Schematic overview of the hNMF algorithm.

3.2 Clustering

3.2.1 Fuzzy C-means clustering

FCM is a clustering method that allows each data point to belong to multiple clusters with varying degrees of membership. These membership grades indicate the degree to which data points belong to each cluster. FCM aims at minimizing the following objective function:

$$\underset{C}{\operatorname{argmin}} \sum_{i=1}^{n} \sum_{j=1}^{k} w_{i,j}^{m} \left\| x_i - c_j \right\|^2 \quad (6)$$

where m is the fuzziness exponent which determines the level of fuzziness, $w_{i,j}$ is the degree of membership of data point x_i to cluster C_j. Here, the data points are the columns of the input matrix X, representing the MP-MRI features per voxel. The cluster centroids c_j and the degrees of membership $w_{i,j}$ are calculated by:

$$c_j = \frac{\sum_{i=1}^{n} w_{i,j}^{m} x_i}{\sum_{i=1}^{n} w_{i,j}^{m}} \quad (7)$$

$$w_{i,j} = \frac{1}{\sum_{j=1}^{k} \left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^{\frac{2}{m-1}}} \quad (8)$$

FCM is carried out through an iterative optimization of the objective function in Equation 6, with the update of cluster centroids c_j and cluster memberships $w_{i,j}$ through equations 7 and 8, respectively. In the current study we have set the fuzziness exponent *m* to 2. Hard segmentation is obtained from FCM by assigning each data point to the cluster for which it has the highest membership value.

3.2.2 Gaussian mixture modelling

GMM aims at finding the maximum likelihood parameters of a mixture of Gaussians fitting the input data. This comes down to maximizing the posterior probability of a parameter set Θ of *k* Gaussian components, given the input data *X*:

$$p(\Theta|X) = \sum_{i=1}^{k} \varphi_i \mathcal{N}(X|\mu_i, \sigma_i) \quad (9)$$

with φ_i the weight of the i^{th} Gaussian component and $\mathcal{N}(x|\mu_i, \sigma_i)$ the probability of a data point *x* (i.e. columns of the data matrix *X* containing MP-MRI features) belonging to a normal distribution with mean μ_i and standard deviation σ_i. The Expectation Maximization (EM) algorithm [52] is the most popular technique to solve this optimization problem. EM alternates between an expectation step (E-step) and a maximization step (M-step) until convergence. In the E-step an estimation of the posterior probability $p(\Theta|X)$ is computed given the current estimation of the model parameters. In the M-step a maximum likelihood update of the model parameters is performed based on the posterior probability computed in the E-step. GMM is a soft clustering technique, i.e. it will provide probabilities of the voxels belonging to the different tissue classes. Hard segmentation is obtained by assigning each voxel to the tissue class with the highest probability. In order to recover from singular covariance matrices, a regularization constant of $1e^{-3}$ was added to the diagonal elements of the covariance matrices [53].

3.2.3 Spectral clustering

Spectral clustering (SC) has become a popular tool for image segmentation. Several studies have applied SC for automated brain tumor and lesion detection [54, 55]. SC refers to a family of graph partitioning methods which make use of the eigenvalue decomposition of a graph Laplacian matrix for dimensionality reduction prior to the actual clustering step [56]. Its success is mainly based on the fact that SC does not make any strong assumptions about the cluster shape nor about the density distribution of the data points. The aim of SC is to find a partitioning of the similarity graph such that the edge weights between the data clusters are very low. The graph Laplacian matrix *L* is derived from the adjacency matrix *W*, which defines the edge weights between the graph nodes. We used a k-nearest neighbour graph to obtain a sparse adjacency matrix, for efficiently performing the eigenvalue decomposition. The Gaussian similarity function $\exp\left(\left(-\|x_i - x_j\|^2/(2\sigma^2)\right)\right)$, where x_i and x_j stand for columns of the data matrix *X*, was used to calculate the weights in *W* [56]. Careful selection of the connectivity parameters is crucial in constructing the adjacency matrix, such that the connected components are well separable. We set the number of *k* nearest neighbours to 50, being in the order of the logarithm of the number of data points for both the UZ Leuven and UZ Gent datasets [57]. For the Gaussian similarity function, σ was set to 0.5, being in the order of the mean distance of the data points to their k-nearest neighbour [56]. Several graph Laplacian matrices have been proposed in literature. We used the normalized Laplacian proposed by Shi and Malik [58]:

$$L = I - D^{-1}W \quad (10)$$

where *I* is the identity matrix and *D* is a diagonal matrix with its diagonal elements equal to the sums of the rows of *W*. Eigenvalue decomposition is then applied to *L* and a matrix *U* is created with the first *k* eigenvectors as its columns. Finally, k-means clustering is applied to the rows of *U* to obtain the *k* clusters.

3.3 Initialization

Except for SC, all of the proposed unsupervised classification methods are non-convex, i.e. the obtained solution will be a local rather than a global optimum, depending on the initialization conditions. A wide range of initialization methods have been developed for NMF and clustering. As NMF and clustering assume different data models, their most common initialization strategies (besides random initialization) also differ. To avoid favouring either method, we have performed segmentation using 2 initialization strategies: the successive projection algorithm (SPA), which has been suggested for NMF source detection, and kmeans++, a popular initialization method for clustering. SPA is a forward selection method which minimizes collinearity of the selected variables in vector space [59]. It has been commonly used as an endmember extraction tool for hyperspectral unmixing. SPA aims at finding the vertices of a convex hull spanning the dataset, under the assumptions of near-separable NMF [60]. SPA iteratively adds data points to the set of initial sources, at each step selecting the data point with the highest l2-norm in the orthogonal subspace of the already selected sources. Kmeans++ is a recently proposed method for initializing the cluster centroids [61]. The intuition behind this approach is that it is favorable to spread out the initial centroids. The first cluster center is chosen randomly from the data points, after which each subsequent centroid is chosen from the remaining data points with probability proportional to its squared distance from the point's closest existing cluster center. As kmeans++ is not a deterministic algorithm, we successively ran it 20 times and withheld the result with the best (= highest or lowest) objective function value for the considered classification method. SC does not require any initialization, but we applied both initialization methods to the k-means clustering step in the final part of the algorithm.

3.4 Validation

The unsupervised classification methods were applied to each patient's MP-MRI dataset individually. Segmentation results obtained from the NMF and clustering methods were validated based on manual expert labelling. Manual delineations of the pathologic tissue regions were obtained at the 2 hospitals from an experienced radiologist, each of them having more than 5 years of experience in neuro-radiology. The number of tissue classes present within each patient's ROI was determined based on visual inspection by the radiologists. The Dice-score [6] was used to quantify the spatial alignment between each method's automated segmentation and the manual segmentation. Dice-scores were calculated for the main pathologic tissue types: actively proliferating tumor (active tumor), necrosis and edema. Furthermore, to be consistent with the results reported for the BRATS challenge [6], Dice-scores were also reported for the tumor core (active tumor + necrosis) and the whole tumor (tumor core + edema). To evaluate the added value of including PWI, DWI and MRSI in terms of segmentation performance, all unsupervised algorithms were also applied to the UZ Gent dataset when considering cMRI data only. Average computation times were calculated for the different methods to compare their computational cost. We also wanted to assess how well the built-in modelling assumptions of the different methods align with the actual data in feature space, e.g. in terms of spatial distribution, the shape of the tissue clusters and the overlap between them. Principal component analysis was applied to the pathologic data of each patient and the data points were projected onto the 2 main principal components to allow for visual inspection of the data distribution.

4 Results

4.1 UZ Gent

A comparison of the segmentation results of the unsupervised algorithms on the UZ Gent MP-MRI dataset is given in Table 2 for kmeans++ initialization and in Table 3 for SPA initialization. The highest mean Dice-score is marked in bold for each tissue class. With both initialization methods, hNMF has the highest mean Dice-score for all tissue classes except for necrosis, for which SC performs best. We also report for each method the number of times it did not detect a pathologic tissue class which was annotated by the radiologist. The number of cases with undetected tissue classes ranged from 0 to 1 out of 21 patients for active tumor, from 0 to 3 out of 12 patients for necrosis and from 2 to 7 out of 15 patients for edema. The high number of cases of undetected edema is also reflected in its low Dice-scores. This is mainly attributed to the relatively high number of grade III patients in the UZ Gent dataset, many of them exhibiting non-enhancing tumor that is hard to differentiate from the peri-tumoral edema. Neither initialization method gave clearly better performance for any of the segmentation algorithms. For edema, we see that kmeans++ achieves Dice-scores which are at least as high as with SPA and the number of undetected edema cases are also generally lower.

		NMF			**Clustering**		
		HALS	**Convex**	**hNMF**	**FCM**	**GMM**	**SC**
Dice [%]	**Tumor**	66±13	61±22	**70±14**	61±21	66±19	69±14
	Necrosis	56±28	54±31	62±25	41±33	58±32	**66±33**
	Edema	33±24	37±27	**46±22**	38±23	36±26	43±24
	Core	76±11	73±15	**79±12**	71±19	77±18	76±16
	Whole	78±13	82±10	**86±8**	78±17	85±10	85±10
Undetected [#]	**Tumor**	0/21	1/21	0/21	0/21	0/21	0/21
	Necrosis	1/12	2/12	0/12	3/12	2/12	2/12
	Edema	4/15	4/15	2/15	3/15	3/15	3/15

Table 2: Segmentation results for the UZ Gent MP-MRI dataset when using kmeans++ initialization. Mean Dice-score ± standard deviation is reported for active tumor, necrosis, edema, the tumor core (active tumor+necrosis) and the whole tumor (core+edema). The number of undetected cases is reported for active tumor, necrosis and edema.

		NMF			**Clustering**		
		HALS	**Convex**	**hNMF**	**FCM**	**GMM**	**SC**
Dice [%]	**Tumor**	65±13	64±18	**71±15**	60±21	66±21	68±14
	Necrosis	55±27	60±28	60±28	41±34	48±33	**65±29**
	Edema	28±23	18±18	**46±23**	38±23	27±26	43±24
	Core	76±11	74±14	**79±12**	70±18	77±18	76±15
	Whole	78±12	84±10	**86±8**	77±15	84±10	85±11
Undetected [#]	**Tumor**	0/21	0/21	0/21	0/21	0/21	0/21
	Necrosis	1/12	1/12	1/12	3/12	3/12	1/12
	Edema	5/15	7/15	2/15	3/15	6/15	3/15

Table 3: Segmentation results for the UZ Gent MP-MRI dataset when using SPA initialization. Mean Dice-score ± standard deviation is reported for active tumor,

necrosis, edema, the tumor core and the whole tumor. The number of undetected cases is reported for active tumor, necrosis and edema.

Figure 2 gives an example of some of the MP-MRI input maps and the segmentation results for a grade III oligo-astrocytoma patient from UZ Gent. No necrotic region was present but the main challenge was the differentiation of non-enhancing tumor from edema. Comparing with the manual segmentation, all classification methods overestimate the active tumor region. hNMF gives the best segmentation for active tumor, slightly better than HALS NMF and Convex NMF. All of the methods obtain a good estimation of the whole tumor region.

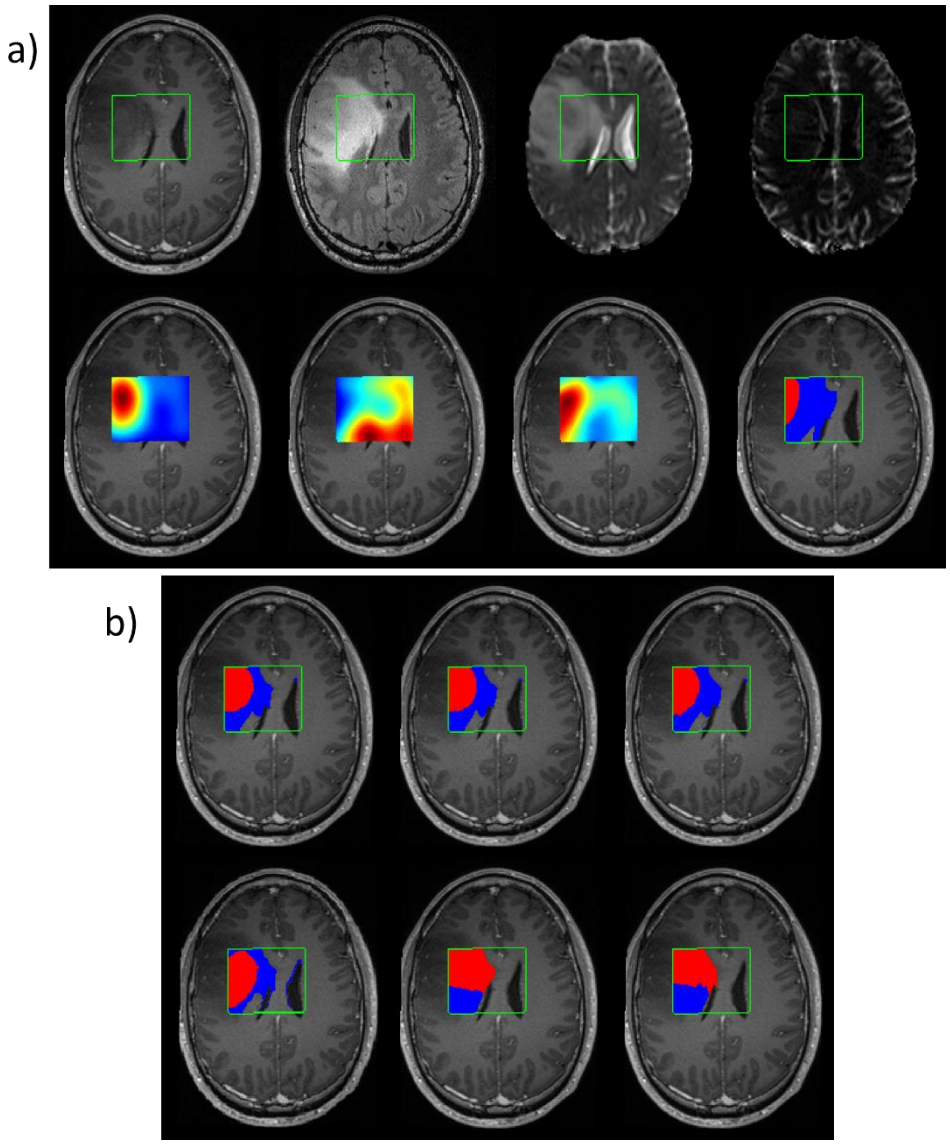


Figure 2: a) Some of the MP-MRI images of a grade III oligo-astrocytoma patient from the UZ Gent dataset. First row, left to right: T1c, FLAIR, ADC, rCBV. Second row, left to right: Cho, NAA, Lac, and manual segmentation of active tumor (red) and edema (blue). The ROI is delineated in green. b) Segmentation results (using kmeans++ initialization), top row left to right: HALS NMF, Convex NMF, hNMF. Second row, left to right: FCM, GMM, SC.

Table 4 compares segmentation performance on the UZ Gent dataset when considering only cMRI data with SPA initialization. hNMF has the highest mean Dice-score for all tissue classes except for

necrosis, for which SC performs better. Compared to the MP-MRI results with SPA initialization in Table 3, Dice-scores are lower or at best equal when using cMRI data only. The reduction in mean Dice-score is between 0 and 4% for active tumor, between 2 and 11% for necrosis, between 0 and 12% for edema, between 1 and 8% for the tumor core and between 0 and 6% for the whole tumor region. The number of cases with undetected tissue classes ranged from 1 to 3 out of 12 patients for necrosis and from 5 to 10 out of 15 patients for edema. None of the methods showed a case of undetected active tumor.

		NMF			**Clustering**		
		HALS	**Convex**	**hNMF**	**FCM**	**GMM**	**SC**
Dice [%]	**Tumor**	64±16	60±21	**68±18**	60±20	64±21	66±20
	Necrosis	51±29	49±31	55±29	39±31	46±30	**58±27**
	Edema	28±25	18±27	**43±25**	31±28	15±28	40±26
	Core	72±17	69±21	**74±16**	69±19	69±20	69±17
	Whole	78±14	78±17	**81±12**	77±17	79±17	80±14
Undetected [#]	**Tumor**	0/21	0/21	0/21	0/21	0/21	0/21
	Necrosis	2/12	2/12	2/12	3/12	2/12	1/12
	Edema	7/15	9/15	5/15	7/15	10/15	5/15

Table 4: Segmentation results for the UZ Gent cMRI data when using SPA initialization. Mean Dice-score ± standard deviation is reported for active tumor, necrosis, edema, the tumor core and the whole tumor. The number of undetected cases is reported for active tumor, necrosis and edema.

4.2 UZ Leuven

The segmentation results for the UZ Leuven dataset are shown in Table 5 for kmeans++ initialization and in Table 6 for SPA initialization. With both initialization methods, hNMF has the highest mean Dice-score for all tissue classes except for the tumor core, for which Convex NMF has a slightly higher mean Dice using kmeans++ initialization. HALS NMF matches hNMF with equal mean Dice-score for active tumor using kmeans++, HALS NMF and SC match hNMF performance for the tumor core using SPA and SC matches hNMF for necrosis with both initialization methods. Globally, GMM has the worst performance on this dataset, with the lowest mean Dice in 6 out of 10 classes from both tables. Neither kmeans++ nor SPA clearly showed better performance for any of the unsupervised methods or for any of the tissue classes. The number of undetected cases ranged from 0 to 2 out of 14 patients for active tumor, from 0 to 3 out of 9 patients for necrosis and from 0 to 2 out of 9 patients for edema. hNMF matches the lowest number of undetected cases for all tissue classes, except for necrosis with SPA initialization, where Convex NMF has no undetected cases and hNMF has one.

		NMF			**Clustering**		
		HALS	**Convex**	**hNMF**	**FCM**	**GMM**	**SC**
Dice [%]	**Tumor**	**73±16**	67±26	**73±15**	66±26	57±26	72±17
	Necrosis	56±28	49±29	**57±29**	45±31	35±28	**57±35**
	Edema	51±23	60±16	**62±21**	60±15	50±29	56±20
	Core	84±9	**85±9**	84±9	83±12	82±10	84±9
	Whole	80±12	83±11	**85±11**	82±10	83±10	83±11
Undetected [#]	**Tumor**	0/14	1/14	0/14	1/14	2/14	0/14
	Necrosis	1/9	2/9	1/9	2/9	3/9	2/9
	Edema	1/9	0/9	0/9	0/9	2/9	0/9

Table 5: Segmentation results for the UZ Leuven MP-MRI dataset when using kmeans++ initialization. Mean Dice-score ± standard deviation is reported for active tumor, necrosis, edema, the tumor core and the whole tumor. The number of undetected cases is reported for active tumor, necrosis and edema.

		NMF			Clustering		
		HALS	**Convex**	**hNMF**	**FCM**	**GMM**	**SC**
Dice [%]	**Tumor**	68±24	72±15	**73±15**	66±26	63±16	72±17
	Necrosis	55±28	55±17	**57±29**	45±31	33±27	**57±35**
	Edema	50±23	40±28	**62±21**	60±15	52±22	56±20
	Core	**84±9**	80±11	**84±9**	82±12	81±9	**84±10**
	Whole	79±13	82±11	**85±11**	82±10	82±9	84±11
Undetected [#]	**Tumor**	1/14	0/14	0/14	1/14	0/14	0/14
	Necrosis	1/9	0/9	1/9	2/9	3/9	2/9
	Edema	1/9	2/9	0/9	0/9	1/9	0/9

Table 6: Segmentation results for the UZ Leuven MP-MRI dataset when using SPA initialization. Mean Dice-score ± standard deviation is reported for active tumor, necrosis, edema, the tumor core and the whole tumor. The number of undetected cases is reported for active tumor, necrosis and edema.

4.3 Computational cost

Table 7 reports the average computation times per patient for the UZ Gent dataset using kmeans++ initialization. These were the computationally most expensive analyses, as the UZ Gent data were 3D (UZ Leuven dataset has 2D ROI) and the analyses are repeated 20 times with kmeans++ to come to the final result. Segmentations were computed in Matlab R2015a (The MathWorks, Natick, MA, USA) with an Intel Core i7-3720QM processor and 8GB of RAM. The input data matrix typically consisted of approximately 30,000 voxels by 24 MRI features. HALS NMF and in particular FCM clustering, both known for their computational efficiency, have relatively low computational costs. SC has a computational cost similar to HALS NMF as we have defined a sparsely connected k-nearest neighbours graph, allowing efficient calculations. Computation times for hNMF and GMM are 2 to 4 times higher compared to HALS NMF and SC. For hNMF, this is partly explained by its hierarchical structure, requiring 3 HALS NMF analyses at each run, but the main expense comes from the NNLS recombining step to calculate the abundance maps at the final stage of the algorithm. Convex NMF has the highest computation time, approximately twice as high as for hNMF, due to its low convergence rate.

	Average computation time [sec]
HALS NMF	80
Convex NMF	429
hNMF	208
FCM	31
GMM	266
SC	74

Table 7: Average computation time (in seconds) per patient on the UZ Gent dataset using kmeans++ initialization.

4.4 Data distribution

To gain some insight into the spatial distribution of the data points in feature space, the shape of the tissue clusters and their overlap, we performed a principal component analysis on the active tumor and necrosis data points of a single GBM patient from each hospital. The projections of the data points onto the first and second principal component are plotted against each other in Figure3, providing insight into the data spread along the 2 main directions of data variance. We can see that the first principal component separates tumor and necrosis, but there is a distinct difference in the data spread of both patients. For the UZ Leuven patient, there is a structured pattern in the data and distinct curves of data points can be observed. Such patterns are not seen in the data of the UZ Gent patient, for which data points appear more randomly distributed. The structure in the UZ Leuven data can be explained by the limited number of voxels within the 2D ROI in combination with the large MRSI voxel size. Often only a few MRSI voxels fully cover the tumor and the necrotic tissue class. The MRSI data have been upsampled to the spatial resolution of $1\times1\times3mm^3$, generating a high number of interpolated and noiseless data points. The structure that can be seen in the data are interpolation patterns within the highly upsampled MRSI data. This explains why SC performs very well and almost as good as hNMF on the UZ Leuven data: it can easily model this kind of structured data as the SC similarity matrix is based on a distance metric among the nearest neighbours of each data point. It also explains why GMM underperforms on the UZ Leuven dataset, as its assumptions of a Gaussian within-class data distribution and ellipsoidal cluster shapes does not hold. For the UZ Gent data, the structured upsampling effect is not relevant as we are analyzing a sufficiently large (3D) ROI, with about 10 times more data points compared to the UZ Leuven patients.

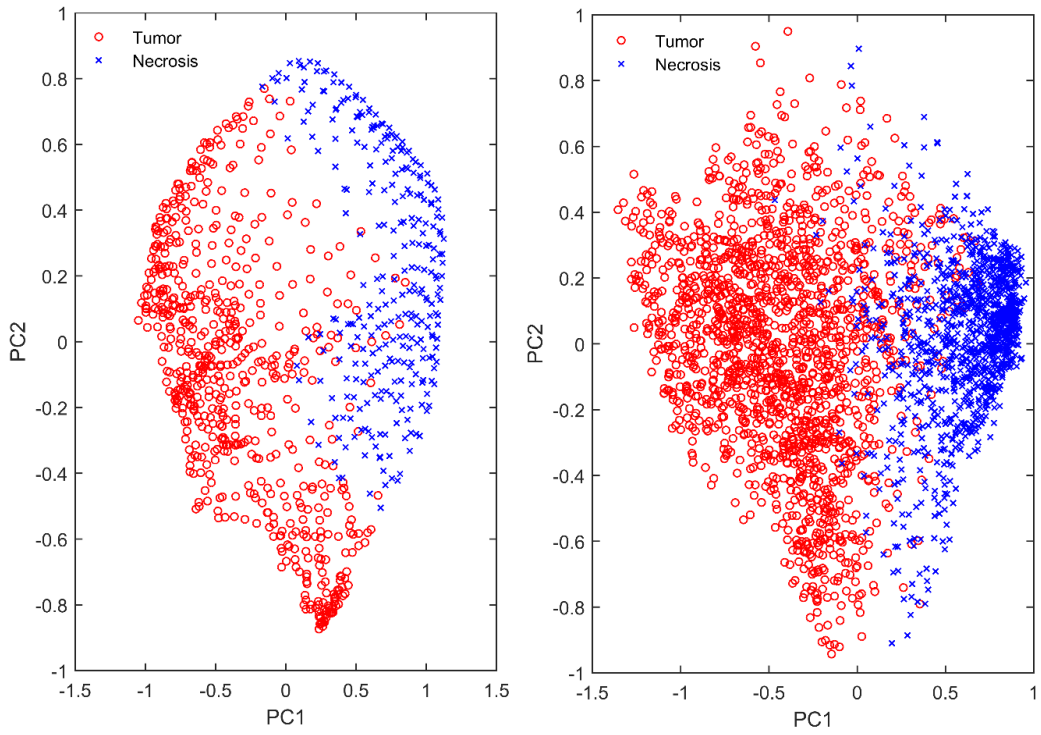


Figure 3: Active tumor and necrosis data points projected onto the plane formed by the first and second principal component for an UZ Leuven GBM patient (left) and for an UZ Gent GBM patient (right).

5 Discussion

5.1 NMF vs clustering

So far only few studies have considered MP-MRI for brain tumor segmentation, although it has been commonly suggested in literature [14, 17, 62]. We have shown in a previous study that significantly higher segmentation accuracy could be achieved by applying hNMF to MP-MRI data compared to using only cMRI data [38]. For the whole tumor region, segmentation algorithms already achieve Dice-scores in the range of inter-rater variability, which has been reported to be between 80 and 85% [6]. Such Dice-scores were also achieved by most algorithms in the current study. Lower Dice-scores are commonly reported for the individual pathologic tissue types. There are several reasons why it is particularly difficult to differentiate the tumor subregions. First of all, HGGs show a high degree of within-tissue heterogeneity. Clustering methods explain the data variability in this way, by incorporating within-class variability into their model. On the other hand, the pathologic region contains a continuum of tissue mixtures rather than absolute tissue boundaries: e.g. active tumor in HGGs contains foci of microscopic necrosis [63] and edema will often be invaded by active tumor cells [3]. These tissue mixtures are further enhanced at the voxel level due to the partial volume effect. Clinical MRI is often acquired in a highly anisotropic way, with relatively low inter-slice resolution. Furthermore, advanced MRI modalities usually have lower spatial resolution than cMRI, causing even more partial volume effects. NMF has this concept of class mixtures built into its model. As such, clustering methods and NMF methods each model only one aspect of the data variability: clustering does not incorporate tissue mixtures and NMF does not consider within-tissue variability. In the current study we did not see either type of methods outperforming the other, suggesting that both aspects of data variability play a considerable role in glioma segmentation.

5.2 hNMF performance

Segmentation performance was compared among the different methods by reporting mean Dice-scores for 5 tissue classes, on 2 independent MP-MRI datasets, using 2 different initialization methods. The hNMF method globally performed best, with the highest mean Dice-score in 17 out of the 20 comparisons on the MP-MRI datasets. When considering only cMRI data of the UZ Gent dataset, hNMF also had the highest mean Dice-score for 4 out of 5 tissue classes. Additionally, hNMF provided robust segmentation, with relatively low standard deviations and the lowest number of undetected cases for nearly all tissue classes. The hierarchical structure of hNMF gives it some advantages over the other methods. At the first (rank-2) NMF level, the tissue classes are divided over 2 sources, with tissues that are most similar in terms of MP-MRI feature vectors being assigned to the same source. At the second level, NMF can focus on the more subtle differences between the similar tissue classes. This property has allowed hNMF to better separate edema from non-enhancing tumor in the UZ Gent dataset, resulting in the highest mean Dice-scores and the lowest number of undetected sources for edema in tables 2 and 3. Another challenging problem for most unsupervised methods is how to cope with data imbalance. Unsupervised methods have to model a large amount of data with a limited number of degrees of freedom, based on a global optimization criterion. This does not favour the detection of small lesions. In the current study this was mainly an issue in detecting small necrotic regions. As the input data are divided into 2 groups after the first NMF level, hNMF might be able to detect small tissue regions more easily at the second NMF level. This has allowed hNMF to detect some small necrotic areas that other methods were not able to differentiate. An important drawback of the original hNMF algorithm as proposed in [38] was its computational expense. This was mainly attributed to the iterative mask selection step after the first NMF level, using hard thresholding to subdivide the voxels over both sources. We have adapted the hNMF algorithm, using k-means clustering after the first NMF level for the voxel assignment. This has

brought the computation time of hNMF within the range of the other methods evaluated in the current study, as shown in Table 7. It must be noted that we did not directly explore the effect of inter-observer variability on the segmentation results. However, all methods were applied to 2 independent datasets with manual segmentation by 2 different raters. Similar results were obtained for both datasets, suggesting that the drawn conclusions are general.

5.3 Spatial regularization

Many classification based segmentation algorithms incorporate some kind of spatial regularization to obtain spatially consistent results. It is to be expected that most neighbouring voxels will belong to the same tissue class, except at the tissue boundaries. Common techniques to encourage spatially consistent segmentation include Markov random fields [12] and spatial smoothness regularization on the NMF abundance matrix *H* [64]. We did not apply spatial regularization to any of the unsupervised methods, as it would bias the comparison to the other methods. In general, we did obtain spatially consistent segmentations, as is illustrated in Figure 2. This is partly explained by the fact that we combined many MP-MRI features, thereby increasing the specificity of the feature set. Secondly, the inclusion of locally averaged MRI features further improved robustness of the segmentation in terms of spatial coherence. Zikic et al. also reported naturally smooth segmentation results by combining cMRI with DWI and including spatial and context-aware features [37]. Nevertheless, the hNMF algorithm might benefit from adding spatial regularization, especially when reducing the number of MRI features or extending the region of interest to full brain slices.

5.4 UZ Gent and UZ Leuven datasets

For the UZ Gent dataset edema is clearly showing the lowest Dice-scores. 8 out of 21 patients in the UZ Gent dataset (6 grade III and 2 GBM) exhibited non-enhancing tumor in combination with vasogenic edema in the ROI. Distinction of both tissue types is often difficult on conventional MRI due to the diffuse infiltrative nature of many gliomas, lacking a well-defined margin. The perilesional environment often contains not only T2/FLAIR hyperintense vasogenic edema but also tumor components which appear less hyperintense on T2/FLAIR weighted images due to hypercellularity. The differences in signal intensity on T2/FLAIR can be subtle making the segmentation in these cases susceptible to subjective interpretation. Even when using MP-MRI, the tissue signatures of both tissue types are similar and proper differentiation remains challenging (see Figure 2). The UZ Leuven dataset only contained 3 (out of 14) grade III patients, and none of the patients showed non-enhancing tumor with edema in the ROI. Due to this clear bias across the patient cohorts, higher Dice-scores were found for edema in the UZ Leuven dataset. Looking at the distribution of the data points of the pathologic tissue classes as shown in Figure 3, we did find a marked difference between both datasets. A structured pattern is seen in the UZ Leuven data, which is explained by the limited extent of the 2D ROI in combination with the low spatial resolution of MRSI compared to the other MRI modalities. SC, which is known to cope well with complex cluster shapes, performs well on the UZ Leuven data, almost as good as hNMF. GMM, due to its strong assumption of Gaussian within-class data distribution, is not able to model the structured data well and has the lowest Dice-scores. However, for the 3D ROI of the UZ Gent dataset, these structured patterns are no longer appearing. Therefore, SC and GMM perform comparably to most of the other methods on the UZ Gent dataset. The added value of including PWI, DWI and MRSI was assessed on the UZ Gent dataset by considering only the cMRI data with SPA initialization (see Table 4). Mean Dice-scores were lower or at best equal for all tissue classes and for all the methods when considering only cMRI data. Sensitivity to this reduction of the MP-MRI dataset was found to be similar across the different methods. The decrease in performance was in general more pronounced for necrosis and edema than for active tumor. On the UZ Leuven dataset, segmentation results for hNMF when considering

only cMRI data have been previously reported in [38]. Significantly lower Dice scores were found for active tumor, the tumor core and the whole tumor region compared to the full MP-MRI dataset.

5.5 Initialization methods

We have performed tissue segmentation using 2 initialization methods: kmeans++ which is commonly used for clustering and SPA which is commonly used for NMF. No bias was seen in the results with both types of initialization: kmeans++ did not improve segmentation results for the clustering methods nor did SPA show better results for the NMF methods. SPA provides a deterministic initialization, such that the obtained segmentation results are reproducible. Kmeans++ on the other hand provides a different initialization at each run. Analyses using kmeans++ were repeated 20 times, withholding the result with the best (= highest or lowest) objective function value for the considered classification method. 20 repetitions was found to be sufficient to give reproducible results, i.e. to obtain Dice-scores differing by less than 1% for each patient upon repeating the analyses. In general, similar performance was found using both types of initialization. The main difference was seen in the segmentation of edema in the UZ Gent dataset: mean Dice-scores for edema were higher with kmeans++ than with SPA for 4 out of 6 classification methods, and were equal for the other 2 methods. The number of undetected cases for edema was also at least as high with SPA as with kmeans++. SPA minimizes collinearity between the initialized sources by successively selecting data points with maximal l2-norm in the orthogonal subspace of the previously selected sources. As such, SPA might have difficulty in correctly initializing tissue classes with similar sources, such as non-enhancing tumor and edema. Indeed, we noticed that SPA often seeded only one source for the combined region of non-enhancing tumor and edema, whereas kmeans++ often obtained at least 2 sources. Kmeans++ is more likely to select data points which are further apart in Euclidian space, but it does not explicitly look for the extremal points in a dataset. As we run kmeans++ a sufficient number of times, it is likely that both non-enhancing tumor and edema get a seeding point in at least one of those runs. At the individual patient level, SC, HALS NMF and hNMF gave very similar results with both initialization methods, which is reflected in the similar mean Dice-scores for most tissue classes. HALS NMF has previously been shown to be insensitive to the initialization strategy, and the same goes for hNMF as it consists of 2 levels of HALS NMF [43]. SC does not require an initialization, but it applies k-means clustering to the matrix of eigenvectors in the final stage of the algorithm. This k-means step however does not seem sensitive to the initialization method, probably because the complexity of the classification problem is already reduced by withholding only the *k* main eigenvectors to find *k* tissue classes.

5.6 Supervised vs unsupervised classification

Recently, supervised classification methods have been receiving most attention for automated brain tumor segmentation [6]. However, unsupervised methods remain competitive, e.g. Juan-Albarracin et al. have recently combined some dedicated cMRI pre-processing steps with several unsupervised methods, including FCM, GMM and hidden Markov random fields. Their results ranked among the best supervised algorithms in the BRATS challenge [14]. A direct comparison between our results and those from the BRATS challenge is not in place, as we are combining cMRI data with PWI, DWI and MRSI in a multi-parametric approach, whereas BRATS only considers cMRI data. Furthermore, we restricted our study to the limited region of interest of the MRSI data, whereas full 3D imaging data of the brain are considered in the BRATS challenge. Nevertheless, several of the unsupervised algorithms reported in this study, and in particular hNMF achieved mean Dice-scores which are higher than the best performing algorithms participating in the BRATS challenge [6], showing the potential of these methods. Unsupervised methods are more flexible as they don't require an extensive set of training data, which is only valid for segmentation under a particular data acquisition

and processing protocol. This limitation might be more restrictive when considering MP-MRI data, as there is a wide variability and fast evolution in acquisition protocols and post-processing methods for PWI, MRSI and DWI [65]. Furthermore, unsupervised methods are not susceptible to overfitting, and they don't require intensity calibration, which can be challenging in the presence of pathologic tissue [17]. On the other hand, an attractive feature of supervised methods is that they don't require a rank estimation step for determining the number of tissue types, and they automatically apply voxel labelling based on the training data. The unsupervised methods presented in this study are not fully automatic: they require the user to specify the number of tissue types to be found within the region of interest. Therefore, further research will focus on automation by incorporating prior knowledge, e.g. tissue probability maps based on an image atlas [14], feature-specific knowledge [8] or user input [6].

6 Conclusion

This study explored the usage of unsupervised classification methods applied to MP-MRI data for segmenting the tumor subregions in HGGs. We have proposed a hierarchical 2-level hNMF method and compared its performance to 5 common unsupervised classification algorithms. hNMF achieved the best segmentation results in terms of mean Dice-scores on 2 independent MP-MRI datasets acquired at different hospitals. This trend was confirmed using both kmeans++ and SPA initialization. Unsupervised methods can be applied to any MP-MRI dataset at the individual patient level, irrespective of the acquisition protocol or image processing methods used. Future work will focus on further automation of unsupervised segmentation methods by incorporating prior knowledge.

7 Acknowledgements

This work has been funded by the following projects: Flemish Government FWO (G.0869.12N); IWT IM (n°135005); Henri Benedictus Fellowship of the Belgian American Educational Foundation; Interuniversity Attraction Poles Program (P7/11) initiated by the Belgian Science Policy Office. European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013): EU MC ITN TRANSACT 2012 (n°316679) and ERC Advanced Grant: BIOTENSORS (n°339804). This paper reflects only the authors' views and the Union is not liable for any use that may be made of the contained information.

References

[1] M. L. Goodenberger and R. B. Jenkins, “Genetics of adult glioma,” Cancer Genetics, vol. 205, no. 12, pp. 613-621, 2012.

[2] Q. T. Ostrom, H. Gittleman, P. Liao, C. Rouse, Y. Chen, J. Dowling, Y. Wolinsky, C. Kruchko, and J. Barnholtz-Sloan, “CBTRUS statistical report: primary brain and central nervous system tumors diagnosed in the United States in 2007-2011,” Neuro-oncology, vol. 16, no. suppl 4, pp. iv1-iv63, 2014.

[3] S. Price, R. Jena, N. Burnet, P. Hutchinson, A. Dean, A. Pena, J. Pickard, T. Carpenter, and J. Gillard, “Improved delineation of glioma margins and regions of infiltration with the use of diffusion tensor imaging: an image-guided biopsy study,” American Journal of Neuroradiology, vol. 27, no. 9, pp. 1969-1974, 2006.

[4] W. Paulus and J. Peiffer, “Intratumoral histologic heterogeneity of gliomas. A quantitative study,” Cancer, vol. 64, no. 2, pp. 442-447, 1989.

[5] C. Weltens, J. Menten, M. Feron, E. Bellon, P. Demaerel, F. Maes, W. Van den Bogaert, and E. van der Schueren, “Interobserver variations in gross tumor volume delineation of brain tumors on computed tomography and impact of magnetic resonance imaging,” Radiotherapy and Oncology, vol. 60, no. 1, pp. 49-59, 2001.

[6] B. Menze, M. Reyes, K. Van Leemput et al., “The multimodal brain tumor image segmentation benchmark (BRATS),” IEEE Transactions on Medical Imaging, vol. 34, no. 10, pp. 1993-2024, 2015.

[7] J. Nie, Z. Xue, T. Liu, G. S. Young, K. Setayesh, L. Guo, and S. T. Wong, “Automated brain tumor segmentation using spatial accuracy-weighted hidden Markov random field,” Computerized Medical Imaging and Graphics, vol. 33, no. 6, pp. 431-441, 2009.

[8] A. F. Kazerooni, M. Mohseni, S. Rezaei, G. Bakhshandehpour, and H. S. Rad, “Multiparametric (ADC/PWI/T2-w) image fusion approach for accurate semi-automatic segmentation of tumorous regions in glioblastoma multiforme,” Magnetic Resonance Materials in Physics, Biology and Medicine, vol. 28, no. 1, pp. 13-22, 2015.

[9] N. Gordillo, E. Montseny, and P. Sobrevilla, “State of the art survey on MRI brain tumor segmentation,” Magnetic resonance imaging, vol. 31, no. 8, pp. 1426-1438, 2013.

[10] W. Phillips, R. Velthuizen, S. Phuphanich, L. Hall, L. Clarke, and M. Silbiger, “Application of fuzzy c-means segmentation technique for tissue differentiation in MR images of a hemorrhagic glioblastoma multiforme,” Magnetic resonance imaging, vol. 13, no. 2, pp.277-290, 1995.

[11] L. M. Fletcher-Heath, L. O. Hall, D. B. Goldgof, and F. R. Murtagh, “Automatic segmentation of non-enhancing brain tumors in magnetic resonance images,” Artificial intelligence in medicine, vol. 21, no. 1, pp. 43-63, 2001.

[12] B. H. Menze, K. Van Leemput, D. Lashkari, M.-A. Weber, N. Ayache, and P. Golland, “A generative model for brain tumor segmentation in multi-modal images,” in Medical Image Computing and Computer-Assisted Intervention-MICCAI 2010. Springer, 2010, vol. 13, no. Pt 2, pp. 151-159.

[13] Y. Zhang, M. Brady, and S. Smith, “Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm,” IEEE Transactions on Medical Imaging, vol. 20, no. 1, pp. 45-57, 2001.

[14] J. Juan-Albarracín, E. Fuster-Garcia, J. V. Manjón, M. Robles, F. Aparici, L. Martí-Bonmatí, and J. M. García-Gomez, “Automated glioblastoma segmentation based on a multiparametric structured unsupervised classification,” PloS One, vol. 10, no. 5, p.e0125143, 2015.

[15] L. Görlitz, B. H. Menze, M.-A. Weber, B. M. Kelm, and F. A. Hamprecht, “Semisupervised tumor detection in magnetic resonance spectroscopic images using discriminative random fields,” in Pattern Recognition. Springer, 2007, pp. 224-233.

[16] V. B. Padole and D. Chaudhari, “Detection of brain tumor in MRI images using mean shift algorithm and normalized cut method,” International Journal of Engineering and Advanced Technology, 2012.

[17] S. Bauer, R. Wiest, L.-P. Nolte, and M. Reyes, “A survey of MRI-based medical image analysis for brain tumor studies,” Physics in Medicine and Biology, vol. 58, no. 13, p.R97, 2013.

[18] S. Cha, “Update on brain tumor imaging: from anatomy to physiology,” American Journal of Neuroradiology, vol. 27, no. 3, pp. 475-487, 2006.

[19] A. R. Padhani and K. A. Miles, “Multiparametric imaging of tumor response to therapy 1,” Radiology, vol. 256, no. 2, pp. 348-364, 2010.

[20] T. Sugahara, Y. Korogi, M. Kochi, I. Ikushima, Y. Shigematu, T. Hirai, T. Okuda, L. Liang, Y. Ge, Y. Komohara et al., “Usefulness of diffusion-weighted MRI with echoplanar technique in the evaluation of cellularity in gliomas,” Journal of Magnetic Resonance Imaging, vol. 9, no. 1, pp. 53-60, 1999.

[21] K. G. Abdullah, D. Lubelski, P. G. Nucifora, and S. Brem, “Use of diffusion tensor imaging in glioma resection,” Neurosurgical focus, vol. 34, no. 4, p. E1, 2013.

[22] T. L. Jones, T. J. Byrnes, G. Yang, F. A. Howe, B. A. Bell, and T. R. Barrick, “Brain tumor classification using the diffusion tensor image segmentation (D-SEG) technique,” Neuro-Oncology, vol. 17, no. 3, pp. 466-476, 2015.

[23] S. Van Cauter, J. Veraart, J. Sijbers, R. R. Peeters, U. Himmelreich, F. De Keyzer, S. W. Van Gool, F. Van Calenbergh, S. De Vleeschouwer, W. Van Hecke, and S. Sunaert, “Gliomas: diffusion kurtosis MR imaging in grading,” Radiology, vol. 263, no. 2, pp.492-501, 2012.

[24] A. J. Steven, J. Zhuo, and E. R. Melhem, “Diffusion kurtosis imaging: an emerging technique for evaluating the microstructural environment of the brain,” American Journal of Roentgenology, vol. 202, no. 1, pp. W26-W33, 2014.

[25] S. Cha, E. A. Knopp, G. Johnson, S. G. Wetzel, A. W. Litt, and D. Zagzag, “Intracranial mass lesions: Dynamic contrast-enhanced susceptibility-weighted echo-planar perfusion MR imaging 1,” Radiology, vol. 223, no. 1, pp. 11-29, 2002.

[26] D. Mukhopadhyay, L. Tsiokas, X.-M. Zhou, D. Foster, J. S. Brugge, and V. P. Sukhatme, “Hypoxic induction of human vascular endothelial growth factor expression through c-Src activation,” Nature, vol. 375, no. 6532, pp. 577-581, 1995.

[27] N. Sadeghi, N. D'Haene, C. Decaestecker, M. Levivier, T. Metens, C. Maris, D. Wikler, D. Balériaux, I. Salmon, and S. Goldman, “Apparent diffusion coefficient and cerebral blood volume in brain gliomas: relation to tumor cell density and tumor microvessel density based on stereotactic biopsies,” American Journal of Neuroradiology, vol. 29, no. 3, pp. 476-482, 2008.

[28] A. Ion-Margineanu, S. Van Cauter, D. M. Sima, F. Maes, S. W. Van Gool, S. Sunaert, U. Himmelreich, and S. Van Huffel, “Tumour relapse prediction using multiparametric MR data recorded during follow-up of GBM patients,” BioMed Research International, vol. 2015, 2015.

[29] L. S. Hu, J. M. Eschbacher, J. E. Heiserman, A. C. Dueck, W. R. Shapiro, S. Liu, J. P. Karis, K. A. Smith, S. W. Coons, P. Nakaji et al., “Reevaluating the imaging definition of tumor progression: perfusion MRI quantifies recurrent glioblastoma tumor fraction, pseudoprogression, and radiation necrosis to predict survival,” Neuro-Oncology, vol. 14, no. 7, pp. 919-930, 2012.

[30] M. Bulik, R. Jancalek, J. Vanicek, A. Skoch, and M. Mechl, “Potential of MR spectroscopy for assessment of glioma grading,” Clinical Neurology and Neurosurgery, vol.115, no. 2, pp. 146-153, 2013.

[31] A. C. Kuesel, G. R. Sutherland, W. Halliday, and I. C. Smith, “1H MRS of high grade astrocytomas: mobile lipid accumulation in necrotic tissue,” NMR in Biomedicine, vol. 7, no. 3, pp. 149-155, 1994.

[32] A. Pirzkall, T. R. McKnight, E. E. Graves, M. P. Carol, P. K. Sneed, W. W. Wara, S. J. Nelson, L. J. Verhey, and D. A. Larson, “MR-spectroscopy guided target delineation for high-grade gliomas,” International Journal of Radiation Oncology* Biology* Physics, vol. 50, no. 4, pp. 915-928, 2001.

[33] A. Di Costanzo, T. Scarabino, F. Trojsi, G. M. Giannatempo, T. Popolizio, D. Catapano, S. Bonavita, N. Maggialetti, M. Tosetti, U. Salvolini, V. d'Angelo, and G. Tedeschi, “Multiparametric 3T MR approach to the assessment of cerebral gliomas: tumor extent and malignancy,” Neuroradiology, vol. 48, no. 9, pp. 622-631, 2006.

[34] R. Verma, E. I. Zacharaki, Y. Ou, H. Cai, S. Chawla, S.-K. Lee, E. R. Melhem, R. Wolf, and C. Davatzikos, “Multiparametric tissue characterization of brain neoplasms and their recurrence using pattern classification of MR images,” Academic Radiology, vol. 15, no. 8, pp. 966-977, 2008.

[35] D. Zikic, B. Glocker, E. Konukoglu, A. Criminisi, C. Demiralp, J. Shotton, O. Thomas, T. Das, R. Jena, and S. Price, “Decision forests for tissue-specific segmentation of high-grade gliomas in multi-channel MR,” in Medical Image Computing and Computer-Assisted Intervention-MICCAI 2012. Springer, 2012, vol. LCNS 7512, pp. 369-376.

[36] Y. Li, D. M. Sima, S. V. Cauter, A. R. Croitor Sava, U. Himmelreich, Y. Pi, and S. Van Huffel, “Hierarchical non-negative matrix factorization (hNMF): a tissue pattern differentiation method for glioblastoma multiforme diagnosis using MRSI,” NMR in Biomedicine, vol. 26, no. 3, pp. 307-319, 2013.

[37] S. Ortega-Martorell, P. J. Lisboa, A. Vellido, M. Juliá-Sapé, and C. Arús, “Non-negative matrix factorisation methods for the spectral decomposition of MRS data from human brain tumours,” BMC Bioinformatics, vol. 13, no. 1, p. 38, 2012.

[38] N. Sauwen, D. M. Sima, S. Van Cauter, J. Veraart, A. Leemans, F. Maes, U. Himmelreich, and S. Van Huffel, “Hierarchical non-negative matrix factorization to characterize brain tumor heterogeneity using multi-parametric MRI,” NMR in Biomedicine, vol. 28, no. 12, pp. 1599-1624, 2015.

[39] D. N. Louis, H. Ohgaki, O. D. Wiestler, W. K. Cavenee, P. C. Burger, A. Jouvet, B. W. Scheithauer, and P. Kleihues, “The 2007 WHO classification of tumours of the central nervous system,” Acta Neuropathologica, vol. 114, no. 2, pp. 97-109, 2007.

[40] J. Veraart, J. Sijbers, S. Sunaert, A. Leemans, and B. Jeurissen, “Weighted linear least squares estimation of diffusion MRI parameters: strengths, limitations, and pitfalls,” NeuroImage, vol. 81, pp.335-346, 2013.

[41] J. Boxerman, K. Schmainda, and R. Weisskoff, “Relative cerebral blood volume maps corrected for contrast agent extravasation significantly correlate with glioma tumor grade, whereas uncorrected maps do not,” American Journal of Neuroradiology, vol. 27, no. 4, pp. 859-867, 2006.

[42] A. R. Croitor Sava, D. M. Sima, J.-B. Poullet, A. J. Wright, A. Heerschap, and S. Van Huffel, “Exploiting spatial information to estimate metabolite levels in two-dimensional MRSI of heterogeneous brain lesions,” NMR in Biomedicine, vol. 24, no. 7, pp. 824-835, 2011.

[43] N. Sauwen, M. Acou, H. Bharath, D. Sima, J. Veraart, F. Maes, U. Himmelreich, E. Achten, and S. Van Huffel, “Initializing nonnegative matrix factorization using the successive projection algorithm for multi-parametric medical image segmentation,” Proceedings of the 24th ESANN European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, 2016.

[44] D. Le Bihan, J.-F. Mangin, C. Poupon, C. A. Clark, S. Pappata, N. Molko, and H. Chabriat, “Diffusion tensor imaging: concepts and applications,” Journal of Magnetic Resonance Imaging, vol. 13, no. 4, pp. 534-546, 2001.

[45] D. H. Poot, A. J. den Dekker, E. Achten, M. Verhoye, and J. Sijbers, “Optimal experimental design for diffusion kurtosis imaging,” IEEE Transactions on Medical Imaging, vol. 29, no. 3, pp. 819-829, 2010.

[46] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, “Multimodality image registration by maximization of mutual information,” IEEE Transactions on Medical Imaging, vol. 16, no. 2, pp. 187-198, 1997.

[47] A. Cichocki, R. Zdunek, and S.-i. Amari, “Hierarchical ALS algorithms for nonnegative matrix and 3D tensor factorization,” in Independent Component Analysis and Signal Separation. Springer, 2007, vol. LCNS 4666, pp. 169-176.

[48] N. Gillis and F. Glineur, “Accelerated multiplicative updates and hierarchical ALS algorithms for nonnegative matrix factorization,” Neural Computation, vol. 24, no. 4, pp.1085-1105, 2012.

[49] C. Ding, T. Li, M. Jordan et al., “Convex and semi-nonnegative matrix factorizations,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 1, pp.45-55, 2010.

[50] D. Kuang and H. Park, “Fast rank-2 nonnegative matrix factorization for hierarchical document clustering,” in Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2013, pp. 739-747.

[51] N. Gillis, D. Kuang, and H. Park, “Hierarchical clustering of hyperspectral images using rank-two nonnegative matrix factorization,” IEEE Transactions on Geoscience and Remote Sensing, vol. 53, no. 4, pp. 2066-2078, 2015.

[52] A. P. Dempster, N. M. Laird, and D. B. Rubin, “Maximum likelihood from incomplete data via the EM algorithm,” Journal of the royal statistical society. Series B (methodological), pp. 1-38, 1977.

[53] F. Diehl, M. J. F. Gales, X. Liu, M. Tomalin, P. C. Woodland et al., “Word boundary modelling and full covariance Gaussians for Arabic speech-to-text systems.” in INTERSPEECH, 2011, pp. 777-780.

[54] F. Yang and P. Grigsby, “Spectral clustering for FDG-PET cervical tumor segmentation,” International Journal of Radiation Oncology* Biology* Physics, vol. 78, no. 3, p. S494, 2010.

[55] M. Manoj and L. Padmasuresh, “An automated multimodal spectral cluster based segmentation for tumor and lesion detection in pet images,” Research Journal of Applied Sciences, Engineering and Technology, vol. 12, no. 5, pp. 522-527, 2016.

[56] U. Von Luxburg, “A tutorial on spectral clustering,” Statistics and Computing, vol. 17, no. 4, pp. 395-416, 2007.

[57] M. Brito, E. Chavez, A. Quiroz, and J. Yukich, “Connectivity of the mutual k-nearest neighbor graph in clustering and outlier detection,” Statistics & Probability Letters, vol. 35, no. 1, pp. 33-42, 1997.

[58] J. Shi and J. Malik, “Normalized cuts and image segmentation,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 8, pp. 888-905, 2000.

[59] M. C. U. Araújo, T. C. B. Saldanha, R. K. H. Galvão, T. Yoneyama, H. C. Chame, and V. Visani, “The successive projections algorithm for variable selection in spectroscopic multicomponent analysis,” Chemometrics and Intelligent Laboratory Systems, vol. 57, no. 2, pp. 65-73, 2001.

[60] N. Gillis, “Successive nonnegative projection algorithm for robust nonnegative blind source separation,” SIAM Journal on Imaging Sciences, vol. 7, no. 2, pp. 1420-1450, 2014.

[61] D. Arthur and S. Vassilvitskii, “k-means++: The advantages of careful seeding,” in Proceedings of the eighteenth annual ACM-SIAM Symposium on Discrete Algorithms. Society for Industrial and Applied Mathematics, 2007, pp. 1027-1035.

[62] F. Dhermain, “Radiotherapy of high-grade gliomas: current standards and new concepts, innovations in imaging and radiotherapy, and new therapeutic approaches,” Chin J Cancer, vol. 33, no. 1, pp. 16-24, 2014.

[63] H. Bell, I. Whittle, M. Walker, H. Leaver, and S. Wharton, “The development of necrosis and apoptosis in glioma: experimental findings using spheroid culture systems,” Neuropathology and Applied Neurobiology, vol. 27, no. 4, pp. 291-304, 2001.

[64] Y. Xie, J. Ho, and B. C. Vemuri, “Nonnegative factorization of diffusion tensor images and its applications,” in Information Processing in Medical Imaging, vol. 22. Springer, 2011, pp. 550-561.

[65] G. S. Young, “Advanced MRI of adult brain tumors,” Neurologic Clinics, vol. 25, no. 4, pp. 947-973, 2007.