

Supplementary Materials

for

Fully automatic Lesion Localization and Characterization: Application to Brain Tumors Using Multiparametric MRI Data

by Alexis Arnaud, Florence Forbes, Nicolas Coquery, Nora Collomb, Benjamin Lemasson, Emmanuel L. Barbier

I. INTRODUCTION

When analyzing brain tumors, two tasks are intrinsically linked, spatial localization and physiological characterization of the lesioned tissues. Automated data-driven solutions exist, based on image segmentation techniques or physiological parameters analysis, but for each task separately, the other being performed manually or with user tuning operations. In this work, the availability of quantitative magnetic resonance (MR) parameters is combined with advanced multivariate statistical tools to design a fully automated method that jointly performs both localization and characterization.

The statistical approach used is presented in Section II. After a presentation of mixture models, Section II-A, and the associated question of the cluster number choice, Section II-B, probabilistic mixture models of generalized Student distributions are then considered and described Section II-C. These generalized distributions provide a larger variety of distributional shapes compared to the more standard Gaussian distributions (e.g. non-elliptical shapes), and can capture non trivial interactions between relevant physiological parameters to account for the different tissue types, to distinguish anomaly (i.e. lesion), and to explore potential heterogeneity of lesions. The proposed procedure is then summed up in Section II-D.

The potential of this generic procedure is demonstrated on a data set of 32 rats, with 26 rats bearing 4 different brain tumors, for which the 5 acquired quantitative MR are presented Section III-A. The localization of the abnormal voxels and a comparison with Gaussian mixture model results are shown in Sections III-B and III-C. The characterization of the abnormal voxels is provided in Sections III-D and III-E, also followed by a comparison with Gaussian mixture model results. Eventually, the localization and characterization on an independent data set of 21 rats is presented in Section III-F.

II. CLUSTERING WITH A PARAMETRIC MIXTURE MODEL

A popular approach for data clustering is the use of a parametric finite mixture model: it relies on the assumption that the data come from several components, or clusters, each one controlled by a parametric distribution. From an observed sample, a parametric finite mixture that best gathers similar observations in clusters is estimated.

A. Parametric mixture model

Let $\mathbf{Y} \in \mathcal{R}^M$ be a real random variable in dimension M , $M \in \mathcal{N}^*$. A parametric mixture model is a statistical model based on a convex combination of $K \in \mathcal{N}^*$ parametric probability distributions $\{f_1, \dots, f_K\}$ described by their respective parameters $\theta = \{\theta_1, \dots, \theta_K\}$. The probability density function p of the mixture is defined by:

$$\forall \mathbf{y} \in \mathcal{R}^M, \quad p(\mathbf{y}; \theta, \pi) = \sum_{k=1}^K \pi_k f_k(\mathbf{y}; \theta_k) \quad (1)$$

where $\pi = \{\pi_1, \dots, \pi_K\}$ are the mixture proportions: $\forall k \in \llbracket 1; K \rrbracket$, $0 < \pi_k < 1$ and $\sum_{k=1}^K \pi_k = 1$.

An equivalent and more convenient representation of a mixture model used an hidden random variable Z which links the random variable \mathbf{Y} to one of the distributions $\{f_1, \dots, f_K\}$ with the following hierarchical representation:

$$\begin{aligned} (\mathbf{Y} | Z = k; \theta_k) &\sim f_k \quad \text{for } k \in \llbracket 1; K \rrbracket \\ (Z; \pi) &\sim \mathcal{M}_1(\pi_1, \dots, \pi_K) \end{aligned} \quad (2)$$

where $\mathcal{M}_1(\pi_1, \dots, \pi_K)$ is the multinomial distribution for one trial, with probability mass function: $p_{\mathcal{M}_1}(Z = k; \pi_1, \dots, \pi_K) = \pi_k$, $k \in \llbracket 1; K \rrbracket$.

Given a sample of independent realizations $\{\mathbf{y}_1, \dots, \mathbf{y}_N\}$ from the mixture model, the likelihood of the sample is defined by:

$$p(\mathbf{y}_1, \dots, \mathbf{y}_N; \theta, \pi) = \prod_{n=1}^N \sum_{k=1}^K \pi_k f_k(\mathbf{y}_n; \theta_k) \quad (3)$$

The estimation of a mixture model from a sample is generally done with an Expectation-Maximization algorithm (EM). Under some standard assumptions, the EM algorithm jointly estimates the mixture proportions π and parameters θ by reaching a (local) maximum of the log-likelihood:

$$\left(\hat{\theta}, \hat{\pi}\right) = \arg \max_{(\theta, \pi)} \log p(\mathbf{y}_1, \dots, \mathbf{y}_N; \theta, \pi) \quad (4)$$

B. Choice of the cluster number using the slope heuristic

The estimation of a mixture model requires the preliminary choice of the number of clusters K . A too small K may not allow a good fit of the data while a too large K may lead to overfitting issues. We treat this choice as a model selection problem by minimizing a penalized log-likelihood criterion to penalize the complexity of the model. The main issue in this standard approach is the need to tune the penalization term. We use the so-called slope-heuristic (Baudry, Maugis and Michel [1]) to choose this penalization and thus the cluster number in a data-driven way (using the data-driven slope estimation method). Baudry, Maugis and Michel list the validated frameworks for this heuristic, as well as promising empirical results for a wide range of application fields. For a collection of mixture models identified by their cluster numbers between 1 and K_{\max} , we extract a subset $S_K = \{K, \dots, K_{\max}\}$ with $K \in \llbracket 1, \dots, K_{\max} - 2 \rrbracket$. Each model of S_K is described by its log-likelihood γ_K and its number of free parameters D_K . On this subset, we compute a robust linear regression on the negative log-likelihoods $\{\gamma_K, \dots, \gamma_{K_{\max}}\}$ with $\{D_K, \dots, D_{K_{\max}}\}$ as regressors, and estimate the regression slope C_K . We then determine the model $k \in \llbracket K, \dots, K_{\max} \rrbracket$ which minimizes the associated penalized log-likelihood:

$$2C_K D_k - \gamma_k$$

For each subset S_K we get an optimal model k_{opt, S_K} , and we choose the optimal model K_{opt} for the whole set $1, \dots, K_{\max}$ as the first value such as $\forall K \geq K_{\text{opt}} k_{\text{opt}, S_K} = K$. This corresponds to the first model from which the log-likelihood is in a linear regime with respect to the number of free parameters. It is also the last model before overfitting.

C. Mixture of multiple scale t -distributions

A standard parametric family for mixture models is the family of Gaussian distributions. In our study, the use of Gaussian distributions is not optimal, because the physiological parameters of interest are observed with large spreads in their statistical distributions [2]. The usual parametric family that can account for heavy tails is the Student distribution family. However, the standard multivariate Student distribution does not allow to adjust different tail thickness to the different dimensions of the data. In this study, we use a mixture of multiple scaled t -distributions (MST), which are a generalization of the multivariate t -distributions [3] that can account for the presence of extreme values in some dimensions. This generalization is based on the expression of the usual t -distribution as an infinite mixture of scaled Gaussian distributions.

There exist quite a few forms of the multivariate t -distribution. Among all the possible multivariate presentations, the most common form is the scale mixture of Gaussians. This leads to the density denoted by $t_M(\mathbf{y}; \mu, \Sigma, \nu)$ of the M -dimensional t -distribution with parameters μ (real location vector), Σ ($M \times M$ real positive definite scale matrix) and ν (positive real degrees of freedom parameter) given by

$$\begin{aligned} t_M(\mathbf{y}; \mu, \Sigma, \nu) &= \int_0^\infty \mathcal{N}_M(\mathbf{y}; \mu, \Sigma/w) \mathcal{G}(w; \nu/2, \nu/2) dw \\ &= \frac{\Gamma((\nu + M)/2)}{|\Sigma|^{1/2} \Gamma(\nu/2) (\pi\nu)^{M/2}} [1 + \delta(\mathbf{y}, \mu, \Sigma)/\nu]^{-(\nu+M)/2} \end{aligned} \quad (5)$$

where $\delta(\mathbf{y}, \mu, \Sigma) = (\mathbf{y} - \mu)^T \Sigma^{-1} (\mathbf{y} - \mu)$ is the Mahalanobis distance between \mathbf{y} and μ (T means transpose) and Γ is the Gamma function: $\Gamma(x) = \int_0^{+\infty} t^{x-1} \exp(-t) dt$, $x \in \mathcal{R}_+^*$. Note that μ is the mean when $\nu > 1$ but Σ is not strictly speaking the covariance matrix of the t -distribution which is $\nu/(\nu - 2)\Sigma$ when $\nu > 2$.

In most of the work on multivariate scale mixture of Gaussians, the weight variable W has been considered as univariate which results in tails with the same heaviness in all dimensions. The extension we propose consists of introducing the parametrization of the scale matrix parameter into $\Sigma = \mathbf{D}\mathbf{A}\mathbf{D}^T$, where \mathbf{D} is the matrix of eigenvectors of Σ and \mathbf{A} is a diagonal matrix with the corresponding eigenvalues of Σ . The matrix \mathbf{D} determines the orientation of the Gaussian and \mathbf{A} its shape. The scaled Gaussian part is then set to $\mathcal{N}_M(\mathbf{y}; \mu, \mathbf{D}\mathbf{\Delta}_w\mathbf{A}\mathbf{D}^T)$, where $\mathbf{\Delta}_w = \text{diag}(w_1^{-1}, \dots, w_M^{-1})$ is the $M \times M$ diagonal matrix whose

diagonal components are the inverse weights $\{w_1^{-1}, \dots, w_M^{-1}\}$. When the weights are all one, a standard multivariate Gaussian case is recovered. The generalization introduced in [3] is therefore to define:

$$p(\mathbf{y}; \mu, \mathbf{D}, \mathbf{A}, \theta) = \int_0^\infty \dots \int_0^\infty \mathcal{N}_M(\mathbf{y}; \mu, \mathbf{D} \Delta_w \mathbf{A} \mathbf{D}^T) f_w(w_1 \dots w_M; \theta) dw_1 \dots dw_M \quad (6)$$

where f_w is now a M -variate density function. The weights are independent so that, *i.e.* with $\theta = \{\theta_1, \dots, \theta_M\}$, $f_w(w_1 \dots w_M; \theta) = f_{W_1}(w_1; \theta_1) \dots f_{W_M}(w_M; \theta_M)$. We can use then the expression below

$$\mathcal{N}_M(\mathbf{y}; \mu, \mathbf{D} \Delta_w \mathbf{A} \mathbf{D}^T) = \prod_{m=1}^M \mathcal{N}_1([\mathbf{D}^T(\mathbf{y} - \mu)]_m; 0, A_m w_m^{-1}) \quad (7)$$

where $[\mathbf{D}^T(\mathbf{y} - \mu)]_m$ denotes the m th component of vector $\mathbf{D}^T(\mathbf{y} - \mu)$ and A_m the m th diagonal element of the diagonal matrix \mathbf{A} (or equivalently the m th eigenvalue of Σ). Using (7), it follows that

$$p(\mathbf{y}; \mu, \mathbf{D}, \mathbf{A}, \theta) = \prod_{m=1}^M \int_0^\infty \mathcal{N}_1([\mathbf{D}^T(\mathbf{y} - \mu)]_m; 0, A_m w_m^{-1}) f_{W_m}(w_m; \theta_m) dw_m. \quad (8)$$

The terms in the product reduce then to standard univariate scale mixtures and in particular, when setting $f_{W_m}(w_m; \theta_m)$ to a Gamma distribution $\mathcal{G}(w_m; \nu_m/2, \nu_m/2)$, it follows a generalization of the multivariate t -distribution. We can use (8) to express easily the density denoted by $p_{\text{MST}}(\mathbf{y}; \mu, \Sigma, \nu)$ with $\nu = \{\nu_1, \dots, \nu_M\}$:

$$p_{\text{MST}}(\mathbf{y}; \mu, \mathbf{D}, \mathbf{A}, \nu) = \prod_{m=1}^M \frac{\Gamma((\nu_m + 1)/2)}{\Gamma(\nu_m/2)(A_m \nu_m \pi)^{1/2}} \left(1 + \frac{[\mathbf{D}^T(\mathbf{y} - \mu)]_m^2}{A_m \nu_m} \right)^{-(\nu_m + 1)/2} \quad (9)$$

We can then rewrite equation (1) as follows to have the density of a mixture of MST (MMST model):

$$\begin{aligned} \forall \mathbf{y} \in \mathcal{R}^M, p_{\text{MMST}}(\mathbf{y}; \psi) &= \sum_{k=1}^K \pi_k p_{\text{MST}}(\mathbf{y}; \mu_k, \mathbf{D}_k, \mathbf{A}_k, \nu_k) \\ &= \sum_{k=1}^K \pi_k \prod_{m=1}^M \frac{\Gamma\left(\frac{\nu_{k,m} + 1}{2}\right)}{\Gamma\left(\frac{\nu_{k,m}}{2}\right) ([\mathbf{A}_k]_{m,m} \nu_{k,m} \pi)^{\frac{1}{2}}} \left(1 + \frac{[\mathbf{D}_k(\mathbf{y} - \mu_k)]_m^2}{[\mathbf{A}_k]_{m,m} \nu_{k,m}} \right)^{-\frac{\nu_{k,m} + 1}{2}} \end{aligned} \quad (10)$$

with: $\psi = \{\psi_1, \dots, \psi_K\}$ where $\psi_k = (\pi_k, \mu_k, \mathbf{D}_k, \mathbf{A}_k, \nu_k)$ for $k \in \llbracket 1 ; K \rrbracket$.

D. Proposed procedure

The approach presented in this paper is a generic and automatic data-driven procedure to detect and characterize an abnormality within data of interest in comparison to reference data. These two main tasks are performed by a mixture of MST distributions, and will be illustrated on quantitative multiparametric MRI data with the goal to segment and diagnose rat brain tumors. The data have to be continuous quantitative measures, and the learning data have to be made of two sets of voxels: one from healthy subjects and one from pathological subjects for which the pathology (here the tumor) type is known. The procedure consists then of five steps: i) a first mixture model f_H is fitted to the healthy subject voxels; ii) this reference model is used to detect abnormal voxels in the healthy and pathological subjects; iii) a second mixture model f_A is fitted to the detected abnormal voxels and yields a clustering of these voxels into several classes; iv) the proportions of these classes in each subject are used as a signature of the pathology and a discriminative (fingerprint) model is learned that can distinguish between different pathology types; v) an additional spatial post-processing is carried out to remove some spatial artifacts and refine the pathology signatures.

The whole procedure is summarized in Figure 1.

III. APPLICATION TO BRAIN TUMORS USING MULTIPARAMETRIC MRI DATA

A. Description of the MRI data

a) *Data acquisition:* The proposed procedure is illustrated on a data set of 53 rats for which 5 quantitative MRI maps are available. MRI was conducted with a horizontal bore 4.7 T Biospec animal imager (Bruker Biospin, Ettlingen, Germany) with an actively decoupled cross-coil setup (body coil for radiofrequency transmission and quadrature surface coil for signal reception) and using Paravision 5.0.1. After second-order shimming, the following MRI protocol was performed. The image acquisition positions were identical for all MRI sequences. 5 slices were acquired with a voxel size of $234 \times 234 \times 800 \mu\text{m}$, unless mentioned otherwise:

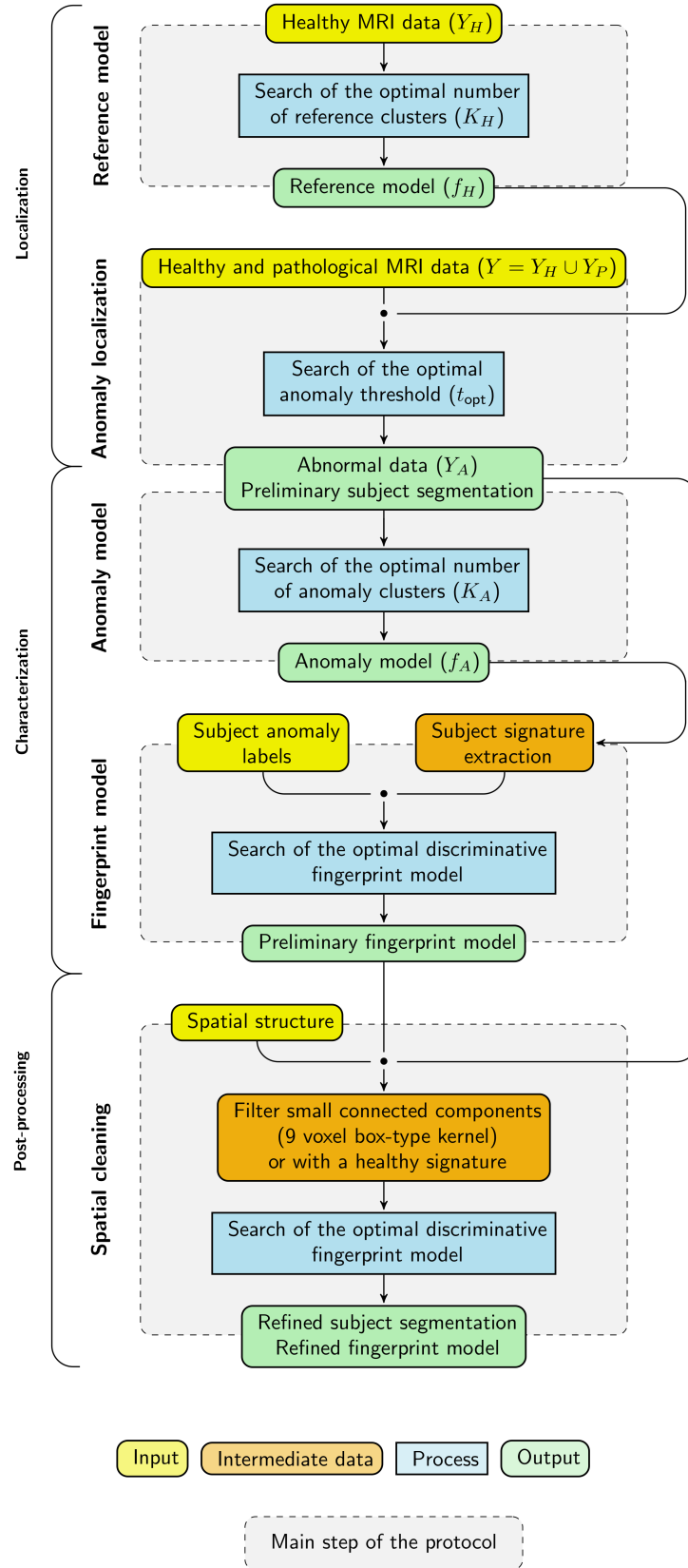


Fig. 1. Construction of a model to automatically localize and characterize lesions. Starting from subjects labeled as healthy or pathological, the procedure is made of 5 main steps.

1) Anatomical T2-weighted (T2w) images were acquired using a spin-echo MRI sequence (repetition time (TR)/effective

echo-time (TE) = 4000/33 *ms*, acceleration factor = 4, NEX = 2, 31 slices with a field of view (FOV) = $30 \times 30 \text{ mm}^2$, acquisition matrix = 256×256 and voxel size = $117 \times 117 \times 800 \text{ }\mu\text{m}^3$.

- 2) Cerebral blood flow (CBF) was determined using pseudo-continuous arterial spin labeling (pCASL; spin-echo echo-planar imaging (EPI), TR/TE = 4500/17.2 *ms*, labeling duration = 4 *s*, postlabeling delay = 0.2 *s*, 20 label/control pairs).
- 3) T1 of brain tissue was determined using an inversion recovery sequence (spin-echo EPI, TR/TE = 10000/20 *ms*, 25 inversion times: TI = 100-3700 *ms*).
- 4) Apparent Diffusion Coefficient (ADC) was mapped using a diffusion-weighted, spin-echo, EPI (TR/TE = 3000/28.6 *ms*, 8 averages). This sequence was applied 6 times; three without diffusion weighting and three times with diffusion weighting ($b = 800 \text{ s/mm}^2$) in three orthogonal directions.
- 5) T2 (resp. T2*) relaxometric maps were acquired using multiple spin (resp. gradient) echo sequences. Multi spin-echo 2D (MSME) (TR = 2000 *ms*, 26 spin-echoes, $\Delta\text{TE} = 12 \text{ ms}$). Multi gradient echo 2D (MGE2D) (TR = 2000 *ms*, 8 gradient-echoes, $\Delta\text{TE} = 4.5 \text{ ms}$).
- 6) Ultrasmall superparamagnetic iron oxide (USPIO) particles were injected via the tail vein in about 20 sec (200 $\mu\text{mol Fe/kg}$ body weight; P904[®], Guerbet, Roissy, France). 3 min after the injection of USPIOs, a post-contrast T2* relaxometric map was acquired.
- 7) The vascular wall integrity was assessed using a dynamic contrast-enhanced MRI approach as previously described in [4]. Briefly, 60 T1-weighted, spin-echo, images (TR/TE = 800/4.2 *ms*, 15.6 *s* per image) were acquired. After acquisition of 10 baseline images, a bolus of Gd-DOTA (200 $\mu\text{mol/kg}$; Guerbet SA, France) was administered through the tail vein and flushed with 250 μL of saline.

b) Data processing for MRI parameter maps: All processing were computed on a voxel basis using custom code developed in the Matlab environment (The MathWorks Inc., Natick, Ma, USA):

- ADC maps were computed as the mean of the ADCs observed in each of three orthogonal directions.
- T1, T2*, and T2 maps were derived using a non-linear fitting algorithm and from the inversion recovery, gradient echoes, and spin-echoes data, respectively.
- The CBF computation was based on the equations described in [5] and assuming a transit delay much shorter than the longitudinal relaxation time of blood.
- Cerebral blood volume (CBV) maps were estimated using the steady-state approach described by [6] and [7]. Changes in transverse relaxation rates due to USPIO (ΔR_2^*) were obtained from the T2* maps collected before and after injection of USPIO. CBV was computed using:

$$\text{CBV} = \frac{3}{4\pi\gamma B_0 \Delta\chi} \Delta R_2^* \quad (11)$$

where $\gamma = 2.67502 \cdot 10^8 \text{ rad/s/T}$, and $B_0 = 4.7\text{T}$. $\Delta\chi$, the susceptibility difference between blood in the presence and in the absence of USPIO, was set to 3.5 *ppm* (SI units).

- The vascular wall integrity was estimated by the area under the signal enhancement curve (AUC) following the Gd-DOTA injection, after baseline removal ([4]).

Voxels were included in the analysis according to the following criteria: $0 < \text{ADC} < 4000 \mu\text{m}^2/\text{s}$, $-2 \cdot 10^5 < \text{AUC} < 1.5 \cdot 10^6 \text{ a.u.}$, $-50 < \text{CBV} < 50\%$, $0 < \text{T1} < 5000 \text{ ms}$ and $0 < \text{T2} < 400 \text{ ms}$. Meeting a single exclusion criterion led to an exclusion of the voxel from all maps. Using these criteria, 0.058% of the voxels were excluded.

B. Healthy subject based reference model

The first steps of our procedure are dedicated to localize all abnormal voxels within the voxel set formed by the pathological subjects (26 rats). This localization is done as an anomaly detection in comparison of a reference model f_H that we fit using a MST mixture on the voxel set from the healthy subjects (6 rats). The slope heuristic selects a number $K_H = 10$ of clusters for the reference model f_H . Figure 2 presents the obtained clustering associated to f_H on a healthy rat, with one color per cluster. Each cluster corresponds to a MST distribution and is described by the distribution parameters $\{\mu_k, \mathbf{D}_k, \mathbf{A}_k, \nu_k\}$ and the proportion parameter π_k (for the k -th cluster). Figure 3 gives a summary of these clusters through the mean parameter μ_k of each distribution, and the cluster size to illustrate the proportion parameter π_k in terms of voxel number. Each hexagonal web-diagram represents a cluster with, at the center, the lowest MR values of these mean parameters: $\text{ADC} = 775 \mu\text{m}^2/\text{s}$, $\text{AUC} = 710.4 \text{ a.u.}$, $\text{CBV} = 1.93\%$, $\text{T1} = 1345 \text{ ms}$, $\text{T2} = 68.3 \text{ ms}$ and $\text{size} = 411$ voxels; and at the border there are the maximal MR values of these mean parameters: $\text{ADC} = 1474 \mu\text{m}^2/\text{s}$, $\text{AUC} = 98588 \text{ a.u.}$, $\text{CBV} = 5.34\%$, $\text{T1} = 2058 \text{ ms}$, $\text{T2} = 126 \text{ ms}$ and $\text{size} = 11779$ voxels. The respective parameter values of each cluster are indicated Table I.

Cluster maps appear spatially structured, with a symmetry between the two hemispheres. One can recognize some main anatomical features: the cortex appears in red, the corpus callosum in green, ventricles in purple, blue and dark blue, the striatum is a mixture of orange (main color), red, and light green. Vascular structures, which exhibit a large AUC, appear in blue (cluster 7).

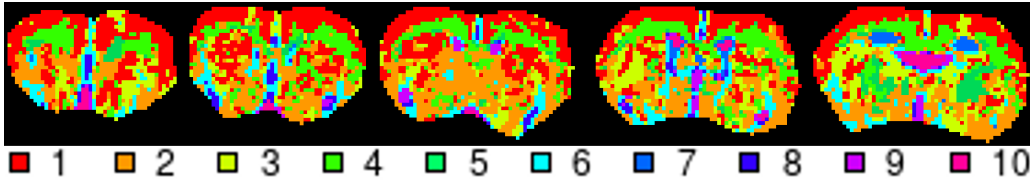


Fig. 2. Clustering using the reference model f_h (MMST model with $K_H = 10$ clusters) on the 5 slices of one healthy rat. From left to right: slices in decreasing order along the vertical axis of the scanner. Each color corresponds to a cluster.

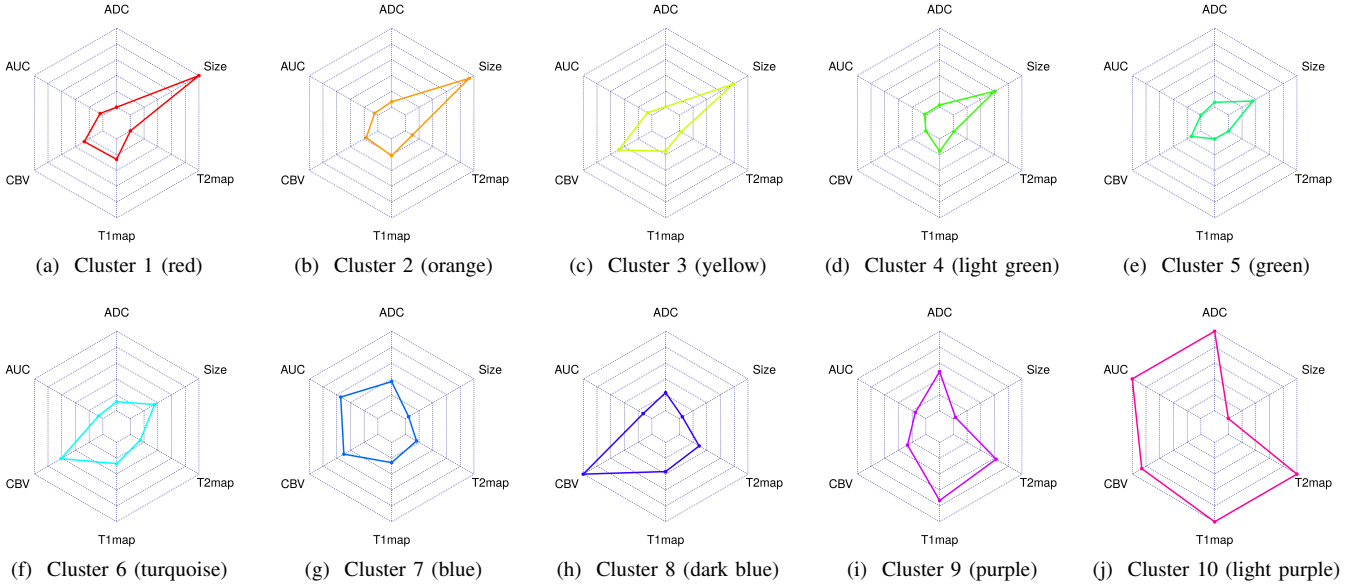


Fig. 3. MR parameter means and size of the reference model clusters (f_H). Range of these parameters between the center and the border of each diagram: $ADC \in [775, 1474] \mu m^2/s$, $AUC \in [710.4, 98588] a.u.$, $CBV \in [1.93, 5.34] \%$, $T1 \in [1345, 2058] ms$, $T2 \in [68.3, 126] ms$ and $Size \in [411, 11779]$ voxels.

TABLE I
MEAN AND SIZE PARAMETERS OF THE REFERENCE MODEL CLUSTERS (MMST MODEL f_h WITH $K_H = 10$ CLUSTERS).

Cluster	1	2	3	4	5	6	7	8	9	10
ADC ($\mu m^2/s$)	775	823	779	795	819	854	1031	934	1120	1474
AUC (<i>a.u.</i>)	4713	5387	6905	2540	710.4	7011	53751	13040	15723	98588
CBV (%)	2.85	2.51	3.56	1.93	2.41	4.01	3.62	5.34	2.84	4.87
T1 (<i>ms</i>)	1530	1499	1458	1457	1345	1536	1527	1607	1867	2058
T2 (<i>ms</i>)	68.3	74.3	69.8	68.9	68.6	76.7	77.8	85.1	104	126
size (voxels)	11779	10025	7155	6863	3185	3422	816	665	730	411

C. Anomaly localization

The reference model f_H is used to detect the voxels which are abnormal compared to f_H in the sense that they are not well statistically explained by this model. The MR parameter values of the voxels from the pathological and healthy subjects form the data set Y on which we want to detect anomaly by computing the log-likelihood of f_H : $\{\log f_H(y), y \in Y\}$. As described in the main article, Section III-C, we adjust another mixture model on this data set to identify the voxel subsets sharing similar log-likelihood values, that is the same abnormality with respect to the reference model f_H . This clustering provides L successive abnormal thresholds $\{t_1, \dots, t_L\}$ and defines nested anomaly segmentations. One of these thresholds is also determined for its ability to best split the data set Y into a healthy part and an abnormal part Y_A . It is called the global threshold t_{opt} .

Figure 4 illustrates these segmentations: the colored voxels are selected as abnormal (most abnormal in red), and the grey ones are considered as healthy (less abnormal in dark grey). A manual delineation (superimposed red line) of the tumor was performed on the anatomical image and the diffusion map. We compare the obtained segmentations using the manual segmentation as ground truth through the associated contingency table (Table II), with the following counts:

- TN (True Negative): the healthy voxels correctly declared as healthy
- FN (False Negative): the tumoral voxels wrongly declared as healthy
- FP (False Positive): the healthy voxels wrongly declared as abnormal

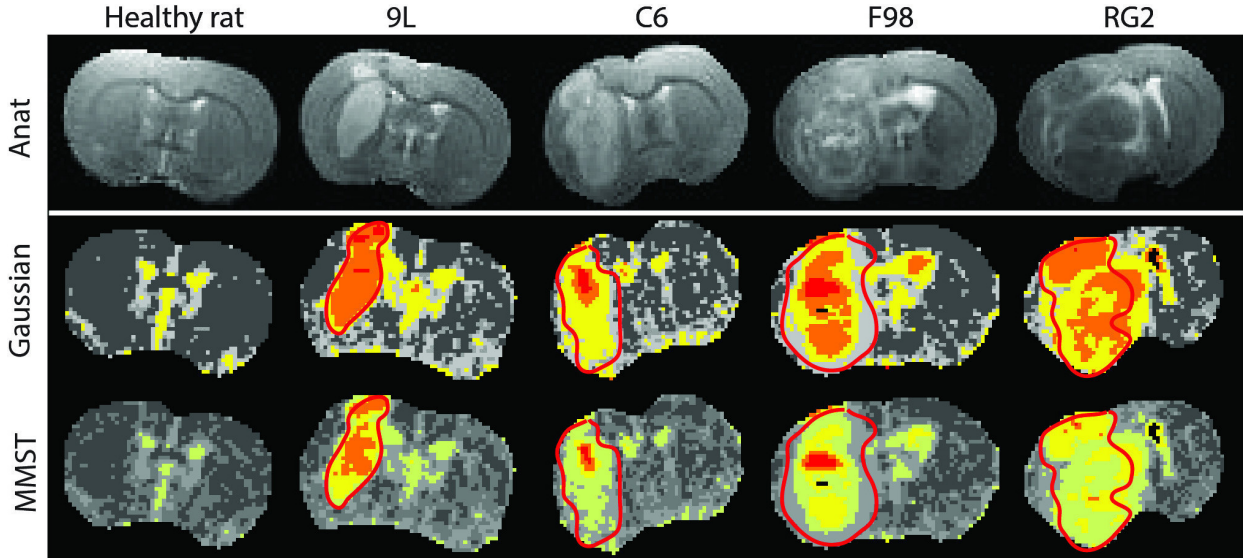


Fig. 4. Nested anomaly segmentations for the different anomaly thresholds with the Gaussian model (center row) and the MMST model (bottom row). The anatomical map (upper row) is displayed as a reference for a visual comparison. The thresholds are ordered from lowest log-likelihood (most abnormal in red) to highest log-likelihood (less abnormal in dark gray) as displayed in the following figures about covering indexes. The colored voxels are those defining the abnormal data set Y_A (i.e. below the global threshold). The gray ones correspond to normal groups (i.e. above the global threshold). The red line superimposed corresponds to the manual segmentation. These segmentations are obtained prior to the spatial post-processing.

- TP (True Positive): the tumoral voxels correctly declared as abnormal

TABLE II
CONTINGENCY TABLE BETWEEN THE AUTOMATIC AND MANUAL SEGMENTATIONS.

		Automatic segmentation		Total
		Normal	Abnormal	
Manual segmentation	Healthy	TN	FP	n_H
	Tumor	FN	TP	n_T
Total		n_{pH}	n_{pT}	n

We use several covering indexes to evaluate the performance of the automatic segmentation:

a) *Adjusted Rand Index (ARI)*: is a covering index defined by Equation (12) [8] which measures the agreement between the manual and automatic segmentations for the 2 subsets: healthy and tumor. These index is derived from the rand index [9] which computes the agreement as the number of voxel pairs which are in the same subset in the manual and in the automatic segmentation, plus the number of voxel pairs which are in different subsets in the manual and in the automatic segmentation. The rand index takes values between 0 and 1, the higher the better: at 0 the segmentations agree on no voxel pairs, at 1 they agree on all pairs (perfect concordance). The ARI is a normalization of the rand index such that the ARI value is still lesser than 1 (perfect concordance) but can take negative values when the segmentation is worse than a random clustering under hypergeometric assumptions. The ARI formula is:

$$ARI = \frac{\binom{TN}{2} + \binom{FN}{2} + \binom{FP}{2} + \binom{TP}{2} - \left[\binom{n_H}{2} + \binom{n_T}{2} \right] \left[\binom{n_{pH}}{2} + \binom{n_{pT}}{2} \right] / \binom{n}{2}}{\frac{1}{2} \left[\binom{n_H}{2} + \binom{n_T}{2} + \binom{n_{pH}}{2} + \binom{n_{pT}}{2} \right] - \left[\binom{n_H}{2} + \binom{n_T}{2} \right] \left[\binom{n_{pH}}{2} + \binom{n_{pT}}{2} \right] / \binom{n}{2}} \leq 1 \quad (12)$$

Figure 5 shows the ARI values for the obtained nested segmentations. The MMST model detects 7 thresholds (Figure 5a, the boxplots are sorted by abnormality threshold: the higher in red, the lower in dark gray) with a global threshold associated to the 4th threshold. We can observe that the global threshold is also the one with the highest median for the ARI values. If we focus on this threshold, the distribution of the ARI values differs per tumor type (Figure 11a): the 9L and RG2 rats present the best agreement with the manual segmentation, and the F98 rats seem to be the most difficult.

b) *DICE index*: is a covering index defined by Equation (13) below [10], also known as the F1 score, which measures the similarity between the manual and automatic segmentations of the tumor area. It is a weighted average of the precision and sensitivity for the recovery of the tumor area. The DICE index takes values between 0 and 1, the higher the better: 0 means that all the tumoral voxels are not correctly predicted as healthy; 1 means that the tumoral voxels and only the tumoral voxels are declared as abnormal (perfect concordance).

$$DICE = \frac{2TP}{2TP + FP + FN} = \frac{2TP}{n_{pT} + n_T} \in [0, 1] \quad (13)$$

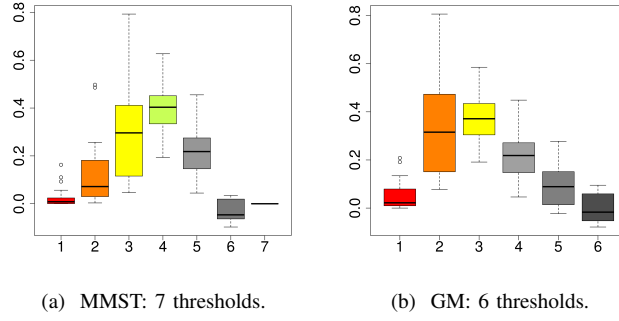


Fig. 5. Adjusted rand index (ARI) of the abnormal voxels associated to the successive abnormal thresholds for the MMST model (a) and the GM model (b), without spatial post-processing. For the MMST model (a), the color code is the same as that of Figure 4.

Figure 6 shows the DICE values for the nested segmentations. As for the ARI, the DICE index is also the highest for the 4th of the 7 detected threshold in terms of median values (Figure 6a). Focusing on this threshold, the distribution per tumor type of the DICE values differs from the one observed with the ARI (Figure 12a): the RG2 rats have the best DICE values. We may remark that the ARI and DICE values give different orders for the tumor types. It can be explained by the fact that the DICE index focuses on the true condition (here the tumor area), e.g. the number of true negatives does not matter, whereas the ARI takes into account all the classes of the segmentations. One common point between ARI and DICE values is the variability of the 9L segmentations, which is high in comparison with the others types, while the one of the C6 rats indices variability is low.

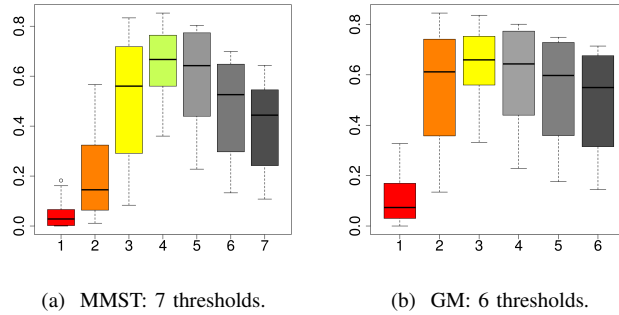


Fig. 6. DICE index of the abnormal voxels associated to the successive abnormal thresholds for the MMST model (a) and the GM model (b), without spatial post-processing. For the MMST model (a), the color code is the same as that of Figure 4.

Table III details additional covering indices. The sensitivity ($\frac{TP}{n_T}$) indicates how many tumoral voxels are correctly predicted as abnormal. The specificity ($\frac{TN}{n_H}$) indicates how many healthy voxels are correctly predicted as healthy. The ROC (Receiver operating characteristic) curve is defined by the pairs $(\frac{TP}{n_T}, \frac{FP}{n_H})$ (false positive rate against sensitivity), and the area under the ROC curve is equivalent to the probability that the procedure gives a higher abnormal level to a tumoral voxel than to a healthy voxel. It takes values between 0 and 1; a value of 0.5 is associated to a random allocation.

Table III shows that the ability to detect abnormality is higher for MMST than GM: the sensitivity is higher (more abnormal voxels are correctly declared as abnormal) with a smaller specificity (more healthy voxels are wrongly declared as abnormal). But globally, the increase of true positive compensates the false positive: the area under the ROC curve is equivalent, the DICE index and the ARI are slightly better. We can notice that the MMST model detects 7 thresholds instead of 6 for the GM model. It thus allows a finer segmentation of the abnormal voxels: the variability of the 2nd GM threshold is equal to the variability of the combined thresholds 2 and 3 of the MMST model (Figures 5b and 6b). For the similarities, the chosen global threshold is the one with the highest median values of ARI and DICE for both the MMST and GM models. When focusing on this threshold, the order of tumor type in terms of ARI and DICE median values is the same for the MMST and GM models, except for the tumor type order based on the DICE values, for which the ranks of 9L and C6 rats are switched (Figures 11c and 12c).

D. Anomaly model

The last steps of our procedure are dedicated to characterize the pathologies using the voxels declared as abnormal. We first adjust a MMST model on the abnormal voxels Y_A , in order to best describe the data using a mixture model before using it to compute the subject fingerprints. The anomaly model f_A is composed of $K_A = 10$ clusters, according to the slope heuristic.

TABLE III
PATHOLOGICAL RATS IN THE TRAINING SET: MEAN COVERING SCORES FOR THE LESION AUTOMATIC SEGMENTATIONS FOR THE GAUSSIAN MIXTURE AND MMST MODELS, WITHOUT SPATIAL POST-PROCESSING.

	Gaussian model					MMST model				
	9L	C6	F98	RG2	All	9L	C6	F98	RG2	All
sensitivity	0.830	0.560	0.553	0.853	0.699	0.862	0.641	0.636	0.880	0.755
specificity	0.922	0.926	0.912	0.829	0.895	0.884	0.884	0.871	0.779	0.852
area under ROC curve	0.928	0.855	0.867	0.894	0.883	0.928	0.852	0.865	0.897	0.883
DICE index	0.510	0.582	0.671	0.801	0.648	0.581	0.582	0.639	0.819	0.661
ARI	0.400	0.372	0.332	0.421	0.381	0.489	0.402	0.322	0.467	0.418

Figure 7 presents the clustering associated to f_A on the pathological rats from the training set, with one color per cluster. Note that the clusters now differ from that obtained from the Y_H data set.

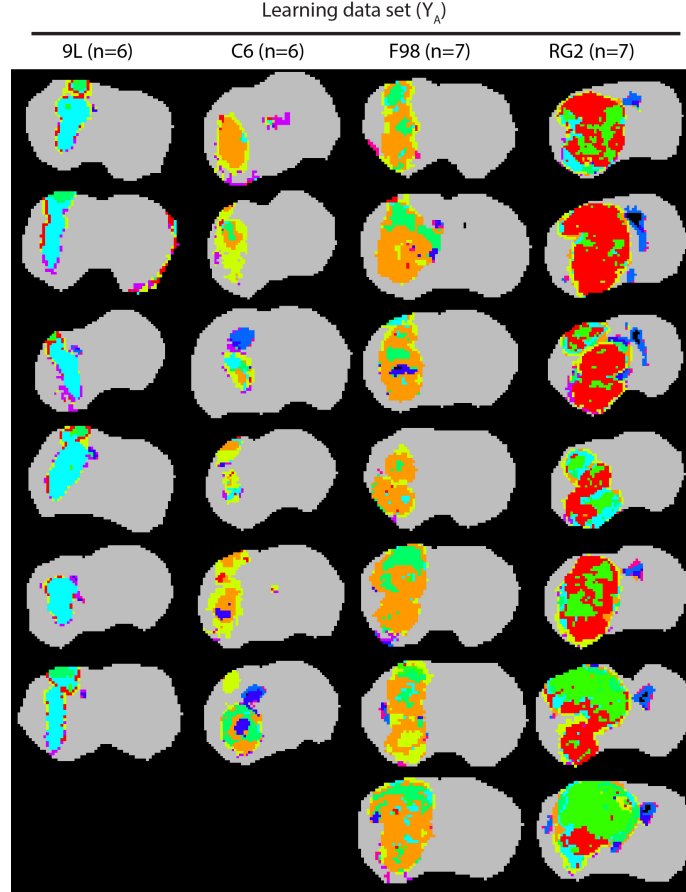


Fig. 7. Anomaly clustering of the pathological rats from the training set Y_A using the anomaly model f_A (MST mixture with $K_A = 10$ clusters) and including spatial post-processing. Each column gathers the rats bearing the same pathology (tumor 9L, C6, F98, or RG2); the central slice is the only one displayed per rat and contains the biggest tumoral area. Note that the color code is different from that of Figure 2. The clusters associated to Y_A are described Figure 8 and Table IV.

As for the reference model, each cluster corresponds to a MST distribution and is described by the distribution parameters $\{\mu_l, \mathbf{D}_l, \mathbf{A}_l, \nu_l\}$ and the proportion parameter π_l (for the l -th cluster). Figure 8 gives a summary of these clusters through the mean parameter μ_l of each distribution, and the cluster size to illustrate the proportion parameter π_l in terms of voxel numbers. Each hexagonal web-diagram represents a cluster with, at the center, the lowest MR values of these mean parameters: $\text{ADC} = 744 \mu\text{m}^2/\text{s}$, $\text{AUC} = 12188 a.u.$, $\text{CBV} = 1.31\%$, $\text{T1} = 1484 \text{ms}$, $\text{T2} = 74.3 \text{ms}$ and $\text{size} = 2814$ voxels; and at the border there are the maximal MR values of these parameters: $\text{ADC} = 1668 \mu\text{m}^2/\text{s}$, $\text{AUC} = 474460 a.u.$, $\text{CBV} = 9.21\%$, $\text{T1} = 2438 \text{ms}$, $\text{T2} = 179 \text{ms}$ and $\text{size} = 11612$ voxels. The mean parameters associated to each cluster are indicated Table IV.

Clusters represent different types of tissues. For example, clusters 7 and 8 show high ADC with limited permeability and could correspond to ventricles/plexus choroid. Cluster 1 shows a normal ADC, a CBV 50% above normal brain values and a high permeability. The highest permeability (cluster 5, green) is associated with an ADC between that of normal tissue and that of ventricles and with a normal CBV.

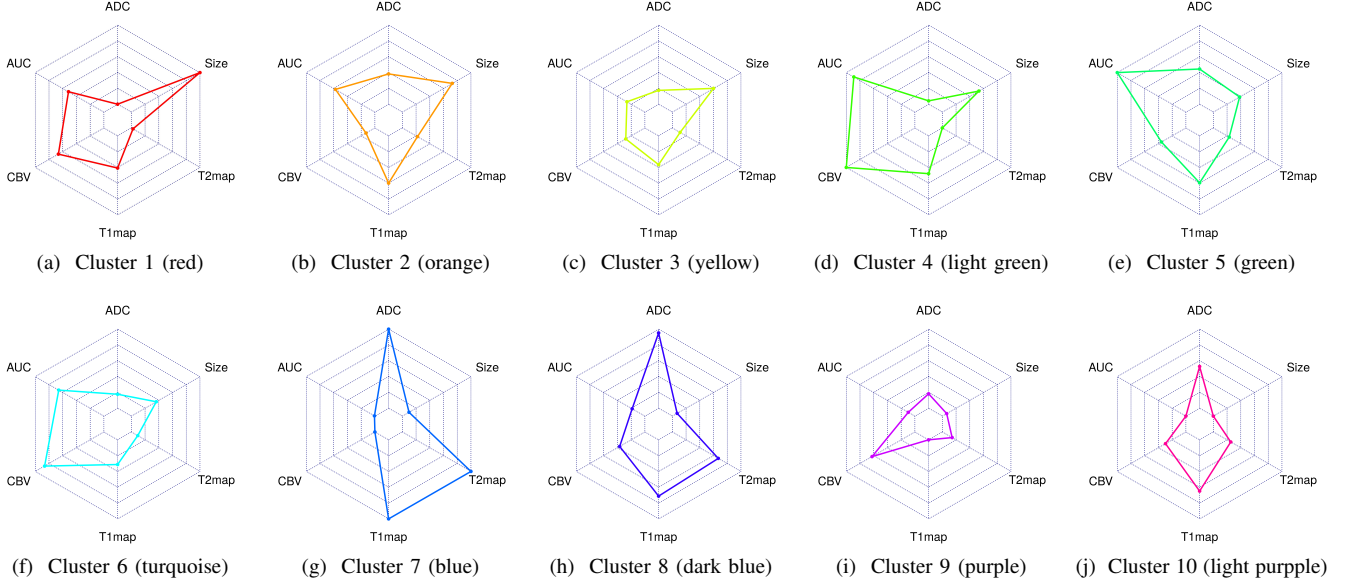


Fig. 8. MR parameter means and size of the anomaly model clusters. Range of these parameters: $ADC \in [744, 1668] \mu m^2/s$, $AUC \in [12188, 474460] a.u.$, $CBV \in [1.31, 9.21] \%$, $T1 \in [1484, 2438] ms$, $T2 \in [74, 179] ms$ and $Size \in [11284, 49732]$ voxels. The color code is similar to that of Figure 7.

TABLE IV
MEAN AND SIZE PARAMETERS OF THE ANOMALY MODEL CLUSTERS (MMST MODEL f_A WITH $K_A = 10$ CLUSTERS).

	1	2	3	4	5	6	7	8	9	10
ADC ($\mu m^2/s$)	744	1098	906	783	1156	908	1668	1624	913	1232
AUC (<i>a.u.</i>)	250599	277915	133332	423702	474460	316320	15018	98009	56680	12188
CBV (%)	6.54	2.33	3.51	9.21	4.12	8.13	1.31	4.21	6.24	3.65
T1 (<i>ms</i>)	1872	2055	1832	1941	2051	1782	2438	2163	1484	2103
T2 (<i>ms</i>)	76.8	97.4	86.2	74.3	98.6	84.1	179	145	89.3	101
size (voxels)	11612	9015	8474	6724	4617	5159	3274	2756	3102	2814

E. Post-processing

We present here the impact of the spatial post-processing on the tumor localization.

Figure 9 presents the anomaly clustering using a Gaussian mixture model based on an automatic localization of the tumor (a variant of the proposed procedure in terms of probability distributions). For this Gaussian mixture also, the slope heuristic selected $K_A = 13$ clusters. As described in Section III-C, the automatic segmentation wrongly declares some voxels from the contralateral part as abnormal. The spatial cleaning allows to remove the isolated voxels and the ones at the border of the skull, but the abnormal voxels in the ventricles often form a subset touching the tumor area. These voxels cannot be removed by our spatial cleaning, which leads to an overestimation of the abnormal part.

Figure 10 presents the anomaly clustering using a MMST model based on an automatic localization of the tumor (the proposed procedure). For the MMST model, the slope heuristic selected $K_A = 10$ clusters. With this model, the abnormal contralateral area is smaller than the one with the Gaussian mixture, which leads to less contacts with the tumor area, and then a better spatial cleaning, as we can observe for the 9L and F98 tumor types.

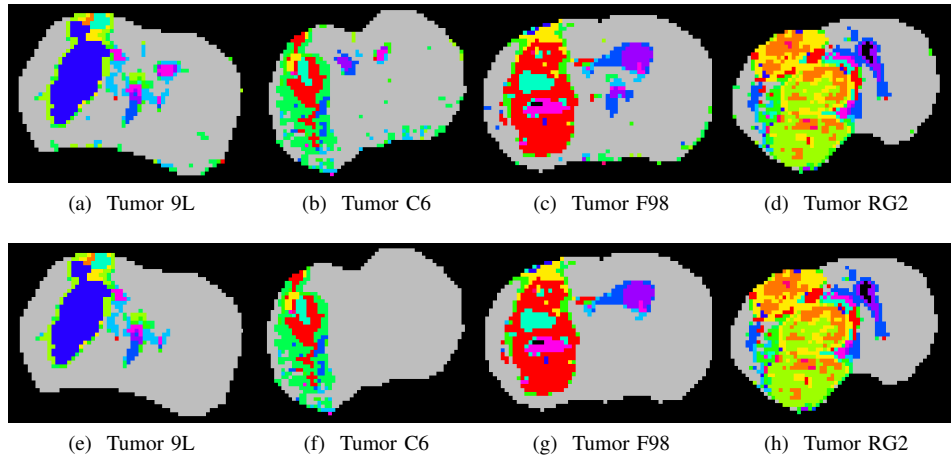


Fig. 9. Gaussian anomaly clustering ($K_A = 13$ clusters): before (upper row) and after (bottom row) the spatial cleaning by removing too small components or components with a healthy fingerprint.

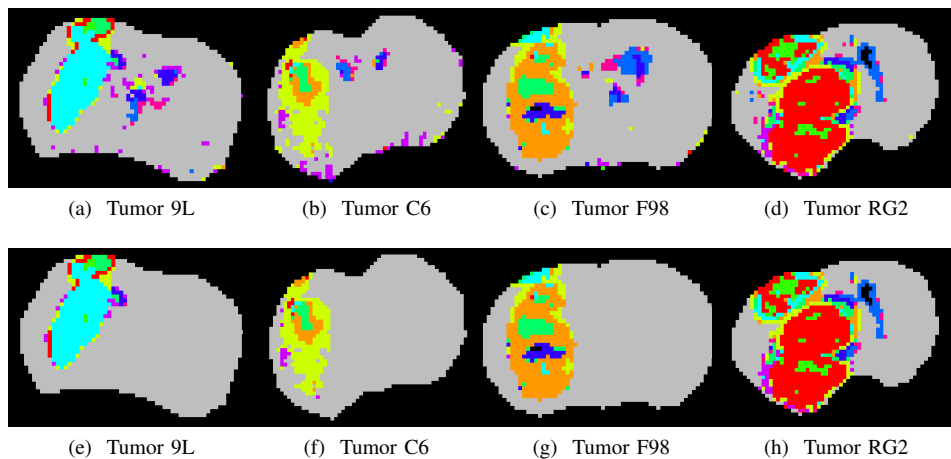


Fig. 10. MMST anomaly clustering ($K_A = 10$ clusters): before (upper row) and after (bottom row) the spatial cleaning by removing too small components or components with a healthy fingerprint.

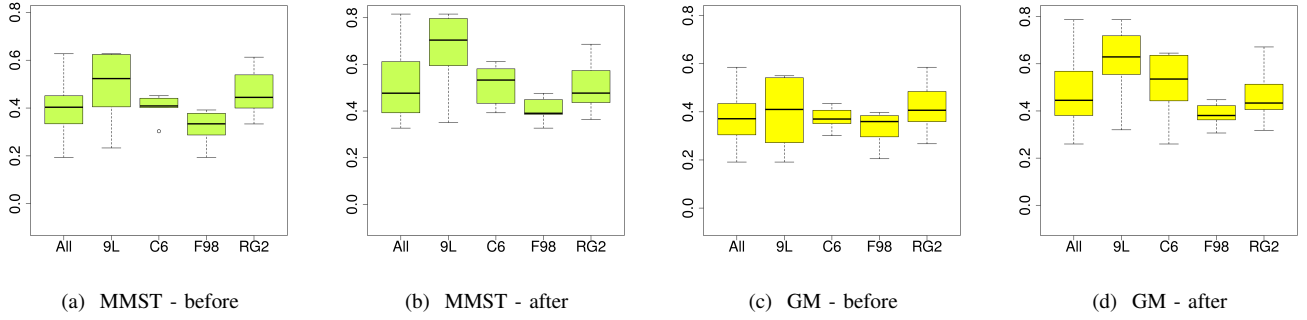


Fig. 11. ARI per tumor type for the segmentation associated to the global threshold, before (MMST: a, GM: c) and after (MMST: b, GM: d) the spatial post-processing.

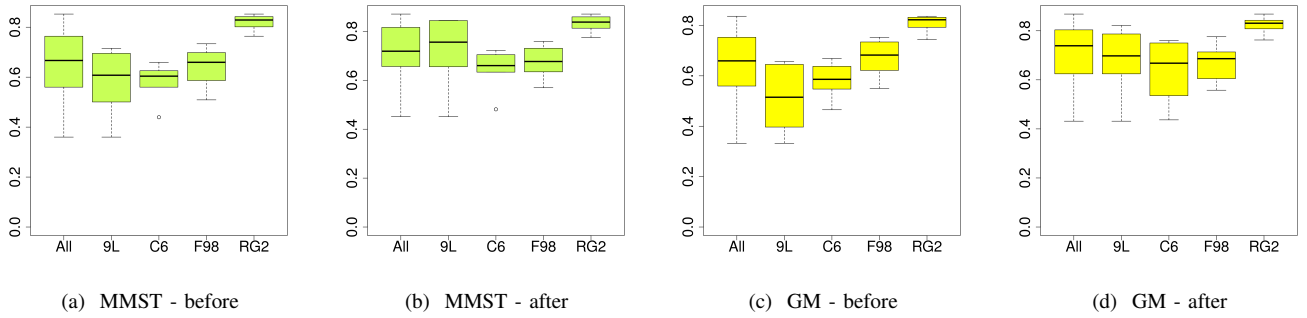


Fig. 12. DICE index per tumor type for the segmentation associated to the global threshold, before (MMST: a, GM: c) and after (MMST: b, GM: d) the spatial post-processing.

TABLE V

PATHOLOGICAL RATS IN THE TRAINING SET: MEAN COVERING SCORES FOR THE LESION AUTOMATIC SEGMENTATIONS FOR THE GAUSSIAN MIXTURE AND MMST MODELS, AFTER SPATIAL POST-PROCESSING.

	Gaussian model					MMST model				
	9L	C6	F98	RG2	All	9L	C6	F98	RG2	All
sensitivity	0.856	0.530	0.541	0.879	0.702	0.827	0.519	0.538	0.851	0.685
specificity	0.949	0.980	0.960	0.809	0.921	0.964	0.984	0.977	0.856	0.943
area under ROC curve	0.916	0.794	0.787	0.882	0.840	0.935	0.867	0.885	0.902	0.894
DICE index	0.677	0.636	0.665	0.823	0.704	0.718	0.644	0.677	0.833	0.721
ARI	0.607	0.509	0.386	0.466	0.487	0.661	0.514	0.409	0.507	0.518

F. Evaluation on an independent test data set

We test our procedure on an independent data set composed of 9L rats ($n = 5$), F98 rats ($n = 5$), and healthy rats ($n = 11$), as described in the manuscript Section III-E. All the results concerning the fingerprint model are presented in the manuscript, Section III-D. We only present here the automatic segmentation after spatial post-processing, Figure 13, for the pathological rats. The healthy test rats present as desired a segmentation with no abnormal voxels, except for the one already shown in the manuscript, Section III-E. For these rats, average ARI and DICE index are 0.49 and 0.69 after spatial post-processing, respectively, values comparable to that obtained during training. These data demonstrate the ability of the proposed procedure to automatically segment tumors from an unseen data set.

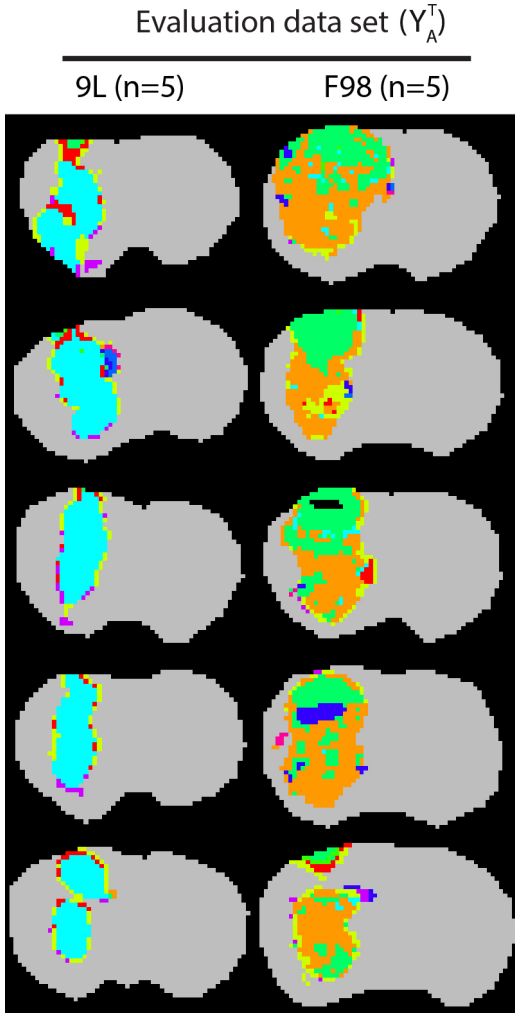
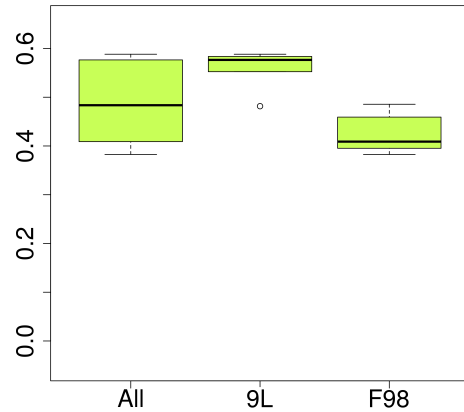
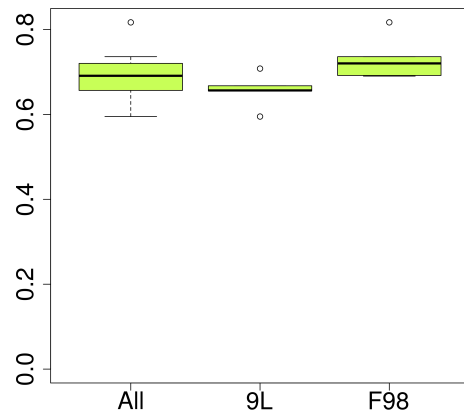


Fig. 13. Anomaly clustering of the pathological rats from the test set Y_A^T using the anomaly model f_A (MST mixture with $K_A = 10$ clusters) and the refined ROI. Each column gathers the rats bearing the same pathology (tumor 9L, or F98); the central slice is the only one displayed per rat and contains the biggest tumor area. The color code is similar to that of Figures 7 and 8.



(a) Adjusted rand index.



(b) DICE index.

Fig. 14. Adjusted rand index (a) and DICE index (b) of the refined segmentation for the test data set Y_A^T , using the reference model and the refined fingerprint model. The results are obtained under the MMST model.

REFERENCES

- [1] J.-P. Baudry, C. Maugis, and B. Michel, “Slope heuristics: overview and implementation,” *Statistics and Computing*, vol. 22, no. 2, pp. 455–470, 2012.
- [2] N. Coquery, O. François, B. Lemasson, C. Debacker, R. Farion, C. Rémy, and E. L. Barbier, “Microvascular MRI and unsupervised clustering yields histology-resembling images in two rat models of glioma,” *Journal of Cerebral Blood Flow & Metabolism*, vol. 34, no. 8, pp. 1354–1362, May 2014.
- [3] F. Forbes and D. Wraith, “A new family of multivariate heavy-tailed distributions with variable marginal amounts of tailweights: Application to robust clustering,” *Statistics and Computing*, vol. 24, no. 6, pp. 971–984, 2014.
- [4] B. Lemasson, R. Serduc, C. Maisin, A. Bouchet, N. Coquery, P. Robert, G. Le Duc, I. Troprès, C. Rémy, and E. L. Barbier, “Monitoring Blood-Brain Barrier Status in a Rat Model of Glioma Receiving Therapy: Dual Injection of Low-Molecular-Weight and Macromolecular MR Contrast Media,” *Radiology*, vol. 257, no. 2, pp. 342–352, 2010.
- [5] C. S. Debacker, A. Daoust, S. Köhler, J. Voiron, J. M. Warnking, and E. L. Barbier, “Impact of tissue T1 on perfusion measurement with arterial spin labeling,” *Magnetic Resonance in Medicine*, vol. 77, no. 4, pp. 1656–1664, 2017.

- [6] I. Troprès, S. Grimault, A. Vaeth, E. Grillon, C. Julien, J.-F. Payen, L. Lamalle, and M. Décorps, "Vessel size imaging," *Magnetic Resonance in Medicine*, vol. 45, no. 3, pp. 397–408, 2001.
- [7] S. Valable, B. Lemasson, R. Farion, M. Beaumont, C. Segebarth, C. Rémy, and E. L. Barbier, "Assessment of blood volume, vessel size, and the expression of angiogenic factors in two rat glioma models: a longitudinal in vivo and ex vivo study," *NMR in Biomedicine*, vol. 21, no. 10, pp. 1043–1056, 2008.
- [8] L. Hubert and P. Arabie, "Comparing partitions," *Journal of Classification*, vol. 2, no. 1, pp. 193–218, 1985.
- [9] W. M. Rand, "Objective Criteria for the Evaluation of Clustering Methods," *Journal of the American Statistical Association*, vol. 66, no. 336, pp. 846–850, December 1971.
- [10] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297–302, 1945.