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

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

Fully Automatic Lesion Localization and Characterization: Application to Brain Tumors using Multiparametric MRI Data

Alexis Arnaud, Florence Forbes
Univ. Grenoble Alpes, INRIA,
Laboratoire Jean Kuntzmann, Grenoble, France
Email: firstname.lastname@inria.fr

Nicolas Coquery, Nora Collomb,
Benjamin Lemasson, Emmanuel L. Barbier
Grenoble Institut des Neurosciences
Inserm U1216 & Univ. Grenoble Alpes, France
Email: firstname.lastname@univ-grenoble-alpes.fr

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. Non trivial interactions between relevant physiological parameters are captured thanks to recent generalized Student distributions that provide a larger variety of distributional shapes compared to the more standard Gaussian distributions. Probabilistic mixtures of the former distributions are then considered to account for the different tissue types and potential heterogeneity of lesions. Discriminative multivariate features are extracted from this mixture modelling and turned into individual lesion signatures. The signatures are subsequently pooled together to build a statistical fingerprint model of the different lesion types that captures lesion characteristics while accounting for inter-subject variability. The potential of this generic procedure is demonstrated on a data set of 53 rats, with 36 rats bearing 4 different brain tumors, for which 5 quantitative MR parameters were acquired.

***Index Terms*—Perfusion imaging, Magnetic resonance imaging (MRI), Animal models and imaging, Computer-aided detection and diagnosis, Probabilistic and statistical methods, Automatic segmentation, Automatic characterization, Brain tumor, Quantitative multiparametric MRI, Mixture model, Anomaly detection, Radiomics, Fingerprint model.**

I. Introduction

In medical imaging, and in particular in magnetic resonance imaging (MRI), several contrasts may be obtained from a single imaging session. A typical MRI exam includes different contrasts such as T1-weighted, T2-weighted, FLAIR, contrast enhanced images, and, depending on the indication, additional contrasts such as perfusion or diffusion MRI. Thereby, MRI yields multiparametric data, *i.e.* several values per pixel or voxel. When diagnosis is required, the radiologist is then left with a potentially large number of information sources but relatively few analysis tools. As regards visual inspection, it is not straightforward to describe or model how a human rater combines information from multiple images when multiple MR data are examined. In addition, as the number and relevance of images increase, the sources of variability, such as inter-operator difference or subjectivity, should be carefully minimized but most current clinical practice lacks quantitative and reproducible assessment (Hectors et al [1], Menze et al [2], Weltens et al [3]).

These limitations are particularly striking in the context of brain tumors, for which MRI is the recommended imaging modality (De Angelis [4], Drevelegas and Papanikolaou [5], Wen et al [6]). Several contrasts are acquired in routine but consensus analysis procedures only propose a limited data exploitation. Most analyses reduce to determining a few global features such as mass diameter, the occurrence of an edema or of contrast enhancement. The heterogeneity of the tumor is not quantitatively assessed but simply used as an indication of which tissue is the most aggressive. This identification can in turn guide a biopsy on which the final diagnosis is usually based. Unfortunately, the limited spatial extent of the biopsy does not guarantee the validity of the subsequent conclusions outside the surged area.

As an alternative for probing the information content of lesions, several attempts have been made to analyze multiparametric MRI data in order to accurately and non invasively characterize the entire tumor. Variability and accuracy of the analyses can be controlled in two main ways, through quantitative feature extraction standardization and through automated region-of-interest (ROI) selection. Most approaches focus on one or the other aspects: segmentation approaches are usually based on a few standard MRI maps, while more advanced feature extraction techniques commit to a preliminary manual ROI delineation, which complicates the analysis of the results due to inter-operator variability. When multivariate features are extracted, it is crucial to account for their interaction as important information may be lost if they are treated independently. It may be that only through the integration of multiple features, highly distinctive and predictive patterns can emerge. Most methods model features using multivariate Gaussian distributions because of their tractability in arbitrary dimensions. Coquery et al [7] propose to analyse 6 MR parameter maps and to identify different tissue types by looking for groups of voxels with similar parameter values. Voxels in manually segmented ROIs are clustered using a 6-dimensional Gaussian mixture model. The number of components (clusters) is chosen according to the Bayesian information criterion. Similarly, Boult et al [8] used manual segmentations followed by a k-means clustering of 3 MRI parameters maps to determine intra-tumoral tissue types whose relevance is again assessed by comparison to histology data. Studies that focus on automated segmentation are generally faced with the issue of automatic lesion classes isolation. Unsupervised segmentation produces a partitioning into several classes with no clear semantic sense. Classes across different segmentations may not always represent the same tissue, complicating its biological inter-

pretation. Typically, Katiyar et al [9] propose an intra-tumoral segmentation technique valid for a single tumor type localized and known in advance. Juan-Albarracín et al [10] augment segmentation with a non pathological tissue detection but requiring the use of an atlas with its lots of registration issues and expert intervention via the use of approximate masks. Moreover, parameters such as the number of tissue classes are arbitrarily fixed and MR features are not quantitative, limiting the study to homogeneous single institution data sets. Regarding features choice, Macyszyn et al [11] address the identification of MRI features that correlate well with patient survival or molecular subtype prediction in glioblastoma. The focus is on extracting prognostic information. Automated segmentations are performed but with a predetermined number of classes whose interpretation is known in advance. The goal is to extract global imaging features rather than to perform refine intra-tumoral analysis. Also, the use of 2-class SVM (Support Vector Machine) classifiers to integrate the features limits the prediction to a small number of different classes, typically short, medium or long survivals and 4 glioblastoma molecular subtypes. Zacharaki et al [12] propose a classification of brain tumors based on extraction of MRI texture and shape features; but their procedure requires ROI delineations which are performed manually. Rasmussen et al [13] also investigate the choice of MR parameters to identify tumor volumes and sub-volumes at best, but they faced the same issue of manual ROI delineation.

These studies among many others highlight the potential of conducting robust analysis from multiparametric data. However, they require a number of tuning operations left to the user. Manual delineations have been already mentioned as an essential step, but most statistical inference procedures also rely on parameters that have to be set in advance such as the number of components in a mixture or the type of mixture distributions. One limitation of the previous studies is that they do not account for dependencies between parameters or use multivariate Gaussian models despite some observed parameter distributions have non Gaussian shapes (Coquery et al [7]). In addition, the diagnosis ability of these approaches is not fully evaluated. Therefore, there is a need for fully automated methods that can analyze multiple MR data in a reproducible way that correlates well with expert analyses.

The approach presented in this paper differs from existing work in one or more of the following ways. First, lesions are not modeled explicitly but detected as abnormal voxels with respect to a statistical model of MR characteristics of normal tissues. Second, the method is fully automated, it performs automatic ROI localization and number of clusters selection. Third, it goes beyond standard Gaussian models for increased robustness to outliers. The method uses generalized Student distributions (Forbes and Wraith [14]) that allow a richer variety of shapes in the multi-dimensional space of the MR parameters and in particular contours that are not necessarily elliptical.

More specifically, in the following developments, a generic and automatic data-driven procedure is described and its performance to segment and diagnose brain tumors without any prior information or spatial knowledge is illustrated. The proposed procedure requires the availability of learning data made of two sets of voxels: one from healthy subjects and one from pathological subjects for which the pathology (*e.g.* tumor) type is known. The procedure consists then of five steps: i) a first mixture model is fitted to the healthy subjects voxels; ii) this reference model is used to detect abnormal voxels in the healthy and pathological subjects; iii) a second mixture model 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 can be carried out to remove some spatial artifacts and refine the pathology signatures.

The whole procedure is described in Section II and summarized in Figure 1. In Section III, its performance is illustrated on an independent evaluation data set.

II. Proposed automatic and data-driven procedure

A FIRST goal is to automatically localize voxels which exhibit abnormal MR features with respect to a learned reference model. The considered features come from multiparametric quantitative MRI data that provides in each voxel a vector with several physiological measures acquired or computed during the MRI session. Quantitative MR measures are considered, in particular, to be as independent as possible of the MRI scanner or the study center (Tofts [15]).

A. Reference model

Starting from a set of reference, typically healthy subjects, the goal is to construct a statistical parametric model of the MR parameters associated to these subjects. Each reference subject is associated to a number M of co-registered MR parameter maps that provide for each voxel v a M-dimensional vector of parameters denoted by y_v. All voxels from all subjects are gathered into a single set of voxels denoted by $\mathcal{V}_H$. The considered data set of M-dimensional vectors, pooling all vectors together, is denoted by $Y_H = \{y_v, v \in \mathcal{V}_H\}$. To characterize the distribution of these MR parameters, we consider a multivariate mixture model to account for the potential heterogeneity in the parameter values due to the presence of different tissue types. This corresponds to cluster the data Y_H into a number of groups (clusters) of similar parameter vectors and to model each group with a parametric distribution. In practice, in each dimension, the data are standardized to avoid scaling effects between MR parameters.

In Coquery et al [7], a Gaussian mixture model is used assuming that each group is distributed according to a Gaussian distribution. However, the observed physiological parameters do not necessarily exhibit a Gaussian shape. Also Gaussian distributions are known to be sensitive to outliers whose occurrence may severely bias the estimation. As a more robust alternative, heavy tail distributions have the ability to accommodate potential outliers. In this paper, we consider such distributions and in particular multiple scale t-distributions (MST) introduced in Forbes and Wraith [14] that generalize the standard Student t-distribution. MST distributions allow the assignment of outlying parameter values to clusters without

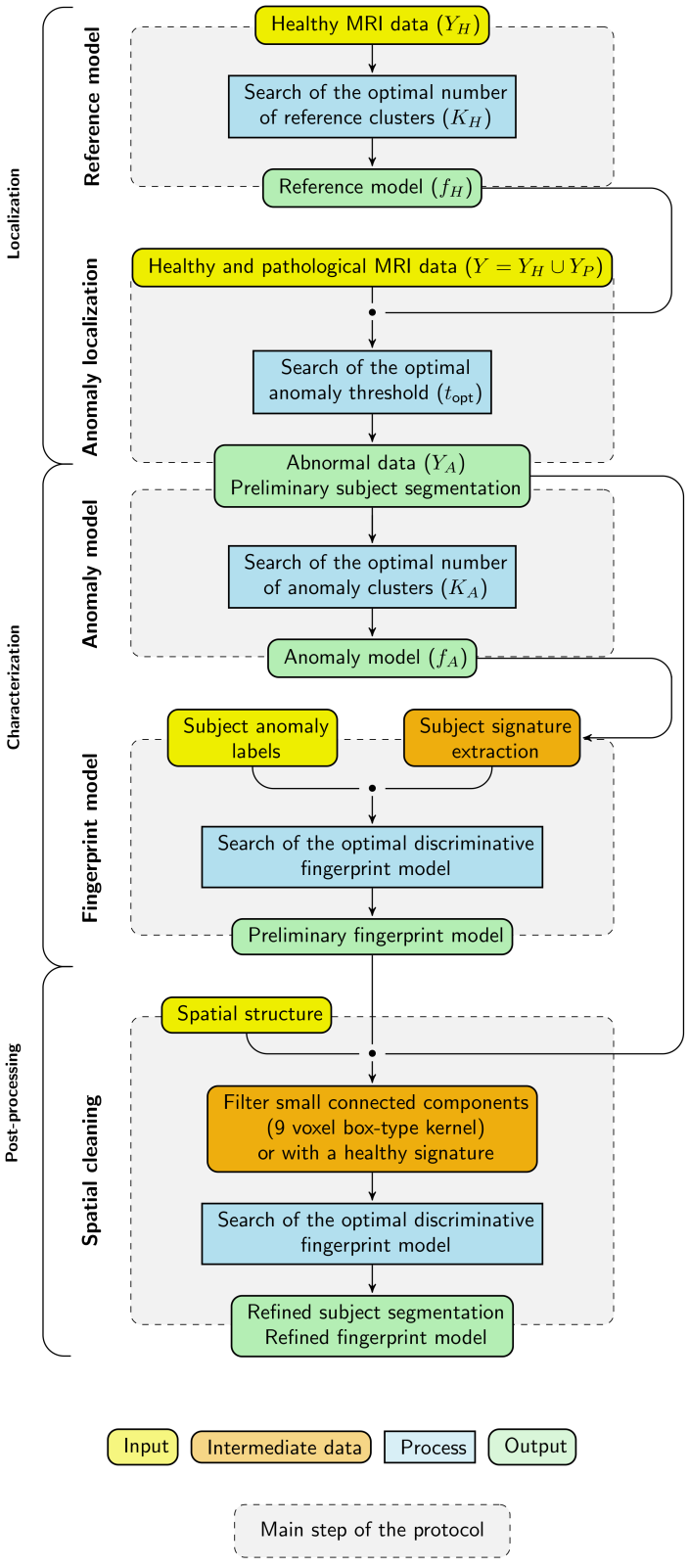


Fig. 1. Construction of a model to automatically localize and characterize lesions. Starting from subjects labeled as healthy or pathological, the procedure is made of 5 main steps.

degrading the location and scale of the clusters. One advantage of the MST distribution over the standard t-distribution is a varying amount of tail weight in each dimension resulting in a much greater variety of distributional shapes. The mixture of MST distributions (MMST) that best fits the observed data Y_H is estimated using an Expectation-Maximization (EM) algorithm as described in Forbes and Wraith [14]. This requires to set the number of groups in the mixture. This number, denoted by K_H, is chosen automatically from the data following a model selection approach. More specifically, we use the so-called slope heuristic (Baudry et al [16]) that generally provides a clear selection (see supplementary materials for more details). The clustering EM algorithm results in K_H clusters, each represented by a multivariate M-dimensional MST distribution. Each distribution summarizes the MR features of a group of voxels and can be seen as a *word* in a *dictionary* of healthy or reference clusters. A by-product of the EM algorithm is that each voxel has a probability to be assigned to each cluster. The set of these membership probabilities can be seen as the signature of the voxel in the coding provided by the dictionary, the whole dictionary being itself summarized by the probability distribution function (pdf) of the estimated MMST model, namely

$$f_H(y) \quad = \quad \sum_{k=1}^{K_H} \pi_k \, \mathcal{MST}(y; \psi_k) \qquad (1)$$

where $\mathcal{MST}(y; \psi_k)$ denotes the MST pdf with parameter ψ_k and proportion π_k learned from Y_H. The pdf in (1) is referred to as the reference model and will be used in the following anomaly detection step.

B. Anomaly localization

Considering a set of pathological subjects, the goal is to identify in each of them the lesioned voxels in order to provide a delineation of potential lesions. As lesions generally exhibit MR parameter values different from healthy tissues, lesion localization and delineation is recast into an anomaly detection task. For a precise localization, mere visual inspection may not be enough and may be tedious when the number of available MR parameters increases. The idea is to use instead the previously built reference model (1) to seek anomaly at the voxel level and in an automated way. The task is to identify voxels whose vectors of MR parameters cannot be well statistically explained by the reference model f_H. As before, the voxels of all pathological subjects are gathered into a set $\mathcal{V}_P$ of voxels and the corresponding set of MR vectors is denoted by $Y_P = \{y_v, v \in \mathcal{V}_P\}$. To deal with comparable values, the normalization applied to the reference observations Y_H is also applied to Y_P. A voxel v is considered as abnormal with regards to the reference model if it corresponds to parameter values y_v which have a low likelihood in the reference model. This likelihood is assessed via the reference pdf value at y_v, *i.e.* $f_H(y_v)$. Since pdf values cannot be interpreted as probabilities, the following key step is to decide on a threshold t_{opt} below which voxels will be declared as outliers, *i.e.* when $f_H(y_v) \leq t_{opt}$. This threshold can be fixed to control the false positive error, *i.e.* when Y_v is a random variable distributed according to f_H (healthy tissue), we seek for t_{opt} so that $p(f_H(Y_v) < t_{opt}) = \alpha$ with a small value of α. However, the α value generally chosen (5%) is arbitrary and is likely not to coincide with lesions present in the data set. A threshold specific to the data under consideration is then preferable and

can be computed as follows. Likelihood scores are computed for all voxels in $\mathcal{V}_P$, *i.e.* $f_H(y_v)$, for all $v \in \mathcal{V}_P$, but also for the voxels that were used to construct f_H, *i.e.* $f_H(y_v)$ for all $v \in \mathcal{V}_H$. Intuitively, among the obtained scores, high $f_H(y_v)$ scores corresponding to parameters close to the reference model should separate from the others. To fix the separation in a data driven way, we fit a MMST model to the log-score data set $\{\log f_H(y_v), v \in \mathcal{V}_H \cup \mathcal{V}_P\}$. The slope heuristic is used to set the number L of components in the mixture. Due to the heterogeneity of lesions, this number is generally greater than 2 indicating the presence of different abnormal tissues with different anomaly levels in the pathological data. Clusters can then be ordered according to their respective mean log-score. The lowest mean corresponds to a group whose departure from the reference model is the highest, while the highest mean should correspond to healthy voxels. The voxels are then partitioned into L groups of successive anomaly levels. Anomaly thresholds are set to the highest likelihood score in each group and denoted by $\{t_1, \ldots, t_L\}$. They are used to provide nested anomaly segmentations that reflect the structure of the lesions (see Figure 4 for an illustration). For a global lesion segmentation, another MMST model is fitted to the log-scores with the number of groups set to 2. The two groups are ordered according to their means. For consistency with the previously computed series of thresholds, we retain as the global threshold t_{opt} the value in the series $\{t_1, \ldots, t_L\}$ which is the closest to the highest score in the first group of the 2-component mixture. The thresholds that correspond to abnormality levels are the ones lower or equal to t_{opt}, that is $\{t_1, \ldots, t_{\text{opt}}\}$. They are associated to colored voxels in Figure 4. For each of them, we compute the false positive probability $\alpha_l = p(f_H(Y_v) < t_l)$ when Y_v is distributed according to f_H. Unfortunately the distribution of $f_H(Y_v)$ as a random variable is usually not known and the α_l's need to be computed empirically using simulations. However, the thresholds t_l correspond in general to extreme quantiles so that standard empirical estimation would lead to $\alpha_l = 0$ in most cases. For a more precise estimation, we propose to use results from extreme value theory that enables a more accurate modelling of the distribution tail (see *e.g.* [17]). Details are given in the Appendix.

At this stage, the successive thresholds can be used to provide lesion delineations into different abnormality levels (see Figure 4) indicative of different tissue types. The global threshold t_{opt} can also be used to separate the input data $Y = Y_H \cup Y_P$ into a set of abnormal values Y_A and the rest:

$$Y_A = \{y_v,\ v \in \mathcal{V}_H \cup \mathcal{V}_P \quad s.t. \quad f_H(y_v) < t_{opt}\}\ .$$

Note that, as the proposed thresholds are quantiles of the reference model pdf, healthy subjects may also exhibit a small fraction of abnormal voxels. Some of them are isolated voxels and can be easily removed using simple morphological operators (see Section II-E). However, these isolated voxels tend to be present in all subjects, for similar reasons (noise, skull stripping, artifacts, etc.), and their removal does not significantly affect the discriminative power of the fingerprint model. In contrast, other voxels may correspond to normal regions or structures (*e.g.* vessels, ventricles) whose physiological characteristics and then MR parameters are close to that of lesioned tissues. Individually, these voxels are therefore correctly detected as deviant from the reference model. However, their global signature is expected to be different from the signature of lesioned tissues and will then be learned by the fingerprint model.

C. Anomaly model

The goal of the two previous steps was essentially to perform automatic ROI localizations. The obtained ROIs provide a set of MR parameter vectors that are referred to as the abnormal data set Y_A. An anomaly model is then constructed following the same procedure as for the reference model (Section II-A). The observations in Y_A are standardized in each dimension and then used to fit a MMST model with a number K_A of clusters selected with the slope heuristic. The fitted mixture is denoted by f_A,

$$f_A(y) \quad = \quad \sum_{k=1}^{K_A} \eta_k\ \mathcal{MST}(y; \phi_k) \tag{2}$$

where $\mathcal{MST}(y; \phi_k)$ denotes the MST pdf with parameter ϕ_k and proportion η_k learned from Y_A. This anomaly model is used in the next section to extract anomaly features from MR maps and construct a signature for each subject under consideration.

D. Fingerprint model

By fingerprint model we mean a model that can correctly characterize and classify a subject into one of a number of classes (*e.g.* different tumor types), based on MR parameters maps. Such a model is built in a supervised manner from pairs associating some chosen features to a class label. This requires the availability of a number of subjects for which the class label is known. The extracted features have then to be as informative as possible so as to allow the correct classification of unlabeled subjects. For each available subject in the learning data set, its anomaly class or label is known. These classes typically include the *healthy* label and different tumor types. For each subject S, features are extracted from a set $\mathcal{V}_S$ of n_S voxels corresponding to the voxels of S detected as abnormal in the previous step (Section II-B). As mentioned in Section II-B, healthy subjects also exhibit a small fraction of abnormal voxels. For each voxel $v \in \mathcal{V}_S$, the anomaly model (Section II-C) provides a probability ρ_k^v that voxel v belongs to cluster k among K_A clusters,

$$\rho_k^v = \frac{\eta_k\ \mathcal{MST}(y_v; \phi_k)}{\sum_{l=1}^{K_A} \eta_l\ \mathcal{MST}(y_v; \phi_l)}.$$

For features at the subject level, we compute for each $k = 1 :\ K_A$, the mean probability over voxels in $\mathcal{V}_S$,

$$\rho_K^S = \frac{\sum_{v \in \mathcal{V}_S} \rho_k^v}{n_S}.$$

The retained feature vector for subject S is then

$$\rho^S \quad = \quad \{\rho_1^S, \ldots, \rho_{K_A-1}^S, n_S\} \tag{3}$$

where the last probability has been removed and replaced with the ROI size n_S to avoid co-linearity. Such a vector captures

the expression level of each anomaly cluster in the ROI of subject S. Intuitively, it is expected to capture the proportions of the different tissue types in the ROI.

As for the supervised learning part, we adjust different discriminant analysis models and compare their ability to correctly predict the label (*e.g.* lesion type) of each subject by a leave-one-out cross-validation procedure. The selected discriminant analysis model is the one providing the highest true positive rate. Further details are given in the application Section III-D.

E. Post-processing

The segmentations obtained in Section II-B using anomaly detection can be further refined by removing connected components which are too small. This is done by applying an erosion-dilatation operator with a 9 voxels box-type kernel on each slice. Note that this could have been done previous to the fingerprint model learning but would not have impacted its discriminative power as already mentioned in Section II-B. In contrast, the fingerprint model can improve the segmentations thanks to its ability to recognize healthy tissue. After erosion-dilatation, the remaining connected components are considered as individual sub-ROIs for which the labels are unknown. Then if a sub-ROI is classified as healthy by the fingerprint model, it is removed from the initial segmentation. This enables the removal of groups of voxels that may individually exhibit MR parameter values close to lesioned tissues but show group anomaly proportions (signature) close to healthy components. For each subject, the remaining sub-ROIs are pulled together and form the refined individual segmentations. The refined segmentations are in turn used to refine the fingerprint model. Going back to the procedure described in Section II-D, new subject signatures are extracted from the refined ROIs and given as input for the estimation of a new fingerprint model referred to below as the refined fingerprint model. After this stage, healthy tissues should correspond to a null signature (*i.e.* with no abnormal pixels) while the lesioned tissues signatures are cleaned from healthy tissues. An illustration is given in Figure 7.

F. Characterization and prediction

To predict the label of a new subject, its ROI is first determined using f_H the reference model (Section II-A) for anomaly detection (Section II-B). The obtained ROI is cleaned from potentially remaining healthy connected components that have not been removed by erosion-dilatation. The initial fingerprint model is used to identify these healthy connected components. The cleaned ROI (see Figure 5) is used to extract a ρ^S signature (see Figure 7) using the anomaly model (Section II-C) as explained in Section II-D. This signature is then given to the refined fingerprint model which provides an associated label (Table III).

III. Application to real multiparametric MRI data from rats with brain tumors

OUR procedure is illustrated on a data set of 53 rats for which 5 quantitative MRI maps are available. Some of the rats were implanted with different tumor types. The study design was approved by the local institutional animal care and use committee (COMETHS). All animal procedures conformed to French government guidelines and were performed under permit 380820 and B3851610008 (for experimental and animal care facilities) from the French Ministry of Agriculture (Articles R214-117 to R214-127 published on 7 February 2013). This study is in compliance with the ARRIVE guidelines (Animal Research: Reporting in Vivo Experiments [18]).

TABLE I
AVAILABLE MR DATA: A LEARNING SET ($Y = Y_H \cup Y_P$) MADE OF HEALTHY (Y_H) AND PATHOLOGICAL (Y_P) MR VALUES IS USED TO LEARNED A FINGERPRINT MODEL. ANOTHER TEST SET (Y^T) INCLUDING BOTH HEALTHY (Y_H^T) AND PATHOLOGICAL (Y_P^T) MR VALUES IS USED FOR VALIDATION. THE SETS OF DETECTED ABNORMAL VOXELS ARE ALSO INDICATED IN BOTH CASES (Y_A AND Y_A^T).

Learning set	voxels	dimension	subjects
$Y = Y_H \cup Y_P$	260405	5	32
Y_H	45051	5	6
Y_P	215354	5	26
Y_A	57547	5	32

Test set	voxels	dimension	subjects
$Y^T = Y_H^T \cup Y_P^T$	150085	5	21
Y_H^T	71340	5	11
Y_P^T	78745	5	10
Y_A^T	20105	5	21

A. Description of the MRI data

a) Rats and tumor types (Table I): The healthy subject group (Y_H) is composed of 6 healthy Fisher rats. The pathological subject group (Y_P) contains 26 subjects with 4 tumor types: 9L (6 Fisher rats), C6 (6 Wistar rats), F98 (7 Fisher rats) and RG2 (7 Fisher rats). The MR parameter maps of all these subjects form the data set Y. For evaluation purpose, another group of subjects is available and contains 5 rats with 9L tumor, 5 rats with F98 tumor, and 11 healthy Fisher rats. The MR parameter maps associated to these subjects are kept as a test set and denoted by Y^T distinguishing the healthy sub-group Y_H^T from the pathological one Y_P^T.

b) MRI parameters (Figure 2): The following 5 quantitative MR maps were acquired on 5 contiguous slices: apparent diffusion coefficient (ADC), T1, T2, cerebral blood volume (CBV), and a vessel permeability map called area under the curve (AUC). All measures are naturally co-localized: all maps were acquired with the same geometry so that each voxel is described by the 5 parameters above. Section III-A in the supplementary materials provides further details about the MRI session. In addition, an anatomical T2-weighted image was acquired to allow automatic skull-striping. A manual delineation (superimposed red line in Figure 2 first row) of the tumor was performed using the anatomical image and the diffusion map. This manual segmentation is used as ground truth for the evaluation of the automatic tumor localization proposed in this study.

B. A healthy subjects based reference model

The reference model f_H is built as a MMST model using an EM algorithm on the healthy subjects maps Y_H. When defining this reference model, the slope heuristic approach selects $K_H = 10$ clusters. An example of this partitioning in $K_H = 10$ clusters is given in Figure 3 for the 5 slices of

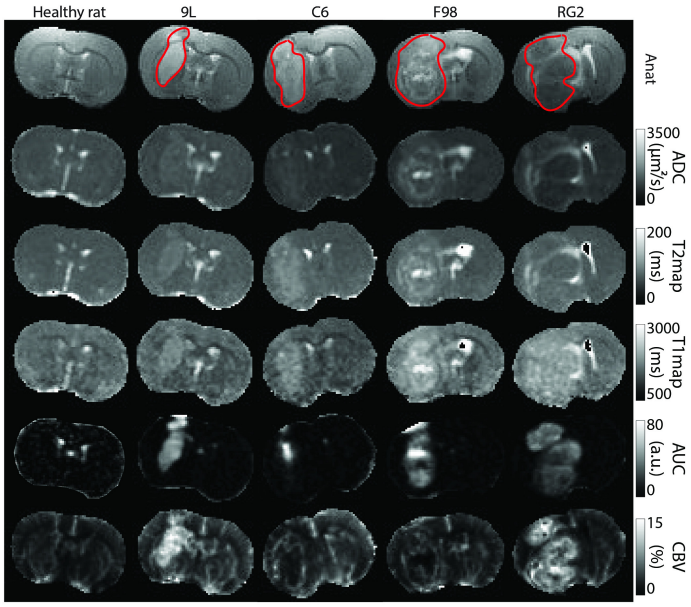


Fig. 2. MRI maps -ADC, T2, T1, AUC, CBV- associated to the central slice for one rat of each group (from the data set Y). The red line superimposed on the anatomical map corresponds to the manual segmentation.

a healthy rat. In the whole set of healthy rats, 4 main clusters (red -1-, orange -2-, yellow -3-, light green -4-) gather 79.5% of the voxels in Y_H. The 6 remaining clusters (green -5-, turquoise -6-, blue -7-, dark blue -8-, purple -9-, light purple -10-) correspond to less represented features such as vessels and ventricles (clusters 9 and 10) or interfaces between the main tissue types (interface between gray matter and cerebro-spinal fluid for clusters 6 and 7). Interestingly, although the model is adjusted without any spatial regularization, the resulting segmentations present spatially homogeneous regions and are rather visually consistent with brain anatomy: clusters 1 and 4 for the cortex and the corpus callosum, clusters 1 to 3 for the striatum, clusters 7 to 10 for the ventricles.

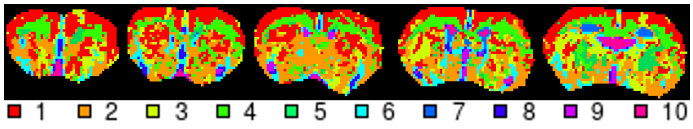


Fig. 3. Reference model clustering (with $K_H = 10$ clusters) on the 5 slices of one healthy rat. From left to right: slices in decreasing order along the vertical axis of the scanner.

C. Anomaly localization

As described in Section II-B, for each MR parameter vector y_v in Y, the pdf $f_H(y_v)$ is evaluated. We found that these values are best partitioned into 7 groups, according to the slope heuristic. Within each group, the voxels have similar likelihoods in the reference model, that is similar degrees of abnormality with respect to the reference model. This partition leads to 7 anomaly thresholds $\{t_1, \ldots, t_7\}$. The optimal threshold that best divides the voxels into 2 subsets, according to a 2-component mixture model (Section II-B), is the 4th threshold $t_{\text{opt}} = t_4$. The abnormal subset $Y_A = \{y_v, f_H(y_v) \leq t_{\text{opt}}\}$ represents 22.1% of the full data set Y with a false positive rate of 1.4%, *i.e.* $p(f_H(Y_v) \leq t_{\text{opt}}) = 0.014$ when the random variable Y_v is distributed according to the reference model. The voxels in Y_A form the subject ROIs.

An illustration is given in Figure 4-A last row. Colored voxels (thresholds t_1 -red-, t_2 -orange-, t_3 -yellow- and t_4 -green-) belong to Y_A while grey ones (thresholds t_5 -light grey-, t_6 -medium grey- and t_7 -dark grey-) are not tagged as abnormal. The concordance with manual segmentation (superimposed red line in Figure 4-A) is indicated in Figure 4-B right via the computation of the Adjusted Rand Index (ARI) for each threshold. The higher the ARI the better, the maximum value being 1 (Rand [19], Hubert and Arabie [20]). For a given threshold t_l, MR parameters y_v in Y such that $f_H(y_v) \leq t_l$ are selected. The corresponding voxels form the segmentations linked to threshold t_l, on which the ARI is computed. More specifically, the ARI is computed at level t_l for each rat. The boxplots in Figure 4-B show the variations of these ARI's independently of the tumor type. Other scores such as the DICE [21] are shown in Supplementary-Table III. Regarding global tumor delineation, it appears that the selected 4th threshold in the MMST case, corresponds to the best all tumor average ARI (0.42). ARI values can also be averaged for rats with the same tumor type. It appears then that some tumors are easier to segment. For instance, in the MMST case, for threshold t_4, the average ARI's are respectively of 0.49, 0.40, 0.32, and 0.47 for 9L, C6, F98 and RG2 tumors. The corresponding boxplots are shown in Supplementary Figure 5a. Inside the tumors, the thresholds also yield some satisfying spatial coherence. The strongest abnormality (red -t_1- and orange -t_2- areas in Figure 4) is mainly located at the center of the tumor area. The delineated regions around (yellow -t_3- and green -t_4-) match with the border of the tumor area, and the highest thresholds ($\{t_5, t_6, t_7\}$ -gray levels-) are mainly for the healthy voxels. However, it also appears that some voxels are tagged as abnormal in healthy subjects (Figure 4-A 1st column) or in the contralateral part of pathological subjects (Figure 4-A 4 last columns). Based on their anatomical location, these voxels mainly correspond to ventricles and possibly blood vessels. As mentioned in Section II-E, most of these wrongly tagged voxels will be easily removed in Section III-D by using their signature after morphological operations. In terms of segmentation for the pathological rats, if we except the contralateral voxels, the 9L rat shows a high concordance with the manual segmentation, while the F98 segmentation is smaller and the RG2 segmentation is bigger. This is consistent with the intrinsic difficulty of delineating tumors of varying visual appearance across MR maps. As a matter of fact, on anatomical (Figure 4, 1st row) and diffusion images, 9L tumors are easier to delineate manually while F98 and RG2 tumors are more diffuse.

For comparison, our protocol is also applied with Gaussian mixture (GM) models. The anomaly localization results are shown in Figures 4-A, second row and B, left. The MMST model provides finer intra-tumoral descriptions with 4 abnormal classes instead of 3 in the Gaussian case, and smoother segmentations. This is quantitatively confirmed by a MMST ARI of 0.42 which is 10.5% higher than the Gaussian ARI of 0.38. If healthy parts are considered instead, MMST and GM ARI are in the same order. Similar conclusions hold for the DICE values given in Supplementary-Table III. Regarding lesion segmentation, the main difference is in the contralateral

areas. With GM models, the contralateral areas detected as abnormal are larger, and more of them are connected to the lesion areas, which leads to a less effective spatial post-processing. Indeed, only connected components not connected to the lesion area can be removed, because in case of contact the ROI signature does not correspond to the healthy one. As shown in Figure 5 for 9L and F98 tumors, the GM case requires more user interpretation than the MMST model to differentiate the lesion from the contralateral area.

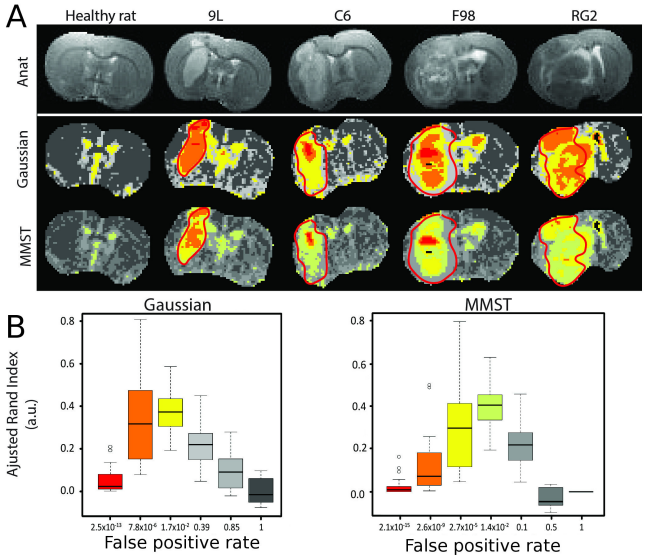


Fig. 4. Nested anomaly segmentations for the different anomaly thresholds (part A), with the associated false positive rates and respective adjusted rand indices (part B) for the overlap with the manual segmentation. The thresholds are ordered from lowest (most abnormal in red) to highest (less abnormal in dark gray). The colored ones are those defining the abnormal data set Y_A. The gray ones correspond to normal groups. The ARI value for a threshold is obtained by concatenating the segmentations associated to the thresholds lower to this threshold. The results are given for both Gaussian mixture and MMST models.

D. Anomaly and fingerprint models

A MMST model f_A is adjusted on the abnormal Y_A data set with a number of clusters $K_A = 10$ selected using the slope heuristic. For each rat S in the learning set, we extract a signature ρ^S as explained in Section II-D representing the proportions of each cluster in the ROI. The signatures are associated to the known subject labels in order to build a discriminative fingerprint model. Three discriminant analysis are compared based on their true positive rate with a leave-one-out procedure: linear discriminant analysis (lda: 90.6%, R package MASS, Venables and Ripley [22]), high dimensional discriminant analysis (hdda: 93.7%, R package HDclassif, Bergé et al [23]), and a discriminant analysis based on a Gaussian finite mixture modeling (mclustda: 90.6%, R package mclust, Fraley et al [24]). The hdda analysis is retained due to its higher score to built a first fingerprint model. This fingerprint model is used in turn to refine the segmentations made in Section III-C. Too small connected components are removed with a 9 voxel box-type kernel and the fingerprint model is used to remove other connected components classified as healthy. The obtained cleaned segmentations are illustrated for some rats in Figure 5. The ARI values for the 26 pathological rats in the training set are summarized in Figure 6, while DICE values are shown in Table II. A signature for each subject is then computed using the cleaned ROI possibly still made of several connected components. The signatures are shown in Figure 7 for the 4 pathological groups and the healthy group. As a result of the post-processing, the healthy subjects present empty signatures. For the other rats, some clusters seem to be specific to some tumor types: cluster 6 (turquoise) for the 9L tumor, clusters 1 (red) and 4 (light green) for the RG2 tumor. The cluster 2 (orange), 3 (yellow) and 5 (green) are mainly present in the signature of the C6 and F98 tumors, and sightly differ in terms of proportions.

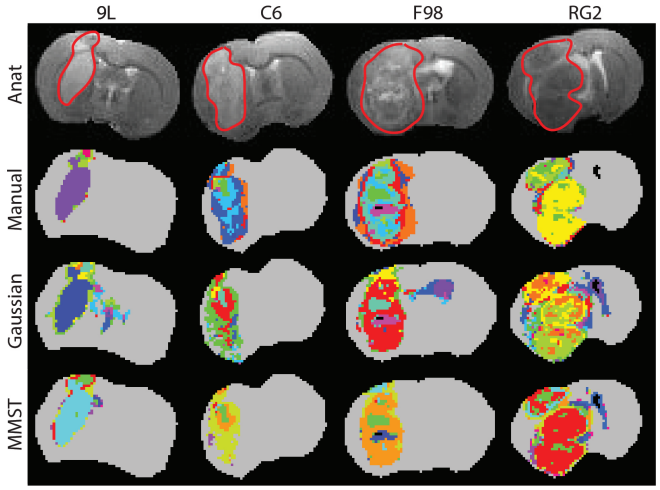
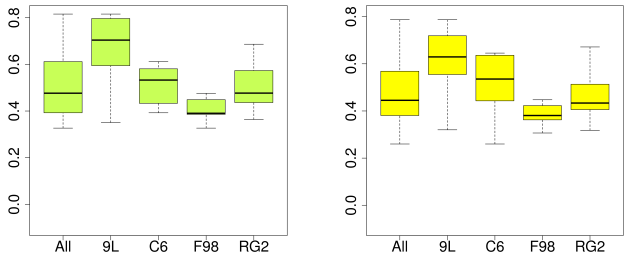


Fig. 5. Intra-tumoral segmentations for some of the pathological rats in the training set. First two rows: manual delineations and Gaussian clustering ($K_A = 13$) as described in [7]. Last two rows: Automatic segmentations and clustering using $K_A = 13$ and $K_A = 10$ with a Gaussian (3rd row) and MST (4th row) mixture model. The ROIs correspond to the refined segmentation (i.e. after spatial post-processing).



(a) MMST (b) Gaussian mixture

Fig. 6. Pathological rats in the training set: Adjusted rand index values per tumor type for the refined segmentations (i.e. after spatial post-processing), using the MMST (a) or Gaussian mixture (b) models.

E. Validation on a test data set

To evaluate the potential of our procedure, a test set different from the learning data set is used. It is composed of 9L rats ($n = 5$), F98 rats ($n = 5$), and healthy rats ($n = 11$). All voxels of these rats are gathered into data set Y^T. After normalization with the normalization values used for the reference model, the pdf $f_H(y_v)$ for each voxel v is computed and values lower than t_{opt} define the abnormal voxels and subset Y_A^T of Y^T. Each vector of parameters in Y_A^T is normalized using the normalization values computed for

TABLE II
PATHOLOGICAL RATS IN THE TRAINING SET: MEAN COVERING SCORES FOR LESION SEGMENTATIONS AFTER POST-PROCESSING USING GAUSSIAN MIXTURE AND MMST MODELS.

	Gaussian mixture model					MMST model				
	9L	C6	F98	RG2	All	9L	C6	F98	RG2	All
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

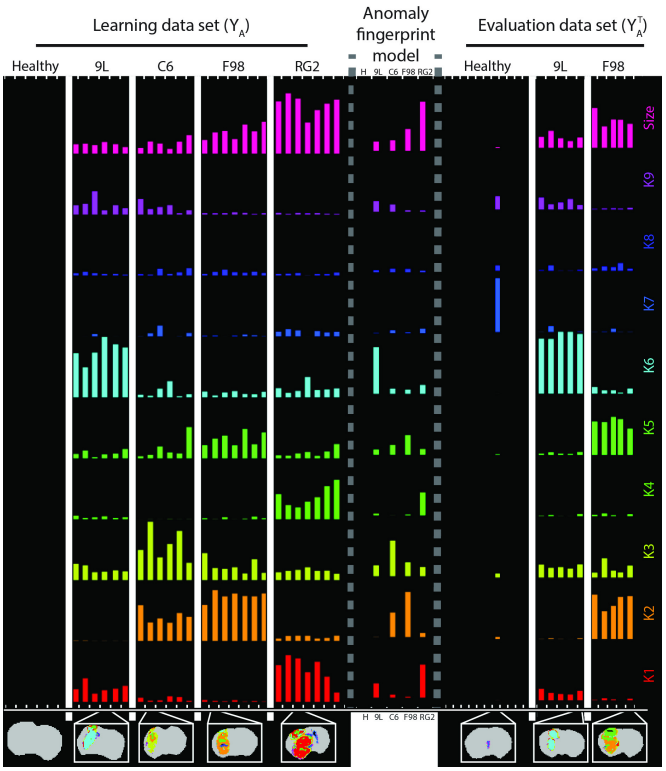


Fig. 7. Refined anomaly signatures by subject. Left: individual subject refined signatures associated to the learning data set Y and grouped by tumor type. Center: mean signatures for each tumor type as provided by the refined fingerprint model. Right: individual refined signatures for each subject in the evaluation data set Y^T. For each tumor type (9L, C6, F98, RG2) and healthy animals (H), an illustration of the segmentation is presented at the bottom.

the anomaly model f_A. Individual subject signatures ρ^S are then first extracted as described in Section II-D, eq. (3), for each rat in the test set. Spatial post-processing is then applied as explained in Section II-E to produce individual refined signatures on which are based the final predictions with the refined fingerprint model. All 9L rats are correctly predicted, but one F98 rat and one healthy rat are misclassified (Table III). The misclassified healthy rat, visible in Figure 7-right part, has only a few number of voxels tagged as abnormal (size, first line). Small differences on these few voxels may lead to very different signatures potentially closer to pathological ones. The misclassification of the F98 rat as a C6 rat may be explained by the C6 and F98 fingerprint models similarity: they share the same clusters but with different proportions. The Gaussian model faces the same difficulties to differentiate the C6 and F98 rats, with 3 miss-classifications. In contrast, all the healthy rats are well recovered (Table III). When manual segmentations are used like in [7], the confusion between F98 and C6 tumors is worse with all F98 rats classified as C6 rats.

The refined segmentations for the evaluation data set test are presented in the Supplementary-Figure 13, with the associated values of ARI (Supplementary-Figure 14a) and DICE (Supplementary-Figure 14b). The average ARI and DICE scores are respectively 0.52 and 0.72.

TABLE III
TUMOR TYPE PREDICTION USING THE REFINED FINGERPRINT MODEL CONSTRUCTED WITH THE TRAINING DATA SET (Y) AND EVALUATED ON A INDEPENDENT TEST DATA SET (Y_T) FOR THE GAUSSIAN MIXTURE (GM) AND MMST MODELS BASED ON MANUAL OR AUTOMATIC SEGMENTATIONS.

Manual segmentation & Gaussian mixture model

	9L	C6	F98	RG2	Healthy
Blind 9L (n=5)	5	.	.	.	.
Blind F98 (n=5)	.	5	.	.	.

Automatic segmentation & Gaussian mixture model

	9L	C6	F98	RG2	Healthy
Blind 9L (n=5)	5	.	.	.	.
Blind F98 (n=5)	.	3	2	.	.
Blind Healthy (n=11)	.	.	.	.	11

Automatic segmentation & MMST model

	9L	C6	F98	RG2	Healthy
Blind 9L (n=5)	5	.	.	.	.
Blind F98 (n=5)	.	1	4	.	.
Blind Healthy (n=11)	.	1	.	.	10

IV. DISCUSSION

IN this study, a modular, fully automatic and data-driven procedure to detect and characterize abnormality within medical images was proposed. This procedure was tested on MRI data collected in rats bearing a brain tumor (4 tumor types). The lesions were localized within the brain as anomaly with respect to a reference probabilistic mixture model built using a dataset collected on healthy animals. Abnormal voxels were then clustered in groups with similar MR parameter values. The proportions of each group in each animal were used to construct a signature of each animal. For each tumor type, the signatures for the animals with this tumor type were used to build a fingerprint model of each tumor type. The specificity and sensitivity of the obtained fingerprints were eventually illustrated on a diagnostic task performed successfully without user interaction on an additional test data set. This first application of a procedure whose purpose is more general (any lesion visible by a radiologist, any type of clinical imaging modalities) relied on data from 4 tumor types and characterized by 5 quantitative parameter maps each. Six healthy rats and 26 pathological rats were used to learn the reference and pathology models and 21 additional rats were used for testing. While this dataset is large compared to other preclinical studies (Coquery et al [7], Boult et al [8], Katiyar et al [9]) and was sufficient to prove our concept (predictive rate greater than 90% in this particular application),

it remains small compare to the volume of data collected daily in patients. Further evaluations are thus required on larger, pre-clinical and clinical data sets to confirm the robustness of the proposed method. One technical point of interest is that MST distributions performed better than Gaussian distributions: they yielded a better spatial agreement with the manual delineation (6.4% higher ARI on the learning set) and a better predictive rate (5.6% higher on the test set). While the quantitative MRI dataset obtained for tumor models and used in this study is probably not the most challenging one to compare the performance of the two distributions, the MST appears promising for its greater ability to accommodate outliers while maintaining a good separation between clusters (Forbes and Wraith [14]). The main evaluation criterion of our procedure was the final diagnostic. Further evaluations of the anomaly detection and of the anomaly model would also be useful to refine the procedure, including the number and type of MR parameter maps used to perform the automated diagnosis. To improve robustness, the results of the intermediate steps could be compared to that of histology. For our specific application (tumors), one could thereby evaluate the ability of MRI to detect abnormal cell densities and differentiate tissue types based on the standard pathologist diagnostic (Louis et al [25]). In addition to improving the quality of the procedure, the results of the intermediate steps may be of interest *per se*. The anomaly detection step, which progressively separates the tumor from the healthy tissue, could help neurosurgeons in planning a tumor resection (Barone, Lawrie and Hart [26]). The anomaly model, which discriminates tissue types within the tumor, could be of interest to pathologists who cannot evaluate functional parameters such as blood flow. Interestingly, the cluster maps show an excellent spatial coherence although no spatial information was used during the procedure: all voxels were considered independent from each other. Moreover, at the lesion detection step, the ARI score reached 0.52 and the DICE 0.72 on the learning set, and respectively 0.49 and 0.69 on the test set (after post-processing in both cases). These scores were obtained from the comparison between a manual delineation performed on two images (anatomical and diffusion-weighted images as prescribed in standard clinical evaluation) and our procedure which used 5 parameter maps to delineate the tumor. As the images used to perform the delineations differed, it was not surprising to find different lesions between the automatic and the manual delineations. Histology could be used as a more reliable ground truth of high cell proliferation areas than manual delineation (De Angelis [4]). As in a clinical setting, it is generally not possible to obtain histological ground truth data on the entire tumor, this validation step has to be performed at a preclinical level. The procedure proposed in this study is limited to the detection of tissues whose signal intensity (or parameter values) differs from that of normal tissue. However, a lesion may also appear as a tissue structure with a different volume (*e.g.* Alzheimer patients exhibit a reduction in the cortical thickness compared to age-matched, healthy, subjects [27]). While a large change in structure volume might be detected with our procedure as a change in the proportions of clusters in normal tissue, it would be of interest to add a spatial detection module based on a priori knowledge (*e.g.* atlas [28]). Moreover, a spatial regularization criterion could be added to the proposed procedure to exploit the fact that spatially close pixels have a higher probability to belong to the same tissue type [9], [10]. Exploiting complementarity between spatial proximity and parametric proximity should strongly reinforce our procedure. Once each subject of the training dataset has been labeled (healthy/pathology, lesion type) and the type of MR parameter maps chosen, the proposed procedure requires no human intervention. The determination of the optimal number of clusters, the anomaly threshold, and the final fingerprint model are data-driven to maximize the contrast between the reference data and the lesions and to best discriminate the tumor fingerprints. In this respect, the proposed procedure differs from that in [7] and is a first attempt to combine both detection and characterization in an all-in-one procedure. It stands as an alternative to texture analysis in the context of radiomics [29]. In the context of multicentre studies, the mixture models at the heart of the procedure could be trained and controlled per center, thereby accounting for inter-center variability. Finally, as each procedure step is based on a statistical model, quality control tests may readily be introduced (*e.g.* data homogeneity, outliers) to obtain robust training datasets or to check the data set quality prior to performing an automated diagnostic. The proposed extensions would help meet the challenge of a human application in which the volume of data and number of tissue types represent an exciting challenge for the proposed computer-aided diagnosis procedure.

APPENDIX

Extreme quantile estimation. Details are given about the computation (Section II-B) of $\alpha = p(f_H(Y) \leq t)$ when Y follows distribution f_H and t is a lower tail quantile. Although f_H is available in close form, the pdf of $f_H(Y)$ is not easy to manipulate. Theoretical quantiles of $f_H(Y)$ are not available in general but empirical estimates are easy to obtain via simulation of *i.i.d.* realizations of $f_H(Y)$. However, the t-values of interest are in general extreme quantiles and very few simulated values of $f_H(Y)$ will be smaller than t in practice so that α will be estimated to 0. This is a standard issue in extreme quantile estimation that can be addressed via extreme value theory (EVT). EVT focuses on distributions of extreme values *i.e.* maxima and minima. Considering a set of *i.i.d.* variables $\{Z_{m,n}, m = 1 : M, n = 1 : N\}$, a set $\{Z_1^*, \ldots, Z_M^*\}$ of M maxima can be defined as $Z_m^* = \max(Z_{m,1}, \ldots, Z_{m,N})$. EVT provides an estimation of the pdf denoted by f of the $\{Z_m^*, m = 1 : M\}$ via the estimation of 3 parameters referred to as the generalized extreme value (GEV) distribution parameters. In particular, most EVT procedures provide good estimations of $p(Z_m^* \leq \eta)$ when η is an extreme value in the upper tail of f. Our task can then be recast as follows,

$$\begin{aligned} & p(f_H(Y) \leq t) &=& \; 1 - p(f_H(Y) > t) \\ \text{with} \quad & p(f_H(Y) > t) &=& \; p(\min(f_H(Y_1), \ldots, f_H(Y_N)) > t)^{1/N} \end{aligned}$$

where $\{Y_1, \ldots, Y_N\}$ are *i.i.d.* random variables with distribution f_H. Then

$$\begin{aligned} p(\min(f_H(Y_1) \ldots f_H(Y_N)) > t) &= p(\max(-f_H(Y_1) \ldots - f_H(Y_N)) \leq -t) \\ &= p(\max(-\log f_H(Y_1) \ldots - \log f_H(Y_N)) \leq -\log t) \; . \end{aligned}$$

This later quantity can be computed by setting $\eta = -\log t$ and $Z_i = -\log f_H(Y_i)$. The log has the advantage to turn the bounded upper tail of $-f_H(Y)$ into an unbounded upper tail in $-\log f_H(Y)$. This provides more stable estimation of the GEV parameters obtained here with the R package [30].

ACKNOWLEDGEMENTS

The authors acknowledge the excellent technical support of the MRI Facility of Grenoble (UMS IRMaGe). IRMaGe is partly funded by the French program *Investissement d'Avenir* run by the French National Research Agency, grant *Infrastructure d'avenir en Biologie Santé* [ANR-11-INBS-0006].

REFERENCES

[1] S. J. C. G. Hectors, I. Jacobs, G. J. Strijkers, and K. Nicolay, "Automatic segmentation of subcutaneous mouse tumors by multiparametric MR analysis based on endogenous contrast," *Magnetic Resonance Materials in Physics, Biology and Medicine*, vol. 28, no. 4, pp. 363–375, 2015.

[2] B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest, L. Lanczi, E. Gerstner, M. A. Weber, T. Arbel, B. B. Avants, N. Ayache, P. Buendia, D. L. Collins, N. Cordier, J. J. Corso, A. Criminisi, T. Das, H. Delingette, C. Demiralp, C. R. Durst, M. Dojat, S. Doyle, J. Festa, F. Forbes, E. Geremia, B. Glocker, P. Golland, X. Guo, A. Hamamci, K. M. Iftekharuddin, R. Jena, N. M. John, E. Konukoglu, D. Lashkari, J. A. Mariz, R. Meier, S. Pereira, D. Precup, S. J. Price, T. R. Raviv, S. M. S. Reza, M. Ryan, D. Sarikaya, L. Schwartz, H. C. Shin, J. Shotton, C. A. Silva, N. Sousa, N. K. Subbanna, G. Szekely, T. J. Taylor, O. M. Thomas, N. J. Tustison, G. Unal, F. Vasseur, M. Wintermark, D. H. Ye, L. Zhao, B. Zhao, D. Zikic, M. Prastawa, M. Reyes, and K. Van Leemput, "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)," *IEEE Transactions on Medical Imaging*, vol. 34, no. 10, pp. 1993–2024, October 2015.

[3] 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.

[4] L. M. De Angelis, "Brain Tumors," *New England Journal of Medicine*, vol. 344, no. 2, pp. 114–123, January 2001.

[5] A. Drevelegas and N. Papanikolaou, *Imaging of Brain Tumors with Histological Correlations*. Springer Berlin Heidelberg, 2011, ch. Imaging Modalities in Brain Tumors, pp. 13–33.

[6] P. Y. Wen, D. R. Macdonald, D. A. Reardon, T. F. Cloughesy, A. G. Sorensen, E. Galanis, J. DeGroot, W. Wick, M. R. Gilbert, A. B. Lassman, C. Tsien, T. Mikkelsen, E. T. Wong, M. C. Chamberlain, R. Stupp, K. R. Lamborn, M. A. Vogelbaum, M. J. van den Bent, and S. M. Chang, "Updated Response Assessment Criteria for High-Grade Gliomas: Response Assessment in Neuro-Oncology Working Group," *Journal of Clinical Oncology*, vol. 28, no. 11, pp. 1963–1972, April 2010.

[7] 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.

[8] J. K. Boult, M. Borri, A. Jury, S. Popov, G. Box, L. Perryman, S. A. Eccles, C. Jones, and S. P. Robinson, "Investigating intracranial tumour growth patterns with multiparametric MRI incorporating Gd-DTPA and USPIO-enhanced imaging," *NMR in Biomedicine*, vol. 29, no. 11, pp. 1608–1617, 2016.

[9] P. Katiyar, M. R. Divine, U. Kohlhofer, L. Quintanilla Martinez, B. Schölkopf, B. J. Pichler, and J. A. Disselhorst, "A Novel Unsupervised Segmentation Approach Quantifies Tumor Tissue Populations Using Multiparametric MRI: First Results with Histological Validation," *Molecular Imaging and Biology*, vol. 19, no. 3, pp. 391–397, 2017.

[10] J. Juan Albarracín, E. Fuster Garcia, J. V. Manjón, M. Robles, F. Aparici, L. Martí Bonmatí, and J. M. García Gómez, "Automated Glioblastoma Segmentation Based on a Multiparametric Structured Unsupervised Classification," *PLoS ONE*, vol. 10, no. 5, pp. 1–20, May 2015.

[11] L. Macyszyn, H. Akbari, J. M. Pisapia, X. Da, M. Attiah, V. Pigrish, Y. Bi, S. Pal, R. V. Davuluri, L. Roccograndi, N. Dahmane, M. Martinez Lage, G. Biros, R. L. Wolf, M. Bilello, D. M. O'Rourke, and C. Davatzikos, "Imaging patterns predict patient survival and molecular subtype in glioblastoma via machine learning techniques," *Neuro-Oncology*, vol. 18, no. 3, pp. 1–9, July 2015.

[12] E. I. Zacharaki, S. Wang, S. Chawla, D. S. Yoo, R. Wolf, E. R. Melhem, and C. Davatzikos, "Classification of Brain Tumor Type and Grade Using MRI Texture and Shape in a Machine Learning Scheme," *Magnetic Resonance in Medicine*, vol. 62, no. 6, pp. 1609–1618, 2009.

[13] J. H. Rasmussen, M. Nørgaard, A. E. Hansen, I. R. Vogelius, M. C. Aznar, H. H. Johannesen, J. Costa, A. M. E. Engberg, A. Kjær, L. Specht, and B. M. Fischer, "Feasibility of Multiparametric Imaging with PET/MR in Head and Neck Squamous Cell Carcinoma," *Journal of Nuclear Medicine*, vol. 58, no. 1, pp. 69–74, 2016.

[14] 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.

[15] P. S. Tofts, *Quantitative MRI of the Brain*. John Wiley & Sons, Ltd, July 2004, ch. Concepts: Measurement and MR, pp. 1–15.

[16] J.-P. Baudry, C. Maugis, and B. Michel, "Slope heuristics: overview and implementation," *Statistics and Computing*, vol. 22, no. 2, pp. 455–470, 2012.

[17] P. Embrechts, C. Klüppelberg, and T. Mikosch, *Modelling Extremal Events: for Insurance and Finance*, 1st ed., ser. Stochastic Modelling and Applied Probability. Springer-Verlag Berlin Heidelberg, 1997, vol. 33.

[18] C. Kilkenny, W. J. Browne, I. C. Cuthill, M. Emerson, and D. G. Altman, "Improving Bioscience Research Reporting: The ARRIVE Guidelines for Reporting Animal Research," *PLOS Biology*, vol. 8, no. 6, pp. 1–5, June 2010.

[19] 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.

[20] L. Hubert and P. Arabie, "Comparing partitions," *Journal of Classification*, vol. 2, no. 1, pp. 193–218, 1985.

[21] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297–302, 1945.

[22] W. N. Venables and B. D. Ripley, *Modern Applied Statistics with S*, 4th ed. Springer-Verlag New York, 2002.

[23] L. Bergé, C. Bouveyron, and S. Girard, "HDclassif : An R Package for Model-Based Clustering and Discriminant Analysis of High-Dimensional Data," *Journal of Statistical Software*, vol. 46, no. 1, pp. 1–29, 2012. [Online]. Available: https://www.jstatsoft.org/index.php/jss/article/view/v046i06

[24] C. Fraley, A. E. Raftery, T. B. Murphy, and L. Scrucca, "mclust Version 4 for R: Normal Mixture Modeling for Model-Based Clustering, Classification, and Density Estimation," Department of Statistics, University of Washington, Box 354322 Seattle, WA 98195-4322 USA, Tech. Rep. 597, 2012. [Online]. Available: http://www.stat.washington.edu/research/reports/2012/tr597.pdf

[25] D. N. Louis, A. Perry, G. Reifenberger, A. von Deimling, D. Figarella Branger, W. K. K. Cavenee, H. Ohgaki, O. D. Wiestler, P. Kleihues, and D. W. Ellison, "The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary," *Acta Neuropathologica*, vol. 131, no. 6, pp. 803–820, 2016.

[26] D. G. Barone, T. A. Lawrie, and M. G. Hart, "Image guided surgery for the resection of brain tumours," *Cochrane Database of Systematic Reviews*, no. 1, 2014.

[27] A.-T. Du, N. Schuff, J. H. Kramer, H. J. Rosen, M. L. Gorno Tempini, K. Rankin, B. L. Miller, and M. W. Weiner, "Different regional patterns of cortical thinning in Alzheimer's disease and frontotemporal dementia," *Brain*, vol. 130, no. 4, pp. 1159–1166, March 2007.

[28] F. Forbes, S. Doyle, D. Garcia Lorenzo, C. Barillot, and M. Dojat, "A Weighted Multi-Sequence Markov Model For Brain Lesion Segmentation," in *13th International Conference on Artificial Intelligence and Statistics, AISTATS 2010, May, 2012*, ser. JMLR Workshop and Conference Proceedings, Neil Lawrence, Ed., vol. 9, Sardinia, Italie, 2010, pp. 225–232.

[29] R. Gillies, P. Kinahan, and Hedvig Hricak, "Radiomics: Images are more than pictures, they are data," *Radiology*, vol. 278, no. 2, pp. 563–577, 2 2016.

[30] A. G. Stephenson, "evd: Extreme Value Distributions," *R News*, vol. 2, no. 2, pp. 31–32, June 2002. [Online]. Available: https://cran.r-project.org/doc/Rnews/Rnews_2002-2.pdf