

Computer-aided detection and diagnosis using Multi-modal MRI for prostate cancer detection

Guillaume Lemaître*

*Parietal team, Inria, CEA, Université Paris-Saclay,
1 Rue Honoré d'Estienne d'Orves, 91120 Palaiseau*

Fabrice Mériadeau

*LE2I UMR6306, CNRS, Arts et Métiers, Univ. Bourgogne Franche-Comté,
12 rue de la Fonderie, 71200 Le Creusot*

Robert Martí

ViCOROB, Universitat de Girona, Campus Montilivi, Edifici P4, 17071 Girona

(Dated: August 20, 2018)

Abstract

Purpose: Prostate cancer (CaP) is the second most diagnosed cancer in men all over the world. In the last decades, new imaging techniques based on magnetic resonance imaging (MRI) have been developed improving the CaP screening. Even if some computer-aided detection and diagnosis (CAD) systems already have been proposed, none of them combined all available MRI modalities.

Methods: In this paper, we propose a CAD system to segment CaP using multiparametric MRI (mp-MRI) and focus on the following steps: (i) investigate the benefit of using all available MRI modalities (i.e. T_2 Weighted (T_2 -W)-MRI, dynamic contrast-enhanced (DCE)-MRI, diffusion weighted (DW)-MRI, and magnetic resonance spectroscopy imaging (MRSI)) and more precisely the effect of including the MRSI modality, (ii) focus on data normalization during the pre-processing stage, (iii) extract computer vision features using the volumetric information, and (iv) overcome the problem of class imbalance.

Results: All experiments are performed on a publicly available dataset of 17 subjects using a leave-one-patient-out cross-validation (LOPO CV) scheme. The best classification result is obtained with a fine-tuned random forest (RF) (i.e. with balancing and feature selection) with a receiver operating characteristic (ROC)-area under the curve (AUC) of 0.836 ± 0.083 .

Conclusion: Combining all available MRI modalities lead to the best classification performance and we show that MRSI has a beneficial effect when it is included in the classification pipeline. In addition, fine tuning the classification pipeline with data balancing and feature selection improves the results.

Keywords: Computer-aided detection, prostate cancer, multi-parametric MRI, feature engineering

I. INTRODUCTION

CaP is the second most frequently diagnosed cancer in men, accounting for 899,000 cases and leading to 258,100 deaths per year⁵. Early detection, diagnosis and accurate risk assessment play a major role in patient treatment. In this regard, Hambrick *et al.*⁷ explore the 5 idea of incorporating CAD systems to assist radiologists in their clinical practice. They show that radiologists assisted with CAD for CaP detection and diagnosis improve their screening performance, especially for inexperienced radiologists, leveling up their performance to the senior radiologists expertise.

Despite recent advances, CaP-CAD remains a young technology mainly because it is 10 based on MRI¹⁰. The two main challenges in CaP-CAD systems, as pointed by the Prostate imaging reporting and data system (PI-RADS) Steering Committee are: (i) improving the detection of clinically significant CaP and (ii) increasing confidence in benign or dormant cases to avoid unnecessary invasive medical exams³⁵.

Lemaitre *et al.*¹⁸ provides an overview and a taxonomy of the developed CAD systems 15 for CaP detection and diagnosis based on MRI modalities. As a conclusion, the following shortcomings have been drawn: (i) not a single study used all available MRI modalities, (ii) little attention has been granted to data normalization, (iii) features have usually been extracted using solely 2D information, and (iv) the problem of data balancing has never been investigated. We refer the reader to¹⁸ for a detailed review of the previous work related to 20 mp-MRI CAD for prostate cancer detection.

Therefore, we here propose a mp-MRI CAD framework using all available MRI modalities, as depicted in Fig. 1. In addition, whenever possible the features are extracted using 3D information of the data acquired. Our framework is also incorporating a data balancing step to reduce the classification bias introduce during the learning phase.

25 The rest of this document is structured as follows: Sect. II details each step of our mp-MRI CAD system while Sect. III reports a set of experiments to give insights about each of these steps. Finally, Sect. IV concludes this work with a discussion and give avenues for future research.

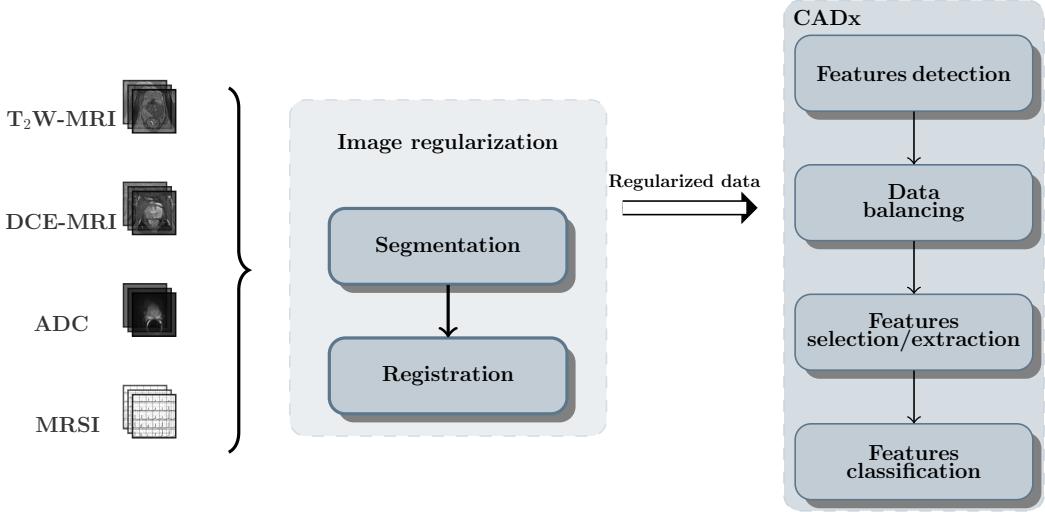


FIG. 1: **Our proposed mp-MRI CAD framework for prostate cancer detection.** The mp-MRI modalities are first regularized by segmenting the prostate organ and registering it. The second step follows a common classification pipeline by engineering features, balancing samples depending of the class occurrences, selecting features, and finally train and test a machine learning model. This model is evaluated through a cross-validation scheme.

II. MATERIALS AND METHODS

We investigate here the different steps of the CAD for CaP pipeline presented in Fig. 1. Indeed, this pipeline is subdivided into 2 steps: (i) image regularization and (ii) computer-aided diagnosis (CADx). The image regularization covers the segmentation of the prostate organ in the different modalities and their realignment. The CADx step is a regular classification pipeline in which features are detected, balanced, selected, and finally classified.

The following sections go into details in each of these steps.

A. Segmentation and registration

We used the manual segmentation provided by our radiologists (prostate gland, cancerous regions, and prostate regions). In addition, we applied different registration methods to align the volumes from the different modalities: (i) the patient motion during the DCE-MRI acquisition, (ii) the patient motion between the T₂-W-MRI and the DCE-MRI acquisitions,

and (iii) the patient motion between the T_2 -W-MRI and the apparent diffusion coefficient (ADC) map acquisition.

- 45 (i) The DCE-MRI acquisition being dynamic, some intra-patient motions might occur during the acquisition. For each serie of this dynamic acquisition, each 3D volume is registered to the first volume acquired to remove the residual motion. The appearance in the DCE-MRI images, however, varies due to the presence or not of the contrast media. Therefore, the metric chosen to be minimized is the mutual information (MI) and the geometric transform has been set to a rigid transform. The optimization is performed using a regular step gradient descent.
- 50 (ii) Once the intra-patient motions is corrected, T_2 -W-MRI and the DCE-MRI are registered. For that matter, the prostate has been segmented in both modalities — T_2 -W-MRI and DCE-MRI — to create two binary masks. These 3D binary masks are subsequently registered using the mean squared error (MSE) metric. Unlike the previous registration, we use a more complex geometric transform by successively finding 55 a rigid transformation, a coarse elastic transformation, and a fine elastic transformation. The B-splines transformation is used as the elastic transform. These successive transformations allow to get a good initialization for the next transformation. The transformation parameters are inferred by minimizing the cost function using a regular step gradient descent.
- 60 (iii) The T_2 -W-MRI and ADC map acquisitions are registered using the same approach as for the registration of the T_2 -W-MRI and the DCE-MRI modalities. Additionally, the CaP, peripheral zone (PZ), and central gland (CG) are segmented on the T_2 -W-MRI and thus T_2 -W-MRI is used as the reference modality.

B. Feature detection

65 Two different types of feature are extracted due to the characteristics of the MRI modalities used in this study: computer vision features are suitable for the T_2 -W-MRI and ADC map while specific feature engineering, embedding domain-specific knowledge, is required for the DCE-MRI and MRSI data.

TABLE I: **Features extracted in T₂-W-MRI and ADC map.** We explore a broad set of image features characterizing intensity, edges, and texture.

Features	Parameters	# dimensions
Intensity		1
DCT decomposition	window: 9 px × 9 px × 3 px	243
Kirsch filter		2
Laplacian filter		1
Prewitt filter		3
Scharr filter		3
Sobel filter		3
Gabor filters	4 frequencies $f \in [0.05, 0.25]$; 4 azimuth angles $\alpha \in [0, \pi]$; 8 elevation angles $\alpha \in [0, 2\pi]$	256
Phase congruency filter	5 orientations; 6 scales	3
Haralick filter	window: 9 px × 9 px × 3 px; # grey levels: 8; distance: 1 px; 13 directions	169
LBP filter	2 radii $r = \{1, 2\}$; 2 neighborhood sizes $N = \{8, 16\}$	6

1. Computer vision features

70 A set of common features reported in¹⁸ are computed. Table I summarizes the different features extracted with their corresponding parameters. Note that all these features are extracted at each voxel of the volume. The voxel intensities are the most common features encoding tumor information. However, those intensities vary between patients. Therefore, T₂-W-MRI modality is normalized using a Rician *apriori* as presented in¹⁹. The ADC maps 75 are standardized using a piecewise linear transformation as proposed in²³. The following set of filters characterizing edges are extracted: (i) Kirsch, (ii) Laplacian, (iii) Prewitt, (iv) Scharr, (v) Sobel, and (vi) Gabor. Apart of the Kirsh filter, all filters are 3D kernels extracting volumetric information instead of solely planar. 3D Gabor filters are not commonly used and we reused the formulation presented in³⁴. Additionally, features based on phase 80 congruency as proposed by Kosevi *et al.* are computed¹⁷. They consist of the orientation image, the local weighted mean phase angle, and the phase angle, which are estimated at each voxel. To characterize the local texture, both second-order gray-level co-occurrence matrix (GLCM)-based features⁹ and rotation invariant and uniform local binary pattern (LBP)²⁴ are extracted. To encode 3D information, the 13 first Haralick features are 85 computed for the 13 possible directions. For the same reason, the LBP codes are computed for the three-orthogonal-planes of each MRI volume.

2. DCE-MRI features

Two family of approaches are commonly used to extract information from the DCE-MRI signals¹⁸: (i) quantitative modeling and (ii) semi-quantitative modeling.

Quantitative approaches uses pharmacokinetic modeling based on a bicompartiment model, namely Brix¹ and Tofts²⁹ models. The parameters of the Brix model are inferred by assuming a linear relationship between the media concentration and the MRI signal intensity. However, this assumption has been shown to lead to inaccurate estimations of the pharmacokinetic parameters¹¹. In contrast, the Tofts model requires the conversion of MRI signal intensity to concentration, which becomes a non-linear relationship using a specific equation of MRI sequences (e.g., FLASH sequence). Tofts modeling, however, is highly complex⁶. Achieving the conversion using the non-linear approach requires the acquisition of a T_1 map which is not always possible during clinical examination. Additionally, computing the parameters require the arterial input function (AIF) which is challenging to measure and influence the estimation.

Semi-quantitative approaches extract a set of parameters related to a mathematical model rather than a pharmacokinetic model, relaxing any pharmacokinetic assumptions regarding the relationship between the MRI signal and the contrast agent^{6,14}. In addition, these methods do not require any knowledge regarding the MRI sequence or any conversion from the signal intensity to the media concentration. However, they present some limitations: the heuristic approach proposed by Huisman *et al.* requires an initial estimate of the standard deviation of the signal noise and some manual tuning¹⁴.

Nevertheless, all of the presented methods suffer from the following two major drawbacks: (i) inter-patient variability and (ii) loss of information. The inter-patient variability is mainly due to the acquisition process and consequently leads to generalization issues in applying a machine learning algorithm. All previous methods extract few discriminative parameters to describe the DCE-MRI signal which might lead to a loss of information. Therefore, we propose a method to normalize DCE-MRI to reduce inter-patient variations. As a consequence, no parametric models is required and the entire DCE-MRI normalized signal can be used as feature. We will later present a set of experiments in Sect. III B comparing each approach and select the most discriminative one. The remainder of this section describes the procedure used to normalize the DCE-MRI sequences.

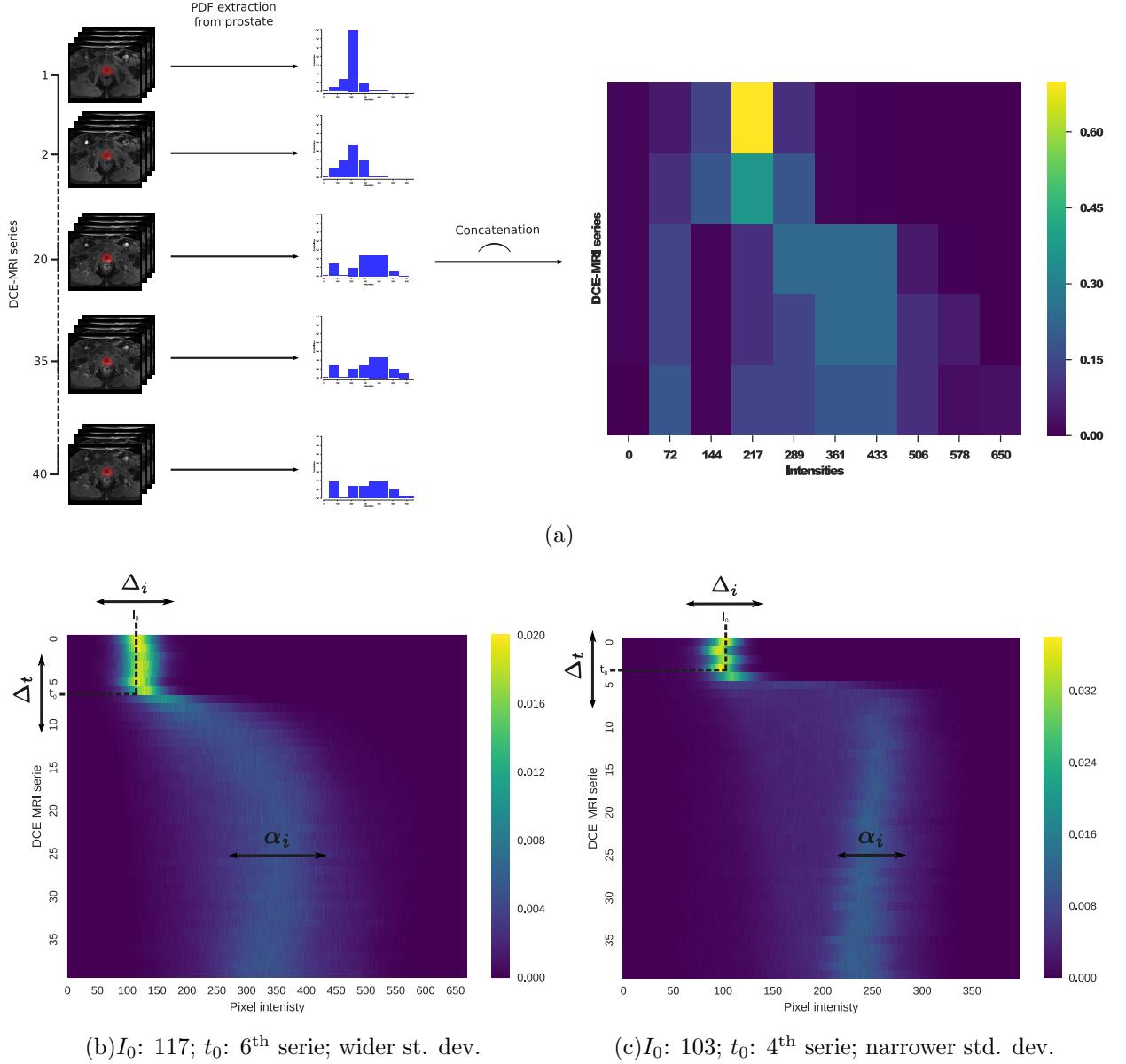


FIG. 2: **Overview of the normalization applied to DCE-MRI modality.** (a) Illustration of the heatmap representation: all probability density function (PDF)s of the prostate gland are concatenated together to build an heatmap; (b)-(c) Heatmap of 2 patients revealing the three types of inter-patient variations: intensity shift (Δ_i), time shift (Δ_t), and intensity scale (α_i).

Normalization of DCE-MRI

In DCE-MRI, the intensity probability density function (PDF) of the prostate gland does not follow a unique type of distribution such as Rician or Gaussian distribution, as shown in Fig. 2(a). Indeed, the inter-patient variations are more complex due to the temporal acquisition. A better means of observing these variations is to represent the intensity PDF of the prostate gland over time—requiring segmentation of the prostate—using a heatmap representation as shown in Fig. 2(a). By analyzing this heatmap

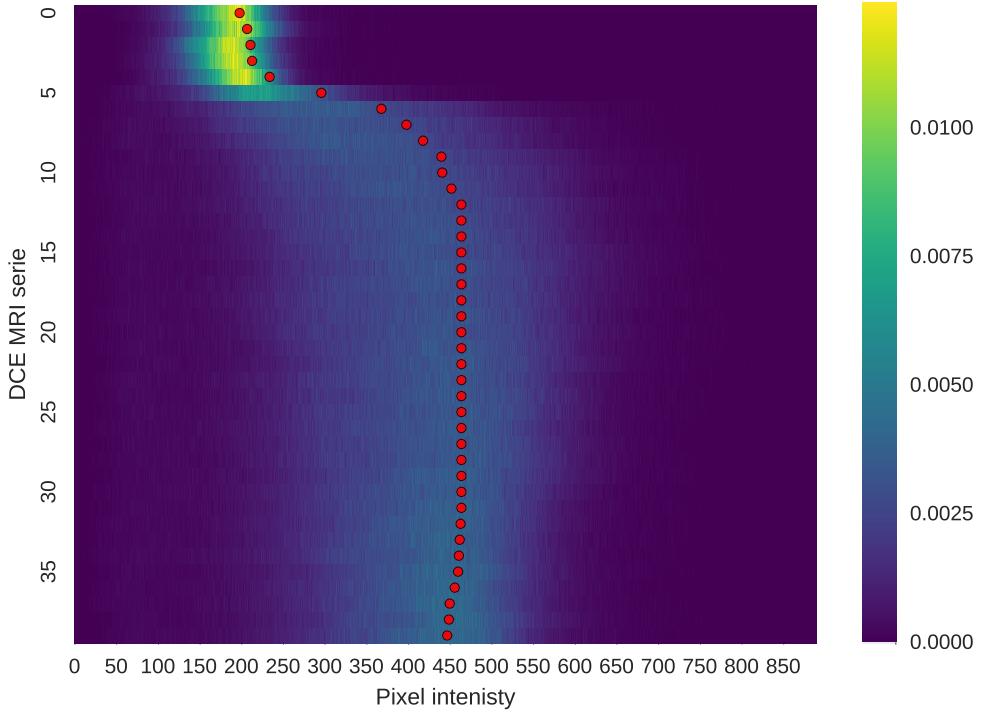


FIG. 3: **Estimation of Δ_i .** This estimate is found by using the shortest-path through the graph \mathcal{G} .

representation across patients (see Fig. 2(c)), the following variations are highlighted: (i) 125 intensity offsets (Δ_i) of the PDF peak, (ii) a time offset (Δ_t) depending on the contrast agent arrival, and (iii) a change of scale (α_i) related to the signal enhancement. Therefore, our normalization method should attenuate all of these variations and be performed globally across the different time sequences rather than for each independent sequence.

Graph-based intensity offsets correction Before standardizing each sequence, the 130 first step of the normalization process is to cancel the intensity shift specific at each patient, which occurs due to the media injection. As previously mentioned, the intensity PDF does not always follow a Rician or a Gaussian distribution over time in DCE-MRI. Therefore, the mean of these distributions cannot be used as a potential estimate for these offsets. Additionally, these offsets should be characterized by a smooth transition between series 135 over time.

Thus, this problem is solved using graph approach: considering the intensity PDF over time as shown in Fig. 2(a), the offsets correspond to the boundary splitting, the heatmap into two partitions such that they are as close as possible to the peak of the intensity PDF

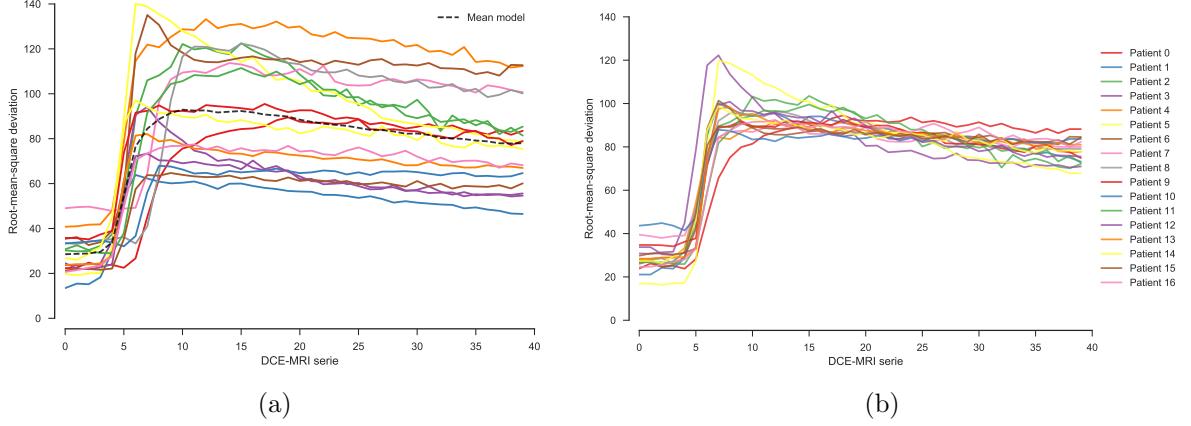


FIG. 4: **Correction of the time offset and the data dispersion.** (a) RMSD computed for each patient of our dataset; (b) RMSD after alignment using the curve parametric model.

(see Fig. 3 for an illustration). Given the heatmap, a directed weighted graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is built by taking each bin—, i.e., the probability for a given time and pixel intensity—, of the heatmap as a node and connecting each pair of bins by an edge. The edge weight w_{ij} between two nodes i and j corresponds to two pixel intensities at positions (x_i, y_i) and (x_j, y_j) , respectively, is defined as in Eq. (1), as follows:

$$w_{ij} = \begin{cases} \alpha \exp(1 - \frac{H(i)}{\max(H)}) & \text{if } x_j = x_i + 1 \text{ and } y_j = y_i, \\ (1 - \alpha) \exp(1 - \frac{H(i)}{\max(H)}) & \text{if } x_j = x_i \text{ and } y_j = y_i + 1, \\ 0 & \text{otherwise,} \end{cases} \quad (1)$$

where H is the heatmap, and α is a smoothing parameter controlling for the partitioning.

Therefore, these offsets related to Δ_i are estimated by finding the shortest-path to cross the graph using Dijkstra's algorithm. The entry and exiting nodes are set to be the bin with the maximum probability for the first value in the DCE-MRI series and the bin corresponding to the median value for the last value of the DCE-MRI series, respectively. To ensure a robust estimation of these offsets, the process of finding the shortest-path is repeated by shifting the data and updating the heatmap as well as the graph \mathcal{G} . The procedure is stopped once the offset found does not change. In general, this process is not repeated more than 3 times. The parameter α is set to 0.9, empirically. Figure 3 illustrates the final estimation of the offsets, Δ_i (i.e., red landmark), found for each value of the DCE-MRI series. Therefore, each intensity offset is subtracted for each DCE-MRI.

155 **Time offset and data dispersion correction** The next variations to correct are the time offset, Δ_t , and the data dispersion, α_i . By computing the root-mean-square deviation (RMSD) of the intensities for each value of the DCE-MRI series, one can observe these two variations as shown in Fig. 4(a). Therefore, these variations are corrected by registering the RMSD of each patient to a mean model computed using all patients and obtaining 160 an estimate for the parameters Δ_t and α_i . An illustration of the correction applied to each RMSD of the patients is shown in Fig. 4(b). Once all of these parameters have been determined, the data are shifted and scaled.

3. MRSI features

165 First, the MRSI signal is pre-processed by correcting the phase³, baseline³⁶, and frequency²⁵ shifts. Additionally, each MRSI spectrum is normalized using the L_2 norm, which has been shown to be the most efficient normalization method in MRSI²⁵. Once the signal are pre-processed, two strategies are commonly used to extract discriminative features from MRSI: (i) relative quantification based on metabolite quantification and (ii) spectra extraction²⁵.

170 In *relative quantification*, the relative concentration of the metabolites of interest (i.e. citrate and choline) is computed by robustly fitting and integrating the peak(s). Therefore, the relative concentrations of citrate and choline are quantified using a Gaussian mixture. We solve this problem with a non-linear curve fitting problem under constraints. The constraints imposed are related to the peak location of the citrate and choline.

175 The latter strategy is based on the work of²⁵. The full MRSI signal from 2 ppm to 4 ppm is used as features, letting the classifier selecting the relevant ppm bandwidth.

Similarly to the DCE-MRI data, a set of experiment is designed in Sect. III B to select the most discriminative approach.

C. Data balancing

180 Imbalanced dataset is a recurrent issue in classification. A dataset is imbalanced when a class is over-represented compared to other classes. In our application, the number of cancerous voxels is under-represented compared to healthy voxels. In classification, imbalanced

datasets are either under- or over-sampled ahead of training, avoiding the classifiers to learn a bias toward the over-represented class. In this section, we present the different methods 185 which we used to alleviate this issue. The reader is referred to Sect. III D 1 presenting the experiments investigating the effect of balancing the dataset during the training phase.

1. *Under-sampling*

A dataset can be balanced by under-sampling (US) samples from the over-represented classes.

190 Nearmiss (NM) offers three different methods to under-sample the majority class²². In nearmiss-1 (NM-1), samples from the majority class are selected such that for each sample, the average distance to the k nearest neighbour (NN) samples from the minority class is minimum. nearmiss-2 (NM-2) diverges from NM-1 by considering the k farthest neighbours samples from the minority class. In nearmiss-3 (NM-3), a subset M containing samples 195 from the majority class is generated by finding the m NN from each sample of the minority class. Then, samples from the subset M are selected such that for each sample, the average distance to the k NN samples from the minority class is maximum. In our experiment, k and m are fixed to 3.

Instance-hardness-threshold (IHT) select samples with a high hardness threshold²⁷. Hardness indicates the likelihood of mis-classification rate for each samples. The notation of instance hardness are drawn through the decomposition of $p(h|t)$ using Bayes' theorem, where h represent the mapping function used to map input features to their corresponding labels and t represents the training set.

$$IH_h(\langle x_i, y_i \rangle) = 1 - p(y_i|x_i, h). \quad (2)$$

Therefore, under-sampling is performed by keeping the most probable samples — i.e, filtering 200 the samples with high hardness value — through k -fold cross-validation (k -CV) training sets while considering specific threshold for filtering.

2. Over-sampling

In contrast to US techniques, a dataset can be balanced by over-sampling (OS) the samples from the under-represented class.

205 Synthetic minority over-sampling techniques (SMOTE) is a method to generate new synthetic samples². Let define x_i as a sample belonging to the minority class. Let define x_{nn} as a randomly selected sample from the k -NN of x_i , with k set to 3. A new sample x_j is generated such that $x_j = x_i + \sigma(x_{nn} - x_i)$, where σ is a random number in the interval $[0, 1]$. SMOTE-borderline1 (SMOTE-b1) over-samples the minority class samples similarly 210 to SMOTE⁸. However, instead of using all the minority samples, it focuses on the borderline samples of minority class. Borderline samples simply indicate the samples that are closer to the other class. First, the borderline samples of minority class are detected. A sample x_i belongs to borderline samples if more than half of its k -NN samples belong to the majority class. Synthetic data is then created based on SMOTE method for borderline samples, by selecting. Then, s -NN of the minority class are selected to generate synthetic sample similarly 215 to SMOTE. SMOTE-borderline2 (SMOTE-b2) performs similarly to SMOTE-b1⁸. However, the s -NN are not computed by only considering the minority class but by considering both classes. The same generation rules as SMOTE is used.

D. Feature selection

220 Tree-based models can be efficiently used to find which set of features is the most discriminative by using the Gini importance. In a tree classifier, the Gini impurity criterion of the child nodes is inferior to the parent node. For each individual feature, summing the decrease of the Gini impurity along the tree gives information about the feature importance: the higher, the better. Therefore, one can sum the relative importance of all features across 225 all the trees of a forest (with an additional normalization depending of the number of trees). Subsequently, the K most important features are selected to perform the feature selection.

Therefore, in addition to use RF as our base classifier, we also use it to identified important features as reported by the experiments in Sect. III D 2.

E. Classification

230 RF serves as our base classifier. The use of RF is motivated since that it leads to the best performance in the state-of-the-art methods^{18,21}. The number of trees in the ensemble is set to 100.

F. Model validation

All experiments use a LOPO CV to report the results of the different models.

235 **G. Computationally aspects**

For reproducibility, we rely on the Python ecosystem: NumPy³³, SciPy¹⁶, scikit-learn²⁶, imbalanced-learn²⁰, scikit-image³², and mahotas⁴. The registration is performed using ITK¹⁵. We further provide all source code of all experiments³⁷ with the associated dataset³⁸.

III. EXPERIMENT AND RESULTS

240 **A. Data**

The mp-MRI data are acquired from a cohort of patients with higher-than-normal level of prostate-specific antigen (PSA). The acquisition is performed using a 3 T whole body MRI scanner (Siemens Magnetom Trio TIM, Erlangen, Germany) using sequences to obtain T₂-W-MRI, DCE-MRI, DW-MRI, and MRSI. Aside of the MRI examination, these patients 245 also have undergone a guided-biopsy. The dataset is composed of a total of 19 patients of which 17 patients have biopsy proven CaP and 2 patients are “healthy” with negative biopsies. From those 17, 12 patients have a CaP in the PZ, 3 patients have CaP in the CG, 2 patients have invasive CaP in both PZ and CG. An experienced radiologist has segmented the prostate organ — on T₂-W-MRI, DCE-MRI, and ADC-MRI — as well as the prostate 250 zones — i.e., PZ and CG —, and CaP on the T₂-W-MRI.

A 3 mm slice fat-suppressed T₂-W fast spin-echo sequence (repetition time (TR): 3400 ms, echo time (TE): 85 ms, echo train length (ETL): 13) is used to acquire images in sagittal and oblique coronal planes, the latter planes being orientated perpendicular or parallel to the

prostate PZ rectal wall axis. Three-dimensional T₂-W fast spin-echo (TR: 3600 ms, TE: 143 ms, ETL: 109, slice thickness: 1.25 mm) images are then acquired in an oblique axial plane. The nominal matrix and field of view (FOV) of the 3D T₂-W fast spin-echo images are 320 mm² × 256 mm² and 280 mm² × 240 mm², respectively, thereby affording sub-millimetric pixel resolution within the imaging plane.

DCE-MRI is performed using a fat suppressed 3D T₁ VIBE sequence (TR: 3.25 ms, TE: 1.12 ms, Flip angle: 10°; Matrix: 256 × 192; FOV: 280 × 210 (with 75 % rectangular FOV); slab of 16 partitions of 3.5 mm thickness; temporal resolution: 6 s/slab over approximately 5 min). A power injector (Medrad, Indianola, USA) is used to provide a bolus injection of Gd-DTPA (Dotarem, Guerbet, Roissy, France) at a dose of 0.2 ml Gd-DTPA/kg of body weight.

DW-MRI images have been acquired using the single-shot spin-echo echo-planar imaging (EPI) technique. As proposed by Stejskal *et al.*²⁸, the diffusion-encoding gradients have been applied using a pulsed gradient spin-echo technique resulting in diffusion images acquired at 2 b-values — i.e., 100 s mm⁻² and 800 s mm⁻² — and in the 3 orthogonal directions. Sequential sampling of the k-space has been used with a TE of 101 ms, a TR of 4200 ms, and a bandwidth of 1180 Hz px⁻¹. Other parameters included a FOV of 240 mm, an acquisition matrix size of 128 × 128 and a slice thickness of 3.5 mm. The ADC map has been directly generated by the Siemens workstation from the raw data on a pixel-by-pixel basis.

MRSI is performed using a water and lipid suppressed double-spin-echo point-resolved spectroscopic (PRESS) sequence optimized for quantification detection of choline and citrate metabolites. Water and lipid have been suppressed using a dual-band spectral spatial pulse technique. In order to eliminate signals from adjacent tissues, especially periprostatic lipids and the rectal wall up to 8 outer voxel saturation pulses have been used. Datasets have been acquired as 16 × 12 × 16 — interpolated to 16 × 16 × 16 phase-encoded spectral arrays, a TE of 140 ms, a TR of 720 ms and 13 min of acquisition time. A spectral bandwidth of 1250 Hz has been used with 512 data points. A combination of an elliptic weighted averaged k-space acquisition scheme 3D filtering of the signal in k-space have been used, the latter in order to reduce intervoxel signal combination. Shimming has been carried out using the Siemenbens 3D Mapshim routine on a voxel adapted to the volume of the entire prostate gland. Additional unsuppressed water acquisitions at TE of 30 ms, 80 ms, and 140 ms of 1.5 min have also been performed in order to allow quantification with respect to prostate

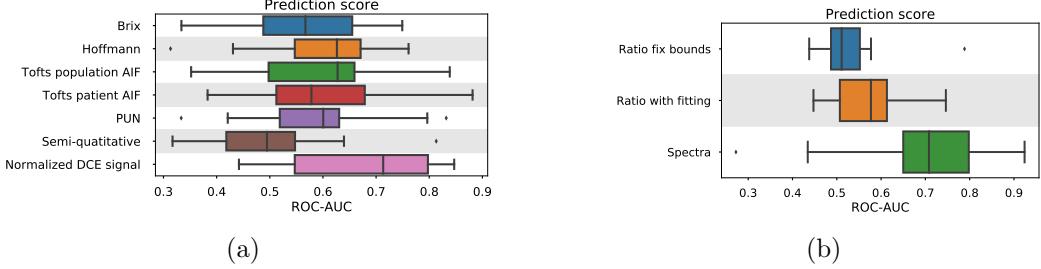


FIG. 5: **Impact of the features used with the DCE-MRI and MRSI.** (a) Comparison of the different quantitative and semi-quantitative models in DCE-MRI; (b) comparison of quantification approaches in MRSI.

water. Systematic verification of the global shim — i.e., over the complete 3D PRESS-selected volume — revealed line widths at half-height of the water peak of the order of 20 Hz to 30 Hz, routinely. The line widths for individual voxels are of the order of 8 Hz to 12 Hz. The total examination time, including the time spent positioning the patient, is approximately 45 minutes.

B. Selection of the feature detection strategies for DCE-MRI and MRSI

As presented in Sect. II B 2 and II B 3, several methods exists to extract information from DCE-MRI and MRSI modalities. These features are redundant and only one of the model/strategy should be selected. Therefore, this experiment intends to compare the induced performance of these features and select the best one.

The methods presented in Sect. II B 2 — i.e. (i) Brix model¹, (ii) Hoffman model¹², (iii) Tofts model²⁹, (iv) phenomenological universalities (PUN) model⁶, (v) semi-quantitative approach¹⁴, and (vi) use of the normalized DCE signal — are extracted. The performance of each strategy is reported in Fig. 5(a) in terms of ROC-AUC. Using the entire normalized DCE signal strategy outperforms the other approaches with an AUC of 0.679 ± 0.136 .

As presented in Sect. II B 3, 3 strategies are used to extract features from the MRSI modality: (i) compute the ratio of the citrate concentration over the choline concentration where the concentrations are determined with fixed ppm bounds, (ii) compute the ratio of the citrate concentration over the choline concentration where the concentrations are determined by fitting the peaks, and (iii) using a part of the MRSI signal containing the choline and citrate peaks. The performance of each strategy is highlighted in the Fig. 5(b).

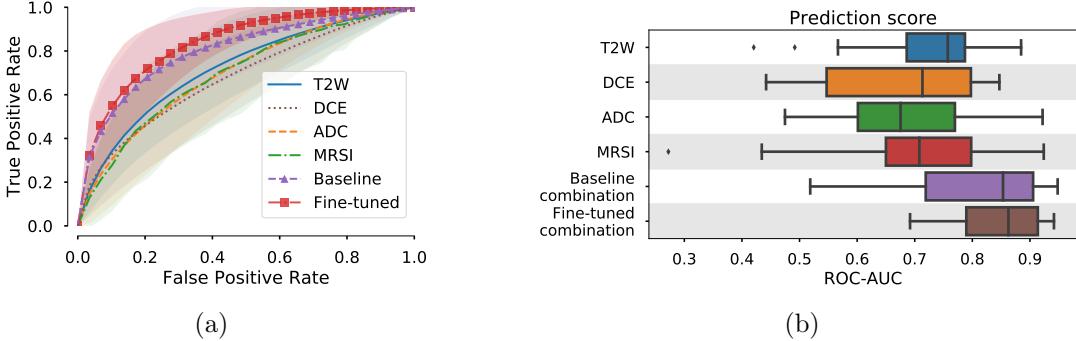


FIG. 6: **Comparison of discriminative power of different strategies.** We investigate the baseline classification performance of each modality. In addition, we explore the combination of those modalities and check the effect of fine tuning the machine learning model learned.

Similarly to the DCE-MRI experiment, using the entire signal lead to the best performance with a ROC-AUC of 0.695 ± 0.159 .

C. Classification baselines

Figure 6 summarizes the classification performance for each individual modality as well as their combinations. Features extracted on the T₂-W-MRI modality are the most discriminative leading to a ROC-AUC of 0.720 ± 0.122 . Features from ADC and MRSI modalities achieve similar performance, i.e. 0.697 ± 0.128 and 0.695 ± 0.159 , respectively. Finally, the DCE-MRI modality is the less discriminative modality with a ROC-AUC of 0.679 ± 0.136 . In a clinical setting, those results would be considered as almost “acceptable” (i.e. ROC-AUC ranging from 0.7 to 0.8¹³). Combining all features together improve the ROC-AUC to 0.802 ± 0.130 which is considered as “excellent” level of discrimination (i.e. ROC-AUC ranging from 0.8 to 0.9¹³).

D. Effect of fine tuning

To further improve the results, we propose to correct the problem of imbalanced classes by applying some over- and under-sampling technique during the training. In addition, we propose to select a subset of the most relevant features.

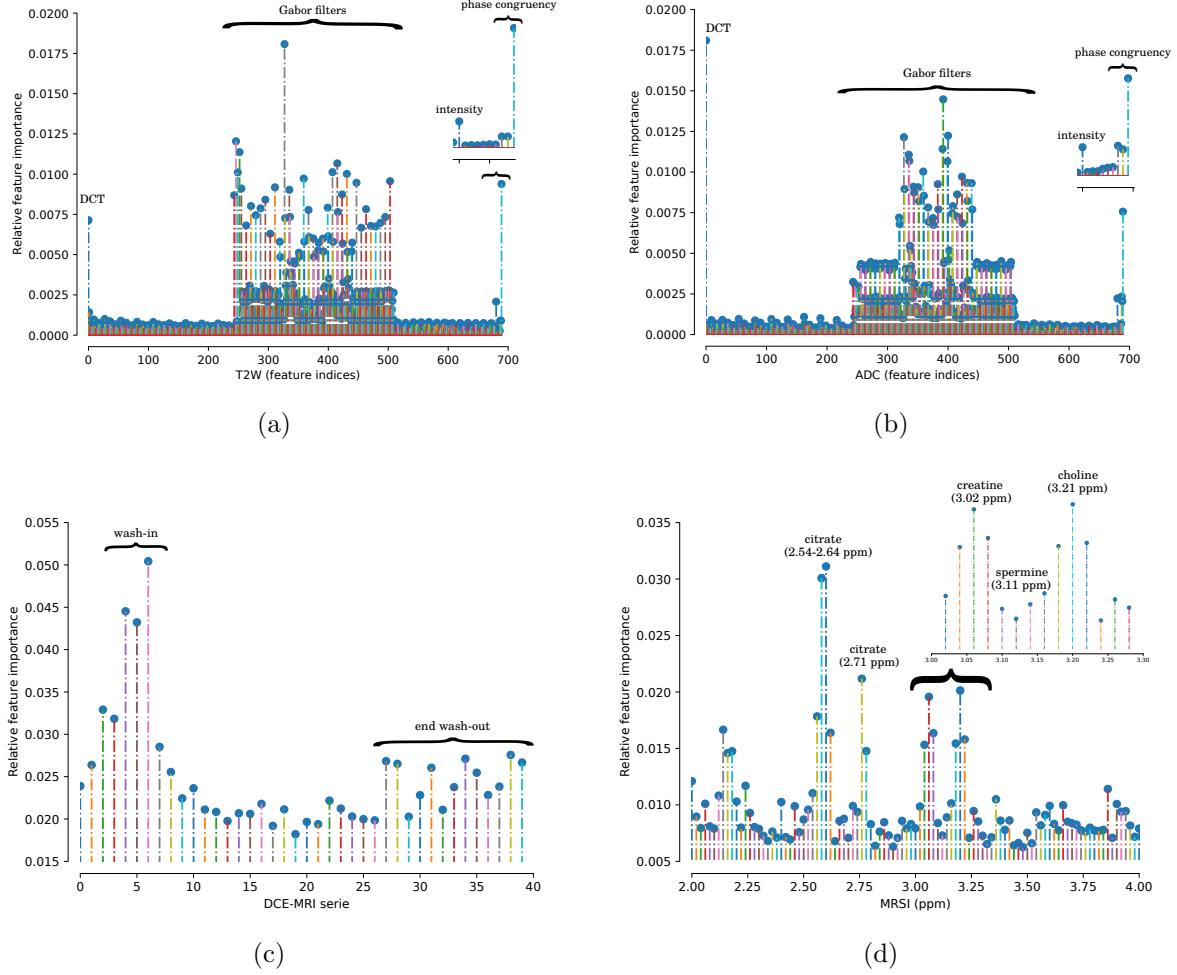


FIG. 7: **Feature importance.** We explore the relative features importance found by a RF classifier on each modality.

1. Data balancing

We empirically check which of the methods presented in Sect. II C leads to the best 325 classification performance. The results are reported in Fig. 8. In conclusion, NM-3 leads to the best performance with a ROC-AUC of 0.824 ± 0.076 .

2. Feature selection

The combination of the features from all modalities lead to a potential correlation between features. In addition, some features might not be discriminative and there is no benefit to 330 include them in the model trained. Therefore, we propose to select a subset of features which are found to be the most important using the approach presented in Sect. II D. Table II

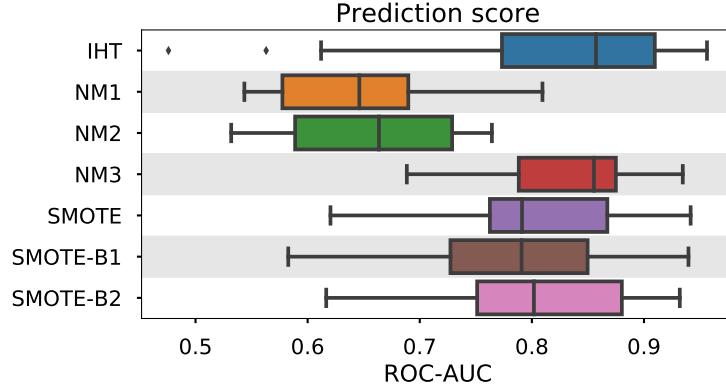


FIG. 8: **Effect of data balancing.** ROC-AUC obtained with the different balancing strategies.

TABLE II: **Feature selection.** Number and type of features selected for each individual modality.

T₂-W-MRI	ADC	DCE-MRI	MRSI
113 Gabor filters	53 Gabor filters	14 samples	78 samples
1 phase congruency	2 phase congruency		
4 edges			
1 intensity			
267 features			

summarizes the number and type of features which are selected. To complement this table, Fig. 7 highlights the relative feature importance of each feature found by the RF classifier.

In the image-based modalities — i.e. T₂-W-MRI and ADC map — the Gabor features, 335 discrete cosine transform (DCT), phase congruency, and intensity are considered particularly relevant. In the MRSI modality, the ppm bands found relevant are associated with the citrate, creatine, and choline while in the DCE-MRI, the wash-in and the end of the wash-out corresponds to the time points considered as important.

3. Fine-tuned classifier performance

340 We finally trained a fine-tuned RF classifier on the subset of features found relevant and by applying the NM-3 under-sampling. The classification performance in terms of ROC-AUC is reported in Fig. 6. The fine tuning of the classifier allows to improve the performance from 0.802 ± 0.130 to 0.836 ± 0.083 . Figure 9 shows some example of CaP segmentation.

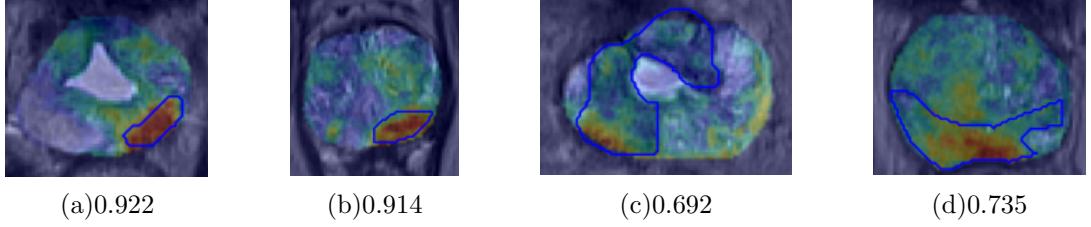


FIG. 9: **Qualitative results.** Detection of our mp-MRI CAD for CaP detection. The blue contours corresponds to the CaP while the jet overlay represents the probability. The score indicated are ROC-AUC.

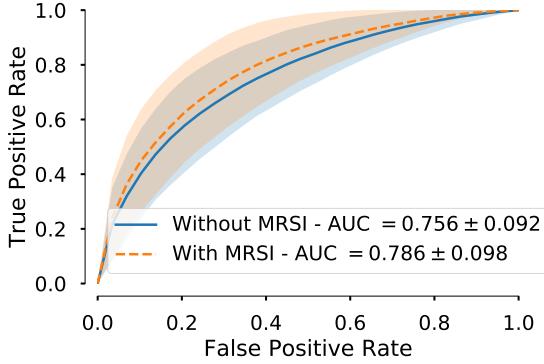


FIG. 10: **Effect of including MRSI modality.** ROC-AUC by plug-in/-out MRSI modality.

E. Effect of including MRSI modality

345 We can recall that MRSI has nearly never been used together with the other modalities — i.e., T₂-W-MRI, DCE-MRI, and ADC map — apart of the recent work of Trigui *et al.*^{30,31}. Therefore, we propose in this experiment to compare the classification performance by removing the MRSI features and observed the effect of including this modality.

350 In this regard, we propose to train 2 models such that one of the model will omit the MRSI modality. Therefore, we use a stacking approach for which the first layer is composed of a classifier (a RF classifier) for each input modality and a second layer with a single classifier (a Gradient Boosting (GB) classifier) to aggregate the predictions of the first layer. Therefore, 2 stacking classifiers are trained: (i) the first model uses 4 base classifiers in its first layer while (ii) the second model uses 3 base classifiers in its first layer, omitting the 355 classifier for the MRSI modality.

As in all other experiments, a LOPO CV validation scheme is used and the results obtained from ROC analysis are depicted in Fig. 10. Thus, including MRSI into the classifi-

cation pipeline increases the AUC from 0.756 ± 0.092 to 0.786 ± 0.098 highlighting the gain of incorporating the MRSI modality in the CAD system.

360 **IV. DISCUSSION**

We designed and fine-tuned a classification pipeline for CaP detection. We would like to stress the following findings drawn during the experiments. The classification of individual modality highlights the weakness of the quantification methods — i.e., pharmokinetic models, semi-quantitative model, and relative quantification of metabolites — which might 365 be due to the loss of information during the quantification procedure. In addition, the normalization proposed in Sect. II B 2 allows to boost the discriminative power of the classifier using the DCE-MRI modality. Furthermore, the features extracted from the T₂-W-MRI are the most discriminative even after features selection. Unlike T₂-W-MRI, DCE-MRI is always the less discriminative method.

370 The experiment link to the feature selection highlights some interesting facts regarding the most discriminative features. On the one hand, the Gabor filters and the phase congruency are always selected, independently of the strategy and modality during the feature selection process. Additionally, edge filters — i.e., Kirsch, Prewitt, Scharr, and Sobel — have been only selected for the T₂-W-MRI. A possible explanation might be due to the fact that T₂-375 W-MRI is the modality with the highest spatial resolution and in which the level of details is the most important. Subsequently, the intensity feature of the T₂-W-MRI modality is selected, implying that our normalization method proposed in¹⁹ is efficient. From those results, approaches based on convolutional neural network should perform well but will require to be tested on a much larger data set.

380 While applying the feature selection on the concatenated set of features, MRSI appeared to be one of the most significant feature by keeping the larger number of features. Along the same line, we show that plug-in/-out the modality from the stacking classifier decreases drastically the classification performance. Finally, we can highlight that the classification performance obtained is worst with patients having a CaP localized in the CG, with a low 385 sensitivity. A possible way to address this issue would be to train 2 classifiers, each of them specialized on a zone of the prostate gland (i.e. CG and PZ) and later combine their predictions.

As avenues for future research, one could switch from voxel-based classification to super-voxel classification such that spatial structure are classified instead of voxel. Furthermore, all
390 features from this chapter can be defined as hand-crafted features. Therefore, an approach with unsupervised learning as convolutional neural network and conjunction with transfer learning should be investigated. It should be also noted that the number of patients included in this research is quite limited and that aggregating patients from multiple sites as well as multiple MRI machines would be required.

395 **V. CONCLUSION**

In this research, we proposed a mp-MRI CAD system for CaP detection. Our CAD system was designed to address to some shortcomings from the literature: additionally to T₂-W-MRI, DCE-MRI, and DW-MRI, our CAD system used the MRSI modality; the set of computer vision features extracted used the volumetric information; we further investigated
400 the effect of fine tuning focusing on data balancing and feature selection. The performance of this fine-tuned classifier in terms of ROC-AUC was 0.836 ± 0.083 .

As future works, this study should be expended on a larger cohort of subjects. In addition, newer approaches such as convolutional neural network could be investigated.

Acknowledgment

405 G. Lemaître was supported by the Generalitat de Catalunya (grant nb. FIDGR2012), the Spanish R+D+I grant TIN2012-37171-C02-01, and the Regional Council of Burgundy (grant nb. 2015-9201AAO050S02760). Calculations were performed using HPC resources from DSI-CCUB (Université de Bourgogne).

* Email: guillaume.lemaitre@inria.fr; Author to whom correspondence should be addressed.;
410 Website: <https://glemaître.github.io/>

¹ Gunnar Brix, Wolfhard Semmler, Rüdiger Port, Lothar R Schad, Günter Layer, and Walter J Lorenz. Pharmacokinetic parameters in cns gd-dtpa enhanced mr imaging. *Journal of computer assisted tomography*, 15(4):621–628, 1991.

- 415 ² N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. Smote: synthetic minority
over-sampling technique. *Journal of artificial intelligence research*, pages 321–357, 2002.
- 420 ³ Li Chen, Zhiqiang Weng, LaiYoong Goh, and Marc Garland. An efficient algorithm for automatic phase correction of {NMR} spectra based on entropy minimization . *Journal of Magnetic Resonance*, 158(12):164 – 168, 2002.
- 425 ⁴ Luis Pedro Coelho. Mahotas: Open source software for scriptable computer vision. *arXiv preprint arXiv:1211.4907*, 2012.
- ⁵ Jacques Ferlay, Hai-Rim Shin, Freddie Bray, David Forman, Colin Mathers, and Donald Maxwell Parkin. Estimates of worldwide burden of cancer in 2008: Globocan 2008. *International journal of cancer*, 127(12):2893–2917, 2010.
- 430 ⁶ AS Gliozzi, S Mazzetti, Pier Paolo Delsanto, D Regge, and M Stasi. Phenomenological universalities: a novel tool for the analysis of dynamic contrast enhancement in magnetic resonance imaging. *Physics in medicine and biology*, 56(3):573, 2011.
- ⁷ T. Hambrock, P. C. Vos, C. A. Hulsbergen-van de Kaa, J. O. Barentsz, and H. J. Huisman. Prostate cancer: computer-aided diagnosis with multiparametric 3-T MR imaging—effect on observer performance. *Radiology*, 266(2):521–530, Feb 2013.
- 435 ⁸ H. Han, W. Y. Wang, and B. H. Mao. Borderline-smote: a new over-sampling method in imbalanced data sets learning. In *International Conference on Intelligent Computing*, pages 878–887. Springer, 2005.
- ⁹ R.M. Haralick, K. Shanmugam, and Its’Hak Dinstein. Textural features for image classification. *Systems, Man and Cybernetics, IEEE Transactions on*, SMC-3(6):610–621, 1973.
- 440 ¹⁰ J. V. Hegde, R. V. Mulkern, L. P. Panych, F. M. Fennessy, A. Fedorov, S. E. Maier, and C. M. Tempany. Multiparametric MRI of prostate cancer: an update on state-of-the-art techniques and their performance in detecting and localizing prostate cancer. *J Magn Reson Imaging*, 37(5):1035–1054, May 2013.
- ¹¹ Melanie Heilmann, Fabian Kiessling, Marta Enderlin, and Lothar R Schad. Determination of pharmacokinetic parameters in dce mri: consequence of nonlinearity between contrast agent concentration and signal intensity. *Investigative radiology*, 41(6):536–543, 2006.
- 445 ¹² Ulf Hoffmann, Gunnar Brix, Michael V Knopp, Thomas Heβ, and Walter J Lorenz. Pharmacokinetic mapping of the breast: a new method for dynamic mr mammography. *Magnetic resonance in medicine*, 33(4):506–514, 1995.

- 445 13 David W Hosmer Jr and Stanley Lemeshow. *Applied logistic regression*. John Wiley & Sons, 2004.
- 14 Henkjan J Huisman, Marc R Engelbrecht, and Jelle O Barentsz. Accurate estimation of pharmacokinetic contrast-enhanced dynamic mri parameters of the prostate. *Journal of Magnetic Resonance Imaging*, 13(4):607–614, 2001.
- 450 15 Luis Ibanez, William Schroeder, Lydia Ng, and Josh Cates. The itk software guide. 2005.
- 16 Eric Jones, Travis Oliphant, and Pearu Peterson. {SciPy}: open source scientific tools for {Python}. 2014.
- 17 Peter Kovesi. Image features from phase congruency. *Videre: Journal of computer vision research*, 1(3):1–26, 1999.
- 455 18 G. Lemaitre, R. Marti, J. Freixenet, J. C. Vilanova, P. M. Walker, and F. Meriaudeau. Computer-aided detection and diagnosis for prostate cancer based on mono and multi-parametric mri: A review. *Computers in Biology and Medicine*, 60:8–31, 2015.
- 19 Guillaume Lemaitre, Mojdeh Rastgo Dastjerdi, Joan Massich, Joan C Vilanova, Paul M Walker, Jordi Freixenet, Anke Meyer-Baese, Fabrice Mériadeau, and Robert Marti. Normalization of t2w-mri prostate images using rician a priori. In *SPIE Medical Imaging*, pages 978529–978529. International Society for Optics and Photonics, 2016.
- 460 20 Guillaume Lemaître, Fernando Nogueira, and Christos K Aridas. Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning. *The Journal of Machine Learning Research*, 18(1):559–563, 2017.
- 465 21 G. Litjens, O. Debats, J. Barentsz, N. Karssemeijer, and H. Huisman. Computer-aided detection of prostate cancer in MRI. *Medical Imaging, IEEE Transactions on*, 33(5):1083–1092, May 2014.
- 22 I. Mani and I. Zhang. knn approach to unbalanced data distributions: a case study involving information extraction. In *Proceedings of Workshop on Learning from Imbalanced Datasets*, 2003.
- 470 23 L. G. Nyul and J. K. Udupa. On standardizing the MR image intensity scale. *Magn Reson Med*, 42(6):1072–1081, Dec 1999.
- 24 Timo Ojala, Matti Pietikainen, and Topi Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on pattern analysis and machine intelligence*, 24(7):971–987, 2002.
- 475 25 S. Parfait, P.M. Walker, G. Crhange, X. Tizon, and J. Mitran. Classification of prostate mag-

- netic resonance spectra using Support Vector Machine . *Biomedical Signal Processing and Control*, 7(5):499 – 508, 2012.
- 26 480 Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *Journal of machine learning research*, 12(Oct):2825–2830, 2011.
- 27 485 M. R. Smith, T. Martinez, and C. Giraud-Carrier. An instance level analysis of data complexity. *Machine learning*, 95(2):225–256, 2014.
- 28 490 Edward O Stejskal and John E Tanner. Spin diffusion measurements: spin echoes in the presence of a time-dependent field gradient. *The journal of chemical physics*, 42(1):288–292, 1965.
- 29 495 Paul S Tofts, Bruce Berkowitz, and Mitchell D Schnall. Quantitative analysis of dynamic gd-dtpa enhancement in breast tumors using a permeability model. *Magnetic Resonance in Medicine*, 33(4):564–568, 1995.
- 30 500 R Trigui, J Mit  ran, PM Walker, L Sellami, and A Ben Hamida. Automatic classification and localization of prostate cancer using multi-parametric mri/mrs. *Biomedical Signal Processing and Control*, 31:189–198, 2017.
- 31 495 Rania Trigui, Johel Miteran, Lamia Sellami, Paul Walker, and Ahmed Ben Hamida. A classification approach to prostate cancer localization in 3t multi-parametric mri. In *Advanced Technologies for Signal and Image Processing (ATSIP), 2016 2nd International Conference on*, pages 113–118. IEEE, 2016.
- 32 505 Stefan Van der Walt, Johannes L Sch  nberger, Juan Nunez-Iglesias, Fran  ois Boulogne, Joshua D Warner, Neil Yager, Emmanuelle Gouillart, and Tony Yu. scikit-image: image processing in python. *PeerJ*, 2:e453, 2014.
- 33 510 St  fan van der Walt, S Chris Colbert, and Gael Varoquaux. The numpy array: a structure for efficient numerical computation. *Computing in Science & Engineering*, 13(2):22–30, 2011.
- 34 515 Yingjie Wang and Chin-Seng Chua. Face recognition from 2d and 3d images using 3d gabor filters. *Image and vision computing*, 23(11):1018–1028, 2005.
- 35 520 Jeffrey C Weinreb, Jelle O Barentsz, Peter L Choyke, Francois Cornud, Masoom A Haider, Katarzyna J Macura, Daniel Margolis, Mitchell D Schnall, Faina Shtern, Clare M Tempany, et al. Pi-rads prostate imaging-reporting and data system: 2015, version 2. *European urology*, 69(1):16–40, 2016.

³⁶ Yuanxin Xi and David M Rocke. Baseline correction for nmr spectroscopic metabolomics data analysis. *BMC bioinformatics*, 9(1):1, 2008.

³⁷ <https://github.com/I2Cvb/mp-mri-prostate>

⁵¹⁰ ³⁸ <https://zenodo.org/record/162231##.W3AAMBh9hhE>