Computer-Aided Diagnosis Systems for Prostate Cancer Detection: Challenges and Methodologies

Guillaume Lemaître, Robert Martí, and Fabrice Meriaudeau

Abstract

Prostate cancer (CaP) is the second most diagnosed cancer in men all over the world. CaP growth is characterized by two main types of evolution: (i) the slow-growing tumours progress slowly and usually remain confined to the prostate gland; (ii) the fast-growing tumours metastasize from prostate gland to other organs, which might lead to incurable diseases. Therefore, early diagnosis and risk assessment play major roles in patient treatment and follow-up. In the last decades, new imaging techniques based on Magnetic Resonance Imaging (MRI) have been developed improving diagnosis. In practise, diagnosis can be affected by multiple factors such as observer variability and visibility and complexity of the lesions. In this regard, computer-aided detection and computer-aided diagnosis systems are being designed to help radiologists in their clinical practice. This current chapter analyzes the current state-of-the-art in the development of computer-aided diagnosis and detection systems for prostate cancer detection.

I. Introduction

A. Prostate anatomy

The prostate is an exocrine gland of the male reproductive system having an inverted pyramidal shape, which is located below the bladder and in front of the rectum as shown in Fig. 1. It measures approximately $3\,\mathrm{cm}$ in height by $2.5\,\mathrm{cm}$ in depth and its weight is estimated from $7\,\mathrm{g}$ to $16\,\mathrm{g}$ for an adult [1]. The prostate size increases at two distinct stages during physical development: initially at puberty to reach its normal size, then again after 60 years of age leading to benign prostatic hyperplasia (BPH) [2].

A zonal classification of the prostate has been suggested by McNeal [3], as depicted in Fig. 2. Subsequently, this categorization has been widely accepted in the literature [2, 4, 5, 6] and is used during all medical examinations (e.g., biopsy, magnetic resonance imaging (MRI) screening). The classification is based on dividing the gland into 3 distinct regions: (i) the central zone (CZ) accounting for $20\,\%$ to $25\,\%$ of the whole prostate gland, (ii) the transitional zone (TZ) standing for $5\,\%$, and (iii) the peripheral zone (PZ) representing the $70\,\%$. In MRI images, tissues of CZ and TZ are difficult to distinguish and are usually merged into a common region, denominated central gland (CG). As part of this classification, the prostate is divided into 3 longitudinal portions depicted in Fig. 2b: base, median gland, and apex.

B. Prostate carcinoma

Prostate cancer prostate cancer (CaP) has been reported on a worldwide scale to be the second most frequently diagnosed cancer of men accounting for $13.6\,\%$ [9]. Statistically, in 2008, the number of new diagnosed cases has been estimated to be $899,000$ with no less than $258,100$ deaths [9]. In United States, aside from skin cancer, CaP is declared to be the most commonly diagnosed cancer among men, implying that approximately 1 in 6 men will be diagnosed with CaP during their lifetime and 1 in 36 will die from this disease, causing CaP to be the second most common cause of cancer death among men [10, 11].

Despite active research to determine the causes of CaP, a fuzzy list of risk factors has arisen [12]. The etiology has been linked to the following factors [12]: (i) family history [13, 14], (ii) genetic factors [15,

G. Lemaître is with CDS Paris-Saclay, Parietal team, Inria, CEA, Université Paris-Saclay, 1 Rue Honoré d'Estienne d'Orves, 91120 Palaiseau, France e-mail: guillaume.lemaitre@inria.fr

R. Martí is with ViCOROB, Universitat de Girona, Campus Montilivi, Edifici P4, 17071 Girona, Spain.

F. Mériaudeau is with CISIR, Electrical & Electronic Engineering Department, Universiti Teknologi Petronas, 32610 Seri Iskandar, Perak, Malaysia.

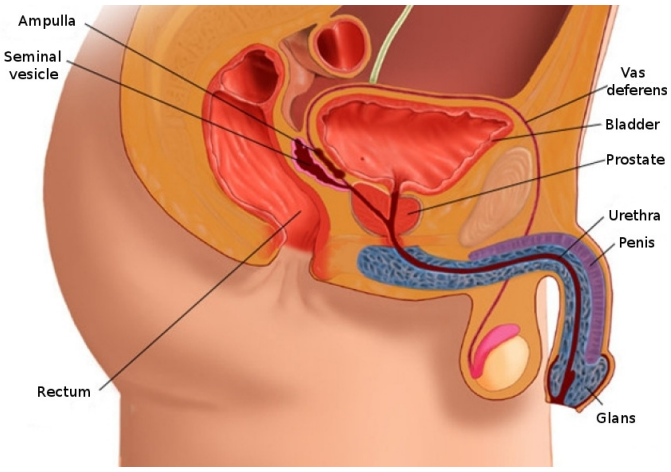
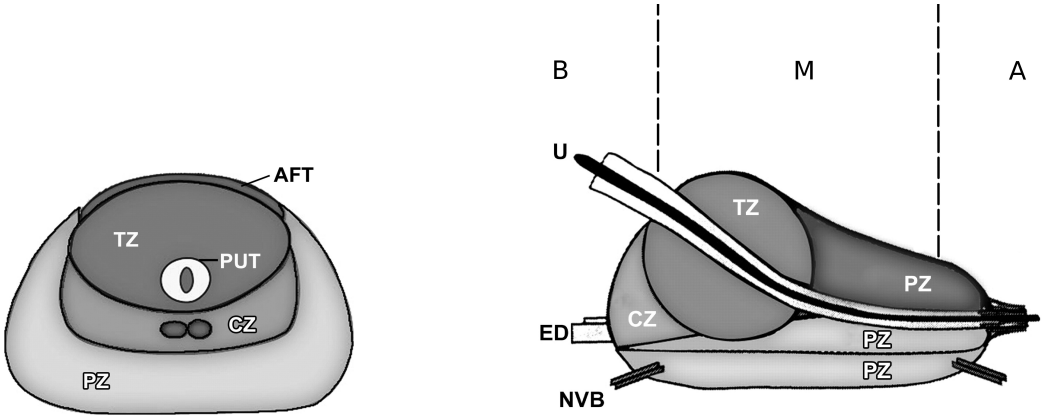


Fig. 1. Sagittal anatomy scheme of the male reproductive system (copyright by [7]).



(a) Transverse anatomy of the prostate.

(b) Sagittal anatomy of the prostate.

Fig. 2. Prostate anatomy with division in different zones. *AFT:* anterior fibromuscular tissue, *CZ:* central zone, *ED:* ejaculatory duct, *NVB:* neurovascular bundle, *PUT:* periurethral tissue, *PZ:* peripheral zone, *U:* urethra, *TZ:* transitional zone, *B:* base, *M:* median, *A:* apex (copyright by [8]).

16, 17], (iii) race-ethnicity [13, 18], (iv) diet [13, 19, 20], and (v) obesity [13, 21]. This list of risk factors alone cannot be used to diagnose CaP and in this way, screening enables early detection and treatment.

CaP growth is characterized by two main types of evolution [22]: slow-growing tumours, accounting for up to $85\,\%$ of all CaPs [23], progress slowly and usually stay confined to the prostate gland. For such cases, treatment can be substituted with active surveillance. In contrast, the second variant of CaPs develops rapidly and metastasises from prostate gland to other organs, primarily the bones [24]. Bone metastases, being an incurable disease, significantly affects the morbidity and mortality rate [25]. Hence, the results of the surveillance have to be trustworthy to distinguish aggressive from slow-growing CaP.

CaP is more likely to come into being in specific regions of the prostate. In that respect, around $70\,\%$ to $80\,\%$ of CaPs originate in PZ whereas $10\,\%$ to $20\,\%$ in TZ [26, 27, 28]. Only about $5\,\%$ of CaPs occur in CZ [27, 29]. However, those cancers appear to be more aggressive and more likely to invade other organs due to their locations [29].

C. CaP screening and imaging techniques

Current CaP screening consists of three different stages. First, prostate-specific antigen (PSA) control is performed to distinguish between low- and high-risk CaP. To assert such diagnosis, samples are taken

during prostate biopsy and finally analyzed to evaluate the prognosis and the stage of CaP. In this section, we present a detailed description of the current screening as well as its drawbacks.

Since its introduction in mid-1980s, PSA is widely used for CaP screening [30]. A higher-than-normal level of PSA can indicate an abnormality of the prostate either as a BPH or a cancer [31]. However, other factors can lead to an increased PSA level such as prostate infections, irritations, a recent ejaculation, or a recent rectal examination [2]. PSA is found in the bloodstream in two different forms: free PSA accounting for about $10\,\%$ and linked to another protein for the remaining $90\,\%$. A level of PSA higher than $10\,\mathrm{ng\,mL^{-1}}$ is considered to be at risk [2]. If the PSA level is ranging from $4\,\mathrm{ng\,mL^{-1}}$ to $10\,\mathrm{ng\,mL^{-1}}$, the patient is considered as suspicious [32]. In that case, the ratio of free PSA to total PSA is computed; if the ratio is higher than $15\,\%$, the case is considered as pathological [2].

A transrectal ultrasound (TRUS) biopsy is carried out for cases which are considered pathological. At least 6 different samples are taken randomly from the right and left parts of the 3 different prostate zones: apex, median, and base. These samples are further evaluated using the Gleason grading system [33]. The scoring scheme to characterize the biopsy sample is composed of 5 different patterns which correspond to grades ranging from 1 to 5. A higher grade is associated with a poorer prognosis [34]. Then, in the Gleason system, 2 scores are assigned corresponding to (i) the grade of the most present tumour pattern, and (ii) the grade of the second most present tumour pattern [34]. A higher Gleason score (GS) indicates a more aggressive tumour [34]. Also, it should be noted that biopsy is an invasive procedure which can result in serious infection or urine retention [35, 36].

Although PSA screening has been shown to improve early detection of CaP [36], its lack of reliability motivates further investigations using MRI-based computer-aided detection and diagnosis (CAD). Two reliable studies — carried out in the United States [37] and in Europe [38, 39] — have attempted to assess the impact of early detection of CaP, with diverging outcomes [36, 40]. The study carried out in Europe[1] concluded that PSA screening reduces CaP-related mortality by $21\,\%$ to $44\,\%$ [38, 39], while the American[2] trial found no such effect [37]. However, both studies agree that PSA screening suffers from low specificity, with an estimated rate of $36\,\%$ [41]. Both studies also agree that over-treatment is an issue: decision making regarding treatment is further complicated by difficulties in evaluating the aggressiveness and progression of CaP [42].

Hence, new screening methods should be developed with improved specificity of detection and more accurate risk assessment (i.e., aggressiveness and progression). Current research is focused on identifying new biological markers to replace PSA-based screening [43, 44, 45]. Until such research comes to fruition, these needs can be met through active-surveillance strategy using multiparametric MRI (mp-MRI) techniques [31, 46]. CAD systems, which is an area of active research and forms the focus of this chapter, can be incorporated into this screening strategy allowing a more systematic and rigorous follow-up.

Another weakness of the current screening strategy is due to the untrustworthy results provided by TRUS biopsy. Due to its "blind" nature, there is a chance of missing aggressive tumours or detecting microfocal "cancers", which influences the aggressiveness-assessment procedure [47]. As a consequence, over-diagnosis is estimated at up to $30\,\%$ [48], while missing clinically significant CaP is estimated at up $35\,\%$ [49]. In an effort to solve both issues, alternative biopsy approaches have been explored. MRI/ultrasound (UTS)-guided biopsy has been shown to outperform standard TRUS biopsy [50]. There, mp-MRI images are fused with UTS images in order to improve localization and aggressiveness assessment to carry out biopsies. Human interaction plays a major role in biopsy sampling which can lead to low repeatability; by reducing potential human errors at this stage, the CAD framework can improve repeatability of examination. CaP detection and diagnosis can benefit from the use of MRI-based CADs.

[1]The European randomized study of screening for prostate cancer (ERSSPC) started in the 1990s in order to evaluate the effect of PSA screening on mortality rate.

[2]The prostate lung colorectal and ovarian (PLCO) cancer screening trial is carried out in the United States and intends to ascertain the effects of screening on mortality rate.

In an effort to improve the current stage of CaP diagnosis and detection, this chapter intends to review the principles of a mp-MRI-CAD system starting by a description of the MRI modalities in Sect. II.

D. CAD systems for CaP

During the last century, physicists have focused on constantly innovating in terms of imaging techniques assisting radiologists to improve cancer detection and diagnosis. However, human diagnosis still suffers from low repeatability, synonymous with erroneous detection or interpretations of abnormalities throughout clinical decisions [51, 52]. These errors are driven by two majors causes [51]: observer limitations (e.g., constrained human visual perception, fatigue or distraction) and the complexity of the clinical cases themselves, for instance due to imbalanced data — the number of healthy cases is more abundant than malignant cases — or overlapping structures.

Computer vision has given rise to many promising solutions, but, instead of focusing on fully automatic computerized systems, researchers have aimed at providing computer image analysis techniques to aid radiologists in their clinical decisions [51]. In fact, these investigations brought about both concepts of computer-aided detection (CADe) and computer-aided diagnosis (CADx) grouped under the acronym CAD. Since those first steps, evidence has shown that CAD systems enhance the diagnosis performance of radiologists. Chan et al. reported a significant $4\,\%$ improvement in breast cancer detection [53], which has been confirmed in later studies [54]. Similar conclusions have been drawn in the case of lung nodule detection [55], colon cancer [56], or CaP as well [52]. Chan et al. also hypothesized that CAD systems will be even more efficient assisting inexperienced radiologists than senior radiologists [53]. That hypothesis has been tested by Hambrock et al. and confirmed in case of CaP detection [52]. In this particular study, inexperienced radiologists obtained equivalent performance to senior radiologists, both using CAD whereas the accuracy of their diagnosis was significantly poorer without CAD's help.

In contradiction with the aforementioned statement, CAD for CaP is a young technology due to the fact that is based on a still young imaging technology: MRI [57]. Indeed, four distinct MRI modalities are employed in CaP diagnosis which have been mainly developed after the mid-1990s: (i) T_2 Weighted (T_2-W)-MRI [58], (ii) dynamic contrast-enhanced (DCE)-MRI [59], (iii) magnetic resonance spectroscopy imaging (MRSI) [60], and (iv) diffusion weighted (DW)-MRI [61]. In addition, the increase of magnetic field strength in clinical settings, from $1.5\,\mathrm{T}$ to $3\,\mathrm{T}$, and the development of endorectal coils, both improved image spatial resolution [62] needed to perform more accurate diagnosis. It is for this matter that the development of CAD for CaP is still lagging behind the fields stated above.

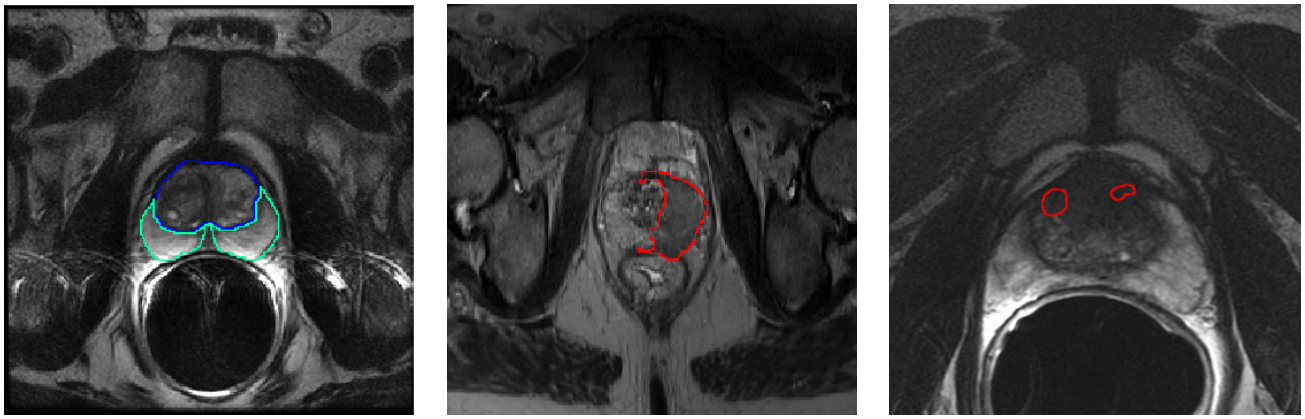
The further sections aim at first, to provide an overview of the current state-of-the-art of CAD for CaP and later, according to the drawn conclusions, to propose a CAD which takes advantages of mp-MRI modalities. A review of the current proposed CAD for CaP is presented in Sect. III.

II. MRI Techniques

MRI provides promising imaging techniques to overcome the drawbacks of current clinical screening techniques mentioned in Sect. I. Unlike TRUS biopsy, MRI examination is a non-invasive protocol and has been shown to be the most accurate and harmless technique currently available [63]. In this section, we review different MRI imaging techniques developed for CaP detection and diagnosis. Features strengthening each modality will receive particular attention together with their drawbacks. Commonly, these features form the basis for developing analytic tools and automatic algorithms. However, we refer the reader to Sect. III-E for more details on automatic feature detection methods since they are part and parcel of the CAD framework.

A. T_2-W-MRI

T_2-W-MRI has been the first MRI-modality used to perform CaP diagnosis using MRI [58]. Nowadays, radiologists make use of it for CaP detection, localization, and staging purposes. This imaging technique is well suited to render zonal anatomy of the prostate [32].



(a) T_2-W-MRI slice of a healthy prostate acquire with a 1.5 T MRI with an endorectal coil. The blue contour represents the CG while the PZ corresponds to the green contour.

(b) T_2-W-MRI slice of a prostate with a CaP highlighted in the PZ using a 3 T MRI scanner without an endorectal coil.

(c) T_2-W-MRI slice of a prostate with a CaP highlighted in the CG using a 1.5 T MRI scanner with an endorectal coil.

Fig. 3. Rendering of T_2-W-MRI prostate image with both 1.5 T and 3 T MRI scanner.

This modality relies on a sequence based on setting a long repetition time (TR), reducing the T_1 effect in nuclear magnetic resonance (NMR) signal measured, and fixing the echo time (TE) to sufficiently large values in order to enhance the T_2 effect of tissues. Thus, PZ and CG tissues are well perceptible in these images. The former is characterized by an intermediate/high-signal intensity (SI) while the latter is depicted by a low-SI [4]. An example of a healthy prostate is shown in Fig. 3a.

In PZ, round or ill-defined low-SI masses are synonymous with CaPs [58] as shown in Fig. 3b. Detecting CaP in CG is more challenging. Both normal CG tissue and malignant tissue, have a low-SI in T_2-W-MRI, reinforcing difficulties to distinguish one among them. However, CaPs in CG appear often as homogeneous mass having ill-defined edges with lenticular or "water-drop" shapes [32, 64] as depicted in Fig. 3c.

CaP aggressiveness has been shown to be inversely correlated with SI. Indeed, CaPs assessed with a GS of 4-5 implied lower SI than the one with a GS of 2-3 [65].

In spite of the availability of these useful and encouraging features, the T_2-W modality lacks reliability [31, 66]. Sensitivity is affected by the difficulties in detecting cancers in CG [66] while specificity rate is highly affected by outliers [32]. In fact, various conditions emulate patterns of CaP such as BPH, post-biopsy hemorrhage, atrophy, scars, and post-treatment [4, 32, 61, 67, 68]. These issues are partly addressed using more innovative and advanced modalities.

B. T_2 map

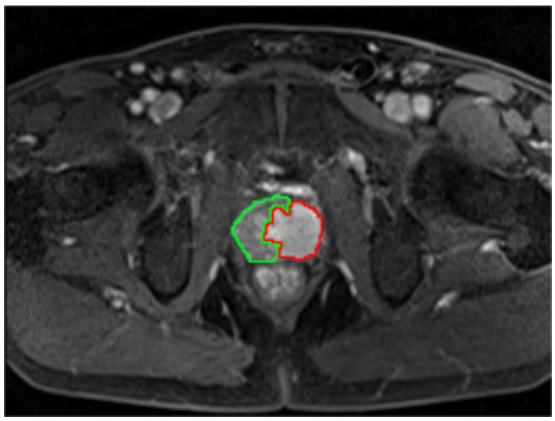
As previously mentioned, T_2-W-MRI modality shows low sensitivity. Moreover, T_2-W-MRI images are a composite of multiple effects [57]. However, T_2 values alone have been shown to be more discriminative [69] and highly correlated with citrate concentration, a biological marker in CaP [70, 71]. T_2 values are computed using the characteristics of transverse relaxation which is formalized as in Eq. (1).

$$M_{xy}(t) = M_{xy}(0) \exp\left(-\frac{t}{\mathrm{T}_2}\right) \ , \tag{1}$$

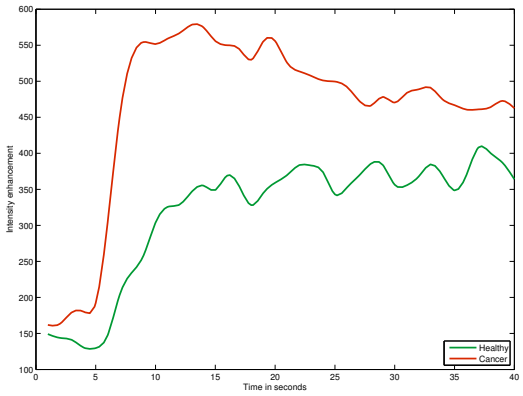
where $M_{xy}(0)$ is the initial value of $M_{xy}(t)$ and T_2 is the relaxation time.

By rearranging Eq. (1), T_2 map is computed by performing a linear fitting on the model presented in Eq. (2) using several TE, $t = \{\mathrm{TE}_1, \mathrm{TE}_2, \ldots, \mathrm{TE}_m\}$.

$$\ln\left[\frac{M_{xy}(t)}{M_{xy}(0)}\right] = -\frac{t}{\mathrm{T}_2} \ . \tag{2}$$



(a) T_1-W-MRI image where the cancer is delimited by the red contour. The green area was still not invaded by the CaP



(b) Enhancement curve computed during the DCE-MRI analysis. The red curve is typical from CaP cancer while the green curve is characteristic of healthy tissue.

Fig. 4. Illustration of typical enhancement signal observed in DCE-MRI analysis collected with a 3 T MRI scanner.

The Fast Spin-Echo (FSE) sequence has been shown to be particularly well suited in order to build a T_2 map and obtain accurate T_2 values [72]. Similar to T_2-W-MRI, T_2 values associated with CaP are significantly lower than those of healthy tissues [70, 73].

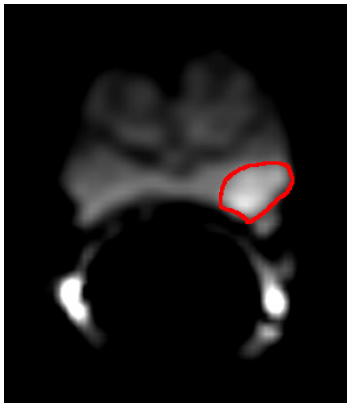
C. DCE-MRI

DCE-MRI is an imaging technique which exploits the vascularity characteristic of tissues. Contrast media, usually gadolinium-based, is injected intravenously into the patient. The media extravasates from vessels to extravascular-extracellular space (EES) and is released back into the vasculature before being eliminated by the kidneys [74]. Furthermore, the diffusion speed of the contrast agent may vary due to several parameters: the permeability of the micro-vessels, their surface area, and the blood flow [75].

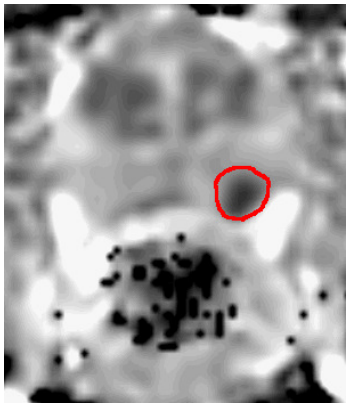
Healthy PZ is mainly made up of glandular tissue, around 70 % [8], which implies a reduced interstitial space restricting exchanges between vessels and EES [76, 77]. Normal CG has a more disorganized structure, composed of mainly fibrous tissue [8, 31], which facilitates the arrival of the contrast agent in EES [78]. To understand the difference between contrast media kinetic in malignant tumours and the two previous behaviours mentioned, one has to focus on the process known as angiogenesis [79]. In order to ensure growth, malignant tumours produce and release angiogenic promoter substances [79]. These molecules stimulate the creation of new vessels towards the tumour [79]. However, the new vessel networks in tumours differ from those present in healthy tissue [74]. They are more porous due to the larger number of "openings" in their capillary walls [8, 74]. In contrast to healthy cases, this increased vascular permeability results in increased contrast agent exchanges between vessels and EES [80].

By making use of the previous aspects, DCE-MRI is based on an acquisition of a set of T_1 Weighted (T_1-W)-MRI images over time. The gadolinium-based contrast agent shortens T_1 relaxation time enhancing contrast in T_1-W-MRI images. The aim is to post-analyze the pharmacokinetic behaviour of the contrast media concentration in prostate tissues [80]. The image analysis is carried out in two dimensions: (i) in the spatial domain on a pixel-by-pixel basis and (ii) in the time domain corresponding to the consecutive images acquired with the MRI. Thus, for each spatial location, a signal linked to contrast media concentration is measured as shown in Fig. 4b [81].

By taking the above remarks into account, CaPs is characterized by a signal having an earlier and faster enhancement and an earlier wash-out — i.e, the rate of the contrast agent flowing out of the tissue — as shown in Fig. 4b [80]. Three different approaches exist to analyze these signals with the aim of labelling them as corresponding to either normal or malignant tissues.



(a) DW-MRI image acquired with a 1.5 T MRI scanner. The cancer corresponds to the high SI region highlighted in red.



(b) ADC map computer after acquisition of DW-MRI images with 1.5 T MRI scanner. The cancer corresponds to the low SI region highlighted in red.

Fig. 5. Illustration of of DW-MRI and ADC map. The signal intensity corresponding to cancer are inversely correlated on these modalities.

Qualitative analysis is based on a qualitative assessment of the signal shape [31]. Quantitative approaches consist of inferring pharmocokinetic parameter values [81]. Those parameters are part of mathematical-pharmacokinetic models which are directly based on physiological exchanges between vessels and EES. Several pharmacokinetic models have been proposed such as the Kety model [82], the Tofts model [83], and mixed models [84, 85]. The last family of methods mixed both approaches and are grouped together under the heading of semi-quantitative methods. They rely on shape characterization using mathematical modelling to extract a set of parameters such as wash-in gradient, wash-out, integral under the curve, maximum signal intensity, time-to-peak enhancement, and start of enhancement [31, 80]. These parameters are depicted in Fig. 18. It has been shown that semi-quantitative and quantitative methods improve localization of CaP when compared with qualitative methods [86]. Sect. III-E2 provides a full description of quantitative and semi-quantitative approaches.

DCE-MRI combined with T_2-W-MRI has shown to enhance sensitivity compared to T_2-W-MRI alone [87, 88, 89, 90]. Despite this fact, DCE-MRI possesses some drawbacks. Due to its "dynamic" nature, patient motions during the image acquisition lead to spatial mis-registration of the image set [80]. Furthermore, it has been suggested that malignant tumours are difficult to distinguish from prostatitis located in PZ and BPH located in CG [31, 80]. These pairs of tissues tend to have similar appearances. Later studies have shown that CaPs in CG do not always manifest in homogeneous fashion. Indeed, tumours in this zone can present both hypo-vascularization and hyper-vascularization which illustrates the challenge of CaP detection in CG [78].

D. DW-MRI

As previously mentioned in the introduction, DW-MRI is the most recent MRI imaging technique aiming at CaP detection and diagnosis [61]. This modality exploits the variations in the motion of water molecules in different tissues [91, 92].

The distinction between healthy and CaP in DW-MRI is based on the following physiological bases. On the one hand, PZ, as previously mentioned, is mainly a glandular and tubular structure allowing water molecules to move freely [8, 31]. On the other hand, CG is made up of muscular or fibrous tissue causing the motion of the water molecules to be more constrained and heterogeneous than in PZ [31]. Then, CaP growth leads to the destruction of normal glandular structure and is associated with an increase in cellular density [31, 92, 93]. Furthermore, these factors both have been shown to be inversely correlated with

water diffusion [92, 93]: higher cellular density implies a restricted water diffusion. Thus, water diffusion in CaP will be more restricted than both healthy PZ and CG [31, 92].

From the NMR principle side, DW-MRI sequence produces contrasted images due to variation of water molecules motion. The method is based on the fact that the signal in DW-MRI images is inversely correlated to the degree of random motion of water molecules [94]. In fact, gradients are used in DW MRI modality to encode spatial location of nuclei temporarily. Simplifying the problem in only one direction, a gradient is applied in that direction, dephasing the spins of water nuclei. Hence, the spin phases vary along the gradient direction depending of the gradient intensity at those locations. Then, a second gradient is applied aiming at cancelling the spin dephasing. Thus, the immobile water molecules will be subject to the same gradient intensity as the initial one while moving water molecules will be subject to a different gradient intensity. Thus, spins of moving water molecules will stay dephased whereas spins of immobile water molecules will come back in phase. As a consequence, a higher degree of random motion results in a more significant signal loss whereas a lower degree of random motion is synonymous with lower signal loss [94]. Under these conditions, the MRI signal is measured as:

$$M_{x,y}\left(t,b\right) = M_{x,y}(0)\exp\left(-\frac{t}{\mathrm{T}_2}\right) S_{\mathrm{ADC}}(b) \ , \tag{3}$$

$$S_{\mathrm{ADC}}(b) = \exp\left(-b \times \mathrm{ADC}\right) \ , \tag{4}$$

where S_{ADC} refers to signal drop due to diffusion effect, ADC is the apparent diffusion coefficient, and b is the attenuation coefficient depending only on of the gradient intensity and the gradient duration [95].

By using this formulation, image acquisition with a parameter b equal to $0\,\mathrm{s\,mm^{-2}}$ corresponds to a T_2-W-MRI acquisition. Then, increasing the attenuation coefficient b — i.e., increase gradient intensity and duration — enhances the contrast in DW-MRI images.

To summarize, in DW-MRI, CaPs are characterized by high-SI compared to normal tissues in PZ and CG as shown in Fig. 5a [32]. However, some tissues in CG looks similar to CaP with higher SI [32].

Diagnosis using DW-MRI combined with T_2-W-MRI has shown a significant improvement compared with T_2-W-MRI alone and provides highly contrasted images [8, 96, 97]. As drawbacks, this modality suffers from poor spatial resolution and specificity due to false positive detection [8]. With a view to eliminate these drawbacks, radiologists use quantitative maps extracted from DW-MRI, which is presented in the next section.

E. ADC map

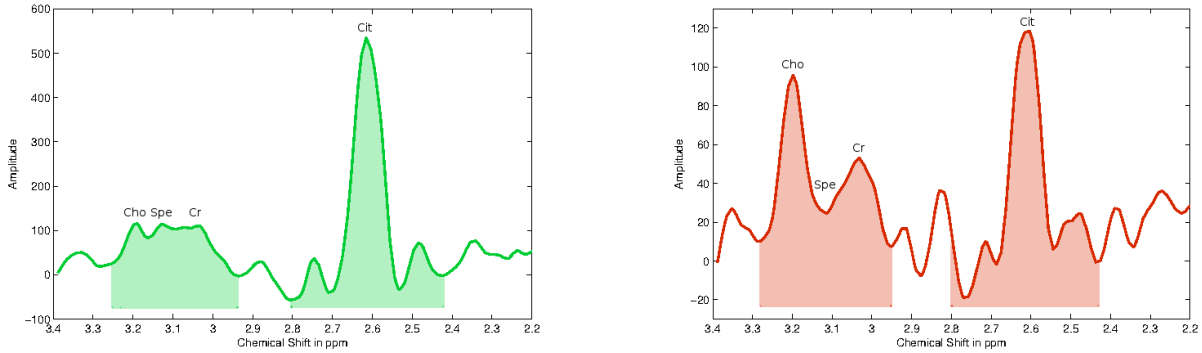
The NMR signal measured for DW-MRI images is not only affected by diffusion as shown in Eq. (3). However, the signal drop — Eq. (4) — is formulated such that the only variable is the acquisition parameter b [95]. The ADC is considered as a "pure" diffusion coefficient and is extracted to build a quantitative map known as the ADC map. From Eq. (3), it is clear that performing multiple acquisitions only varying b will not have any effect on the term $M_{x,y}(0)\exp\left(-\frac{t}{\mathrm{T}_2}\right)$. Thus, Eq. (3) can be rewritten as:

$$S(b) = S_0 \exp\left(-b \times \mathrm{ADC}\right) \ . \tag{5}$$

To compute the ADC map, a minimum of two acquisitions are necessary: (i) for b equal to $0\,\mathrm{s\,mm^{-2}}$ where the measured signal is equal to S_0, and (ii) b_1 greater than $0\,\mathrm{s\,mm^{-2}}$, typically $1000\,\mathrm{s\,mm^{-2}}$. Then, the ADC map can be computed as:

$$\mathrm{ADC} = -\frac{\ln\left(\dfrac{S(b_1)}{S_0}\right)}{b_1} \ . \tag{6}$$

More accurate ADC maps are computed by acquiring a set of images with different values for the parameter b and fitting linearly a semi-logarithm function using the model presented in Eq. (5).



(a) Illustration of an MRSI spectrum of a healthy voxel acquired with a 3 T MRI.

(b) Illustration of an MRSI spectrum of a cancerous voxel acquired with a 3 T MRI.

Fig. 6. Illustration of an MRSI spectrum for both healthy and cancerous voxels with a 3 T MRI. The highlighted areas correspond to the related concentration of the metabolites which is computed by integrating the area under each peak. Acronyms: choline (Cho), spermine (Spe), creatine (Cr) and citrate (Cit).

Regarding the appearance of the ADC maps, it has been previously stated that by increasing the value of b, the signal of CaP tissue increases significantly. Considering Eq. (6), the tissue appearance in the ADC map is the inverse of DW-MRI images. Then, CaP tissue is associated with low-SI whereas healthy tissue appears brighter as depicted in Fig. 5b [32].

Similar to the gain achieved by DW-MRI, diagnosis using ADC map combined with T_2-W-MRI significantly outperforms T_2-W-MRI alone [8, 98]. Moreover, it has been shown that ADC coefficient is correlated with GS [99, 100, 101].

However, some tissues of the CG mimic CaP with low-SI [66] and image distortion can arise due to hemorrhage [8]. It has also been noted that a high variability of the ADC occurs between different patients making it difficult to define a static threshold to distinguish CaP from non-malignant tumours [8].

F. MRSI

CaP induces metabolic changes in the prostate compared with healthy tissue. Thus, CaP detection can be carried out by tracking changes of metabolite concentration in prostate tissue. MRSI is an NMR-based technique which generates spectra of relative metabolite concentration in a region of interest (ROI).

In order to track changes of metabolite concentration, it is important to know which metabolites are associated with CaP. To address this question, clinical studies identified three biological markers: (i) citrate, (ii) choline, and (iii) polyamines composed mainly of spermine, and in less abundance of spermidine and putrescine [102, 103, 104].

Citrate is involved in the production and secretion of the prostatic fluid, and the glandular prostate cells are associated with a high production of citrate enabled by zinc accumulation by these same cells [103]. However, the metabolism allowing the accumulation of citrate requires a large amount of energy [103]. In contrast, malignant cells do not have high zinc levels leading to lower citrate levels due to citrate oxidization [103]. Furthermore, this change results in a more energy-efficient metabolism enabling malignant cells to grow and spread [103].

An increased concentration of choline is related to CaP [102]. Malignant cell development requires epigenetic mechanisms resulting in metabolic changes and relies on two mechanisms: deoxyribonucleic acid (DNA) methylation and phospholid metabolism which both result in choline uptake, explaining its increased level in CaP tissue [102]. Spermine is also considered as a biological marker in CaP [104, 105]. In CaP, reduction of the ductal volume due to shifts in polyamine homeostasis might lead to a reduced spermine concentration [105].

To determine the concentration of these biological markers, one has to focus on the MRSI modality. In theory, in presence of a homogeneous magnetic field, identical nuclei precesses at the same operating frequency known as the Lamor frequency [106]. However, MRSI is based on the fact that identical nuclei will slightly precess at different frequencies depending on the chemical environment in which they are immersed [106], a phenomenon known as the chemical shift effect (CSE) [2]. Given this property, metabolites are identified and their concentrations are determined. In this regard, the Fourier transform is used to obtain the frequency spectrum of the NMR signal [2, 106]. In this spectrum, each peak is associated with a particular metabolite and the area under each peak corresponds to the relative concentration of this metabolite, as illustrated in Fig. 6 [2].

Two different quantitative approaches are used to decide whether or not the spectra of a ROI is associated with CaP: (i) relative quantification or (ii) absolute quantification [107]. In relative quantification, the ratio of choline-polyamines-creatine to citrate is computed. The integral of the signal is computed from choline to creatine — i.e., from $3.21\,\mathrm{ppm}$ to $3.02\,\mathrm{ppm}$ — because the peaks in this region can be merged at clinical magnetic field strengths [31, 105], as depicted in Fig. 6). Considering the previous assumptions that choline concentration rises and citrate concentration decreases in the presence of CaP, the ratio computed should be higher in malignant tissue than in healthy tissue.

In contrast with relative quantification, absolute quantification measures molar concentrations by normalizing relative concentrations using water as reference [107]. In this case, "true" concentrations are directly used to differentiate malignant from healthy tissue. However, this method is not commonly used as it requires an additional step of acquiring water signals, inducing time and cost acquisition constraints.

MRSI allows examination with high specificity and sensitivity compared to other MRI modalities [8]. Furthermore, it has been shown that combining MRSI with MRI improves detection and diagnosis performance [108, 109, 110]. Citrate and spermine concentrations are inversely correlated with the GS allowing us to distinguish low- from high- grade CaPs [104]. However, choline concentration does not provide the same properties [104].

Unfortunately, MRSI also presents several drawbacks. First, MRSI acquisition is time consuming which prevents this modality from being used in daily clinical practise [32]. In addition, MRSI suffers from low spatial resolution due to the fact that signal-to-noise (SNR) is linked to the voxel size. However, this issue is addressed by developing new scanners with higher magnetic field strengths such as $7.5\,\mathrm{T}$ [104]. Finally, a high variability of the relative concentrations between patients has been observed [8]. The same observation has been made depending on the zones studied (ie., PZ, CG, base, mid-gland, apex) [107, 111]. Due to this variability, it is difficult to use a fixed threshold to differentiate CaP from healthy tissue.

G. Summary and conclusions

Table I provides an overview of the different modalities presented in the previous section. Indeed, each MRI modality alone provides a different discriminative level to distinguish CaP from healthy tissue. A recurrent statement in the literature is, however, the ability to combine these MRI modalities would lead to the best diagnosis performance. In this regard, we will present in the next chapter automatic tools which have been developed to design mp-MRI CAD systems for the detection of CaP.

III. Review of CAD sytems for CaP

As previously mentioned Sect. I-D, CADs are developed to advise and backup radiologists in their tasks of CaP detection and diagnosis, but not to provide fully automatic decisions [51]. CADs can be divided into two different sub-groups: either as CADe, with the purpose to highlight probable lesions in MRI images, or CADx, which focuses on differentiating malignant from non-malignant tumours [51]. Moreover, an intuitive approach, motivated by developing a framework combining detection-diagnosis, is to mix both CADe and CADx by using the output of the former mentioned as a input of the latter named. Although the outcomes of these two systems should differ, the framework of both CAD systems is similar. A general CAD work-flow is presented in Fig. 7.

TABLE I
OVERVIEW OF THE FEATURES ASSOCIATED WITH EACH MRI MODALITY USED FOR MEDICAL DIAGNOSIS BY RADIOLOGISTS. ACRONYMS: PROSTATE CANCER (CAP) - SIGNAL INTENSITY (SI) - GLEASON SCORE (GS).

Modality	Significant features	CaP	Healthy tissue	GS correlation
T_2-W-MRI	SI	low-SI in PZ [4]	intermediate to high-SI in PZ [4]	+ [65]
	Shape	round or ill-defined mass in PZ [58]		0
	SI	low-SI in CG [32, 64]	low-SI in CG [32, 64]	0
	Shape	homogeneous mass with ill-defined edges in CG [32, 64]		0
T_2 map	SI	low-SI [70, 73]	intermediate to high-SI [70, 73]	+ [69, 70, 71]
DCE MRI	Semi-quantitative features [80]:			
	• wash-in	faster	slower	0
	• wash-out	faster	slower	0
	• integral under the curve	higher	lower	0
	• maximum signal intensity	higher	lower	0
	• time-to-peak enhancement	faster	slower	0
	Quantitative features (Tofts' parameters [81]):			
	• k_{ep}	higher	lower	0
	• K^{trans}	higher	lower	0
DW MRI	SI	higher-SI [32, 94]	lower-SI [32, 94]	+
ADC map	SI	low-SI [32]	high-SI [32]	+ [99, 100, 101]
MRSI	Metabolites:			
	• citrate (2.64 ppm) [112]	lower concentration [102, 103, 105]	higher concentration [102, 103, 105]	+ [104]
	• choline (3.21 ppm) [112]	higher concentration [102, 103, 105]	lower concentration [102, 103, 105]	0 [104]
	• spermine (3.11 ppm) [112]	lower concentration [102, 103, 105]	higher concentration [102, 103, 105]	+ [104]

Notes:
+ = significantly correlated;
0 = no correlation.

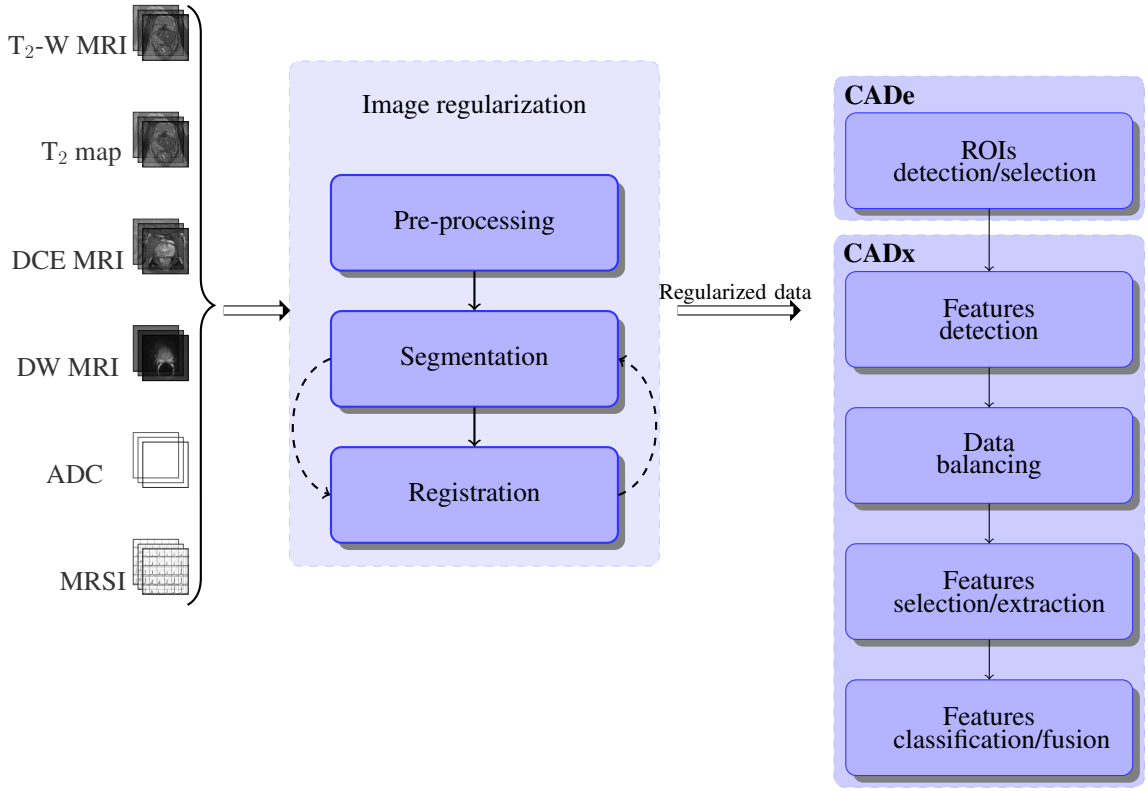


Fig. 7. Common CAD framework based on MRI images used to detect CaP.

MRI modalities mentioned in Sect. II are used as inputs of CAD for CaP. These images acquired from the different modalities show a large variability between patients: the prostate organ can be located at different positions in images — due to patient motion, variation of acquisition plan — and the SI can be corrupted with noise or artifacts during the acquisition process caused by the magnetic field non-homogeneity or the use of endorectal coil. To address these issues, the first stage of CAD is to pre-process mp-MRI images to reduce noise, remove artifacts, and standardize the SI. Subsequently, most of the later processes are only focusing on the prostate organ; therefore it is necessary to segment the prostate in each MRI modality to define it as a ROI. However, data may suffer from misalignment due to patient motions or different acquisition parameters. Therefore, a registration step is usually performed so that all the previously segmented MRI images are in the same reference frame. Registration and segmentation can be swapped depending on the strategy chosen.

Some studies do not fully apply the methodology depicted in Fig. 7. Details about those can be found in Table II. Some studies bypass the pre-processing stages to proof the robustness of their approaches to noise or other artifacts, by using directly the raw data as inputs of their CAD systems. In some cases, prostate segmentation is performed manually as well as registration. Sometimes, it is also assumed that no patient motions occur during the acquisition, removing the need of registering the mp-MRI images.

Once the data are regularized, it becomes possible to extract features and classify the data to obtain the location of possible lesions (i.e., CADe) or/and the malignancy nature of these lesions (i.e., CADx).

In a CADe framework, *possible lesions are segmented automatically* and further used as input of a CADx. Nevertheless, some works also used a fusion of CADe-CADx framework in which a voxel-based features are directly used, in which the location of the malignant lesions are obtained as results. On the other hand, manual lesions segmentation is not considered to be part of CADe.

CADx is composed of the processes allowing to *distinguish malignant from non-malignant tumours*. Here, CaP malignancy is defined using the grade of the GS determined after post biopsy or prostatectomy. As presented in Fig. 7, CADx is usually composed of the three common steps used in a classification framework: (i) features detection, (ii) feature extraction/selection, and (iii) feature classification.

This chapter is organized using the methodology presented in Fig. 7. Methods embedded in the image regularization framework are presented initially to subsequently focus on the image classification framework, being divided into CADe and CADx. Finally, we present a summary of the results reported in the state-of-the-art as well as a discussion that follows. Table II summarizes the 56 different CAD studies reviewed in this section. The first set of information reported is linked to the data acquisition such as the number of patients included in the study, the modalities acquired as well as the strength of the field of the scanner used. Subsequently, information about the prostate zones considered in the CAD analysis — i.e. PZ or CG — are reported since that detecting CaP in the CG is a more challenging problem and has received particular attention only in the recent publications.

The papers have been selected by investigating referenced international peer-reviewed journals as well as international peer-reviewed conferences. Additionally, a breadth-search first (or snowball sampling) have been used to refine missing publications. Only studies proposing CAD systems specifically for CaP have been reviewed.

TABLE II: Overview of the different studies reviewed with their main characteristics. Acronyms: number (#) - image regularization (Reg.). Notes: ✗: not used or not implemented; ✓[!]: partially implemented; ✓: used or implemented.

Index	Study	#	MRI-modality				Strength of field		Studied zones		CAD stages		
		Cases	T_2-W	DCE	DW	MRSI	1.5 T	3 T	PZ	CG	Reg.	CADe	CADx
[113, 114]	Ampeliotis et al.	25	✓	✓	✗	✗	✓	✗	✓	✗	✓[!]	✗	✓
[115]	Antic et al.	53	✓	✗	✓	✗	✓	✗	✓	✓	✗	✗	✓
[116]	Artan et al.	10	✓	✓	✓	✗	✓	✗	✓	✗	✗	✓	✓
[117]	Artan et al.	21	✓	✓	✓	✗	✓	✗	✓	✗	✓[!]	✓	✓
[118, 119]	Cameron et al.	5/13	✓	✗	✓	✗	✗	✓	✓	✓	✗	✓	✓
[120]	Chan et al.	15	✓	✗	✓	✗	✓	✗	✓	✗	✗	✗	✓
[121]	Chung et al.	20	✓	✗	✓	✗	✗	✓	✓	✓	✗	✓	✓
[122, 123]	Giannini et al.	10/56	✓	✓	✓	✗	✓	✗	✓	✗	✓	✓	✓
[124]	Kelm et al.	24	✗	✗	✗	✓	✓	✗	✓	✓	✓[!]	✓	✓
[125]	Khalvati et al.	20	✓	✗	✓	✗	✗	✓	✓	✓	✗	✓	✓
[126]	Langer et al.	25	✓	✓	✓	✗	✓	✗	✓	✗	✓[!]	✗	✓
[127]	Lehaire et al.	35	✓	✓	✓	✗	✓	✗	✓	✗	✓[!]	✗	✓
[128]	Lemaitre	17	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓	✓
[129]	Litjens et al.	188	✓	✓	✓	✗	✗	✓	✓	✗	✓[!]	✓	✓
[130]	Litjens et al.	288	✓	✓	✓	✗	✗	✓	✓	✓	✓[!]	✓	✓
[131]	Litjens et al.	347	✓	✓	✓	✗	✗	✓	✓	✓	✓[!]	✓	✓
[132]	Liu et al.	11	✓	✓	✓	✗	✓	✗	✓	✗	✓[!]	✓	✓
[133]	Liu et al.	54	✓	✓	✓	✗	✗	✓	✓	✓	✓[!]	✗	✓
[134]	Lopes et al.	27	✓	✗	✗	✗	✓	✗	✓	✗	✓[!]	✓	✓
[135]	Lv et al.	55	✓	✗	✗	✗	✓	✗	✓	✗	✓[!]	✗	✓
[136]	Matulewicz et al.	18	✗	✗	✗	✓	✗	✓	✓	✓	✗	✓	✓
[137]	Mazzetti et al.	10	✗	✓	✗	✗	✓	✗	✓	✗	✓[!]	✓	✓
[138, 139]	Niaf et al.	23/30	✓	✓	✓	✗	✓	✗	✓	✗	✓[!]	✗	✓
[140, 141]	Ozer et al.	20	✓	✓	✓	✗	✓	✗	✓	✗	✓[!]	✓	✓
[142]	Parfait et al.	22	✗	✗	✗	✓	✗	✓	✓	✓	✓[!]	✓	✓
[101]	Peng et al.	48	✓	✓	✓	✗	✗	✓	✓	✓	✗	✗	✓
[143]	Puech et al.	100	✗	✓	✗	✗	✓	✗	✓	✓	✗	✗	✓
[144, 145]	Rampun et al.	45	✓	✗	✗	✗	✓	✗	✓	✗	✗	✓	✓
[144, 146, 147]	Rampun et al.	45	✓	✗	✗	✗	✓	✗	✓	✗	✗	✓	✓
[148]	Samarasinghe et al.	40	✗	✓	✗	✗	✗	✓	✓	✗	✓[!]	✗	✓
[149]	Sung et al.	42	✗	✓	✗	✗	✗	✓	✓	✓	✗	✓	✓
[150]	Tiwari et al.	14	✗	✗	✗	✓	✓	✗	✓	✓	✓[!]	✓	✓
[151]	Tiwari et al.	18	✗	✗	✗	✓	✓	✗	✓	✓	✓[!]	✓	✓
[152]	Tiwari et al.	18	✗	✗	✗	✓	✓	✗	✓	✓	✓[!]	✓	✓
[153]	Tiwari et al.	15	✓	✗	✗	✓	✓	✗	✓	✓	✓[!]	✓	✓
[154]	Tiwari et al.	19	✓	✗	✗	✓	✓	✗	✓	✓	✓[!]	✓	✓
[155]	Tiwari et al.	36	✓	✗	✗	✓	✓	✗	✓	✓	✗	✓	✓

[156]	Tiwari et al.	29	✓	✗	✗	✓	✓	✗	✓	✓	✓!	✓	✓
[157, 158]	Trigui et al.	34	✓	✓	✓	✓	✗	✓	✓	✓	✓!	✓	✓
[159]	Viswanath et al.	16	✓	✗	✗	✓	✓	✗	✓	✓	✗	✓	✓
[160]	Viswanath et al.	6	✓	✓	✗	✗	✗	✓	✓	✓	✓!	✓	✓
[161]	Viswanath et al.	6	✓	✓	✗	✗	✗	✓	✓	✓	✓	✓	✓
[162]	Viswanath et al.	12	✓	✓	✓	✗	✗	✓	✓	✓	✓!	✓	✓
[163]	Viswanath et al.	22	✓	✗	✗	✗	✗	✓	✓	✓	✓	✓	✓
[164]	Vos et al.	29	✓	✓	✗	✗	✓	✗	✓	✗	✓!	✗	✓
[165]	Vos et al.	29	✗	✓	✗	✗	✓	✗	✓	✗	✓!	✗	✓
[166]	Vos et al.	29	✓	✓	✗	✗	✓	✗	✓	✗	✓!	✗	✓
[167]	Vos et al.	NA	✓	✓	✓	✗	✗	✓	✓	✗	✓!	✓	✓

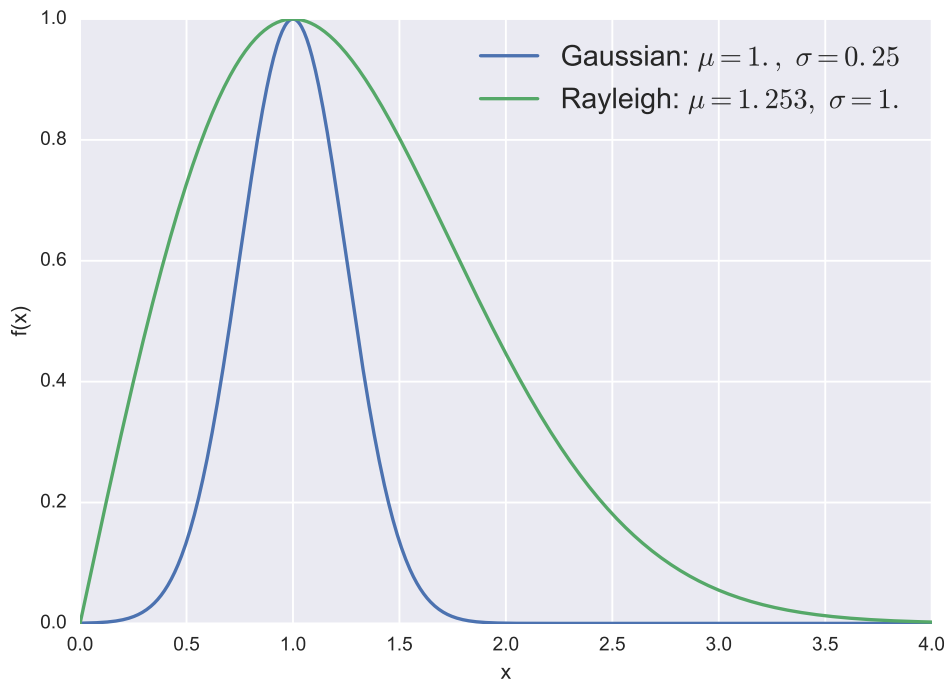


Fig. 8. Illustration of a Gaussian and Rayleigh distribution. Although the mode of these distributions are identical, it can be noted that the Rayleigh distribution ($\mu = 1.253$) is suffering of a bias term when compared with the Gaussian distribution ($\mu = 1$).

A. Pre-processing

Three different groups of pre-processing methods are commonly applied to images as initial stage in CADs for CaP. These methods are explained for both MRI and MRSI modalities

1) MRI modalities:

Noise filtering The NMR signal, measured and acquired in the k-space, is affected by noise. This noise obeys a complex Gaussian white noise mainly due to thermal noises in the patient [168]. Furthermore, MRI images visualized by radiologists are in fact the magnitude images resulting from the complex Fourier transform of the k-space data. The complex Fourier transform does not affect the Gaussian noise characteristics since this is a linear and orthogonal transform [168]. However, the calculation of the magnitude is a non-linear transform — i.e., the square root of the sum of squares of real and the imaginary parts — implying that the noise distribution is no longer Gaussian; it indeed follows a Rician distribution making the denoising task more challenging. Briefly, a Rician distribution is characterized as follows: in low-SI region (low-SNR), it can be approximated with a Rayleigh distribution while in high-SI region (high-SNR), it is similar to a Gaussian distribution [169]. Refer to Fig. 8 to observe the difference between a Gaussian and a Rayleigh distribution. Comprehensive reviews regarding denoising methods can be found in [170, 171].

Median filtering is the simplest approach used to address the denoising issue in MRI images [140, 141]. In both studies, Ozer et al. [140, 141] used a square-shaped kernel of size $5\,\mathrm{px} \times 5\,\mathrm{px}$.

More recently, Rampun et al. used a combination of median and anisotropic diffusion filter [144, 145, 146, 147], proposed in [172]. In low-SNR images, the gradient generated by an edge and noise can be similar, making the denoising by diffusion more challenging. In this condition, the threshold allowing to locally differentiate a noise gradient from an edge gradient needs to be increased, at the cost of blurring edges after filtering. Therefore, Ling and Bovik [172] proposed to apply a standard anisotropic diffusion filter with a low threshold followed by a median filtering to remove spikes.

Samarasinghe et al. filtered DCE-MRI images with a sliding 3D Gaussian filter [148]. However, from a theoretical point of view, this simple filtering method is not well formalized to address the

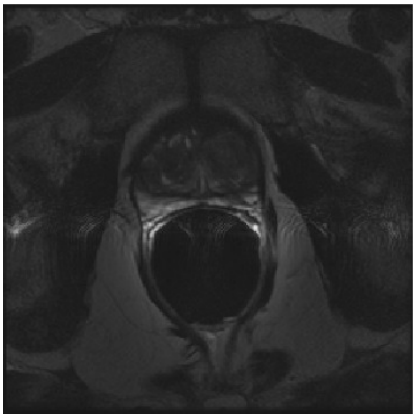


Fig. 9. Example of artifacts with high SI due to perturbation from the endorectal coil which create non-homogeneity.

noise distribution in MRI images. That is why more complex approaches have been proposed to overcome this problem. Another common method used to denoise MRI images is based on wavelet decomposition and shrinkage. This filtering exploits the sparsity property of the wavelet decomposition. The projection of a noisy signal from the spatial-domain to the wavelet-domain implies that only few wavelet coefficients contribute to the "signal-free noise" while all wavelet coefficients contribute to the noise [173]. Therefore, insignificant wavelet coefficients are thresholded/attenuated to enforce the sparsity in the wavelet-domain, which results to a denoising process in the spatial domain. Investigations focus on the strategies to perform the most adequate coefficient shrinkage (e.g., thresholding, singularity property, or Bayesian framework) [174]. Ampeliotis et al. denoised the magnitude MRI images [113, 114] — i.e., T_2-W-MRI and DCE-MRI — by wavelet shrinkage, using thresholding techniques [175]. However, since the wavelet transform is an orthogonal transform, the Rician distribution of the noise is preserved in the wavelet-domain. Hence, for low-SNR, the wavelet and scaling coefficients still suffer from a bias due to this specific noise distribution [168]. That is why, Lopes et al. filtered T_2-W-MRI images [134], using the method proposed in [176] based on joint detection and estimation theory. In this approach, the wavelet coefficients "free-of-noise" are estimated from the noisy wavelet coefficients using a maximum *a posteriori* (MAP) estimate. Furthermore, the designed estimator takes spatial context into account by including both local and global information in the prior probabilities. The different probabilities needed by the MAP are empirically estimated by using mask images, representing the locations of the significant wavelet coefficients. These mask images are computed by thresholding the detail images obtained from the wavelet decomposition. To remove the bias from the wavelet and scaling coefficients, the squared magnitude MRI image is computed instead of the magnitude MRI image as proposed in [168]. This involves changing the Rician distribution to a scaled non-central Chi-squared distribution. It implies that the wavelet coefficients are also unbiased estimators and the scaling coefficients are unbiased estimators but up to a constant C as defined in Eq. (7) which needs to be subtracted from each scaling coefficient such as:

$$C = 2^{(J+1)}\hat{\sigma}^2 \ , \tag{7}$$

where J is the number of levels of the wavelet decomposition and $\hat{\sigma}$ is an estimate of the noise standard deviation.

Bias correction Besides being corrupted by noise, MRI images are also affected by the inhomogeneity of the MRI field commonly referred to as bias field [177]. This bias field results in a smooth variation of the SI through the image. When an endorectal coil is used, a resulting artifact of an hyper-intense signal is observed around the coil as depicted in Fig. 9. As a consequence, the SI of identical tissues varies depending on their spatial location in the image making further processes such as segmentation,

registration, or classification more challenging [178, 179]. A comprehensive review of bias correction methods is proposed in [179].

The model of image formation is usually formalized as:

$$s(\mathbf{x}) = o(\mathbf{x})b(\mathbf{x}) + \eta(\mathbf{x}) \ , \tag{8}$$

where $s(\mathbf{x})$ is the corrupted SI at the pixel for the image coordinates $\mathbf{x} = \{x, y\}$, $o(\mathbf{x})$ is the "noise-free signal", $b(\mathbf{x})$ is the bias field function and $\eta(\mathbf{x})$ is an additive white Gaussian noise.

Hence, the task of bias correction involves estimating the bias function $b(\mathbf{x})$ in order to infer the "signal-free bias" $o(\mathbf{x})$.

Viswanath et al. corrected this artifact on T_2-W-MRI images [161], using the model proposed in [177], in which Styner et al. model the bias field function by using a linear combination of Legendre polynomials f_i as:

$$\begin{aligned} \hat{b}(\mathbf{x}, \mathbf{p}) &= \sum_{i=0}^{m-1} p_i f_i(\mathbf{x}) \\ &= \sum_{i=0}^{l} \sum_{j=0}^{l-i} p_{ij} P_i(x) P_j(y) \ , \end{aligned} \tag{9}$$

where $\hat{b}(\cdot)$ is the bias estimation with the image coordinates $\mathbf{x} = \{x, y\}$ and the m coefficients of the linear combination $\mathbf{p} = p_{11}, \ldots, p_{ij}$; m can be defined as $m = (l+1)\frac{(l+2)}{2}$ where l is the degree of Legendre polynomials chosen and $P_i(\cdot)$ denotes a Legendre polynomial of degree i.

This family of functions offers to model the bias function as a smooth inhomogeneous function across the image. To estimate the set of parameters $\mathbf{p}$, a cost function is defined which relies on the following assumptions: (i) an image is composed of k regions with a mean μ_k and a variance σ_k^2 for each particular class, and (ii) each noisy pixel belongs to one of the k regions with its SI value close to the class mean μ_k. Hence, the cost function is defined as:

$$C(\mathbf{p}) = \sum_{\mathbf{x}} \prod_{k} \rho_k(s(\mathbf{x}) - \hat{b}(\mathbf{x}, \mathbf{p}) - \mu_k) \ , \tag{10}$$

$$\rho_k(x) = \frac{x^2}{x^2 + 3\sigma_k^2} \ , \tag{11}$$

where $\rho_k(\cdot)$ is a M-estimator allowing estimations to be less sensitive to outliers than the usual squared distance [180].

Finally, the parameters $\mathbf{p}$ are estimated by finding the minimum of the cost function $C(\mathbf{p})$, which was optimized using the non-linear $(1+1)$ Evolution Strategy (ES) optimizer [181].

In a later publication, Viswanath et al. [163] as well as Giannini et al. [123] corrected T_2-W-MRI using the well known N3 algorithm [182] in which Sled et al. infer the bias function using the probability density functions (PDFs) of the signal and bias. Taking advantage of the logarithm property, the model in Eq. (8) becomes additive as expressed in Eq. (12).

$$\begin{aligned} \log s(\mathbf{x}) &= \log b(\mathbf{x}) + \log\left(o(\mathbf{x}) + \frac{\eta(\mathbf{x})}{b(\mathbf{x})}\right) \ , \\ &\approx \log b(\mathbf{x}) + \log \hat{o}(\mathbf{x}) \ , \end{aligned} \tag{12}$$

where $\hat{o}(\mathbf{x})$ is the signal only degraded by noise. Sled et al. show that Eq. (12) is related to PDFs such that:

$$S(s) = B(s) * O(s) \ , \tag{13}$$

where $S(\cdot)$, $B(\cdot)$, and $O(\cdot)$ are the PDFs of $s(\cdot)$, $b(\cdot)$, and $o(\cdot)$, respectively.

The corrupted signal s is restored by finding the multiplicative field b which maximizes the frequency content of the distribution O. Sled et al. [182] argued that a brute-force search through all possible fields b and selecting the one which maximizes the high frequency content of O is possible but far too complex. By assimilating the bias field distribution to be a near Gaussian distribution as *a priori*, it is then possible to infer the distribution O using the Wiener deconvolution given B and S and later estimate the corresponding smooth field b.

Lv et al. corrected the non-homogeneity in T_2-W-MRI images [135] by using the method proposed in [183]. Madabhushi et al. [183] proposed to correct the MRI images by detecting the image foreground via generalized scale (g-scale) in an iterative manner and estimating a bias field function based on a 2nd order polynomial model. First, the background of the MRI image is eliminated by thresholding, in which the threshold value is commonly equal to the mean SI of the considered image. Then, a seeded region growing algorithm is applied in the image foreground, considering every thresholded pixel as a potential seed. However, pixels already assigned to a region are not considered any more as a potential seed. As in seeded region growing algorithm [184], two criteria are taken into account to expand a region. First, the region grows using a connected-neighbourhood, initially defined by the user. Then, the homogeneity of SI is based on a fuzzy membership function taking into account the absolute difference of two pixel SI. Depending on the membership value — corresponding to a threshold which needs to be defined — the pixel considered is merged or not to the region. Once this segmentation is performed, the largest region R is used as a mask to select pixels of the original image and the mean SI, μ_R, is computed. The background variation $b(\mathbf{x})$ is estimated as:

$$b(\mathbf{x}) = \frac{s(\mathbf{x})}{\mu_R}, \ \forall \mathbf{x} \in R \ , \tag{14}$$

where $s(\mathbf{x})$ is the original MRI image.

Finally, a 2nd order polynomial $\hat{b}_\Theta(\mathbf{x})$ is fitted in a least-squares sense as in Eq. (15),

$$\hat{\Theta} = \arg\min_{\Theta} |b(\mathbf{x}) - \hat{b}_\Theta(\mathbf{x})|^2, \ \forall \mathbf{x} \in R \ . \tag{15}$$

Finally, the whole original MRI image is corrected by dividing it by the estimated bias field function $\hat{b}_\Theta(\mathbf{x})$. The convergence is reached when the number of pixels in the largest region R does not change significantly between two iterations.

SI normalization/standardization As discussed in the later section, segmentation or classification tasks are usually composed of a learning stage using a set of training patients. Hence, one can emphasize the desire to perform automatic diagnosis with a high repeatability or in other words, one would ensure to obtain consistent SI of tissues across patients of the same group — i.e., healthy patients *vs.* patients with CaP — for each MRI modality. However, it is a known fact that variability between patients occurs during the MRI examinations even using the same scanner, protocol or sequence parameters [185]. Hence, the aim of normalization or standardization of the MRI data is to remove the variability between patients and enforce the repeatability of the MRI examinations. These standardization methods are categorized either as statistical-based standardization or organ SI-based standardization.

Artan et al. [116, 117], Ozer et al. [140, 141], and Rampun et al. [144, 145, 146, 147, 186] standardized T_2-W-MRI, DCE-MRI, and DW-MRI images by computing the *standard score* (also called *z-score*) of the pixels of the PZ as:

$$I_s(\mathbf{x}) = \frac{I_r(\mathbf{x}) - \mu_{pz}}{\sigma_{pz}}, \ \forall \mathbf{x} \in \text{PZ} \ , \tag{16}$$

where $I_s(\mathbf{x})$ is the standardized SI with the image coordinates $\mathbf{x} = \{x, y\}$, $I_r(\mathbf{x})$ is the raw SI, μ_{pz} is the mean SI of the PZ and σ_{pz} is the SI standard deviation in the PZ. This transformation enforces the image PDF to have a zero mean and a unit standard deviation. In a similar way, Liu et al. normalized T_2-W-MRI by making use of the median and inter-quartile range for all the pixels [133].

Lemaitre et al. proposes to use the a priori that the underlying data distribution follows a Rician distribution instead of a Gaussian distribution [187]. The data are standardized by removing the mean and scaled it by dividing by the standard-deviation which are empirically computed as follows:

$$\mu_r = \sigma\sqrt{\frac{\pi}{2}}\ L_{1/2}(-\frac{\nu^2}{2\sigma^2})\ , \tag{17}$$

$$\sigma_r = 2\sigma^2 + \nu^2 - \frac{\pi\sigma^2}{2} L_{1/2}^2\left(\frac{-\nu^2}{2\sigma^2}\right)\ , \tag{18}$$

where ν and σ are the distance between the reference point and the center of the bivariate distribution and the scale, respectively; $L_{1/2}$ denotes a Laguerre polynomial.

Lv et al. scaled the SI of T_2-W-MRI images using the method proposed in [188] based on PDF matching [135]. This approach is based on the assumption that MRI images from the same sequence should share the same PDF appearance. Hence, one can approach this issue by transforming and matching the PDFs using some statistical landmarks such as quantiles. Using a training set, these statistical landmarks — such as minimum, 25th percentile, median, 75th percentile, and maximum — are extracted for N training images:

$$\begin{aligned}
\Phi_0 &= \{\phi_0^1, \phi_0^2, \cdots, \phi_0^N\}\ , \\
\Phi_{25} &= \{\phi_{25}^1, \phi_{25}^2, \cdots, \phi_{25}^N\}\ , \\
\Phi_{50} &= \{\phi_{50}^1, \phi_{50}^2, \cdots, \phi_{50}^N\}\ , \\
\Phi_{75} &= \{\phi_{75}^1, \phi_{75}^2, \cdots, \phi_{75}^N\}\ , \\
\Phi_{100} &= \{\phi_{100}^1, \phi_{100}^2, \cdots, \phi_{100}^N\}\ ,
\end{aligned} \tag{19}$$

where $\phi_{n^{\text{th}}}^{i^{\text{th}}}$ is the nth percentile of the ith training image.

Lemaitre et al. extended the use of non-parametric transformation using an elastic transformation instead of a piecewise-linear transformation [187]. They proposed to use a generic method to register functional data based on the square-root slope function (SRSF) representation [189] which transforms the Fisher-Rao metric into the conventional $\mathbb{L}^2$ metric, and thus allows to define a cost function corresponding to an Euclidean distance between two functions in this new representation.

Then, the mean of each statistical landmarks $\{\bar{\Phi}_0, \bar{\Phi}_{25}, \bar{\Phi}_{50}, \bar{\Phi}_{75}, \bar{\Phi}_{100}\}$ is also calculated. Once this training stage is performed, a piecewise linear transformation $\mathcal{T}(\cdot)$ is computed as in Eq. (20). For each test image t, this transformation maps each statistical landmark ${\varphi_{(\cdot)}}^t$ of the image t to the pre-learned statistical landmarks $\bar{\Phi}_{(\cdot)}$. An example of such piecewise linear function is depicted in Fig. 10.

$$\mathcal{T}(s(\mathbf{x})) = \begin{cases}
\lceil \bar{\Phi}_0 + (s(\mathbf{x}) - \varphi_0^t)\left(\frac{\bar{\Phi}_{25}-\bar{\Phi}_0}{\varphi_{25}^t-\varphi_0^t}\right)\rceil\ , & \text{if } \varphi_0^t \leq s(\mathbf{x}) < \varphi_{25}^t)\ , \\
\lceil \bar{\Phi}_{25} + (s(\mathbf{x}) - \varphi_{25}^t)\left(\frac{\bar{\Phi}_{50}-\bar{\Phi}_{25}}{\varphi_{50}^t-\varphi_{25}^t}\right)\rceil\ , & \text{if } \varphi_{25}^t \leq s(\mathbf{x}) < \varphi_{50}^t)\ , \\
\lceil \bar{\Phi}_{50} + (s(\mathbf{x}) - \varphi_{50}^t)\left(\frac{\bar{\Phi}_{75}-\bar{\Phi}_{50}}{\varphi_{75}^t-\varphi_{50}^t}\right)\rceil\ , & \text{if } \varphi_{50}^t \leq s(\mathbf{x}) < \varphi_{75}^t)\ , \\
\lceil \bar{\Phi}_{75} + (s(\mathbf{x}) - \varphi_{75}^t)\left(\frac{\bar{\Phi}_{100}-\bar{\Phi}_{75}}{\varphi_{100}^t-\varphi_{75}^t}\right)\rceil\ , & \text{if } \varphi_{75}^t \leq s(\mathbf{x}) \leq \varphi_{100}^t)\ ,
\end{cases} \tag{20}$$

Viswanath et al. used a variant of the piecewise linear normalization presented in [190], to standardize T_2-W-MRI images [161, 162, 163]. Instead of computing the PDF of an entire image, a pre-segmentation of the foreground is carried out via g-scale which has been discussed in the bias

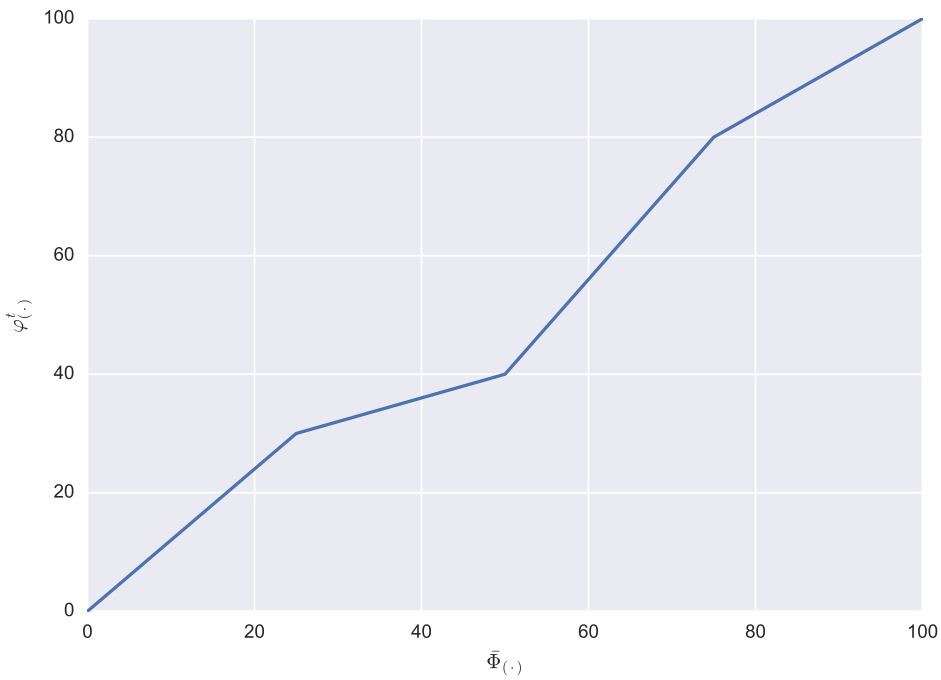
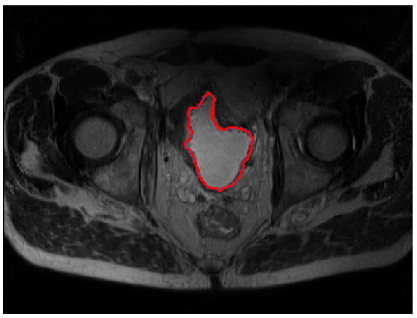


Fig. 10. Example of piecewise linear normalization as proposed in [188].



(a) Illustration and location of the bladder on a T_2-W-MRI image acquired with a 3 T MRI scanner

(b) Illustration and location of the femoral arteries on a T_1-W-MRI image acquired with a 3 T MRI scanner

Fig. 11. Illustration of the two organs used in [138, 139] to normalize T_2-W-MRI and T_1-W-MRI images.

correction section. Once the foreground is detected, the largest region is extracted, and the regular piecewise linear normalization is applied.

The standardization problem can be tackled by normalizing the MRI images using the SI of some known organs present in these images. Niaf et al. and Lehaire et al. normalized T_2-W-MRI images by dividing the original SI of the images by the mean SI of the bladder [127, 138, 139], which is depicted in Fig. 11a. Giannini et al. also normalized the same modality but using the signal intensity of the obturator muscle [123]. Likewise, Niaf et al. standardized the T_1-W-MRI images using the arterial input function (AIF) [138]. They computed the AIF by taking the mean of the SI in the most enhanced part of the common femoral arteries — refer to Fig. 11b — as proposed in [191]. Along the same line, Samarasinghe et al. normalized the SI of lesion regions in T_1-W-MRI using the mean intensity of the prostate gland in the same modality [148].

2) MRSI modalities: As presented in Sect. II-F, MRSI is a modality related to a one dimensional signal. Hence, specific pre-processing steps for this type of signals have been applied instead of standard signal

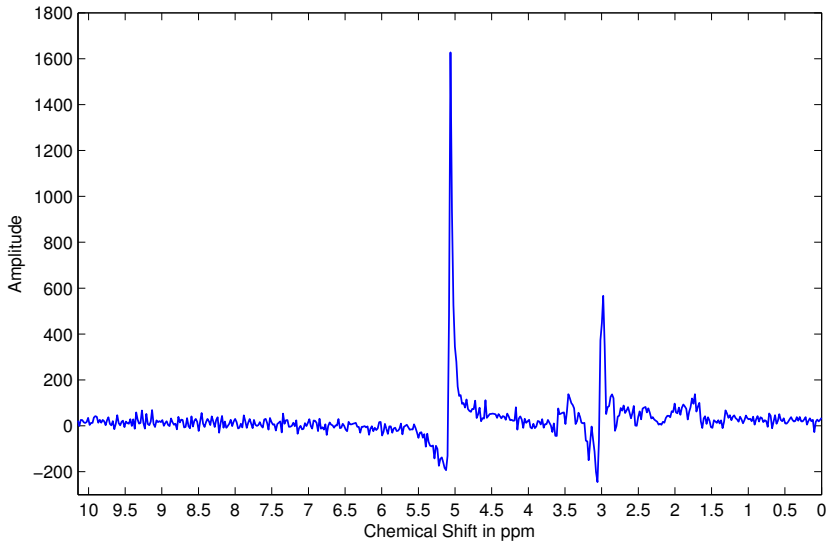


Fig. 12. Illustration of phase misalignment in an MRSI spectra acquired with a 3 T MRSI scanner. Note the distortion of the signal specially visible for the water and citrate peaks visible at $5\,\mathrm{ppm}$ and $3\,\mathrm{ppm}$, respectively

processing methods.

Phase correction Acquired MRSI spectra suffer from zero-order and first-order phase misalignment [192, 193] as depicted in Fig. 12. Parfait et al. and Trigui et al. used a method proposed by Chen et al. where the phase of MRSI signal is corrected based on entropy minimization in the frequency domain [142, 157, 158]. The corrected MRSI signal $o(\xi)$ can be expressed as:

$$\begin{aligned} \Re(o(\xi)) &= \Re(s(\xi))\cos(\Phi(\xi)) - \Im(\xi)\sin(\Phi(\xi)) \ , \\ \Im(o(\xi)) &= \Im(s(\xi))\cos(\Phi(\xi)) + \Re(\xi)\sin(\Phi(\xi)) \ , \\ \Phi(\xi) &= \phi_0 + \phi_1 \frac{\xi}{N} \ , \end{aligned} \tag{21}$$

where $\Re(\cdot)$ and $\Im(\cdot)$ are the real and imaginary part of the complex signal, respectively, $s(\xi)$ is the corrupted MRSI signal, ϕ_0 and ϕ_1 are the zero-order and first-order phase correction terms respectively and N is the total number of samples of the MRSI signal.

Chen et al. tackled this problem as an optimization in which ϕ_0 and ϕ_1 have to be inferred. Hence, the simplex Nelder-Mead optimizer [194] is used to minimize the following cost function based on the *Shannon entropy* formulation:

$$\hat{\Phi} = \arg\min_{\Phi} \left[-\sum \Re(s'(\xi)) \ln \Re(s'(\xi)) + \lambda \|\Re(s(\xi))\|_2 \right] \ , \tag{22}$$

where $s'(\xi)$ is the first derivative of the corrupted signal $s(\xi)$ and λ is a regularization parameter. Once the best parameter Φ vector is obtained, the MRSI signal is corrected using Eq. (21).

Water and lipid residuals filtering The water and lipid metabolites occur in much higher concentrations than the metabolites of interest, namely choline, creatine, and citrate [193, 195]. Fortunately, specific MRSI sequences have been developed in order to suppress water and lipid metabolites using pre-saturation techniques [195]. However, these techniques do not perfectly remove water and lipids peaks and some residuals are still present in the MRSI spectra as illustrated in Fig. 13. Therefore, different post-processing methods have been proposed to enhance the quality of the MRSI spectra

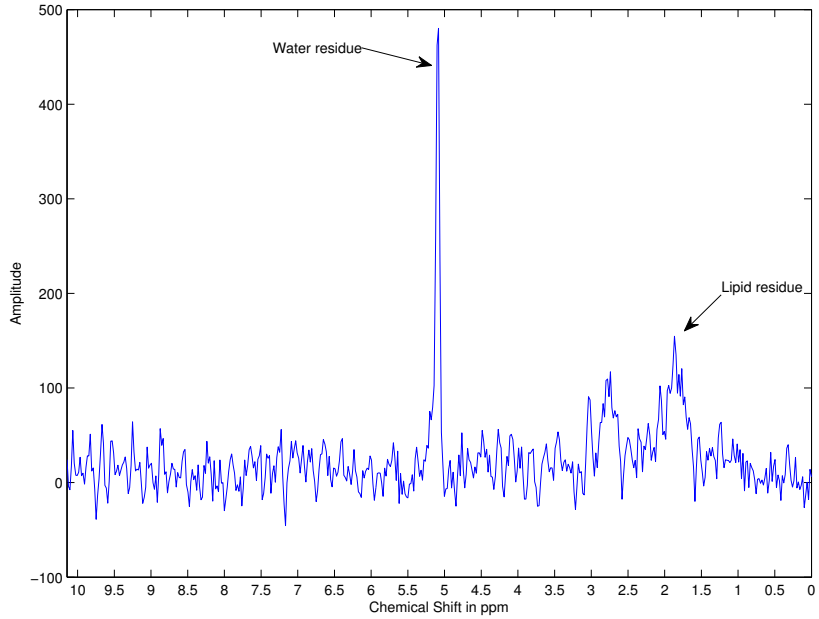


Fig. 13. Illustration of the residues of water and fat even after their suppression during the acquisition protocol. The acquisition has been carried out with a 3 T MRI.

by removing these residuals. For instance, Kelm et al. [124] used the HSVD algorithm proposed by Pijnappel et al. [196] which models the MRSI signal by a sum of exponentially damped sine waves in the time domain as Eq. (23).

$$s(t) = \sum_{k=1}^{K} a_k \exp(i\phi_k) \exp(-d_k + i2\pi f_k)t + \eta(t) \ , \tag{23}$$

where a_k is the amplitude proportional to the metabolite concentration with a resonance frequency f_k, d_k represents the damping factor of the exponential, ϕ_k is the first-order phase, and $\eta(t)$ is a complex white noise.

The "noise-free signal" can be found using the singular value decomposition (SVD) decomposition [196]. Therefore, the noisy signal is reorganized inside a Hankel matrix H. It can be shown that the signal is considered as a "noise-free signal" if the rank of H is equal to rank K. However, due to the presence of noise, H is in fact a full rank matrix. Thus, to recover the "noise-free signal", the rank of H is truncated to K using its SVD decomposition. Hence, knowing the cut off frequencies of water — i.e., $4.65\,\mathrm{ppm}$ — and lipid — i.e., $2.2\,\mathrm{ppm}$ — metabolites, their corresponding peaks are reconstructed and subtracted from the original signal [197].

Baseline correction Sometimes, the problem discussed in the above section regarding the lipid molecules is not addressed simultaneously with water residuals suppression. Lipids and macro-molecules are known to affect the baseline of the MRSI spectra, causing errors while quantifying metabolites, especially the citrate metabolite.

Parfait et al. made the comparison of two different methods to detect the baseline and correct the MRSI spectra [142] which are based on [198, 199]. Lieber and Mahadevan-Jansen corrected the baseline in the frequency domain by fitting a low degree polynomial $p(x)$ — e.g., 2nd or 3rd degree — to the MRSI signal $s(x)$ in a least-squares sense [198]. Then, the values of the fitted polynomial are re-assigned as:

$$p_f(x) = \begin{cases} p(x) \ , & \text{if } p(x) \leq s(x) \ , \\ s(x) \ , & \text{if } p(x) > s(x) \ . \end{cases} \tag{24}$$

Finally, this procedure of fitting and re-assignment is repeated on $p_f(x)$ until a stopping criterion is reached. The final polynomial function is subtracted from the original signal $s(x)$ to correct it. Parfait et al. [142] modified this algorithm by convolving a Gaussian kernel to smooth the MRSI signal instead of fitting a polynomial function, keeping the rest of the algorithm identical. Unlike Lieber and Mahadevan-Jansen [198], Devos et al. [199] corrected the baseline in the time domain by multiplying the MRSI signal by a decreasing exponential function as:

$$c(t) = \exp(-\beta t) \ , \tag{25}$$

with a typical β value of 0.15. However, Parfait et al. concluded that the method proposed in [198] outperformed the one in [199]. The later study of Trigui et al. used this conclusion and adopted the same method [157, 158].

The previous baseline correction methods does not provide an optimal solution since the iterative low-pass filter enforces too much the smoothness of the baseline. Xi and Rocke proposed a baseline detection derived from a parametric smoothing model [200]. The NMR signal is formalized as a sum of a pure signal, the baseline function, and an additive Gaussian noise such as:

$$y_i = b_i + \mu_i e^{n_i} + \varepsilon_i \ , \tag{26}$$

where y_i is the NMR signal, b_i is the baseline, μ_i is the true signal, and n_i and ε_i are Gaussian noises.

Xi and Rocke propose to find the baseline function through an iterative optimization by maximizing the following cost function:

$$F(b) = \sum_{i=1}^{N} b_i - \frac{A^* N^4}{\sigma} \sum_{i=1}^{N} (b_{i+1} + b_{i-1} - 2b_i)^2 - \frac{1.25 B^*}{\sigma} \sum_{i=1}^{N} (b_i - \gamma_i)^2 g(b_i - \gamma_i) \ , \tag{27}$$

where $g(b_i - \gamma_i)$ is the Heaviside function, A^* and B^* are the terms controlling the smoothness and negative penalties, respectively, σ is an estimation of the standard deviation of the noise, and N is the total number of points in the MRSI signal.

The standard deviation of the noise σ is estimated as in [200], and the A^* and B^* are empirically set to 5×10^{-6} and 100, respectively, for all the MRSI signal. This method was used in the work of Lemaitre [128].

In the contemporary work of Tiwari et al. [155], the authors detected the baseline using a local non-linear fitting method avoiding regions with significant peaks, which have been detected using an experimentally parametric signal-to-noise ratio set to a value larger than $5\,\mathrm{dB}$.

Frequency alignment Due to variations of the experimental conditions, a frequency shift is commonly observed in the MRSI spectra [192, 193] as depicted in Fig. 14. Tiwari et al. [155] corrected this frequency shift by first detecting known metabolite peaks such as choline, creatine, or citrate and minimizing the frequency error between the experimental and theoretical values for each of these peaks [155].

Normalization The NMR spectra is subject to variations due to intra-patient variations and non homogeneity of the magnetic field. Parfait et al. as in [199] compared two methods to normalize MRSI signal [142]. In each method, the original MRSI spectra is divided by a normalization factor, similar to the intensity normalization described earlier. The first approach consists in estimating the water concentration from an additional MRSI sequence where the water has not been suppressed. The

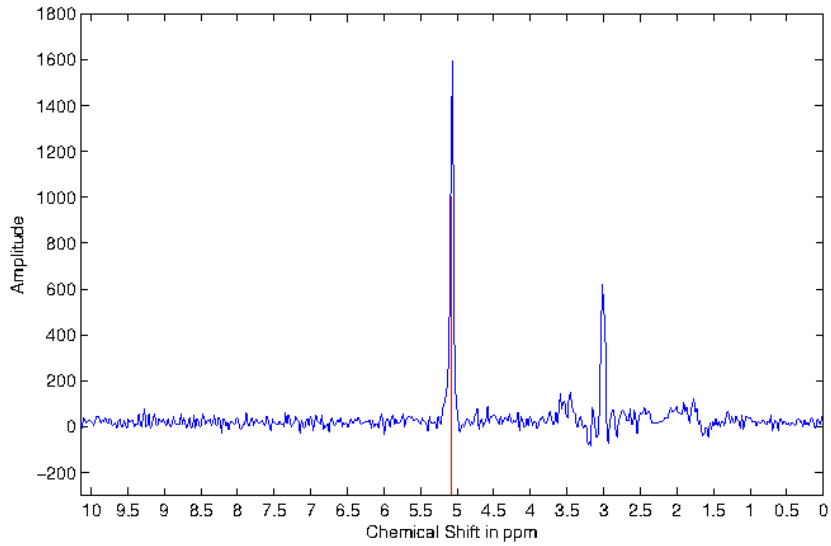


Fig. 14. Illustration of frequency misalignment in a MRSI spectra acquired with a 3 T MRSI scanner. The water peak is known to be aligned at $4.65\,\mathrm{ppm}$. However, it can be seen that the peak on this spectra is aligned at around $5.1\,\mathrm{ppm}$.

TABLE III
OVERVIEW OF THE PRE-PROCESSING METHODS USED IN CAD SYSTEMS.

Pre-processing operations	**References**
MRI pre-processing:	
Noise filtering:	
• Anisotropic median-diffusion filtering	[144, 145, 146, 147, 186]
• Gaussian filtering	[148]
• Median filtering	[140, 141]
• Wavelet-based filtering	[113, 114, 134]
Bias correction:	
• Parametric methods	[123, 135, 161]
• Non-parametric methods	[162]
Standardization:	
• Statistical-based normalization:	[116, 117, 128, 135, 140, 141, 144, 145, 146, 147, 161, 162, 163, 186]
• Organ SI-based normalization	[127, 138, 139, 148]
MRSI pre-processing:	
Phase correction	[128, 142, 157, 158]
Water and lipid residuals filtering	[124]
Baseline correction	[128, 142, 155, 157, 158]
Frequency alignment	[128, 155, 157, 158]
Normalization	[128, 142, 157, 158]

estimation is performed using the previously HSVD algorithm. The second approach does not require any additional acquisition and is based on the L_2 norm of the MRSI spectra $\|s(\xi)\|_2$. It should be noted that both Parfait et al. and Devos et al. concluded that the L_2 normalization is the most efficient method [142]. Lately, Trigui et al. used the L_2 normalization in their framework [157, 158].

The different pre-processing methods are summarized in Table III.

B. Segmentation

The segmentation task consists in delineating the prostate boundaries in the MRI and is of particular importance for focusing the posterior processing on the organ of interest [201]. In this section, only the

segmentation methods used in CAD for CaP are presented. An exhaustive review of prostate segmentation methods in MRI is available in [201].

Manual segmentation To highlight the importance of prostate segmentation task in CAD systems, it is interesting to note the large number of studies which manually segment the prostate organs [116, 117, 127, 136, 138, 139, 140, 141, 143, 157, 158, 164, 165]. In all the cases, the boundaries of the prostate gland are manually defined in order to limit further processing to only this area. This approach ensures the right delineation of the organ, although is subjective and prone to the rater variability; nevertheless this procedure is highly time consuming and should be performed by a radiologist.

Region-based segmentation Litjens et al. used a multi-atlas-based segmentation using multi-modal images — i.e., T_2-W-MRI and ADC map — to segment the prostate with an additional pattern recognition method to differentiate CG and PZ [130], as proposed in [202]. This method consists in three different steps: (i) the registration between each atlas and the multi-modal images, (ii) the atlas selection, and finally (iii) the classification of the prostate voxels into either CG or PZ classes. Each atlas and the MRI images are registered through two successive registrations: a rigid registration to roughly align the atlases and the MRI images followed by an elastic registration using a B-spline transformation. The cost function driving the registration is defined as the weighted sum of the mutual information (MI) of both T_2-W-MRI and ADC map. The final atlas is selected using either a majority voting or the simultaneous truth and performance level estimation (STAPLE) approach [203]. Subsequently, each voxel within the prostate is classified either as CG or PZ using a linear discriminant analysis (LDA) classifier. Three types of features are considered to characterize the voxels: (i) anatomy, (ii) intensity, and (iii) texture. The relative position and the relative distance from the voxel to the border of the prostate encode the anatomical information. The intensity features consist in the intensity of the voxel in the ADC coefficient and the T_2 map. The texture features are composed of 5 different features: homogeneity, correlation [204], entropy, texture strength [205], and local binary pattern (LBP) [206]. Finally, the final segmentation is obtained by removing artifacts and smoothing the contour between the zones using the thin plate spline (TPS) [207].

Litjens et al. used an almost identical algorithm in [131], initially proposed for the PROMISE12 challenge [208]. Their segmentation method is also based on multi-atlas multi-modal images, but the SIMPLE method [209] is used instead, to combine labels after the registration of the different atlas to obtain the final segmentation.

Finally, Rampun et al. recurrently used a method to segment the PZ [144, 145, 146, 147, 186], which is proposed in [210]. The PZ is modelled using a quadratic function driven by the centre of the prostate, the left-most, and the right-most coordinates of the prostate boundaries.

Model-based segmentation: Viswanath et al. [160, 161] used the MANTRA method [211]. Multi-attribute non-initializing texture reconstruction based active shape model (MANTRA) [211] is closely related to the active shape model (ASM) from [212]. This algorithm consists of two stages: (i) a training stage where a shape and an appearance model are generated and (ii) the actual segmentation based on the learned model. For the training stage, a set of landmarks is defined and the shape model is generated as in the original ASM method [212]. Then, to model the appearance, a set of K texture images $\{I_1, I_2, \cdots, I_k\}$ based on first and second order statistical texture features is computed. For a given landmark l with its given neighbourhood $\mathcal{N}(l)$, its feature matrix extracted is expressed as:

$$f_l = \{I_1(\mathcal{N}(l)), I_2(\mathcal{N}(l)), \cdots, I_k(\mathcal{N}(l))\} \ , \tag{28}$$

where $I_k(\mathcal{N}(l))$ represents a feature vector obtained by sampling the kth texture map using the neighbourhood $\mathcal{N}(l)$. Therefore, multiple landmarks are generated followed by a decomposition using principal components analysis (PCA) [213] to learn the appearance variations as in ASM.

For the segmentation stage, the mean shape learned previously is initialized in the test image. The same associated texture images as in the training stage are computed. For each landmark l, a

neighbourhood of patches are used to sample the texture images and a reconstruction is obtained using the appearance model previously trained. The new landmark location will be defined as the position where the MI is maximal between the reconstructed and original values. This scheme is performed in a multi-resolution manner as in [212].

Subsequently, Viswanath et al. in [163], used the weighted ensemble of regional image textures for active shape model segmentation (WERITAS) method also proposed in Toth et al.. Similarly to MANTRA, WERITAS is also based on the ASM formulation [214]. It differs in the last stage of the algorithm in which the Mahalanobis distance is used instead of the MI metric, to adapt the positions of new landmarks. In the training stage, the Mahalanobis distance is computed between landmarks and neighbour patches for each of the features. Subsequently, a new metric is proposed as a linear weighted combination of those Mahalanobis distances which maximizes the correlation with the Euclidean distance between the patches and the true landmarks. In the segmentation step, this metric is then computed between the initialized landmarks and neighbouring patches in order to update landmark positions, in a similar fashion to other active contour model (ACM) models.

Litjens et al. as well as Vos et al. used an approach proposed in [215] in which the bladder, the prostate, and the rectum are segmented [129, 167]. The segmentation task is performed as an optimization problem taking 3 parameters into account linked to organ characteristics such as: (i) the shape (i.e., an ellipse), (ii) the location, and (iii) the respective angles between them. Furthermore, Litjens et al. used only the ADC map to encode the appearance [129] whereas Vos et al. used both ADC and T_2 maps [167]. The cost function, defined as the sum of the deviations, is minimized using a quasi-Newton optimizer. This rough segmentation is then used inside a Bayesian framework to refine the segmentation.

Giannini et al. segmented the prostate with a multi-Otsu thresholding [216] in ADC images [123]. Further morphological operations are applied to improve the segmentation.

Only the work of Tiwari et al. used the MRSI modality to segment the prostate organ [152]. The prostate is segmented based on an unsupervised hierarchical spectral clustering. First, each MRSI spectrum is projected into a lower-dimensional space using graph embedding [217]. To proceed, a similarity matrix W is computed using a Gaussian similarity measure from Euclidean distance [218] such that:

$$W(\mathbf{x},\mathbf{y}) = \begin{cases} \exp\left(\frac{\|s(\mathbf{x})-s(\mathbf{y})\|_2^2}{\sigma^2}\right) \ , & \text{if } \|\mathbf{x}-\mathbf{y}\|_2 < \epsilon \ , \\ 0 \ , & \text{if } \|\mathbf{x}-\mathbf{y}\|_2 > \epsilon \ . \end{cases} \tag{29}$$

where $s(\mathbf{x})$ and $s(\mathbf{y})$ are the MRSI spectra for the voxels $\mathbf{x}$ and $\mathbf{y}$, respectively, σ is the standard deviation of the Gaussian similarity measure, and ϵ is the parameter to defined an ϵ-neighbourhood.

The projection can be performed as a generalized eigenvector problem such that:

$$\begin{aligned} Lu &= \lambda D u \ , \\ D(\mathbf{x},\mathbf{x}) &= \sum_{\mathbf{y}} W(\mathbf{x},\mathbf{y}) \ , \\ L &= D - W \ , \end{aligned} \tag{30}$$

where D is the diagonal weight matrix, L is the Laplacian matrix, λ and u represent the eigenvalues and eigenvectors. Once the MRSI spectra are projected into the lower-dimensional space, a replicate k-means clustering method is used to define 2 clusters. Subsequently, the data corresponding to the largest cluster is assumed to belong to the non-prostate voxels and thus these voxels are eliminated from the processing. The full procedure is repeated until the total number of voxels left is inferior to a given threshold experimentally set.

The segmentation algorithms used in CAD system for the detection of CaP are summarized in Table IV.

TABLE IV
OVERVIEW OF THE SEGMENTATION METHODS USED IN CAD SYSTEMS.

Segmentation methods	References
MRI-based segmentation:	
Manual segmentation	[116, 117, 127, 128, 136, 138, 139, 140, 141, 143, 157, 158, 164, 165, 166, 167]
Region-based segmentation	[130, 131, 144, 145, 146, 147, 186]
Model-based segmentation	[123, 129, 160, 161, 162, 167]
MRSI-based segmentation:	
Clustering	[152]

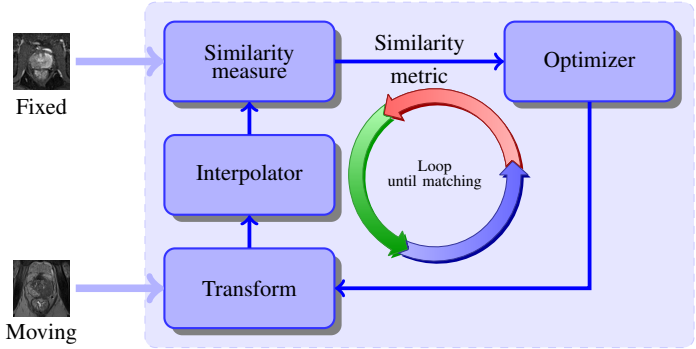


Fig. 15. Typical framework involved to solve the registration problem.

C. Registration

Image registration plays a vital role in CAD systems using mp-MRI images. The features detected in each modality are grouped depending of their spatial location, requiring a perfect alignment of the mp-MRI ahead of the classification.

Image registration is the procedure consisting of aligning an unregistered image — also called moving image — into a template image — also called fixed image — via a geometric transformation. This problem is usually addressed as depicted in Fig. 15. An iterative procedure takes place to infer the geometric transformation, parametric or non-parametric, via an optimizer which maximizes the similarity between the two images. In the following, a review of the different components of a typical registration framework: transformation model, similarity metric, optimizer, and interpolation are presented. To conclude a summary is given focusing on the registration approaches applied in CAD for CaP systems. Exhaustive reviews covering all registration methods in computer science and medical fields can be found in [219, 220].

Geometric transformation models As previously mentioned, the registration process is equivalent to find a geometric transformation which minimizes the difference between two images. From all CAD systems reviewed, only parametric methods have been implemented. Three different groups of parametric transformation models have been used — i.e., rigid, affine, and elastic — each of them characterized by a specific degree of freedom.

The simplest transformation used in terms of degrees of freedom is usually referred to as rigid transformation. This type of transformation is only composed of a rotation and a translation. Therefore, for the 2D case where $\mathbf{x} = (x, y) \in \mathbb{R}^2$, a rigid transformation $\mathcal{T}_R$ is formalized as:

$$\begin{aligned}\mathcal{T}_R(\mathbf{x}) &= \begin{bmatrix} R & \mathbf{t} \\ \mathbf{0^T} & 1 \end{bmatrix} \mathbf{x} \ , \\ &= \begin{bmatrix} \cos\theta & -\sin\theta & t_x \\ \sin\theta & \cos\theta & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \ , \end{aligned} \tag{31}$$

where θ is the rotation angle and $\{t_x, t_y\}$ represents the translation along $\{x, y\}$ respectively. In the case of 3D registration using volume, an additional component z is introduced such that $\mathbf{x} = (x, y, z)$. Thus, the rotation matrix $\mathbf{R}$ becomes of size 3×3 whereas the translation vector $\mathbf{t}$ consists of a vector of 3 variables. The geometric transformation $\mathcal{T}_R(\cdot)$ is embedded into a matrix of size 4×4.

The affine transformation provides additional degrees of freedom, providing rotation, translation, — as with the rigid transformations — and also shearing and scaling. Hence, for a 2D space where $\mathbf{x} = (x, y) \in \mathbb{R}^2$, an affine transformation $\mathcal{T}_A$ is formalized as:

$$\begin{aligned}\mathcal{T}_A(\mathbf{x}) &= \begin{bmatrix} A & \mathbf{t} \\ \mathbf{0^T} & 1 \end{bmatrix} \mathbf{x} \ , \\ &= \begin{bmatrix} a_{11} & a_{12} & t_x \\ a_{21} & a_{22} & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \ . \end{aligned} \tag{32}$$

where the 4 parameters $\{a_{11}, a_{12}, a_{21}, a_{22}\}$ of the affine matrix and $\{t_x, t_y\}$ of the translation encode the deformation. As in the rigid registration case, in 3D the affine transformation $\mathcal{T}_A(\cdot)$ is of size 4×4 but now with 12 parameters involved.

Finally, the last group of transformations is known as elastic transformations and offers the advantage to handle local distortions. In the reviewed CAD systems, the radial basis functions are used to formalize the local distortions such as:

$$\mathcal{T}_E(\mathbf{x}) = \begin{matrix} a_{11}x - a_{12}y + t_x + \sum_i c_i g(\|\mathbf{x} - p_i\|) \\ a_{21}x + a_{22}y + t_y + \sum_i c_i g(\|\mathbf{x} - p_i\|) \end{matrix} \ , \tag{33}$$

where $\mathbf{x}$ are the control points in both images and $g(\cdots)$ is the actual radial basis function.

Two radial basis functions are used: (i) the TPS and (ii) the B-splines. Apart from the formalism, these two approaches have a main difference: with B-splines, the control points are usually uniformly and densely placed on a grid whereas with TPS, the control points correspond to some detected or selected key points. By using TPS, Mitra et al. obtained more accurate and time efficient results than with the B-splines strategy [222].

It is reasonable to point out that usually only rigid or affine registrations are used to register mp-MRI from a same protocol. Elastic registration methods are more commonly used to register multi-protocol images such as histopathology with MRI images [211, 214].

Similarity measure The most naive similarity measure used in reviewed registration framework is the mean squared error (MSE) of the SI of MRI images. For a pair of images I and J, the MSE is formalized as:

$$\text{MSE} = \frac{1}{N} \sum_x \sum_y \left[I(x, y) - J(x, y) \right]^2 \ , \tag{34}$$

where N is the total number of pixels. This metric is not well suited when mp-MRI images are involved due to the tissue appearance variations between the different modalities.

(a) Illustration of a joint histogram between two aligned image.

(b) Illustration of a joint histogram between two misaligned image.

Fig. 16. Difference observed in joint histogram between aligned and misaligned images. The joint measure will be more concentrated of the histogram in the case that the images are aligned and more randomly distributed in the case that both images are more misaligned.

In this regard, MI was introduced as a similarity measure in registration framework in the late 1990's by Pluim et al. [223]. The MI measure finds its foundation in the assumption that a homogeneous region in the first modality image should also appear as a homogeneous region in the second modality, even if their SIs are not identical. Thus, those regions share information and the registration task is achieved by maximizing this common information. Hence, MI of two images A and B is defined as:

$$MI(A;B) = S(A) + S(B) - S(A,B) \ , \tag{35}$$

where $S(A)$, $S(B)$, and $S(A,B)$ are the marginal entropies of A and B and the joint entropy, respectively. Therefore, maximizing the MI is the equivalent of minimizing the joint entropy. The joint entropy measure is related to the degree of uncertainty or dispersion of the data in the joint histogram of the images A and B. As shown in Fig. 16, the data in the joint histogram are concentrated in the case of aligned images (see Fig. 16a) while it is more randomly distributed in the case of misaligned images (see Fig. 16b). The entropy is computed based on an estimation of the PDF of the images and thus histogram or Parzen window methods are a common way to estimate these PDFs.

A generalized form of MI, combined mutual information (CMI), has been proposed by Chappelow et al. [224]. CMI encompasses interdependent information such as texture and gradient information into the metric. Hence, for both of images A and B, the image ensembles ϵ_n^A and ϵ_m^B are generated and composed of n and m images based on the texture and gradient. Then, the CMI is formulated such as:

$$CMI(\epsilon_n^A;\epsilon_m^B) = S(\epsilon_n^A) + S(\epsilon_m^B) - S(\epsilon_n^A,\epsilon_m^B) \ . \tag{36}$$

From Eq. (36), note that CMI is estimated from high-dimensional data and as a consequence the histogram-based methods to estimate the PDFs are not suitable anymore [224]. However, other alternative approaches are used such as the one employed in [225] to compute the α-MI [226].

Optimization methods Registration is usually regarded as an optimization problem where the parameters of the geometric transformation model have to be inferred by minimizing/maximizing the similarity measure. Iterative optimization methods are commonly used, where the most common methods used are the L-BFGS-B quasi-Newton method [227] and the gradient descent [228]. During our review, we noticed that authors do not usually linger over optimizer choice.

Interpolation The registration procedure involves transforming an image and pixels mapped to non-integer points must be approximated using interpolation methods. As for the optimization methods, we notice that little attention has been paid on the choice of those interpolations methods. However, com-

TABLE V
CLASSIFICATION OF THE DIFFERENT REGISTRATION METHODS USED IN THE CAD SYSTEMS REVIEWED. ACRONYMS: MEAN SQUARED ERROR (MSE), MUTUAL INFORMATION (MI), COMBINED MUTUAL INFORMATION (CMI), GRADIENT DESCENT (GD), LIMITED-MEMORY BROYDEN-FLETCHER-GOLDFARB-SHANNON BOX CONSTRAINTS (L-BFGS-B).

Study	**Modality**	**Type**	**Geometric model**		**Similarity measure**			**Optimizer**	
index	**registered**		Affine	Elastic	MSE	MI	CMI	GD	L-BFGS-B
[113, 114]	T_2-W - DCE	2D	✓	–	✓	–	–	–	–
[122, 123]	T_2-W - DW	2D	✓	✓	–	–	–	–	–
[122, 123]	T_2-W - DCE	2D	✓	✓	–	✓	–	✓	–
[160, 161]	T_2-W - DCE	2D	✓	–	–	✓	–	–	–
[162]	T_2-W - DCE - DW	3D	✓	–	–	–	✓	✓	–
[164]	T_2-W - DCE	3D	✓	–	–	✓	–	–	–
[166]	T_2-W - DCE	3D	✓	✓	–	✓	–	–	✓
[128]	T_2-W - DCE - DW	3D	✓	✓	✓	✓	–	✓	–

Notes:
–: not used or not mentioned.
✓: used or implemented.

monly used methods are bi-linear, nearest-neighbour, bi-cubic, spline, and inverse-distance weighting method [229].

Registration methods used in CAD systems Table V summarizes the framework used to register mp-MRI images in CAD for CaP.

Ampeliotis et al. in [113, 114] did not use the framework as presented in Fig. 15 to register 2D T_2-W-MRI and DCE-MRI images. By using image symmetries and the MSE metric, they found the parameters of an affine transformation but without using a common objective function. The scale factor, the rotation, and the translation are independently and sequentially estimated.

Giannini et al. used also a in-house registration method for 2D T_2-W-MRI and DW-MRI images using an affine model [122, 123]. The bladder is first segmented in both modalities in order to obtain its contours which are then used as a metric function (i.e. distance between contours) for registration.

Giannini et al. and also Vos et al. used a framework based on finding an affine transformation to register the T_2-W-MRI and DCE-MRI images using MI [122, 166, 230]. Then, an elastic registration using B-spline takes place using the affine parameters to initialize the geometric model with the same similarity measure. However, the two approaches differ regarding the choice of the optimizer since a gradient descent is used in [122] and a quasi-Newton method in [166]. Moreover, Giannini et al. applied a 2D registration whereas Vos et al. registered 3D volumes.

Viswanath et al. as well as Vos et al. registered T_2-W-MRI and DCE-MRI images using an affine registration and a MI metric [160, 161, 164]. However, the choice of the optimizer has not been specified. Furthermore, Viswanath et al. focused on 2D registration [160, 161] while Vos et al. performed 3D registration [164].

Finally, Viswanath et al. performed a 3D registration with the three modalities, T_2-W-MRI, DCE-MRI, and DW-MRI, using an affine transformation model combined with the CMI similarity measure [162]. Moreover, in this latter work, the authors employed a gradient descent approach [224] to solve this problem but suggested that the Nelder-Mead simplex and the quasi-Newton methods are other possible solutions.

Lemaitre registered T_2-W-MRI, DCE-MRI, and DW-MRI images [128]. The intra-patient motions occurring during DCE-MRI is corrected by rigidly registering all series to the first serie using MI and a gradient descent optimizer. Once the intra-patient motions are corrected, T_2-W-MRI and the DCE-MRI are co-registered. For that matter, the prostate has been segmented in both modalities — T_2-W-MRI and DCE-MRI — to create two binary masks. Therefore, these 3D binary masks are directly registered using the MSE metric and a gradient descent with an elastic registration based on B-splines. The T_2-W-MRI and ADC map acquisitions are registered using the same approach as for the registration of

TABLE VI
OVERVIEW OF THE CADE STRATEGIES EMPLOYED IN CAD SYSTEMS.

CADe: ROIs selection strategy	References
All voxels-based approach	[116, 117, 122, 124, 125, 127, 128, 132, 134, 136, 137, 140, 141, 142, 144, 145, 146, 147, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 186]
Lesions candidate detection	[118, 119, 129, 130, 131, 167]

the T_2-W-MRI and the DCE-MRI modalities.

D. CADe: ROIs detection/selection

As discussed in the introduction and shown in Fig. 7, the image classification framework is often composed of a CADe and a CADx. In this section, we focus on studies which embed a CADe in their framework. Two approaches are considered to define a CADe: (i) voxel-based delineation and (ii) lesion segmentation. These methods are summarized in Table VI. The first strategy is in fact linked to the nature of the classification framework and concerns the majority of the studies reviewed [116, 117, 122, 124, 125, 127, 128, 132, 134, 136, 137, 140, 141, 142, 144, 145, 146, 147, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 186]. Each voxel is a possible candidate and will be classified as cancer or healthy. The second group of methods is composed of method implementing a lesion segmentation algorithm to delineate potential candidates to further obtain a diagnosis through the CADx. This approach is borrowed from other application areas such as breast cancer. These methods are in fact very similar to the classification framework used in CADx later.

Regarding lesion candidate detection, Vos et al. highlighted lesion candidates by detecting blobs in the ADC map [167]. These candidates are filtered using some *a priori* criteria such as SI or diameter. As mentioned in Sect. II-F and Table I, low SI in ADC map can be linked to potential CaP. Hence, blob detectors are suitable to highlight these regions. Blobs are detected in a multi-resolution scheme, by computing the three main eigenvalues $\{\lambda_{\sigma,1}, \lambda_{\sigma,2}, \lambda_{\sigma,3}\}$ of the Hessian matrix, for each voxel location of the ADC map at a specific scale σ [231]. The probability p of a voxel $\mathbf{x}$ being a part of a blob at the scale σ is given by:

$$P(\mathbf{x}, \sigma) = \begin{cases} \frac{\|\lambda_{\sigma,3}(\mathbf{x})\|^2}{\|\lambda_{\sigma,1}(\mathbf{x})\|} \ , & \text{if } \lambda_{\sigma,k}(\mathbf{x}) > 0 \text{ with } k = \{1,2,3\} \ , \\ 0 \ , & \text{otherwise} \ . \end{cases} \tag{37}$$

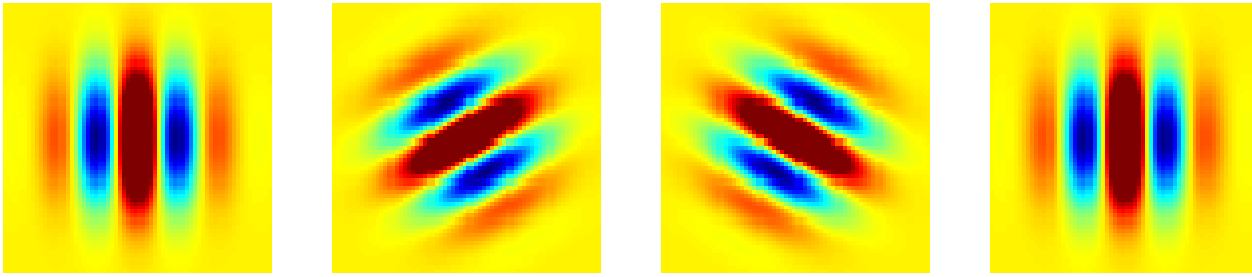
The fusion of the different scales is computed as:

$$L(\mathbf{x}) = \max P(\mathbf{x}, \sigma), \forall \sigma \ . \tag{38}$$

The candidate blobs detected are then filtered depending on their appearances — i.e., maximum of the likelihood of the region, diameter of the lesion — and their SI in ADC and T_2-W-MRI images. The detected regions are then used as inputs for the CADx. Cameron et al. used a similar approach by automatically selecting low SI connected regions in the ADC map with a size larger than $1\,\mathrm{mm}^2$ [118, 119].

Litjens et al. used a pattern recognition approach in order to delineate the ROIs [129]. A blobness map is computed in the same manner as in [166] using the multi-resolution Hessian blob detector on the ADC map, T_2-W, and pharmacokinetic parameters maps (see Sect. III-E for details about those parameters). Additionally, the position of the voxel $\mathbf{x} = \{x, y, z\}$ is used as a feature as well as the Euclidean distance of the voxel to the prostate center. Hence, each feature vector is composed of 8 features and a support vector machines (SVM) classifier is trained using a radial basis function (RBF) kernel (see Sect. III-H for more details).

Subsequently, Litjens et al. modified this approach by including only features related to the blob detection on the different maps as well as the original SIs of the parametric images [130]. Two new



(a) $\theta = 0°$. (b) $\theta = 60°$. (c) $\theta = 120°$. (d) $\theta = 180°$.

Fig. 17. Illustration of 4 different Gabor filters varying their orientations θ.

maps are introduced based on texture and a k-nearest neighbour (k-NN) classifier is used instead of a SVM classifier. The candidate regions are then extracted by performing a local maxima detection followed by post-processing region-growing and morphological operations.

E. CADx: Feature detection

Discriminative features which help to recognize CaP from healthy tissue need to be first detected. This processing is known in computer vision as feature extraction. However, feature extraction also refers to the name given in pattern recognition to some types of dimension reduction methods which are later presented. In order to avoid confusion between these two aspects, in this survey, the procedure "detecting" or "extracting" features from images and signals is defined as feature detection. This section summarizes the different features used in CAD for CaP.

1) Image-based features: This section focuses on image-based features which can be categorized into two categories: (i) voxel-wise detection and (ii) region-wise detection.

Voxel-wise detection in MRI This strategy refers to the fact that a feature is extracted at each voxel location. As discussed in Sect. II, CaP has an influence on the SI in mp-MRI images. Therefore, intensity-based feature is the most commonly used feature [113, 114, 116, 117, 118, 119, 120, 121, 122, 123, 125, 126, 127, 128, 129, 130, 131, 132, 138, 139, 140, 141, 144, 145, 157, 158, 164]. This feature consists in the extraction of the intensity of the MRI modality of interest.

Edge-based features have also been used to detect SI changes but bring additional information regarding the SI transition. Each feature is computed by convolving the original image with an edge operator. Three operators are commonly used: (i) Prewitt operator [232], (ii) Sobel operator [233], and (iii) Kirsch operator [234]. These operators differ due to the kernel used which attenuates more or less the noise. Multiple studies used the resulting magnitude and orientation of the edges computed in their classification frameworks [121, 125, 127, 128, 138, 139, 146, 147, 153, 154, 156, 159, 162, 186].

Gabor filters [235, 236] offer an alternative to the usual edge detector, with the possibility to tune the direction and the frequency of the filter to encode a specific pattern. A Gabor filter is defined by the modulation of a Gaussian function with a sine wave which can be further rotated and is formalized as in Eq. 39.

$$g(x, y; \theta, \psi, \sigma, \gamma) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \cos\left(2\pi \frac{x'}{\lambda} + \psi\right) \ , \tag{39}$$

with

$$\begin{aligned} x' &= s\left(x\cos\theta + y\sin\theta\right) \ , \\ y' &= s\left(-x\sin\theta + y\cos\theta\right) \ , \end{aligned}$$

where λ is the wavelength of the sinusoidal factor, θ represents the orientation of the Gabor filter, ψ is the phase offset, σ is the standard deviation of the Gaussian envelope, γ is the spatial aspect ratio,

and s is the scale factor. In an effort to characterize pattern and texture, a bank of Gabor filters is usually created with different angles, scale, and frequency — refer to Fig. 17 — and then convolved with the image. Viswanath et al. [163], Tiwari et al. [155] and more recently Khalvati et al. [125] and Chung et al. [121] have designed a bank of Gabor filters to characterized texture and edge information in T_2-W-MRI and DW-MRI modalities.

Lemaitre [128] used a 3D Gabor filters bank. 3D Gabor filters [237] are not commonly used and we recall their formulation in Eq. (40).

$$g(\mathbf{x};\boldsymbol{\sigma},f,\theta,\phi)=\hat{g}(\mathbf{x};\boldsymbol{\sigma})\exp(j2\pi f\left(x\sin\theta\cos\phi+y\sin\theta\sin\phi+z\cos\theta\right))\ , \tag{40}$$

where,

$$\hat{g}(\mathbf{x};\boldsymbol{\sigma})=\frac{1}{(2\pi)^{\frac{3}{2}}}\exp\left(-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}+\frac{z^2}{\sigma_z^2}\right)\right)\ , \tag{41}$$

where $\mathbf{x}$ is the position vector $\{x,y,z\}$, $\boldsymbol{\sigma}$ is the standard deviation vector $\{\sigma_x,\sigma_y,\sigma_z\}$ of the 3D Gaussian envelope, f is the radial center frequency of the sine wave, θ is the elevation angle, and ϕ is the azimuth angle.

Additionally, Lemaitre [128] features based on phase congruency as proposed byKovesi are computed [238]. Therefore, from a set of Log-Gabor filter bank, the orientation image, the local weighted mean phase angle, and the phase angle are estimated at each voxel.

Texture-based features provide other characteristics discerning CaP from healthy tissue. The most common texture analysis for image classification is based on the gray-level co-occurence matrix (GLCM) with their related statistics which have been proposed by Haralick et al. in [239]. In a neighborhood around a central voxel, a GLCM is build considering each voxel pair defined by a specific distance and angle. Then, using the GLCM, a set of statistical features is computed as defined in Table VII and assigned to the location of the central voxel. Therefore, N — up to 14 — statistical maps are derived from the GLCM analysis, one per statistics presented in Table VII. GLCM is commonly used in CAD systems, on the different MRI modalities, namely T_2-W-MRI, DCE-MRI, or DW-MRI [115, 118, 119, 121, 125, 127, 128, 138, 139, 146, 147, 153, 154, 156, 157, 159, 161, 162, 163, 186]. However, the statistics extracted from the GLCM across studies vary. Along the same line, Rampun et al. extracted from T_2-W-MRI [146, 186] Tamura features [240] composed of three features to characterize texture: (i) coarseness, (ii) contrast, and (iii) directionality.

Lopes et al. used fractal analysis and more precisely a local estimation of the fractal dimension [241], to describe the texture roughness at a specific location. The fractal dimension is estimated through a wavelet-based method in multi-resolution analysis. They showed that cancerous tissues have a higher fractal dimension than healthy tissue.

Chan et al. described texture using the frequency signature via the discrete cosine transform (DCT)[242] defining a neighbourhood of $7\,\mathrm{px}\times 7\,\mathrm{px}$ for modalities used, namely T_2-W-MRI and DW-MRI. The DCT allows to decompose a portion of an image into a coefficient space, where few of these coefficients encode the significant information. The DCT coefficients are computed such as:

$$C_{k_1,k_2}=\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}p_{m,n}\cos\left[\frac{\pi}{M}\left(m+\frac{1}{2}\right)k_1\right]\cos\left[\frac{\pi}{N}\left(n+\frac{1}{2}\right)k_2\right]\ , \tag{42}$$

where C_{k_1,k_2} is the DCT coefficient at the position k_1,k_2, M and N are the dimension of the neighbourhood and $p_{m,n}$ is the pixel SI at the position $\{m,n\}$.

Lemaitre [128] incorporated those features in their CAD system.

Regarding other features, Viswanath et al. projected T_2-W-MRI images into the wavelet space, using the Haar wavelet, and used the resulting coefficients as features [163]. Litjens et al. computed the texture map based on T_2-W-MRI images using a Gaussian filter bank [129]. Likewise, Rampun

TABLE VII
THE 14 STATISTICAL FEATURES FOR TEXTURE ANALYSIS COMMONLY COMPUTED FROM THE GLCM p AS PRESENTED BY [239].

Statistical features	**Formula**
Angular second moment	$\sum_i \sum_j p(i,j)^2$
Contrast	$\sum_{n=0}^{N_g-1} n^2 \left[\sum_{i=1}^{N_g-1} \sum_{j=1}^{N_g-1} p(i,j)\right] \ , \lvert i-j \rvert = n$
Correlation	$\frac{\sum_i \sum_j (ij)p(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y}$
Variance	$\sum_i \sum_j (i-\mu)^2 p(i,j)$
Inverse difference moment	$\sum_i \sum_j \frac{1}{1+(i-\mu)^2} p(i,j)$
Sum average	$\sum_{i=2}^{2N_g} i p_{x+y}(i)$
Sum variance	$\sum_{i=2}^{2N_g} (i-f_s)^2 p_{x+y}(i)$
Sum entropy	$-\sum_{i=2}^{2N_g} p_{x+y}(i) \log p_{x+y}(i)$
Entropy	$-\sum_i \sum_j p(i,j) \log p(i,j)$
Difference variance	$\sum_{i=0}^{N_g-1} i^2 p_{x-y}(i)$
Difference entropy	$-\sum_{i=0}^{N_g-1} p_{x-y}(i) \log p_{x-y}(i)$
Info. measure of corr. 1	$\frac{S(X;Y)-S_1(X;Y)}{\max(S(X),S(Y))}$
Info. measure of corr. 2	$\sqrt{(1-\exp\left[-2(H_2(X;Y)-H(X;Y))\right])}$
Max. corr. coeff.	$\sqrt{\lambda_2}$, of $Q(i,j) = \sum_k \frac{p(i,k)p(j,k)}{p_x(i)p_y(k)}$

et al. employed a rotation invariant filter bank proposed in [243]. The bank is composed of 48 filters including Gaussian filters, first and second derivatives of Gaussian filters as well as Laplacian of Gaussian.

Region-wise detection in MRI Unlike the previous section, another strategy is to study a region instead of each pixel independently. Usually, the feature maps are computed using the method presented in voxel-based approach followed by a step in which features are computed in some specific delineated regions to characterize them.

The most common feature type is based on statistics and more specifically the statistic-moments such as mean, standard deviation, kurtosis, and skewness [101, 113, 114, 115, 118, 119, 121, 125, 127, 129, 130, 131, 138, 139, 146, 147, 153, 154, 156, 159, 161, 162, 163, 186]. Additionally, some studies extract additional statistical landmarks based on percentiles [101, 115, 127, 129, 130, 131, 138, 139, 165, 166, 167] The percentiles to use are manually determined by observing the PDF of the features and checking which values allow the best to differentiate malignant from healthy tissue.

Further statistics are computed through the use of histogram-based features. Liu et al. introduced 4 different types of histogram-based features to characterize hand-delineated lesions [133]. The first type corresponds to the histogram of the SI of the image. The second type is the histogram of oriented gradient (HOG) [244] which encodes the local shape of the object of interest by using the distribution of the gradient directions. This descriptor is extracted mainly in three steps. First, the gradient image and its corresponding magnitude and direction are computed. Then, the ROI is divided into cells and an oriented-based histogram is generated for each cell. At each pixel location, the orientation of the gradient votes for a bin of the histogram and this vote is weighted by the magnitude of the same gradient. Finally, the cells are grouped into blocks and each block is normalized. The third histogram-based type used in [133] is the shape context introduced in [245]. The shape context is also a way to describe the shape of an object of interest. First, a set of points defining edges have to be detected and for each point of each edge, a log-polar-based histogram is computed using the relative points distribution. The last set of histogram-based feature extracted is based on the framework

TABLE VIII
PARAMETERS USED AS FEATURES FOR A DCE SEMI-QUANTITATIVE ANALYSIS IN CAD SYSTEMS.

Semi-quantitative features	Explanations
Amplitude features:	
S_0	Amplitude at the onset of the enhancement
$S_{\max}$	Amplitude corresponding to 95% of the maximum amplitude
S_p	Amplitude corresponding to the maximum amplitude
S_f	Amplitude at the final time point
Time features:	
t_0	Time at the onset of the enhancement
$t_{\max}$	Time corresponding to 95% of the maximum amplitude
t_p	Time corresponding to the maximum amplitude
t_f	Final time
t_{tp}	Time to peak which is the time from t_0 to t_p
Derivatives and integral features:	
WI	Wash-in rate corresponding to the signal slope from t_0 to t_m or t_p
WO	Wash-out rate corresponding to the signal slope from t_m or t_p to t_p
$IAUC$	Initial area under the curve which is the area between t_0 to t_f

described in [246] which is using the Fourier transform of the histogram created via local binary pattern (LBP) [206]. LBP is generated by comparing the value of the central pixel with its neighbours, defined through a radius and the number of connected neighbours. Then, in the ROI, the histogram of the LBP distribution is computed. The discrete fourier transform (DFT) of the LBP histogram is used to make the feature invariant to rotation.

Another subset of features are anatomical-based features and have been used in [118, 119, 130, 131, 136]. Litjens et al. computed the volume, compactness, and sphericity related to the given region [130, 131]. Additionally, Litjens et al. also introduced a feature based on symmetry in which they compute the mean of a candidate lesion as well as its mirrored counter-part and compute the quotient as feature [131]. Lemaitre [128] incorporated those features in their CAD system. Matulewicz et al. introduced 4 features corresponding to the percentage of tissue belonging to the regions PZ, CG, periurethral region, or outside the prostate region for the considered ROI [136]. Finally, Cameron et al. defined 4 features based on morphology and asymmetry: (i) the difference of morphological closing and opening of the ROI, (ii) the difference of the initial perimeter and the one after removing the high-frequency components, (iii) the difference between the initial ROI and the one after removing the high-frequency components, and (iv) the asymmetry by computing the difference of the two areas splitting the ROI by its major axes [118, 119].

The last group of region-based feature is based on fractal analysis. This group of features is based on estimating the fractal dimension which is a statistical index representing the complexity of the analyzed texture. Lv et al. proposed two features based on fractal dimension: (i) texture fractal dimension and (ii) histogram fractal dimension [135]. The first feature is based on estimating the fractal dimension on the SI of each image and thus this feature is a statistical characteristic of the image roughness. The second fractal dimension is estimated using the PDF of each image and characterizes the complexity of the PDF. Lopes et al. proposed a 3D version to estimate the fractal dimension of a volume using a wavelet decomposition [134].

2) DCE-based features: DCE-MRI is more commonly based on a SI analysis over time as presented in Sect. II-C. In this section, the specific features extracted for DCE-MRI analysis are presented.

Whole-spectra approach Some studies are using the whole DCE time series as feature vector [113, 114, 128, 155, 159, 160]. In some cases, the high-dimensional feature space is reduced using dimension reduction methods as it will be presented in the Sect. III-G.

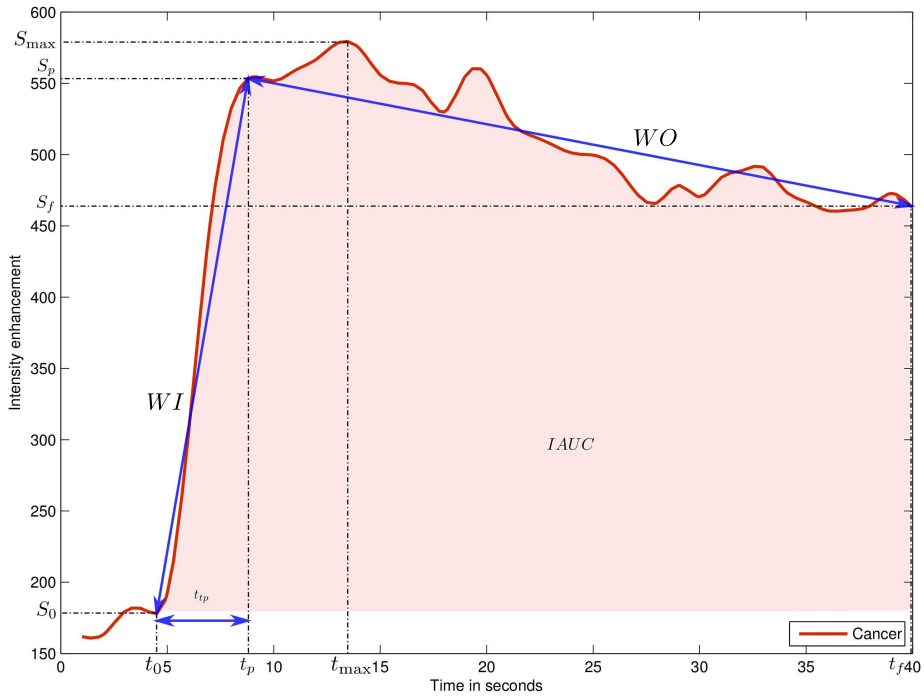


Fig. 18. Graphical representation of the different semi-quantitative features used for DCE-MRI analysis.

Semi-quantitative approach Semi-quantitative approaches are based on mathematically modelling the DCE time series. The parameters modelling the signal are commonly used, mainly due to the simplicity of their computation [123, 127, 128, 137, 138, 139, 143, 148, 149, 157, 158]. Parameters included in semi-quantitative analysis are summarized in Table VIII and also graphically depicted in Fig. 18. A set of time features corresponding to specific amplitude level (start, maximum, and end) are extracted. Then, derivative and integral features are also considered as discriminative and are commonly computed.

Quantitative approach As presented in Sect. II, quantitative approaches correspond to mathematical-pharmacokinetic models based on physiological exchanges. Four different models have been used in CAD for CaP systems. The most common model reviewed is the *Brix model* [116, 117, 128, 132, 140, 141, 149]. This model is formalized such as:

$$\frac{S(t)}{S(0)} = 1 + Ak_{ep}\left(\frac{\exp(-k_{ep}t) - \exp(-k_{el}t)}{k_{el} - k_{ep}}\right) \ , \tag{43}$$

where $S(\cdot)$ is the DCE signal, A is the parameter simulating the tissue properties, k_{el} is the parameter related to the first-order elimination from the plasma compartment, and k_{ep} is the parameter of the transvascular permeability. The parameters k_{ep}, k_{el}, and A are computed from the MRI data and used as features.

Another model is Tofts model [83] which has been used in [122, 123, 126, 127, 128, 137, 138, 139]. In this model, the DCE signal relative to the concentration is presented as:

$$C_t(t) = v_p C_p(t) + K_{trans}\int_0^t C_p(\tau)\exp(-k_{ep}(t-\tau))\ d\tau \ , \tag{44}$$

where $C_t(\cdot)$ is the concentration of the medium, $C_p(\cdot)$ is the AIF which has to be estimated independently, K_{trans} is the parameter related to the diffuse transport of media across the capillary endothelium, k_{ep} is the parameter related to the exchanges back into the vascular space, and v_e is the extravascular-extracellular space fraction defined such that $v_e = 1 - v_p$. In this model, parameters K_{trans}, k_{ep}, and v_e are computed and used as features.

Mazzetti et al., Giannini et al., and Lemaitre used the Weibull function [122, 123, 128, 137] which is formalized as:

$$S(t) = At\exp(-t^B) \ , \tag{45}$$

where A and B are the two parameters which have to be inferred.

They also used another empirical model which is based on the West-like function and named the phenomenological universalities (PUN) [247], formalized as:

$$S(t) = \exp\left[rt + \frac{1}{\beta}a_0 - r\left(\exp(\beta t) - 1\right)\right] \ , \tag{46}$$

where the parameters β, a_0 and r are inferred. For all these models, the parameters are inferred using an optimization curve fitting approach.

3) MRSI-based features:

Whole spectra approach in MRSI As in the case of DCE analysis, one common approach is to incorporate the whole MRSI spectra in the feature vector for classification [124, 128, 136, 142, 150, 152, 153, 154, 156, 157, 158, 160]. Sometimes post-processing involving dimension reduction methods is performed to reduce the complexity during the classification as it will be presented in Sect. III-G.

Quantification approach in MRSI We can reiterate that in MRSI only few biological markers — i.e., choline, creatine, and citrate metabolites — are known to be useful to discriminate CaP and healthy tissue. Therefore, only the concentrations of these metabolites are considered as a feature prior to classification. In order to perform this quantification, 4 different approaches have been used. Kelm et al. used the following models [124]: QUEST [248], AMARES [249], and VARPRO [250]. They are all time-domain quantification methods varying by the type of pre-knowledge embedded and the optimization approaches used to solve the quantification problem. Unlike the time-domain quantification approaches, Parfait et al. used the LcModel approach proposed in [251] which solves the optimization problem in the frequency domain. Although Parfait et al. and Lemaitre used each metabolite relative concentration individually [128, 142], other authors such as Kelm et al. proposed to compute relative concentrations as the ratios of metabolites as shown in Eq. 47 and Eq. 48.

$$R_1 = \frac{[\text{Cho}] + [\text{Cr}]}{[\text{Cit}]} \ . \tag{47}$$

$$R_2 = \frac{[\text{Cit}]}{[\text{Cho}] + [\text{Cr}] + [\text{Cit}]} \ , \tag{48}$$

where Cit, Cho and Cr are the relative concentration of citrate, choline, and creatine, respectively.

Recently Trigui et al. used an absolute quantification approach from which water sequences are acquired to compute the absolute concentration of the metabolites [157, 158]. Absolute quantification using water as reference is based on the fact that the fully relaxed signal from water or metabolites is proportional to the number of moles of the molecules in the voxel [252].

Wavelet decomposition approach in MRSI Tiwari et al. performed a wavelet packet decomposition [253] of the spectra using the Haar wavelet basis function and use its coefficients as features.

The feature detection methods used in CAD are summarized in Table IX.

TABLE IX

OVERVIEW OF THE FEATURE DETECTION METHODS USED IN CAD SYSTEMS. NOTES: (✓|- ✓|- ✓|-): TRIPLET STATING THE IMPLEMENTATION OR NOT OF THE FEATURE FOR RESPECTIVELY T_2-W-MRI IMAGES, DCE-MRI IMAGES, DW-MRI IMAGES; ✓: USED OR IMPLEMENTED; ✓[!]: PARTIALLY IMPLEMENTED.

Feature detection methods	**Indexes**
MRI image:	
Voxel-wise detection	
Intensity-based	✓- -[113, 114, 144, 145, 164] - - ✓[122] ✓- ✓[116, 117, 118, 119, 120, 121, 123, 125, 126, 129, 130, 131, 132, 140, 141, 157, 158] ✓✓✓[127, 128, 138, 139]
Edge-based	
• Prewitt operator	✓- -[153, 154, 156, 159] ✓-✓[128]
• Sobel operator	✓- -[146, 147, 153, 154, 156, 159, 161, 162, 163, 186] ✓-✓[128] ✓✓✓[127, 138, 139]
• Kirsch operator	✓- -[153, 154, 156, 159, 161, 162, 163] ✓-✓[121, 125, 128] ✓✓✓[127, 138, 139]
• Gabor filtering	✓- -[155, 159, 163] ✓-✓[121, 125, 128]
• Phase congruency	✓-✓[128]
Texture-based	
• Haralick features	✓- -[115, 146, 147, 153, 154, 156, 157, 159, 161, 163, 186] ✓✓-[162] ✓-✓[118, 119, 121, 125] ✓✓✓[127, 130, 138, 139]
• Tamura features	✓- -[146, 147, 186]
• Fractal analysis	✓- -[134, 135]
• DCT	✓✓✓[120]
• Wavelet-based features	✓- -[163]
• Gaussian filter bank	✓- -[131, 146, 147, 186]
• Laplacian of Gaussian filter bank	✓- -[146, 147, 186]
Position-based	[120, 129, 130, 131]
Region-wise detection	
Statistical-based	
• Percentiles	- ✓-[165] - - ✓[101, 115] ✓✓-[166] ✓✓✓[127, 129, 130, 131, 138, 139, 167]
• Statistical-moments	✓- -[113, 114, 146, 147, 153, 154, 156, 159, 161, 163, 186] - - ✓[115] ✓✓-[162] ✓- ✓[101, 118, 119, 121, 125] ✓✓✓[127, 129, 130, 131, 138, 139]
Histogram-based	
• PDF	✓✓✓[133]
• HOG	✓✓✓[133]
• Shape context	✓✓✓[133]
• LBP	✓✓✓[128, 133]
Anatomical-based	[118, 119, 130, 131, 136]
Fractal-based	[134, 135]
DCE signal:	
Whole spectra approach	[113, 114, 128]
Semi-quantitative approach	✓[!][143] [123, 127, 128, 137, 138, 139, 148, 149, 157, 158]
Quantitative approach	
• Toft model	✓[!][101, 133] [122, 123, 126, 127, 128, 129, 130, 131, 137, 138, 139]
• Brix model	✓[!][116, 117, 140, 141] [128, 132, 149]
• Weibull function	[122, 123, 128, 137]
• PUN	[122, 123, 128, 137]
MRSI signal:	
Whole spectra approach	[124, 128, 136, 142, 150, 151, 152, 153, 154, 156, 159]
Quantification approach	[124, 128, 142, 157, 158]
Wavelet-based approach	[155]

F. Feature balancing

Data imbalanced is a recurrent issue in classification, notably in medical data. The problem of imbalanced dataset lies in the fact that one of the class has a smallest number of data — i.e., in medical data, the class corresponding to patients with a disease — compared with the other classes. Therefore, solving the problem of imbalanced is equivalent to under- or over-sampling part of the dataset to obtain equal number of samples in the different classes. Recently, Lemaitre used balancing methods during the learning stage to tackle this issue [254].

1) Under-sampling: Techniques that reduce the number of samples of the majority class to be equal to the number of samples of minority class are referred to as under-sampling (US) techniques.

Nearmiss (NM) offers three different methods to under-sample the majority class [255]. In nearmiss-1 (NM-1), samples from the majority class are selected such that for each sample, the average distance to the k neareast 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 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.

Instance-hardness-threshold (IHT) select samples with a high hardness threshold [256]. 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.

$$IH_h(\langle x_i, y_i \rangle) = 1 - p(y_i|x_i, h). \tag{49}$$

Therefore, under-sampling is performed by keeping the most probable samples — i.e, filtering 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, data can be balanced by over-sampling (OS) in which the new samples belonging to the minority class are generated, aiming at equalizing the number of samples in both classes.

Synthetic minority over-sampling techniques (SMOTE) is a method to generate new synthetic samples [257]. 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 $x_j = x_i + \sigma\left(x_{nn} - x_i\right)$, where σ is a random number in the interval $[0, 1]$.

SMOTE-borderline1 (SMOTE-b1) over-samples the minority class samples similarly to SMOTE [258]. 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 them. Then, s-NN of the minority class are selected to generate synthetic sample similarly to SMOTE.

SMOTE-borderline2 (SMOTE-b2) performs similarly to SMOTE-b1 [258]. 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.

The data balancing used in CAD systems are summarized in Table X.

G. CADx: Feature selection and feature extraction

As presented in the previous section, it is a common practise to extract a wide variety of features. While dealing with mp-MRI, the feature space created is a high-dimensional space which might mislead

TABLE X
OVERVIEW OF THE DATA BALANCING METHODS USED IN CAD SYSTEMS.

Balancing methods	References
Under-sampling:	
Nearmiss-1 & Nearmiss-2 & Nearmiss-3	[128]
Instance-hardness-threshold	[128]
Over-sampling:	
SMOTE & SMOTE-b1 & SMOTE-b2	[128]

or corrupt the classifier during the training phase. Therefore, it is of interest to reduce the number of dimensions before proceeding to the classification task. The strategies used can be grouped as: (i) feature selection and (ii) feature extraction. In this section only the methods used in CAD for CaP systems are presented.

1) Feature selection: The feature selection strategy is based on selecting the most discriminative feature dimensions of the high-dimensional space. Thus, the low-dimensional space is then composed of a subset of the original features detected. In this section, methods employed in CAD for CaP detection are presented. A more extensive review specific to feature selection is available in [259].

Niaf et al. make use of the p-value by using the independent two-sample t-test with equal mean for each feature dimension [138, 139]. In this statistical test, there are 2 classes: CaP and healthy tissue. Hence, for each particular feature, the distribution of each class is characterized by their means $\bar{X}_1$ and $\bar{X}_2$ and standard deviation s_{X_1} and s_{X_2}. Therefore, the null hypothesis test is based on the fact that these both distribution means are equal. The t-statistic used to verify the null hypothesis is formalized such that:

$$\begin{aligned} t &= \frac{\bar{X}_1 - \bar{X}_2}{s_{X_1X_2} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \ , \\ s_{X_1X_2} &= \sqrt{\frac{(n_1-1)s_{X_1}^2 + (n_2-1)s_{X_2}^2}{n_1+n_2-2}} \ , \end{aligned} \tag{50}$$

where n_1 and n_2 are the number of samples in each class. From Eq. (50), more the means of the class distribution diverge, the larger the t-statistic t will be, implying that this particular feature is more relevant and able to make the distinction between the two classes.

The p-value statistic is deduced from the t-test and corresponds to the probability of obtaining such an extreme test assuming that the null hypothesis is true [260]. Hence, smaller the p-value, the more likely the null hypothesis to be rejected and more relevant the feature is likely to be. Finally, the features are ranked and the most significant features are selected. However, this technique suffers from a main drawback since it assumes that each feature is independent, which is unlikely to happen and introduces a high degree of redundancy in the features selected.

Vos et al. in [167] employed a similar feature ranking approach but make use of the Fisher discriminant ratio to compute the relevance of each feature dimension. Taking the aforementioned formulation, the Fisher discriminant ratio is formalized as the ratio of the interclass variance to the intraclass variance as:

$$F_r = \frac{(\bar{X}_1 - \bar{X}_2)^2}{s_{X_1}^2 + s_{X_2}^2} \ . \tag{51}$$

Therefore, a relevant feature dimension is selected when the interclass variance is maximum and the intraclass variance in minimum. Once the features are ordered, the authors select the feature dimensions with the largest Fisher discriminant ratio.

Lemaitre used the one-way analysis of variance (ANOVA) test [128]. This test is based on computing the F-test which is the ratio of the between-group variability over the with-in group variability. The F-value is computed for each pair of features and the K feature dimensions corresponding to the largest F-values are kept.

MI is a possible metric to use for selecting a subset of feature dimensions. This method has previously been presented in Sect. III-C and expressed in Eq. (35). Peng et al. [261] introduced two main criteria to select the feature dimensions based on MI: (i) maximal relevance and (ii) minimum redundancy. Maximal relevance criterion is based on the paradigm that the classes and the feature dimension which has to be selected have to share a maximal MI and is formalized as:

$$\arg\max Rel(\mathbf{x}, c) = \frac{1}{|\mathbf{x}|} \sum_{x_i \in \mathbf{x}} MI(x_i, c) \ , \tag{52}$$

where $\mathbf{x} = \{x_i; i = 1, \cdots, d\}$ is a feature vector of d dimensions and c is the class considered. As in the previous method, using maximal relevance criterion alone imply an independence between each feature dimension. The minimal redundancy criterion enforce the selection of a new feature dimension which shares as little as possible MI with the previously selected feature dimensions such that:

$$\arg\min Red(\mathbf{x}) = \frac{1}{|\mathbf{x}|^2} \sum_{x_i, x_j \in \mathbf{x}} MI(x_i, x_j) \ . \tag{53}$$

Combination of these two criteria is known as the minimum redundancy maximum relevance (mRMR) algorithm [261]. Two combinations are usually used: (i) the difference or (ii) the quotient. This method has been used at several occasions for the selecting a subset of features prior to classification [121, 125, 127, 138, 139, 163].

Random forest (RF) provides information regarding the importance of each feature using the metric which is based on the Gini index. The feature importance in RF is linked with 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, adding the decrease of the Gini impurity along the tree gives information about the feature importance: the higher, the better. Therefore, one can add the decrease of the Gini impurity across all the trees of a forest and obtain the importance of a specific feature for this forest. Subsequently, the K most important features are selected to perform the feature selection. Lemaitre used this approach to make a feature selection when training their RF [128].

2) Feature extraction: The feature extraction strategy is related to dimension reduction methods but not selecting discriminative features. Instead, these methods aim at mapping the data from the high-dimensional space into a low-dimensional space to maximize the separability between the classes. As in the previous sections, only methods employed in CAD system are reviewed in this section. We refer the reader to [262] for a full review of feature extraction techniques.

PCA is the most commonly used linear mapping method in CAD systems. PCA is based on finding the orthogonal linear transform mapping the original data into a low-dimensional space. The space is defined such that the linear combinations of the original data with the kth greatest variances lie on the kth principal components [263]. The principal components are computed by using the eigenvectors-eigenvalues decomposition of the covariance matrix. Let $\mathbf{x}$ denote the data matrix. Then, the covariance matrix and eigenvectors-eigenvalues decomposition are defined as in Eq. (54), and Eq. (55), respectively. The eigenvectors-eigenvalues decomposition can be formalized as:

$$\Sigma = \mathbf{x}^{\mathrm{T}} \mathbf{x} \ . \tag{54}$$

$$\mathbf{v}^{-1} \Sigma \mathbf{v} = \Lambda \ , \tag{55}$$

where $\mathbf{v}$ are the eigenvectors matrix and Λ is a diagonal matrix containing the eigenvalues.

It is then possible to find the new low-dimensional space by sorting the eigenvectors using the eigenvalues and finally select the eigenvectors corresponding to the largest eigenvalues. The total variation that is the sum of the principal eigenvalues of the covariance matrix [262], usually corresponds to the $95\,\%$ to $98\,\%$ of the cumulative sum of the eigenvalues. Tiwari et al. used PCA in order to reduce the complexity of feature space [128, 151, 152, 155].

Sparse-PCA is another approach for feature extraction and dimension reduction [264]. Similarly to PCA, this approach projects the data as a linear combination of input data. However, instead of using original data, it uses a sparse representation of the data, and therefore projects them as linear combination of few input components rather than all of them. Referring to Eq. (55), the cost function of sparse-PCA is formulated to maximize the variance while maintaining the sparsity constraint:

$$\begin{aligned} &\arg\max \quad \mathbf{v}^{-1}\Sigma\mathbf{v} \ , \\ &\text{subject to } \|\mathbf{v}\|_2 = 1 \ , \\ &\|\mathbf{v}\|_0 \leq k \ . \end{aligned} \tag{56}$$

where k indicates that number of non-zero elements in $\mathbf{v}$.

Lemaitre used this approach to decompose the signal obtained from the DCE-MRI and MRSI acquisitions.

Similarly to PCA decomposition, independent components analysis (ICA) is projecting data on independent components [265]. However, it does not require orthogonality of the space and does not assume Gaussian distribution for each independent source. Therefore, opposite to PCA it can recover uniquely the signals themselves rather than linear subspace in which the signals lie [266]. This method has been used by Lemaitre [128].

Non-linear mapping has been also used for dimension reduction and is mainly based on Laplacian eigenmaps and locally linear embedding (LLE) methods. Laplacian eigenmaps also referred as spectral clustering in computer vision, aim to find a low-dimensional space in which the proximity of the data should be preserved from the high-dimensional space [217, 218]. Therefore, two adjacent data points in the high-dimensional space should also be close in the low-dimensional space. Similarly, two distant data points in the high-dimensional space should also be distant in the low-dimensional space. To compute this projection, an adjacency matrix is defined as:

$$W(i,j) = \exp \|\mathbf{x}_i - \mathbf{x}_j\|_2 \ , \tag{57}$$

where $\mathbf{x}_i$ and $\mathbf{x}_j$ are the two samples considered. Then, the low-dimensional space is found by solving the generalized eigenvectors-eigenvalues problem:

$$(D - W)\mathbf{y} = \lambda D\mathbf{y} \ , \tag{58}$$

where D is a diagonal matrix such that $D(i,i) = \sum_j W(j,i)$. Finally the low-dimensional space is defined by the k eigenvectors of the k smallest eigenvalues [218]. Tiwari et al. [150, 152, 153] and Viswanath et al. [159] used this spectral clustering to project their feature vector into a low-dimensional space. The feature space in these studies is usually composed of features extracted from a single or multiple modalities and then concatenated before applying the Laplacian eigenmaps dimension reduction technique.

Tiwari et al. used a slightly different approach by combining the Laplacian eigenmaps techniques with a prior multi-kernel learning strategy [152, 156]. First, multiple features are extracted from multiple modalities. The features of a single modality are then mapped to a higher-dimensional space via the Kernel trick [267], namely a Gaussian kernel. Then, each kernel is linearly combined to obtain a combined kernel K and the adjacency matrix W is computed. Finally, the same scheme as in the Laplacian eigenmaps is applied. However, in order to use the combined kernel, Eq. (58) is rewritten as:

$$K(D - W)K^{\mathrm{T}}\mathbf{y} = \lambda KDK^{\mathrm{T}}\mathbf{y} \ , \tag{59}$$

which is solved as a generalized eigenvectors-eigenvalues problem as previously. Viswanath et al. used Laplacian eigenmaps inside a bagging framework in which multiple embeddings are generated by successively selecting feature dimensions [162].

LLE is another common non-linear dimension reduction technique widely used, first proposed in [268]. LLE is based on the fact that a data point in the feature space is characterized by its neighbourhood. Thus, each data point in the high-dimensional space is transformed to represent a linear combination of its k-nearest neighbours. This can be expressed as:

$$\hat{\mathbf{x}}_i = \sum_j W(i,j)\mathbf{x}_j \ , \tag{60}$$

where $\hat{\mathbf{x}}_i$ are the data points estimated using its neighbouring data points $\mathbf{x}_j$, and W is the weight matrix. The weight matrix W is estimated using a least square optimization as in Eq. (61).

$$\begin{aligned} \hat{W} &= \arg\min_W \sum_i |\mathbf{x}_i - \sum_j W(i,j)\mathbf{x}_j|^2 \ , \\ &\text{subject to } \sum_j W(i,j) = 1 \ , \end{aligned} \tag{61}$$

Then, the essence of LLE is to project the data into a low-dimensional space, while retaining the data spatial organization. Therefore, the projection into the low-dimensional space is tackled as an optimization problem as:

$$\hat{\mathbf{y}} = \arg\min_{\mathbf{y}} \sum_i |\mathbf{y}_i - \sum_j W(i,j)\mathbf{y}_j|^2 \ . \tag{62}$$

This optimization is solved as an eigenvectors-eigenvalues problem by finding the kth eigenvectors corresponding to the kth smallest eigenvalues of the sparse matrix $(I-W)^{\mathrm{T}}(I-W)$.

Tiwari et al. used a modified version of the LLE algorithm in which they applied LLE in a bagging approach with multiple neighbourhood sizes [151]. The different embeddings obtained are then fused using the maximum likelihood (ML) estimation.

Another way of reducing the complexity of high-dimensional feature space is to use the family of so-called dictionary-based methods. Sparse coded features (SCF) representation has become very popular in other computer vision application and has been used by Lehaire et al. in [127]. The main goal of sparse modeling is to efficiently represent the images as a linear combination of a few typical patterns, called atoms, selected from a dictionary. Sparse coding consists of three main steps: sparse approximation, dictionary learning, and low-level features projection [269].

Sparse approximation - Given a dictionary $\mathbf{D} \in \mathbb{R}^{n\times K}$ composed of K atoms and an original signal $\mathbf{y} \in \mathbb{R}^n$ — i.e., one feature vector —, the sparse approximation corresponds to find the sparest vector $\mathbf{x} \in \mathbb{R}^K$ such that:

$$\arg\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2 \qquad \text{s.t. } \|\mathbf{x}\|_0 \leq \lambda \ , \tag{63}$$

where λ is a specified sparsity level.

Solving the above optimization problem is an NP-hard problem [270]. However, approximate solutions are obtained using greedy algorithms such as Matching Pursuit (MP) [271] or orthogonal matching pursuit (OMP) [272, 273].

Dictionary learning - As stated previously, the sparse approximation is computed given a specific dictionary $\mathbf{D}$, which involves a learning stage from a set of training data. This dictionary is learned using K-SVD which is a generalized version of K-means clustering and uses SVD. The dictionary is built, in

TABLE XI
OVERVIEW OF THE FEATURE SELECTION AND EXTRACTION METHODS USED IN CAD SYSTEMS.

Dimension reduction methods	References
Feature selection:	
Statistical test	[128, 138, 139, 167]
MI-based methods	[121, 125, 127, 128, 138, 139, 164]
Correlation-based methods	[146, 186]
Feature extraction:	
Linear mapping	
PCA	[151, 152]
Sparse-PCA	[128]
ICA	[128]
Non-linear mapping	
Laplacian eigenmaps	[150, 152, 153, 154, 159, 162]
LLE and LLE-based	[151, 152, 159, 160]
Dictionary-based learning	
Sparse coding	[127]
BoW	[144, 145]

an iterative manner by solving the optimization problem of Eq. (64), by alternatively computing the sparse approximation of $\mathbf{X}$ and the dictionary $\mathbf{D}$.

$$\arg\min_{\mathbf{D},\mathbf{X}} \|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_2 \qquad \text{s.t. } \|\mathbf{x}_i\|_1 \leq \lambda \ , \tag{64}$$

where $\mathbf{Y}$ is a training set of low-level descriptors, $\mathbf{X}$ is the associated sparse coded matrix — i.e., set of high-level descriptors — with a sparsity level λ, and $\mathbf{D}$ is the dictionary with K atoms. Given $\mathbf{D}$, $\mathbf{X}$ is computed using the batch-OMP algorithm, while given $\mathbf{X}$, $\mathbf{D}$ is sequentially updated, one atom at a time using SVD.

Low-level features projection - Once the dictionary is learned, each set of low-level features $\mathbf{F}_I$ previously extracted is encoded using the dictionary $\mathbf{D}$, solving the optimization problem presented in Eq. (63) such that $\mathbf{F}_I \simeq \mathbf{D}\mathbf{X}_I$.

The bag of words (BoW) approach offers an alternative method [274] for feature extraction. BoW was used by Rampun et al. in [144, 145]. This model represents the features by creating a codebook or visual dictionary, from the set of low-level features. The set of low-level features are clustered using *k*-means to create the dictionary with *k* clusters known as visual words. Once the codebook is created from the training set, the low-level descriptors are replaced by their closest word within the codebook. The final descriptor is a histogram of size *k* which represents the codebook occurrences for a given mapping.

The feature selection and extraction used in CAD systems are summarized in Table XI.

H. CADx: Classification

Once the feature vector has been extracted and eventually the complexity reduced, it is possible to make a decision and classify this feature vector to belong to CaP or healthy tissue. A full review of classification methods used in pattern recognition is available in [275].

Rule-based method Lv et al. make use of a decision stump classifier to distinguish CaP and healthy classes [135]. Puech et al. detect CaP by implementing a given set of rules and scores based on a medical support approach [143]. During the testing, the feature vector goes through these different rules, and a final score is computed resulting to a final decision.

Clustering methods k-nearest neighbour (k-NN) is one of the simplest supervised machine learning classification methods. In this method, a new unlabelled vector is assigned to the most represented class from its k nearest-neighbours in the feature space. The parameter k is usually an odd number

in order to avoid any tie case. k-NN has been one of the methods used in [138, 139, 145] mainly to make a comparison with different machine learning techniques. Litjens et al. used this method to roughly detect potential CaP voxels before performing a region-based classification [130].

The k-means algorithm is an unsupervised clustering method in which the data is partitioned into k clusters in an iterative manner. First, k random centroids are defined in the feature space and each data point is assigned to the nearest centroid. Then, the centroid position for each cluster is updated by computing the mean of all the samples belonging to this particular cluster. Both assignment and updating are repeated until the centroids are stable. The number of clusters k is usually defined as the number of classes. This algorithm can also be used for "on-line" learning. In case that new data has to be incorporated, the initial centroid positions correspond to the results of a previous k-means training and is followed by the assignment-updating stage previously explained. Tiwari et al. used k-means in an iterative procedure [150, 152]. Three clusters were defined corresponding to CaP, healthy, and non-prostate. k-means is repeatedly applied and at each iteration, the voxels corresponding to the largest cluster are excluded under the assumption that it is assigned to "non-prostate" cluster. The algorithm stopped when the number of voxels in all remaining clusters were smaller than a given threshold. Tiwari et al. and Viswanath et al. used k-means in a repetitive manner in order to be less sensitive to the centroids initialization [151, 159, 160]. Thus, k clusters are generated T times and the final assignment is performed by majority voting using a co-association matrix as proposed in [276].

Linear model classifiers Linear discriminant analysis (LDA) is used as a classification method in which the optimal linear separation between 2 classes is found by maximizing the inter-class variance and minimizing the intra-class variance [277]. The linear discriminant function is defined as:

$$\delta_k(\mathbf{x}_i) = \mathbf{x}_i^{\mathrm{T}}\Sigma^{-1}\mu_k - \frac{1}{2}\mu_k^{\mathrm{T}}\Sigma^{-1}\mu_k + \log(\pi_k) \ , \tag{65}$$

where $\mathbf{x}_i = \{x_1, x_2, \ldots, x_n\}$ is an unlabelled feature vector of n features, Σ is the covariance matrix of the training data, μ_k is the mean vector of the class k, and π_k is the prior probability of class k. To perform the classification, a sample $\mathbf{x}_i$ is assigned to the class which maximizes the discriminant function as in Eq. (66).

$$C(\mathbf{x}_i) = \arg\max_k \delta_k(\mathbf{x}_i) \ . \tag{66}$$

LDA has been used in [115, 120, 138, 139, 167].

Logistic regression is also used to perform binary classification and provides the probability of an observation to belong to a class. The posterior probability of one of the classes, c_1 is written as:

$$p(c_1|\mathbf{x}_i) = \frac{1}{1 + \exp(-\mathbf{w}^{\mathrm{T}}\mathbf{x}_i)} \ , \tag{67}$$

with $p(c_2|\mathbf{x}_i) = 1 - p(c_1|\mathbf{x}_i)$ and where $\mathbf{w}$ is the vector of the regression parameters allowing to obtain a linear combination of the input feature vector $\mathbf{x}_i$. Thus, an unlabelled observation $\mathbf{x}_i$ is assigned to the class which maximizes the posterior probability as shown in Eq. (68).

$$C(\mathbf{x}_i) = \arg\max_k p(C = k|\mathbf{x}_i) \ . \tag{68}$$

From Eq. (67), one can see that the key to classification using logistic regression model is to infer the set of parameters $\mathbf{w}$ through a learning stage using a training set. This vector of parameters $\mathbf{w}$ is inferred by estimating the maximum likelihood. This step is performed through an optimization scheme, using a quasi-Newton method [227], which seeks in an iterative manner for the local minimum in the derivative of Eq. (67). This method has been used to create a linear probabilistic model in [124, 127, 143, 186].

Non-linear model classifier Viswanath et al. used quadratic discriminant analysis (QDA) instead of LDA [163]. Unlike in LDA in which one assumes that the class covariance matrix Σ is identical for all classes, a covariance matrix Σ_k specific to each class is computed. Thus, Eq. (65) becomes:

$$\delta_k(\mathbf{x}_i) = \mathbf{x}_i^{\mathrm{T}}\Sigma_k^{-1}\mu_k - \frac{1}{2}\mu_k^{\mathrm{T}}\Sigma_k^{-1}\mu_k + \log(\pi_k) \ , \tag{69}$$

where $\mathbf{x}_i$ has additional terms corresponding to the pairwise products of individual features such as $\{x_1, x_2, \ldots, x_n, x_1^2, x_1x_2, \ldots x_n^2\}$. The classification scheme in the case of the QDA is identical to Eq. (66).

Probabilistic classifiers The most commonly used classifier is the naive Bayes classifier which is a probabilistic classifier assuming independence between each feature dimension [278]. This classifier is based on Bayes' theorem:

$$p(C = k|\mathbf{x}) = \frac{p(C)p(\mathbf{x}|C)}{p(\mathbf{x})} \ , \tag{70}$$

where $p(C = k|\mathbf{x})$ is the posterior probability, $p(C)$ is the prior probability, $p(\mathbf{x}|C)$ is the likelihood, and $p(\mathbf{x})$ is the evidence. However, the evidence term is usually discarded since it is not class dependent and plays the role of a normalization term. Hence, in a classification scheme, an unlabelled observation is classified to the class which maximizes the posterior probability as:

$$C(\mathbf{x}_i) = \arg\max_k p(C = k|\mathbf{x}_i) \ , \tag{71}$$

$$p(C = k|\mathbf{x}_i) = p(C = k)\prod_{j=1}^{n} p(x_{ij}, |C = k) \ , \tag{72}$$

where d is the number of dimensions of the feature vector $\mathbf{x}_i = \{x_{i1}, \cdots, x_{id}\}$. Usually, a model includes both the prior and likelihood probabilities and it is common to use an equal prior probability for each class or eventually a value based on the relative frequency derived from the training set. Regarding the likelihood probability, it is common to choose a Gaussian distribution to characterize each class. Thus, each class is characterized by two parameters: (i) the mean and (ii) the standard deviation. These parameters are inferred from the training set by using the maximum likelihood estimation (MLE) approach. The naive Bayes classifier has been used in [118, 119, 122, 137, 138, 139, 139, 144, 145, 146, 186].

Ensemble learning classifiers AdaBoost (AdB) is an adaptive method based on an ensemble learning method and initially proposed in [279]. AdB linearly combines several weak learners resulting into a final strong classifier. A weak learner is defined as a classification method performing slightly better than a random classifier. Popular choices regarding the weak learner classifiers are: decision stump or decision tree learners such as iterative dichotomiser 3 (ID3) [280], C4.5 [281], and classification and regression tree (CART) [282].

AdB is considered as an adaptive method in the way that the weak learners are selected. The selection is performed in an iterative manner. At each iteration t, the weak learner selected h_t corresponds to the one minimizing the classification error on a distribution of weights D_t, that is associated with the training samples. Each weak learner is assigned a weight α_t as:

$$\alpha_t = \frac{1}{2}\ln\frac{1-\epsilon_t}{\epsilon_t} \ , \tag{73}$$

where ϵ_t corresponds to the classification error rate of the weak learner on the distribution of weight D_t.

Before performing a new iteration, the distribution of weights D_t is updated such that the weights associated with the misclassified samples by h_t increase and the weights of well classified samples decrease as shown in Eq. (74).

$$D_{t+1}(i) = \frac{D_t(i) \exp\left(-\alpha_t y_i h_t(\mathbf{x}_i)\right)}{Z_t} \ , \tag{74}$$

where $\mathbf{x}_i$ is the ith sample corresponding to class y_i and Z_t is a normalization factor forcing D_{t+1} to be a probability distribution. This procedure allows to select a weak learner at the next iteration $t+1$ which will classify in priority the previous misclassified samples. Thus, after T iterations, the final strong classifier corresponds to the linear combination of the weak learners selected and the classification is performed such that:

$$C(\mathbf{x}_i) = \text{sign}\left(\sum_{t=1}^{T} \alpha_t h_t(\mathbf{x}_i)\right) \ . \tag{75}$$

Lopes et al. make use of the AdB classifier to perform their classification [134] while Litjens et al. used the GentleBoost variant [283] which provides a modification of the function affecting the weight at each weak classifier [131].

Gradient Boosting (GB) is an ensemble classifier [284] which similarly to AdB is formulated as in Eq. (76).

$$F(x) = \sum_{m=1}^{M} \gamma_m h_m(x) \ , \tag{76}$$

where $h_m(x)$ is a weak learner with its associated weight γ_m. As with AdB, $h_m(x)$ is chosen to minimize a loss function using the additive model F_{m-1}. The difference between AdB and GB lie in the fact the this minimization is tackle as a numerical optimization problem using the steepest descent. Lemaitre used the GB as a meta-classifier to fusion several RF classifiers [128].

RF is a classification method which is based on creating an ensemble of decision trees and was introduced in [285]. In the learning stage, multiple decision tree learners [282] are trained. However, each decision tree is trained using a different dataset. Each of these datasets corresponds to a bootstrap sample generated by randomly choosing n samples with replacement from the initially N samples available [286]. Then, randomization is also part of the decision tree growth. At each node of the decision tree, from the bootstrap sample of D dimensions, a number of $d \ll D$ dimensions will be randomly selected. Finally, the dth dimension in which the classification error is minimum is used. This best "split" classifier is often evaluated using MI or Gini index. Finally, each tree is grown as much as possible without using any pruning procedure. In the prediction stage, the unlabelled sample is introduced in each tree and each of them assign a class. Finally, it is common to use a majority voting approach to choose the final class label. The RF classifier has been used in [124, 128, 131, 144, 145, 146, 148, 155, 156, 157, 158, 161, 186].

Probabilistic boosting-tree is another ensemble learning classifier which shares principles with AdB but using them inside a decision tree [287]. In the training stage, the probabilistic boosting-tree method grows a decision tree and at each node, a strong classifier is learned in an almost comparable scheme to AdB. Once the strong learner is trained, the training set is split into two subsets which are used to train the next strong classifiers in the next descending nodes. Thus, three cases are conceivable to decide in which branch to propagate each sample training $\mathbf{x}_i$:

- if $q(+1, \mathbf{x}_i) - \frac{1}{2} > \epsilon$ then $\mathbf{x}_i$ is propagated to the right branch set and a weight $w_i = 1$ is assigned.
- if $q(-1, \mathbf{x}_i) - \frac{1}{2} > \epsilon$ then $\mathbf{x}_i$ is propagated to the left branch set and a weight $w_i = 1$ is assigned.
- else $\mathbf{x}_i$ will be propagated in both branches with $w_i = q(+1, \mathbf{x}_i)$ in the right branch and $w_i = q(-1, \mathbf{x}_i)$ in the left branch.

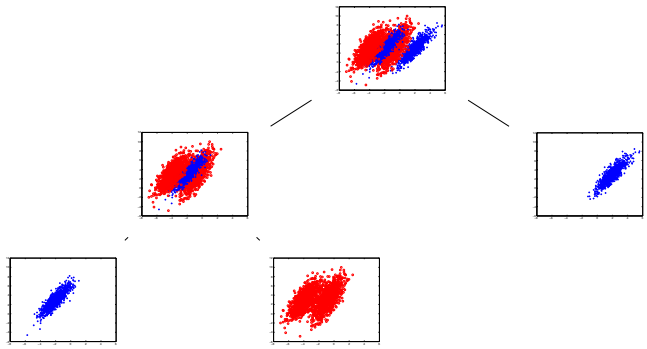


Fig. 19. Representation of the capabilities of the probabilistic boosting-tree algorithm to split at each node of the tree the positive and negative samples.

with $\mathbf{w} = w_i, i = \{1, \cdots, N\}$ corresponding to distribution of weights, N the number of samples as in AdB and $q(\cdot)$ is defined as:

$$q(+1, \mathbf{x}_i) = \frac{\exp(2H(\mathbf{x}_i))}{1 + \exp(2H(\mathbf{x}_i))} , \tag{77}$$

$$q(-1, \mathbf{x}_i) = \frac{\exp(-2H(\mathbf{x}_i))}{1 + \exp(-2H(\mathbf{x}_i))} . \tag{78}$$

Employing such a scheme tends to divide the data in such a way that positive and negative samples are naturally split as shown in Eq. 19. In the classification stage, the unlabelled sample $\mathbf{x}$ is propagated through the tree, where at each node, it is classified by each strong classifier previously learned and where an estimation of the posterior distribution is computed. The posterior distribution corresponds to the sum of the posterior distribution at each node of the decision tree. The probabilistic boosting-tree classifier has been used in [153, 154, 155, 162].

Kernel method A Gaussian process for classification is a kernel method in which it is assumed that the data can be represented by a single sample from a multivariate Gaussian distribution [288]. In the case of linear logistic regression for classification, the posterior probability is expressed as:

$$\begin{aligned} p(y_i|\mathbf{x}_i, \mathbf{w}) &= \sigma(y_i f(\mathbf{x}_i)) , \\ f(\mathbf{x}_i) &= \mathbf{x}_i^{\mathrm{T}} \mathbf{w} , \end{aligned} \tag{79}$$

where $\sigma(\cdot)$ is the logistic function and $\mathbf{w}$ are the parameters vector of the model. Thus, the classification using Gaussian processes is based on assigning a Gaussian process prior over the function $f(\mathbf{x})$ which is characterized by a mean function $\bar{f}$ and covariance function K. Therefore, in the training stage, the best mean and covariance functions have to be inferred in regard to our training data using a Newton optimization and a Laplacian approximation. The prediction stage is performed in two stages. First, for a new observation $\mathbf{x}_*$, the corresponding probability $p(f(\mathbf{x}_*)|f(\mathbf{x}))$ is computed such that:

$$\begin{aligned} p(f(\mathbf{x}_*)|f(\mathbf{x})) &= \mathcal{N}(K_* K^{-1} \bar{f}, K_{**} - K_*(K')^{-1} K_*^{\mathrm{T}}) , \\ K' &= K + W^{-1} , \\ W &= \nabla\nabla \log p(\mathbf{y}|f(\mathbf{x})) , \end{aligned} \tag{80}$$

where K_{**} is the covariance function $k(\mathbf{x}_*, \mathbf{x}_*)$ the testing sample $\mathbf{x}_*$, K_* is the covariance function $k(\mathbf{x}, \mathbf{x}_*)$ of training-testing samples $\mathbf{x}$ and $\mathbf{x}_*$. Then, the function $f(\mathbf{x}_*)$ is squashed using the sigmoid function and the probability of the class membership is defined such that:

$$C(\mathbf{x}_*) = \sigma\left(\frac{\bar{f}(\mathbf{x}_*)}{\sqrt{1+var(f(\mathbf{x}_*))}}\right) \ . \tag{81}$$

Only Kelm et al. used Gaussian process for classification of MRSI data [124].

Sparse kernel methods In a classification scheme using Gaussian processes, when a prediction is performed, the whole training data are used to assign a label to the new observations. That is why this method is also called kernel method. Sparse kernel category is composed of methods which rely only on a few labelled observations of the training set to assign the label of new observations [275].

Support vector machines (SVM) is a sparse kernel method aiming at finding the best linear hyper-plane — non-linear separation is discussed further — which separates 2 classes such that the margin between the two classes is maximized [289]. The margin is in fact the region defined by 2 hyper-planes splitting the 2 classes, such that there is no points lying in between. The distance between these 2 hyper-planes is equal to $\frac{2}{\|\mathbf{w}\|}$ where $\mathbf{w}$ is the normal vector of the hyper-plane splitting the classes. Thus, maximizing the margin is equivalent to minimizing the norm $\|\mathbf{w}\|$. Hence, this problem is solved by an optimization approach and formalized as:

$$\begin{aligned} &\arg\min_{\mathbf{w}} \quad \frac{1}{2}\|\mathbf{w}^2\| \ , \\ &\text{subject to} \quad y_i(\mathbf{w}.\mathbf{x}_i - b) \geq 1, \ i = \{1, \ldots, N\} \ , \end{aligned} \tag{82}$$

where $\mathbf{x}_i$ is a training sample with is corresponding class label y_i. From Eq. (82), it is important to notice that only few points from the set of N points are selected which later define the hyper-plane. This constraint is imposed in the optimization problem using Lagrange multipliers $\boldsymbol{\alpha}$. All points which are not lying on the margin are assigned a corresponding $\alpha_i = 0$, which is formalized as Eq. (83).

$$\arg\min_{\mathbf{w},b}\max_{\boldsymbol{\alpha}\geq 0}\left\{\frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{n}\alpha_i[y_i(\mathbf{w}\cdot\mathbf{x_i} - b) - 1]\right\} \ . \tag{83}$$

The different parameters are inferred using quadratic programming. This version of SVM is known as hard-margin since no points can lie in the margin area. However, it is highly probable not to find any hyper-plane splitting the classes such as specified previously. Thus, a soft-margin optimization approach has been proposed [290], where points have the possibility to lie on the margin but at the cost of a penalty ξ_i which is minimized in the optimization process such that:

$$\arg\min_{\mathbf{w},\xi,b}\max_{\boldsymbol{\alpha},\boldsymbol{\beta}}\left\{\frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n}\xi_i - \sum_{i=1}^{n}\alpha_i[y_i(\mathbf{w}\cdot\mathbf{x_i} - b) - 1 + \xi_i] - \sum_{i=1}^{n}\beta_i\xi_i\right\} \ . \tag{84}$$

The decision to assign the label to a new observation $\mathbf{x}_i$ is taken such that:

$$C(\mathbf{x}_i) = \text{sign}\left(\sum_{n=1}^{N}\alpha_n(\mathbf{x}_n.\mathbf{x}_i) + b_0\right) \ , \tag{85}$$

where $\mathbf{x}_n|n = \{1, \cdots, S\}$, S being the support vectors.

SVM can also be used as a non-linear classifier by performing a Kernel trick [291]. The original data $\mathbf{x}$ is projected to a high-dimensional space in which it is assumed that a linear hyper-plane splits the 2 classes. Different kernels are popular such as the RBF kernel, polynomial kernels, or sigmoid kernels. In CAD for CaP systems, SVM is the most popular classification method and has been used in a multitude of research works [101, 116, 117, 120, 121, 123, 125, 127, 129, 130, 133, 134, 138, 139, 140, 141, 142, 149, 155, 158, 164, 165, 166, 167].

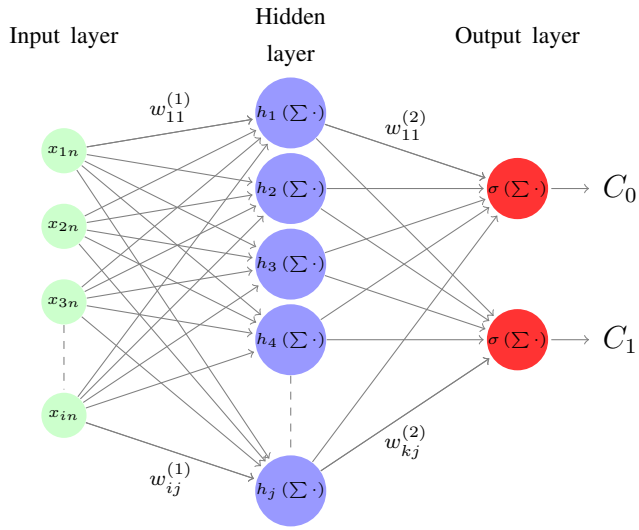


Fig. 20. Representation of a neural network of the multilayer perceptron family.

Relevant vector machine (RVM) is a sparse version of Gaussian process previously presented, proposed in [292]. RVM is identical to a Gaussian process with the following covariance function [293]:

$$K_{RVM}(\mathbf{x}_p, \mathbf{x}_q) = \sum_{j=1}^{M} \frac{1}{\alpha_j} \Phi_j(\mathbf{x}_p)\Phi_j(\mathbf{x}_q) \ , \tag{86}$$

where $\phi(\cdot)$ is a Gaussian basis function, $\mathbf{x}_i | i = \{1, \cdots, N\}$ are the N training points, and $\boldsymbol{\alpha}$ are the weights vector. As mentioned in [293], the sparsity regarding the relevance vector arises for some j, the weight $\alpha_j^{-1} = 0$. The set of weights $\boldsymbol{\alpha}$ is inferred using the expectation maximization algorithm. Ozer et al. used of RVM and make a comparison with SVM for the task of CaP detection [140, 141].

Neural network Multilayer perceptron is a feed-forward neural network considered as the most successful model of this kind in pattern recognition [275]. The most well known model used is based on 2 layers where a prediction of an observation is computed as:

$$C(\mathbf{x}_n, w_{ij}^{(1)}, w_{kj}^{(2)}) = \sigma\left[\sum_{j=0}^{M} w_{kj}^{(2)} \ h\left(\sum_{i=0}^{D} w_{ij}^{(1)} x_{in}\right)\right] \ , \tag{87}$$

where $h(\cdot)$ and $\sigma(\cdot)$ are 2 activation functions usually non-linear, $w_{ij}^{(1)}$ and $w_{kj}^{(2)}$ are the weights associated with the linear combination with the input feature $\mathbf{x}_n$ and the hidden unit.

A graphical representation of this network is presented in Eq. 20. Relating Fig. 20 with Eq. (87), it can be noted that this network is composed of some successive non-linear mapping of the input data. First, a linear combination of the input vector $\mathbf{x}_n$ is mapped into some hidden units through a set of weights $w_{ij}^{(1)}$. This combination becomes non-linear by the use of the activation function $h(\cdot)$ which is usually chosen to be a sigmoid function. Then, the output of the networks consists of a linear combination of the hidden units and the set of weights $w_{kj}^{(2)}$. This combination is also mapped non-linearly using an activation function $\sigma(\cdot)$ which is usually a logistic function. Thus, the training of such a network resides in finding the best weights $w_{ij}^{(1)}$ and $w_{kj}^{(2)}$ which model the best the training data. The error of this model is computed as:

$$E(w_{ij}^{(1)}, w_{kj}^{(2)}) = \frac{1}{2} \sum_{n=1}^{N} \left(C(\mathbf{x}_n, w_{ij}^{(1)}, w_{kj}^{(2)}) - y(\mathbf{x}_n)\right)^2 \ , \tag{88}$$

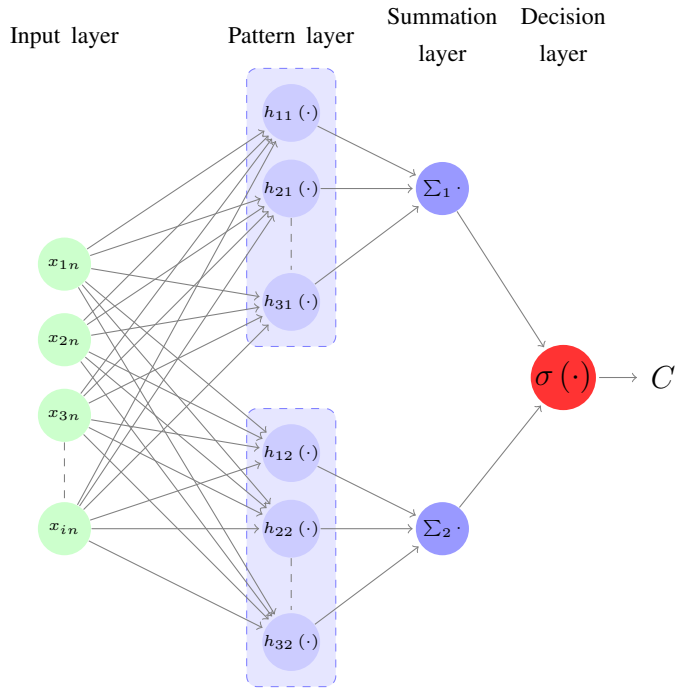


Fig. 21. Representation of a neural network of the probabilistic neural network family.

where $\mathbf{x}_n | n = \{1, \cdots, N\}$ are the N training vectors with their corresponding class label $y(\mathbf{x}_n)$.

Therefore, the best set of weights is inferred in an optimization framework where the error $E(\cdot)$ needs to be minimized. This optimization is performed using a gradient descent method where the derivative of Eq. (88) is computed using the back-propagation algorithm proposed by [294]. This type of network has been used multiple times [136, 142, 146, 157, 158].

Probabilistic neural networks are another type of feed-forward networks which is derived from the multilayer perceptron case and has been proposed by [295]. This classifier is modelled by changing the activation function $h(\cdot)$ in Eq. (87) to an exponential function such that:

$$h(\mathbf{x}_n) = \exp\left(-\frac{(\mathbf{w}_j - \mathbf{x})^{\mathrm{T}}(\mathbf{w}_j - \mathbf{x})}{2\sigma^2}\right) , \tag{89}$$

where σ is a free parameter set by the user.

The other difference of the probabilistic neural networks compared with the multilayer perceptron networks resides in the architecture as shown in Fig. 21. This network is formed by 2 hidden layers. The first hidden layer consists of the pattern layer, in which the mapping is done using Eq. (89). This pattern layer is sub-divided into a number of groups corresponding to the number of classes. The second hidden layer corresponds to the summation layer which simply sums the output of each sub-group of the pattern layer. This method is used in [113, 114, 162].

Graphical model classifiers Markov random field (MRF) is used as a lesion segmentation method to detect CaP. First, let define s as a pixel which belongs to a certain class denoted by ω_s. The labelling process is defined as $\omega = \{\omega_s, s \in I\}$ where I is the set of all the pixels inside the image. The observations corresponding to SI in the image are noted $\mathcal{F} = \{f_s | s \in I\}$. Thus, the image process $\mathcal{F}$ represents the deviation from the labelling process ω [296]. Hence, lesion segmentation is equivalent to estimating the best $\hat{\omega}$ which maximizes the posterior probability $p(\omega|\mathcal{F})$. Thus, using a Bayesian approach, this problem is formulated such that:

$$p(\omega|\mathcal{F}) = \arg\max_{\omega} \prod_{s \in I} p(f_s|\omega_s)p(\omega) \ . \tag{90}$$

It is generally assumed that $p(f_s|\omega_s)$ follows a Gaussian distribution and that the pixels classes $\lambda = \{1, 2\}$ for a binary classification are characterized by their respective mean μ_λ and standard deviation σ_λ. Then, ω is a Markov random field, thus:

$$p(\omega) = \frac{1}{Z} \exp\left(-U(\omega)\right) \ , \tag{91}$$

where Z is a normalization factor to obtain a probability value, $U(\cdot)$ is the energy function.

Thus, the segmentation problem is solved as an optimization problem where the energy function $U(\cdot)$ has to be minimized. There are different possibilities to define the energy function $U(\cdot)$. However, it is common to define the energy function such that it combines two types of potential function: (i) a local term relative to the pixel itself and (ii) a smoothing prior which embeds neighbourhood information which penalizes the energy function affecting the region homogeneity. This optimization of such a function can be performed using an algorithm such as iterated conditional modes [296]. Liu et al. and Ozer et al. used MRF as an unsupervised method to segment lesions in mp-MRI images [132, 141]. Artan et al. and Chung et al. used conditional random fields instead of MRF for MRI segmentation [116, 117, 121]. The difference between these 2 methods resides in the fact that conditional probabilities are defined such as:

$$p(\omega|\mathcal{F}) = \frac{1}{Z} \exp\left[-\sum_{s \in I} V_{C1}(\omega_s|\mathcal{F}) - \sum_{\{s,r\} \in C} V_{C2}(\omega_s, \omega_r|\mathcal{F})\right] \ . \tag{92}$$

$V_{C1}(\cdot)$ is the state (or partition) feature function and $V_{C2}(\cdot)$ is the transition (or edge) feature function [297].

Classification methods used to distinguish CaP from healthy tissue in in CAD systems are summarized in Table XII.

I. Model validation

In pattern recognition, the use of model validation techniques to assess the performance of a classifier plays an important role for reporting results. Two techniques are broadly used in the development of CAD systems and are summarized in Table XIII. The most popular technique used in CAD systems is the leave-one-out cross-validation (LOOCV) technique. From the whole data, one patient is kept for validation and the other cases are used for training. This manipulation is repeated until each patient has been used for validation. This technique is popular when working with a limited number of patients, allowing to train on representative number of cases even with a small dataset. However, LOOCV cross-validation suffers from a large variance and is considered as an unreliable estimate [298].

The other technique is the k-CV technique which is based on splitting the dataset into k subsets where the samples are randomly selected. Then, one fold is kept for testing and the remaining subsets are used for training. The classification is then repeated as in the LOOCV technique. In fact LOOCV is a particular case of k-CV when k equals the number of patients. In the reviewed papers, the typical values used for k has been set to three and five. k-CV is regarded as more appropriate than LOOCV, but the number of patients in the dataset needs to be large enough for the results to be meaningful.

TABLE XII
OVERVIEW OF THE CLASSIFIERS USED IN CAD SYSTEMS.

Classifier	References
Rule-based method:	[135, 143]
Clustering methods:	
k-means clustering	[150, 151, 152]
k-NN	[130, 138, 139, 145]
Linear model classifiers:	
LDA	[115, 120, 131, 138, 139, 167]
Logistic regression	[124, 126, 127, 186]
Non-linear classifier:	
QDA	[163]
Probabilistic classifier:	
Naive Bayes	[118, 119, 122, 137, 138, 139, 144, 145, 146, 186]
Ensemble learning classifiers:	
AdB	[131, 134]
GB	[128]
RF	[124, 128, 131, 144, 145, 146, 148, 155, 156, 157, 158, 161, 186]
Probabilistic boosting tree	[152, 154, 155]
Kernel method:	
Gaussian processes	[124]
Sparse kernel methods:	
SVM	[101, 116, 117, 120, 121, 123, 125, 127, 129, 130, 133, 134, 138, 139, 140, 141, 142, 149, 155, 158, 164, 165, 166, 167]
RVM	[140, 141]
Neural network:	
Multiple layer perceptron	[136, 142, 146, 157, 158]
Probabilistic neural network	[113, 114, 162]
Graphical model classifiers:	
Markov random field	[132, 141]
Conditional random field	[116, 117, 121]

TABLE XIII
OVERVIEW OF THE MODEL VALIDATION TECHNIQUES USED IN CAD SYSTEMS.

Model validation techniques	References
LOOCV	[101, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 124, 125, 127, 130, 131, 137, 138, 139, 140, 141, 143, 156, 162, 164, 164, 166]
k-CV	[129, 142, 144, 145, 146, 147, 152, 153, 154, 155, 157, 158, 161, 163, 167, 186]

TABLE XIV
OVERVIEW OF THE EVALUATION METRICS USED IN CAD SYSTEMS.

Evaluation metrics	References
Accuracy	[116, 117, 132, 149, 155]
Sensitivity - Specificity	[101, 116, 117, 118, 119, 122, 125, 132, 134, 137, 140, 141, 142, 148, 151, 152, 157, 158, 159, 160]
ROC - AUC	[101, 114, 115, 120, 122, 123, 124, 126, 127, 133, 134, 135, 136, 137, 138, 139, 144, 145, 146, 147, 153, 154, 155, 156, 161, 162, 163, 164, 165, 166, 186]
FROC	[129, 130, 167]
Dice's coefficient	[116, 117, 132, 140]

J. Evaluation measures

Several metrics are used in order to assess the performance of a classifier and are summarized in Table XIV. Voxels in the MRI image are classified into healthy or malign tissue and compared with a ground-truth. This allows to compute a confusion matrix by counting true positive (TP), true negative (TN), false positive (FP), and false negative (FN) samples. From this analysis, different statistics are extracted.

The first statistic used is the accuracy which is computed as the ratio of true detection to the number of samples. However, depending on the strategy employed in the CAD work-flow, this statistic is highly biased by a high number of true negative samples which boost the accuracy score overestimating the actual performance of the classifier. That is why, the most common statistics computed are sensitivity and specificity defined in Eq. (93) and Eq. (94), respectively. The metrics give a full overview of the performance of the classifier.

$$\mathrm{SE} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \ , \tag{93}$$

$$\mathrm{SP} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}} \ . \tag{94}$$

These statistics are also used to compute the receiver operating characteristic (ROC) curves [299], which give information about voxel-wise classification. This analysis represents graphically the sensitivity as a function of $(1-\text{specificity})$, which is in fact the false positive rate, by varying the discriminative threshold of the classifier. By varying this threshold, more true negative samples are found but often at the cost of detecting more false negatives. However, this fact is interesting in CAD since it is possible to obtain a high sensitivity and to ensure that no cancers are missed even if more false alarms have to be investigated or the opposite. A statistic derived from ROC analysis is the area under the curve (AUC) which corresponds to the area under the ROC and is a measure used to make comparisons between models.

The free-response receiver operating characteristic (FROC) extends the ROC analysis but to a lesion-based level. The same confusion matrix is computed where the sample are not pixels but lesions. However, it is important to define what is a true positive sample in that case. Usually, a lesion is considered as a true positive sample if the region detected by the classifier overlaps "sufficiently" the one delineated in the ground-truth. However, "sufficiently" is a subjective measure defined by each researcher and can correspond to one pixel only. However, an overlap of $30\,\%$ to $50\,\%$ is usually adopted. Finally, in addition to the overlap measure, the Dice's coefficient is often computed to evaluate the accuracy of the lesion localization. This coefficient consists of the ratio between twice the number of pixels in common and the sum of the pixels of the lesions in the ground-truth GT and the output of the classifier S, defined as shown in Eq. (95).

$$Q_D = \frac{2|\mathrm{GT} \cap \mathrm{S}|}{|\mathrm{GT}| + |\mathrm{S}|} \ . \tag{95}$$

IV. Discussion

A. Results reported

As discussed previously in Sect. III-J, different metrics have been used to report results. A comparison of the different methods reviewed is given depending on the metric used in field of research and also the type of MRI scanner used, i.e. $1.5\,\mathrm{T}$ or $3\,\mathrm{T}$. For each field, the *best classification performance* obtained in each study have been reported in these figures. The results in terms of AUC-ROC are depicted in Fig. 23. The results vary from $71\,\%$ to $97\,\%$ for some experiments with a $1.5\,\mathrm{T}$ MRI scanner and from $77\,\%$ to $95\,\%$ with a $3\,\mathrm{T}$ MRI scanner.

The results in regard of sensitivity and specificity are reported in Fig. 24. In the case that the data have been collected with a $1.5\,\mathrm{T}$ MRI scanner, the sensitivity ranges from $74\,\%$ to $100\,\%$ and the specificity

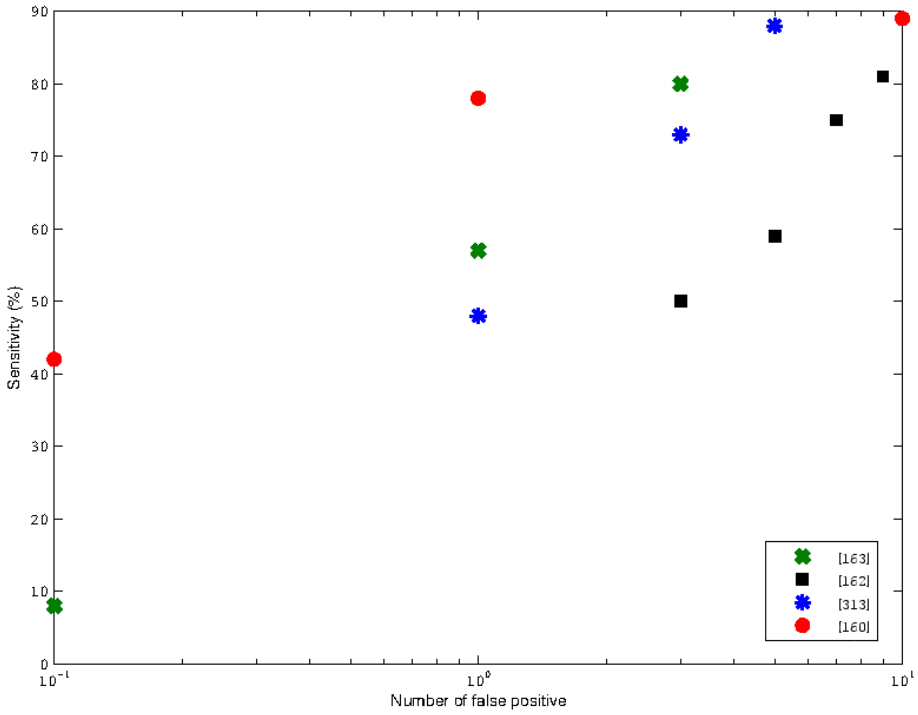


Fig. 22. Comparison in terms of FROC of the methods using data from 3 T MRI scanner.

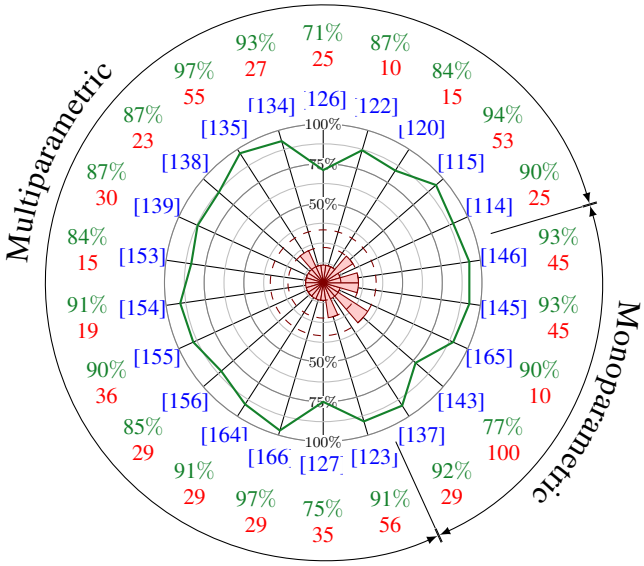
from 43 % to 93 %. For the experiments carried out with a 3 T MRI scanner, the sensitivity varies from 60 % to 99 % and the specificity from 66 % to 100 %. Four studies also use FROC analysis to report their results and are reported in Fig. 22.

B. Comparison

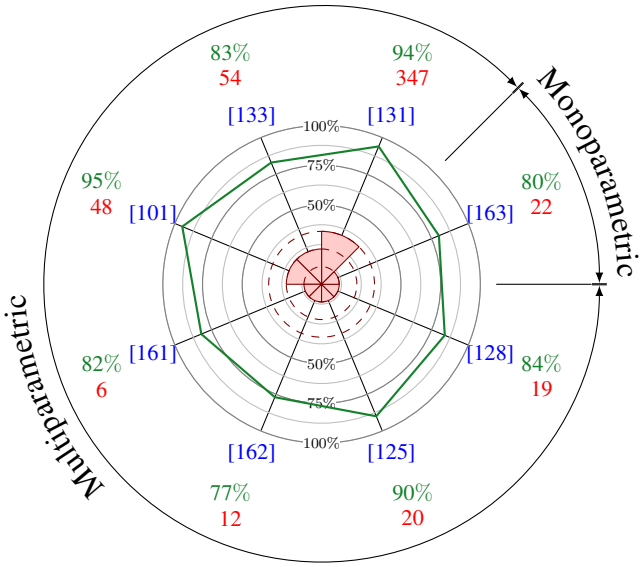
We would like to stress the following findings drawn during the review of the different studies:

1) Quantitatively, it is difficult to make a fair comparison between the different studies reviewed. Different factors come into play to elucidate this fact. Mainly a lack of standardization has to be pointed out in regard to experimental evaluation: (i) different datasets are used during the evaluation of the frameworks developed hindering an inter-study comparison. The same conclusion has been recently drawn by [131] supporting this argument; (ii) the experimental results are not reported with a common metric which leads to the inability to compare the different studies.

2) However, multiple studies reported some performance improvements using mp-MRI techniques instead of mono-parametric imaging techniques. Considering only the most recent studies proposing CADe-CADx frameworks, the following results can be highlighted. Viswanath et al. obtained an AUC of 77 % using an ensemble learning approach combining the features from the three MRI modalities — i.e., T_2-W-MRI, DCE–MRI, and DW-MRI, while the results obtained as standalone modality range from 62 % to 65 % [162]. Tiwari et al. drawn similar conclusions by using T_2-W-MRI and MRSI modalities as both in standalone and multi-parametric frameworks with an improved AUC ranging from 57 %-76 % to 85 % [156]. The most recent work of Litjens et al. obtained an improved AUC metric from 71 %-76 % considering each modality separately — i.e., T_2-W-MRI, DCE-MRI, and DW-MRI — to 89 % in their mp-MRI framework.



(a)



(b)

Fig. 23. Numerical and graphical comparison of the results in terms of AUC for 1.5 T (a) and 3 T (b) MRI scanners. The green value represents the metric and are graphically reported in the green curve in the center of the figure. The red value and areas correspond to the number of patients in the dataset. The numbers between brackets in blue correspond to the reference as reported in Table II.

3) The studies comparing particular combination of more than a single modality give rise to the same fact [129, 131, 133, 141]: using 3 modalities lead to better performances than using any combination of 2 modalities.

4) Unlike the previous remark 2, no straightforward conclusions can be given regarding the classification performance using each modality in a standalone framework. The modality being processed by different methods, it does not allow us to conclude if a modality by itself is more suited than another. However, we are able to distinguish some interesting trends which deserve the attention of the community. Tiwari et al. in [153, 155, 156] observed that MRSI is a more suitable modality than

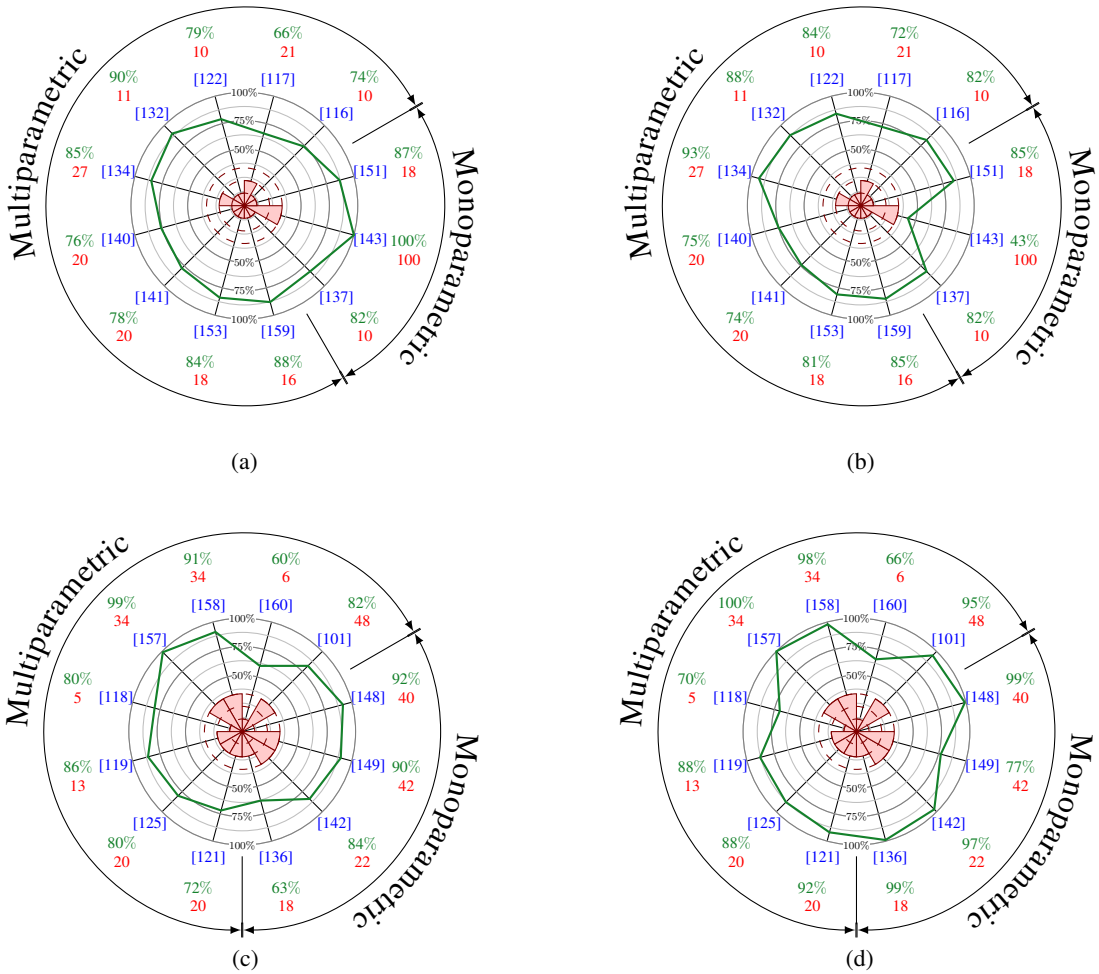


Fig. 24. Numerical and graphical comparison of the results in terms of SE (a), (c) and SP (b), (d) for 1.5 T and 3 T MRI scanners. The value in green represents the metric and are graphically reported in the green curve in the center of the figure. The red value and areas correspond to the number of patients in the dataset. The numbers between brackets in blue correspond to the reference as reported in Table II.

T_2-W to highlight CaP. Moreover, ADC maps have shown a better discriminative power than T_2-W as well [101, 126, 162]. Lately, Litjens et al. observed that DW-MRI modality is more suitable than both DCE-MRI and T_2-W-MRI to distinguish CaP in their CADx system [131]. Recently, Rampun et al. showed, however, some promising results using T_2-W-MRI only in conjunction with textons and BoW; this study should be transposed to other MRI modalities [145].

5) Furthermore, mp-MRI has attracted the attention of both radiologists and computer vision researchers. Indeed, pioneer research groups included new modalities over years when at the same time, new research groups directly introduced mp-MRI CAD systems. These facts lead us to think that CaP researches will benefit from mp-MRI techniques.

6) When focusing on the different modalities used, it can be pointed out that only Trigui et al. reported the use of all modalities in a single framework by incorporating the MRSI modality [157, 158]. Although the results reported are promising, the detection has been performed at MRSI scale. Lemaitre performed a similar study but at a finer resolution (i.e., T_2-W-MRI) [128]. They concluded that MRSI was significantly improving the classification performance for CaP detection.

7) Lately, 3 studies focused on developing a region-based classification in which PZ and CG will be

analyzed separately [130, 131, 163]. The promising obtained results indicate that this strategy should be further investigated.

8) Recent studies are using quantitative features in addition to SI. It seems that these quantitative features provide uncorrelated information with respect to SI features and should lead to better classification performance when combined all together.

9) Regarding the methods used in the "image regularization" — i.e., pre-processing, segmentation, and registration — it is particularly difficult to distinguish the benefit of a method over another since none of the studies focus on making comparison of these processing stages. The focus is usually entirely based on the "image classification" framework where different methods are directly compared. Note that the performance of a classifier is highly linked with the features vector extracted from particular data. Hence, one can not conclude that a machine learning method is more appropriate than another, but we can identify a trend in which SVM as well as ensemble learning classifiers — i.e., AdaBoost, GentleBoost, and RF — seem to perform better than neural network, LDA, or Naive Bayes.

10) We would like to draw the attention of the reader on the feature extraction/selection stage. This processing could reduce the complexity and also allow to find a better feature space for classification. However, few studies are performing such approaches. Niaf et al., Khalvati et al., Chung et al., Rampun et al., and Lemaitre are successfully applying a scheme to reduce the number of dimensions by selecting the most discriminative features [121, 125, 128, 138, 139, 146, 186]. It allows them to obtain improved performances compared with a classification performed with their initial feature vector. Another group of studies also applied different feature extraction methods [127, 128, 144, 145, 150, 151, 152, 154, 155, 156, 159, 160, 163]. In these specific cases, no comparison is performed against the original data.

11) Currently, only the work of Lemaitre tackled the problem of balancing in dataset. They found that balancing can have a positive effect on the classification performance. In our humble opinion, classification performance would benefit from further investigation in this area.

V. Conclusion

This chapter leads to some general discussions which could direct to future avenues for research. As previously mentioned, no open mp-MRI is currently available. This fact leads to an impossibility to fairly compare the different algorithms designed over years. Also, the availability of a full mp-MRI dataset, could lead to the development of algorithms which use all the different modalities currently available. Recalling Table II, it can be noted that a single research work provides a solution using at the same time the 4 different modalities. Also, all the algorithms are focused on one type of scanner only, either $1.5\,\mathrm{T}$ and $3\,\mathrm{T}$. A dataset including both these types of imaging could allow development of more generic algorithms.

Analyzing the different stages of the CAD work-flow, it is seen that the current CAD systems do not include all the pre-processing steps. It could be interesting to evaluate the improvement using these pre-processing steps on the final results. Regarding segmentation and registration of the prostate, CAD systems could greatly benefit from specific research in these areas which could lead to a better automation of those systems.

Additionally, few research works focuses on the problem of imbalanced dataset. While classifying at the voxel-level, the medical dataset are highly imbalanced regarding the frequencies of CaP against healthy samples. Imbalanced data substantially compromises the learning process since most of the standard machine learning algorithms expect balanced class distribution or an equal misclassification cost [300]. Therefore, it seems important to investigate this field of pattern recognition to improve the classification performance while developing CAD systems.

References

[1] K. H. Leissner and L. E. Tisell, "The weight of the human prostate," *Scand. J. Urol. Nephrol.*, vol. 13, no. 2, pp. 137–142, 1979.

[2] S. Parfait, "Classification de spectres et recherche de biomarqueurs en spectroscopie par résonqnce magnétique nulcléaire du proton dans les tumeurs prostatiques," Ph.D. dissertation, Université de Bourgogne, 2010.

[3] J. E. McNeal, "The zonal anatomy of the prostate," *Prostate*, vol. 2, pp. 35–49, 1981.

[4] H. Hricak, G. C. Dooms, J. E. McNeal, A. S. Mark, M. Marotti, A. Avallone, M. Pelzer, E. C. Proctor, and E. A. Tanagho, "MR imaging of the prostate gland: normal anatomy," *AJR Am J Roentgenol*, vol. 148, pp. 51–58, Jan 1987.

[5] A. Villers, A. Steg, and L. Boccon-Gibod, "Anatomy of the prostate: review of the different models," *Eur. Urol.*, vol. 20, pp. 261–268, 1991.

[6] F. V. Coakley and H. Hricak, "Radiologic anatomy of the prostate gland: a clinical approach," *Radiol. Clin. North Am.*, vol. 38, pp. 15–30, Jan 2000.

[7] Geckomedia. (2011, 06) Natom Anatomy. [Online]. Available: http://www.natomshop.com/

[8] Y. J. Choi, J. K. Kim, N. Kim, K. W. Kim, E. K. Choi, and K. S. Cho, "Functional MR imaging of prostate cancer," *Radiographics*, vol. 27, pp. 63–75, 2007.

[9] J. Ferlay, H. R. Shin, F. Bray, D. Forman, C. Mathers, and D. M. Parkin, "Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008," *Int. J. Cancer*, vol. 127, no. 12, pp. 2893–2917, Dec 2010.

[10] R. Siegel, D. Naishadham, and A. Jemal, "Cancer statistics, 2013," *CA Cancer J Clin*, vol. 63, no. 1, pp. 11–30, Jan 2013.

[11] A. C. American Cancer Society, "Cancer Facts and Figures 2013," http://www.cancer.org/research/cancerfactsfigures, Accessed the 06/2013., 2013, accessed: 2013-08-01.

[12] ——, "Cancer Facts and Figures 2010," http://www.cancer.org/research/cancerfactsfigures, 2010, accessed: 2013-08-01. [Online]. Available: http://www.cancer.org/research/cancerfactsfigures

[13] E. Giovannucci, Y. Liu, E. A. Platz, M. J. Stampfer, and W. C. Willett, "Risk factors for prostate cancer incidence and progression in the health professionals follow-up study," *Int. J. Cancer*, vol. 121, no. 7, pp. 1571–1578, Oct 2007.

[14] G. D. Steinberg, B. S. Carter, T. H. Beaty, B. Childs, and P. C. Walsh, "Family history and the risk of prostate cancer," *Prostate*, vol. 17, no. 4, pp. 337–347, 1990.

[15] M. L. Freedman, C. A. Haiman, N. Patterson, G. J. McDonald, A. Tandon, A. Waliszewska, K. Penney, R. G. Steen, K. Ardlie, E. M. John, I. Oakley-Girvan, A. S. Whittemore, K. A. Cooney, S. A. Ingles, D. Altshuler, B. E. Henderson, and D. Reich, "Admixture mapping identifies 8q24 as a prostate cancer risk locus in African-American men," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 103, no. 38, pp. 14 068–14 073, Sep 2006.

[16] L. T. Amundadottir, P. Sulem, J. Gudmundsson, A. Helgason, A. Baker, B. A. Agnarsson, A. Sigurdsson, K. R. Benediktsdottir, J. B. Cazier, J. Sainz, M. Jakobsdottir, J. Kostic, D. N. Magnusdottir, S. Ghosh, K. Agnarsson, B. Birgisdottir, L. Le Roux, A. Olafsdottir, T. Blondal, M. Andresdottir, O. S. Gretarsdottir, J. T. Bergthorsson, D. Gudbjartsson, A. Gylfason, G. Thorleifsson, A. Manolescu, K. Kristjansson, G. Geirsson, H. Isaksson, J. Douglas, J. E. Johansson, K. Balter, F. Wiklund, J. E. Montie, X. Yu, B. K. Suarez, C. Ober, K. A. Cooney, H. Gronberg, W. J. Catalona, G. V. Einarsson, R. B. Barkardottir, J. R. Gulcher, A. Kong, U. Thorsteinsdottir, and K. Stefansson, "A common variant associated with prostate cancer in European and African populations," *Nat. Genet.*, vol. 38, no. 6, pp. 652–658, Jun 2006.

[17] I. Agalliu, R. Gern, S. Leanza, and R. D. Burk, "Associations of high-grade prostate cancer with BRCA1 and BRCA2 founder mutations," *Clin. Cancer Res.*, vol. 15, no. 3, pp. 1112–1120, Feb 2009.

[18] R. M. Hoffman, F. D. Gilliland, J. W. Eley, L. C. Harlan, R. A. Stephenson, J. L. Stanford, P. C. Albertson, A. S. Hamilton, W. C. Hunt, and A. L. Potosky, "Racial and ethnic differences in advanced-stage prostate cancer: the Prostate Cancer Outcomes Study," *J. Natl. Cancer Inst.*, vol. 93, no. 5, pp. 388–395, Mar 2001.

[19] R. W. Ma and K. Chapman, "A systematic review of the effect of diet in prostate cancer prevention and treatment," *J Hum Nutr Diet*, vol. 22, no. 3, pp. 187–199, Jun 2009.

[20] D. D. Alexander, P. J. Mink, C. A. Cushing, and B. Sceurman, "A review and meta-analysis of prospective studies of red and processed meat intake and prostate cancer," *Nutr J*, vol. 9, p. 50, 2010.

[21] C. Rodriguez, S. J. Freedland, A. Deka, E. J. Jacobs, M. L. McCullough, A. V. Patel, M. J. Thun, and E. E. Calle, "Body mass index, weight change, and risk of prostate cancer in the Cancer Prevention Study II Nutrition Cohort," *Cancer Epidemiol. Biomarkers Prev.*, vol. 16, no. 1, pp. 63–69, Jan 2007.

[22] S. Strum and D. Pogliano, "What every doctor who treats male patients should know," PCRI Insights vol. 8, no. 2, May 2005.

[23] G. L. Lu-Yao, P. C. Albertsen, D. F. Moore, W. Shih, Y. Lin, R. S. DiPaola, M. J. Barry, A. Zietman, M. O'Leary, E. Walker-Corkery, and S. L. Yao, "Outcomes of localized prostate cancer following conservative management," *JAMA*, vol. 302, no. 11, pp. 1202–1209, Sep 2009.

[24] G. Oster, L. Lamerato, A. G. Glass, K. E. Richert-Boe, A. Lopez, K. Chung, A. Richhariya, T. Dodge, G. G. Wolff, A. Balakumaran, and J. Edelsberg, "Natural history of skeletal-related events in patients with breast, lung, or prostate cancer and metastases to bone: a 15-year study in two large US health systems," *Support Care Cancer*, vol. 21, no. 12, pp. 3279–3286, Dec 2013.

[25] L. Ye, H. G. Kynaston, and W. G. Jiang, "Bone metastasis in prostate cancer: molecular and cellular mechanisms (Review)," *Int. J. Mol. Med.*, vol. 20, no. 1, pp. 103–111, Jul 2007.

[26] C. L. Carrol, F. G. Sommer, J. E. McNeal, and T. A. Stamey, "The abnormal prostate: MR imaging at 1.5 T with histopathologic correlation," *Radiology*, vol. 163, no. 2, pp. 521–525, May 1987.

[27] J. E. McNeal, E. A. Redwine, F. S. Freiha, and T. A. Stamey, "Zonal distribution of prostatic adenocarcinoma. Correlation with histologic pattern and direction of spread," *Am. J. Surg. Pathol.*, vol. 12, no. 12, pp. 897–906, Dec 1988.

[28] T. A. Stamey, A. N. Donaldson, C. E. Yemoto, J. E. McNeal, S. Sozen, and H. Gill, "Histological and clinical findings in 896 consecutive prostates treated only with radical retropubic prostatectomy: epidemiologic significance of annual changes," *J. Urol.*, vol. 160, no. 6 Pt 2, pp. 2412–2417, Dec 1998.

[29] R. J. Cohen, B. A. Shannon, M. Phillips, R. E. Moorin, T. M. Wheeler, and K. L. Garrett, "Central zone carcinoma of the prostate gland: a distinct tumor type with poor prognostic features," *J. Urol.*, vol. 179, no. 5, pp. 1762–1767, May 2008.

[30] R. Etzioni, D. F. Penson, J. M. Legler, D. di Tommaso, R. Boer, P. H. Gann, and E. J. Feuer, "Overdiagnosis due to prostate-specific antigen screening: lessons from U.S. prostate cancer incidence trends," *J. Natl. Cancer Inst.*, vol. 94, no. 13, pp. 981–990, Jul 2002.

[31] C. M. Hoeks, J. O. Barentsz, T. Hambrock, D. Yakar, D. M. Somford, S. W. Heijmink, T. W. Scheenen, P. C. Vos, H. Huisman, I. M. van Oort, J. A. Witjes, A. Heerschap, and J. J. Futterer, "Prostate cancer: multiparametric MR imaging for detection, localization, and staging," *Radiology*, vol. 261, no. 1, pp. 46–66, Oct 2011.

[32] J. O. Barentsz, J. Richenberg, R. Clements, P. Choyke, S. Verma, G. Villeirs, O. Rouviere, V. Logager, and J. J. Futterer, "ESUR prostate MR guidelines 2012," *Eur Radiol*, vol. 22, no. 4, pp. 746–757, Apr 2012.

[33] D. F. Gleason, *Urologic pathology: The prostate*. Lea and Febiger., 1977, ch. The Veteran's Administration Cooperative Urologic Research Group: histologic grading and clinical staging of prostatic carcinoma, p. 171198.

[34] J. I. Epstein, W. C. Allsbrook, M. B. Amin, and L. L. Egevad, "The 2005 International Society of

Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma," *Am. J. Surg. Pathol.*, vol. 29, no. 9, pp. 1228–1242, Sep 2005.

[35] N. Hara, M. Okuizumi, H. Koike, M. Kawaguchi, and V. Bilim, "Dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) is a useful modality for the precise detection and staging of early prostate cancer," *Prostate*, vol. 62, no. 2, pp. 140–147, Feb 2005.

[36] R. Chou, J. M. Croswell, T. Dana, C. Bougatsos, I. Blazina, R. Fu, K. Gleitsmann, H. C. Koenig, C. Lam, A. Maltz, J. B. Rugge, and K. Lin, "Screening for prostate cancer: a review of the evidence for the U.S. Preventive Services Task Force," *Ann. Intern. Med.*, vol. 155, no. 11, pp. 762–771, Dec 2011.

[37] G. L. Andriole, E. D. Crawford, R. L. Grubb, S. S. Buys, D. Chia, T. R. Church, M. N. Fouad, E. P. Gelmann, P. A. Kvale, D. J. Reding, J. L. Weissfeld, L. A. Yokochi, B. O'Brien, J. D. Clapp, J. M. Rathmell, T. L. Riley, R. B. Hayes, B. S. Kramer, G. Izmirlian, A. B. Miller, P. F. Pinsky, P. C. Prorok, J. K. Gohagan, and C. D. Berg, "Mortality results from a randomized Prostate-cancer screening trial," *New England Journal of Medicine*, vol. 360, no. 13, pp. 1310–1319, 2009.

[38] F. H. Schröder, J. Hugosson, M. J. Roobol, T. L. Tammela, S. Ciatto, V. Nelen, M. Kwiatkowski, M. Lujan, H. Lilja, M. Zappa, L. J. Denis, F. Recker, A. Pez, L. Määttänen, C. H. Bangma, G. Aus, S. Carlsson, A. Villers, X. Rebillard, T. van der Kwast, P. M. Kujala, B. G. Blijenberg, U.-H. Stenman, A. Huber, K. Taari, M. Hakama, S. M. Moss, H. J. de Koning, and A. Auvinen, "Prostate-cancer mortality at 11 years of follow-up," *New England Journal of Medicine*, vol. 366, no. 11, pp. 981–990, 2012.

[39] J. Hugosson, S. Carlsson, G. Aus, S. Bergdahl, A. Khatami, P. Lodding, C. G. Pihl, J. Stranne, E. Holmberg, and H. Lilja, "Mortality results from the Göteborg randomised population-based prostate-cancer screening trial," *Lancet Oncol.*, vol. 11, no. 8, pp. 725–732, Aug 2010.

[40] A. Heidenreich, P. A. Abrahamsson, W. Artibani, J. Catto, F. Montorsi, H. Van Poppel, M. Wirth, and N. Mottet, "Early detection of prostate cancer: European Association of Urology recommendation," *Eur. Urol.*, vol. 64, no. 3, pp. 347–354, Sep 2013.

[41] F. H. Schroder, H. B. Carter, T. Wolters, R. C. van den Bergh, C. Gosselaar, C. H. Bangma, and M. J. Roobol, "Early detection of prostate cancer in 2007. Part 1: PSA and PSA kinetics," *Eur. Urol.*, vol. 53, no. 3, pp. 468–477, Mar 2008.

[42] C. Delpierre, S. Lamy, M. Kelly-Irving, F. Molinie, M. Velten, B. Tretarre, A. S. Woronoff, A. Buemi, B. Lapotre-Ledoux, S. Bara, A. V. Guizard, M. Colonna, and P. Grosclaude, "Life expectancy estimates as a key factor in over-treatment: the case of prostate cancer," *Cancer Epidemiol*, vol. 37, no. 4, pp. 462–468, Aug 2013.

[43] A. Bourdoumis, A. G. Papatsoris, M. Chrisofos, E. Efstathiou, A. Skolarikos, and C. Deliveliotis, "The novel prostate cancer antigen 3 (PCA3) biomarker," *Int Braz J Urol*, vol. 36, no. 6, pp. 665–668, 2010.

[44] R. Morgan, A. Boxall, A. Bhatt, M. Bailey, R. Hindley, S. Langley, H. C. Whitaker, D. E. Neal, M. Ismail, H. Whitaker, N. Annels, A. Michael, and H. Pandha, "Engrailed-2 (EN2): a tumor specific urinary biomarker for the early diagnosis of prostate cancer," *Clin. Cancer Res.*, vol. 17, no. 5, pp. 1090–1098, Mar 2011.

[45] J. Brenner, A. Chinnaiyan, and S. Tomlins, "ETS fusion genes in prostate cancer," in *Prostate Cancer*, ser. Protein Reviews, D. J. Tindall, Ed. Springer New York, 2013, vol. 16, pp. 139–183.

[46] C. M. Moore, A. Ridout, and M. Emberton, "The role of MRI in active surveillance of prostate cancer," *Curr Opin Urol*, vol. 23, no. 3, pp. 261–267, May 2013.

[47] M. Noguchi, T. A. Stamey, J. E. McNeal, and C. M. Yemoto, "Relationship between systematic biopsies and histological features of 222 radical prostatectomy specimens: lack of prediction of tumor significance for men with nonpalpable prostate cancer," *J. Urol.*, vol. 166, no. 1, pp. 104–109, Jul 2001.

[48] G. P. Haas, N. B. Delongchamps, R. F. Jones, V. Chandan, A. M. Serio, A. J. Vickers, M. Jumbelic, G. Threatte, R. Korets, H. Lilja, and G. de la Roza, "Needle biopsies on autopsy prostates: sensitivity

of cancer detection based on true prevalence," *J. Natl. Cancer Inst.*, vol. 99, no. 19, pp. 1484–1489, Oct 2007.

[49] A. V. Taira, G. S. Merrick, R. W. Galbreath, H. Andreini, W. Taubenslag, R. Curtis, W. M. Butler, E. Adamovich, and K. E. Wallner, "Performance of transperineal template-guided mapping biopsy in detecting prostate cancer in the initial and repeat biopsy setting," *Prostate Cancer Prostatic Dis.*, vol. 13, no. 1, pp. 71–77, Mar 2010.

[50] N. B. Delongchamps, M. Peyromaure, A. Schull, F. Beuvon, N. Bouazza, T. Flam, M. Zerbib, N. Muradyan, P. Legman, and F. Cornud, "Prebiopsy magnetic resonance imaging and prostate cancer detection: comparison of random and targeted biopsies," *J. Urol.*, vol. 189, no. 2, pp. 493–499, Feb 2013.

[51] M. L. Giger, H. P. Chan, and J. Boone, "Anniversary paper: History and status of CAD and quantitative image analysis: the role of Medical Physics and AAPM," *Med Phys*, vol. 35, no. 12, pp. 5799–5820, Dec 2008.

[52] 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*, vol. 266, no. 2, pp. 521–530, Feb 2013.

[53] H. P. Chan, B. Sahiner, M. A. Helvie, N. Petrick, M. A. Roubidoux, T. E. Wilson, D. D. Adler, C. Paramagul, J. S. Newman, and S. Sanjay-Gopal, "Improvement of radiologists' characterization of mammographic masses by using computer-aided diagnosis: an ROC study," *Radiology*, vol. 212, no. 3, pp. 817–827, Sep 1999.

[54] J. C. Dean and C. C. Ilvento, "Improved cancer detection using computer-aided detection with diagnostic and screening mammography: prospective study of 104 cancers," *AJR Am J Roentgenol*, vol. 187, no. 1, pp. 20–28, Jul 2006.

[55] F. Li, M. Aoyama, J. Shiraishi, H. Abe, Q. Li, K. Suzuki, R. Engelmann, S. Sone, H. Macmahon, and K. Doi, "Radiologists' performance for differentiating benign from malignant lung nodules on high-resolution CT using computer-estimated likelihood of malignancy," *AJR Am J Roentgenol*, vol. 183, no. 5, pp. 1209–1215, Nov 2004.

[56] N. Petrick, M. Haider, R. M. Summers, S. C. Yeshwant, L. Brown, E. M. Iuliano, A. Louie, J. R. Choi, and P. J. Pickhardt, "CT colonography with computer-aided detection as a second reader: observer performance study," *Radiology*, vol. 246, no. 1, pp. 148–156, Jan 2008.

[57] 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*, vol. 37, no. 5, pp. 1035–1054, May 2013.

[58] H. Hricak, R. D. Williams, D. B. Spring, K. L. Moon, M. W. Hedgcock, R. A. Watson, and L. E. Crooks, "Anatomy and pathology of the male pelvis by magnetic resonance imaging," *AJR Am J Roentgenol*, vol. 141, no. 6, pp. 1101–1110, Dec 1983.

[59] R. A. Huch Boni, J. A. Boner, U. M. Lutolf, F. Trinkler, D. M. Pestalozzi, and G. P. Krestin, "Contrast-enhanced endorectal coil MRI in local staging of prostate carcinoma," *J Comput Assist Tomogr*, vol. 19, no. 2, pp. 232–237, 1995.

[60] J. Kurhanewicz, D. B. Vigneron, H. Hricak, P. Narayan, P. Carroll, and S. J. Nelson, "Three-dimensional H-1 MR spectroscopic imaging of the in situ human prostate with high (0.24-0.7-cm3) spatial resolution," *Radiology*, vol. 198, no. 3, pp. 795–805, Mar 1996.

[61] J. Scheidler, R. Petsch, U. Muller-Lisse, A. Heuck, and M. Reiser, "Echo-planar diffusion-weighted MR imaging of the prostate," in *Proceedings of the 7th Annual Meeting of ISMRM Philadelphia*, 1999, p. 1103.

[62] M. G. Swanson, D. B. Vigneron, T. K. Tran, N. Sailasuta, R. E. Hurd, and J. Kurhanewicz, "Single-voxel oversampled J-resolved spectroscopy of in vivo human prostate tissue," *Magn Reson Med*, vol. 45, no. 6, pp. 973–980, Jun 2001.

[63] B. Turkbey and P. L. Choyke, "Multiparametric MRI and prostate cancer diagnosis and risk

stratification," *Curr Opin Urol*, vol. 22, no. 4, pp. 310–315, Jul 2012.

[64] O. Akin, E. Sala, C. S. Moskowitz, K. Kuroiwa, N. M. Ishill, D. Pucar, P. T. Scardino, and H. Hricak, "Transition zone prostate cancers: features, detection, localization, and staging at endorectal MR imaging," *Radiology*, vol. 239, no. 3, pp. 784–792, Jun 2006.

[65] L. Wang, Y. Mazaheri, J. Zhang, N. M. Ishill, K. Kuroiwa, and H. Hricak, "Assessment of biologic aggressiveness of prostate cancer: correlation of MR signal intensity with Gleason grade after radical prostatectomy," *Radiology*, vol. 246, no. 1, pp. 168–176, Jan 2008.

[66] A. P. Kirkham, M. Emberton, and C. Allen, "How good is MRI at detecting and characterising cancer within the prostate?" *Eur. Urol.*, vol. 50, no. 6, pp. 1163–1174, Dec 2006.

[67] L. E. Quint, J. S. Van Erp, P. H. Bland, S. H. Mandell, E. A. Del Buono, H. B. Grossman, G. M. Glazer, and P. W. Gikas, "Carcinoma of the prostate: MR images obtained with body coils do not accurately reflect tumor volume," *AJR Am J Roentgenol*, vol. 156, no. 3, pp. 511–516, Mar 1991.

[68] M. Cruz, K. Tsuda, Y. Narumi, Y. Kuroiwa, T. Nose, Y. Kojima, A. Okuyama, S. Takahashi, K. Aozasa, J. O. Barentsz, and H. Nakamura, "Characterization of low-intensity lesions in the peripheral zone of prostate on pre-biopsy endorectal coil MR imaging," *Eur Radiol*, vol. 12, no. 2, pp. 357–365, Feb 2002.

[69] W. Liu, B. Turkbey, J. Senegas, S. Remmele, S. Xu, J. Kruecker, M. Bernardo, B. J. Wood, P. A. Pinto, and P. L. Choyke, "Accelerated T2 mapping for characterization of prostate cancer," *Magn Reson Med*, vol. 65, no. 5, pp. 1400–1406, May 2011.

[70] G. P. Liney, M. Lowry, L. W. Turnbull, D. J. Manton, A. J. Knowles, S. J. Blackband, and A. Horsman, "Proton MR T2 maps correlate with the citrate concentration in the prostate," *NMR Biomed*, vol. 9, no. 2, pp. 59–64, Apr 1996.

[71] G. P. Liney, L. W. Turnbull, M. Lowry, L. S. Turnbull, A. J. Knowles, and A. Horsman, "In vivo quantification of citrate concentration and water T2 relaxation time of the pathologic prostate gland using 1H MRS and MRI," *Magn Reson Imaging*, vol. 15, no. 10, pp. 1177–1186, 1997.

[72] G. P. Liney, A. J. Knowles, D. J. Manton, L. W. Turnbull, S. J. Blackband, and A. Horsman, "Comparison of conventional single echo and multi-echo sequences with a fast spin-echo sequence for quantitative T2 mapping: application to the prostate," *J Magn Reson Imaging*, vol. 6, no. 4, pp. 603–607, 1996.

[73] P. Gibbs, D. J. Tozer, G. P. Liney, and L. W. Turnbull, "Comparison of quantitative T2 mapping and diffusion-weighted imaging in the normal and pathologic prostate," *Magn Reson Med*, vol. 46, no. 6, pp. 1054–1058, Dec 2001.

[74] I. Gribbestad, K. Gjesdal, G. Nilsen, S. Lundgren, M. Hjelstuen, and A. Jackson, "An introduction to dynamic contrast-enhanced MRI in oncology," in *Dynamic Contrast-Enhanced Magnetic Resonance Imaging in Oncology*, ser. Medical Radiology, A. Jackson, D. Buckley, and G. Parker, Eds. Springer Berlin Heidelberg, 2005, pp. 1–22.

[75] A. R. Padhani, "Dynamic contrast-enhanced MRI in clinical oncology: current status and future directions," *J Magn Reson Imaging*, vol. 16, no. 4, pp. 407–422, Oct 2002.

[76] D. L. Buckley, C. Roberts, G. J. Parker, J. P. Logue, and C. E. Hutchinson, "Prostate cancer: evaluation of vascular characteristics with dynamic contrast-enhanced T1-weighted MR imaging–initial experience," *Radiology*, vol. 233, no. 3, pp. 709–715, Dec 2004.

[77] C. G. van Niekerk, J. A. van der Laak, M. E. Borger, H. J. Huisman, J. A. Witjes, J. O. Barentsz, and C. A. Hulsbergen-van de Kaa, "Computerized whole slide quantification shows increased microvascular density in pT2 prostate cancer as compared to normal prostate tissue," *Prostate*, vol. 69, no. 1, pp. 62–69, Jan 2009.

[78] C. G. van Niekerk, J. A. Witjes, J. O. Barentsz, J. A. van der Laak, and C. A. Hulsbergen-van de Kaa, "Microvascularity in transition zone prostate tumors resembles normal prostatic tissue," *Prostate*, vol. 73, no. 5, pp. 467–475, Apr 2013.

[79] P. Carmeliet and R. K. Jain, "Angiogenesis in cancer and other diseases," *Nature*, vol. 407, no. 6801, pp. 249–257, Sep 2000.

[80] S. Verma, B. Turkbey, N. Muradyan, A. Rajesh, F. Cornud, M. A. Haider, P. L. Choyke, and M. Harisinghani, "Overview of dynamic contrast-enhanced MRI in prostate cancer diagnosis and management," *AJR Am J Roentgenol*, vol. 198, no. 6, pp. 1277–1288, Jun 2012.

[81] P. Tofts, "T1-weighted DCE imaging concepts: modelling, acquisition and analysis," in *Magneton Flash*. Siemens, 2010.

[82] S. Kety, "The theory and applications of the exchange of inert gas at the lungs and tissues," *Pharmacol. Rev.*, vol. 3, no. 1, pp. 1–41, Mar 1951.

[83] P. S. Tofts, "Modeling tracer kinetics in dynamic Gd-DTPA MR imaging," *J Magn Reson Imaging*, vol. 7, no. 1, pp. 91–101, 1997.

[84] H. B. Larsson, T. Fritz-Hansen, E. Rostrup, L. Sondergaard, P. Ring, and O. Henriksen, "Myocardial perfusion modeling using MRI," *Magn Reson Med*, vol. 35, no. 5, pp. 716–726, May 1996.

[85] K. S. St Lawrence and T. Y. Lee, "An adiabatic approximation to the tissue homogeneity model for water exchange in the brain: I. Theoretical derivation," *J. Cereb. Blood Flow Metab.*, vol. 18, no. 12, pp. 1365–1377, Dec 1998.

[86] A. B. Rosenkrantz, A. Sabach, J. S. Babb, B. W. Matza, S. S. Taneja, and F. M. Deng, "Prostate cancer: comparison of dynamic contrast-enhanced MRI techniques for localization of peripheral zone tumor," *AJR Am J Roentgenol*, vol. 201, no. 3, pp. W471–478, Sep 2013.

[87] G. J. Jager, E. T. Ruijter, C. A. van de Kaa, J. J. de la Rosette, G. O. Oosterhof, J. R. Thornbury, S. H. Ruijs, and J. O. Barentsz, "Dynamic TurboFLASH subtraction technique for contrast-enhanced MR imaging of the prostate: correlation with histopathologic results," *Radiology*, vol. 203, no. 3, pp. 645–652, Jun 1997.

[88] J. K. Kim, S. S. Hong, Y. J. Choi, S. H. Park, H. Ahn, C. S. Kim, and K. S. Cho, "Wash-in rate on the basis of dynamic contrast-enhanced MRI: usefulness for prostate cancer detection and localization," *J Magn Reson Imaging*, vol. 22, no. 5, pp. 639–646, Nov 2005.

[89] H. P. Schlemmer, J. Merkle, R. Grobholz, T. Jaeger, M. S. Michel, A. Werner, J. Rabe, and G. van Kaick, "Can pre-operative contrast-enhanced dynamic MR imaging for prostate cancer predict microvessel density in prostatectomy specimens?" *Eur Radiol*, vol. 14, no. 2, pp. 309–317, Feb 2004.

[90] B. Zelhof, M. Lowry, G. Rodrigues, S. Kraus, and L. Turnbull, "Description of magnetic resonance imaging-derived enhancement variables in pathologically confirmed prostate cancer and normal peripheral zone regions," *BJU Int.*, vol. 104, no. 5, pp. 621–627, Sep 2009.

[91] D. Le Bihan, E. Breton, D. Lallemand, M. L. Aubin, J. Vignaud, and M. Laval-Jeantet, "Separation of diffusion and perfusion in intravoxel incoherent motion MR imaging," *Radiology*, vol. 168, no. 2, pp. 497–505, Aug 1988.

[92] D. M. Koh and D. J. Collins, "Diffusion-weighted MRI in the body: applications and challenges in oncology," *AJR Am J Roentgenol*, vol. 188, no. 6, pp. 1622–1635, Jun 2007.

[93] D. M. Somford, J. J. Futterer, T. Hambrock, and J. O. Barentsz, "Diffusion and perfusion MR imaging of the prostate," *Magn Reson Imaging Clin N Am*, vol. 16, no. 4, pp. 685–695, Nov 2008.

[94] T. A. Huisman, "Diffusion-weighted imaging: basic concepts and application in cerebral stroke and head trauma," *Eur Radiol*, vol. 13, no. 10, pp. 2283–2297, Oct 2003.

[95] D. Le Bihan, E. Breton, D. Lallemand, P. Grenier, E. Cabanis, and M. Laval-Jeantet, "MR imaging of intravoxel incoherent motions: application to diffusion and perfusion in neurologic disorders," *Radiology*, vol. 161, no. 2, pp. 401–407, Nov 1986.

[96] R. Shimofusa, H. Fujimoto, H. Akamata, K. Motoori, S. Yamamoto, T. Ueda, and H. Ito, "Diffusion-weighted imaging of prostate cancer," *J Comput Assist Tomogr*, vol. 29, no. 2, pp. 149–153, 2005.

[97] A. R. Padhani, "Integrating multiparametric prostate MRI into clinical practice," *Cancer Imaging*, vol. 11 Spec No A, pp. 27–37, 2011.

[98] K. W. Doo, D. J. Sung, B. J. Park, M. J. Kim, S. B. Cho, Y. W. Oh, Y. H. Ko, and K. S. Yang, "Detectability of low and intermediate or high risk prostate cancer with combined T2-weighted and diffusion-weighted MRI," *Eur Radiol*, vol. 22, no. 8, pp. 1812–1819, Aug 2012.

[99] T. Hambrock, D. M. Somford, H. J. Huisman, I. M. van Oort, J. A. Witjes, C. A. Hulsbergen-van de Kaa, T. Scheenen, and J. O. Barentsz, "Relationship between apparent diffusion coefficients at 3.0-T MR imaging and Gleason grade in peripheral zone prostate cancer," *Radiology*, vol. 259, no. 2, pp. 453–461, May 2011.

[100] Y. Itou, K. Nakanishi, Y. Narumi, Y. Nishizawa, and H. Tsukuma, "Clinical utility of apparent diffusion coefficient (ADC) values in patients with prostate cancer: can ADC values contribute to assess the aggressiveness of prostate cancer?" *J Magn Reson Imaging*, vol. 33, no. 1, pp. 167–172, Jan 2011.

[101] Y. Peng, Y. Jiang, C. Yang, J. Brown, T. Antic, I. Sethi, C. Schmid-Tannwald, M. Giger, S. Eggener, and A. Oto, "Quantitative analysis of multiparametric prostate MR images: differentiation between prostate cancer and normal tissue and correlation with Gleason score–a computer-aided diagnosis development study," *Radiology*, vol. 267, no. 1, pp. 787–796, June 2013.

[102] H. M. Awwad, J. Geisel, and R. Obeid, "The role of choline in prostate cancer," *Clin. Biochem.*, vol. 45, no. 18, pp. 1548–1553, Dec 2012.

[103] L. C. Costello and R. B. Franklin, "The clinical relevance of the metabolism of prostate cancer; zinc and tumor suppression: connecting the dots," *Mol. Cancer*, vol. 5, p. 17, 2006.

[104] G. F. Giskeodegard, H. Bertilsson, K. M. Selnaes, A. J. Wright, T. F. Bathen, T. Viset, J. Halgunset, A. Angelsen, I. S. Gribbestad, and M. B. Tessem, "Spermine and citrate as metabolic biomarkers for assessing prostate cancer aggressiveness," *PLoS ONE*, vol. 8, no. 4, p. e62375, 2013.

[105] M. van der Graaf, R. G. Schipper, G. O. Oosterhof, J. A. Schalken, A. A. Verhofstad, and A. Heerschap, "Proton MR spectroscopy of prostatic tissue focused on the detection of spermine, a possible biomarker of malignant behavior in prostate cancer," *MAGMA*, vol. 10, no. 3, pp. 153–159, Jul 2000.

[106] E. Haacke, R. Brown, M. Thompson, and R. Venkatesan, *Magnetic resonance imaging: Physical principles and sequence design*. Wiley, 1999.

[107] G. Lemaître, "Absolute quantification at 3 T," Master's thesis, Université de Bourgogne, Heriot-Watt University, Universitat de Girona, 2011.

[108] J. Scheidler, H. Hricak, D. B. Vigneron, K. K. Yu, D. L. Sokolov, L. R. Huang, C. J. Zaloudek, S. J. Nelson, P. R. Carroll, and J. Kurhanewicz, "Prostate cancer: localization with three-dimensional proton MR spectroscopic imaging–clinicopathologic study," *Radiology*, vol. 213, no. 2, pp. 473–480, Nov 1999.

[109] Y. Kaji, J. Kurhanewicz, H. Hricak, D. L. Sokolov, L. R. Huang, S. J. Nelson, and D. B. Vigneron, "Localizing prostate cancer in the presence of postbiopsy changes on MR images: role of proton MR spectroscopic imaging," *Radiology*, vol. 206, no. 3, pp. 785–790, Mar 1998.

[110] J. C. Vilanova, J. Comet, C. Barceló-Vidal, J. Barceló, E. López-Bonet, A. Maroto, M. Arzoz, À. Moreno, and J. Areal, "Peripheral zone prostate cancer in patients with elevated PSA levels and low free-to-total PSA ratio: detection with MR imaging and MR spectroscopy," *Radiology*, vol. 253, no. 1, pp. 135–143, 2009.

[111] P. Walker, G. Crehange, S. Parfait, A. Cochet, P. Maignon, L. Cormier, and F. Brunotte, "Absolute quantification in 1H MRSI of the prostate at 3T," in *ISMRM Annual Meeting 2010*, 2010.

[112] S. Verma, A. Rajesh, J. J. Futterer, B. Turkbey, T. W. Scheenen, Y. Pang, P. L. Choyke, and J. Kurhanewicz, "Prostate MRI and 3D MR spectroscopy: how we do it," *AJR Am J Roentgenol*, vol. 194, no. 6, pp. 1414–1426, Jun 2010.

[113] D. Ampeliotis, A. Anonakoudi, K. Berberidis, and E. Z. Psarakis, "Computer aided detection of prostate cancer using fused information from dynamic contrast enhanced and morphological magnetic resonance images," in *IEEE International Conference on Signal Processing and Communications*, 2007, pp. 888–891.

[114] D. Ampeliotis, A. Anonakoudi, K. Berberidis, E. Z. Psarakis, and A. Kounoudes, "A computer-aided system for the detection of prostate cancer based on magnetic resonance image analysis," in *International Symposium on Communications, Control and Signal Processing*, 2008.

[115] T. Antic, Y. Peng, Y. Jiang, M. L. Giger, S. Eggener, and A. Oto, "A study of T2-weighted MR image texture features and diffusion-weighted MR image features for computer-aided diagnosis of prostate cancer," in *Proc. SPIE 8670, Medical Imaging 2013: Computer-Aided Diagnosis*, 2013, pp. 86 701H–86 701H–6.

[116] Y. Artan, D. Langer, M. Haider, T. H. Van der Kwast, A. Evans, M. Wernick, and I. Yetik, "Prostate cancer segmentation with multispectral MRI using cost-sensitive Conditional Random Fields," in *Biomedical Imaging: From Nano to Macro, 2009. ISBI '09. IEEE International Symposium on*, 2009, pp. 278–281.

[117] Y. Artan, M. A. Haider, D. L. Langer, T. H. van der Kwast, A. J. Evans, Y. Yang, M. N. Wernick, J. Trachtenberg, and I. S. Yetik, "Prostate cancer localization with multispectral MRI using cost-sensitive support vector machines and conditional random fields," *IEEE Trans Image Process*, vol. 19, no. 9, pp. 2444–2455, Sep 2010.

[118] A. Cameron, A. Modhafar, F. Khalvati, D. Lui, M. J. Shafiee, A. Wong, and M. Haider, "Multiparametric mri prostate cancer analysis via a hybrid morphological-textural model," in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2014, pp. 3357–3360.

[119] A. Cameron, F. Khalvati, M. A. Haider, and A. Wong, "Maps: a quantitative radiomics approach for prostate cancer detection," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 6, pp. 1145–1156, 2016.

[120] I. Chan, W. Wells, R. V. Mulkern, S. Haker, J. Zhang, K. H. Zou, S. E. Maier, and C. M. Tempany, "Detection of prostate cancer by integration of line-scan diffusion, T2-mapping and T2-weighted magnetic resonance imaging; a multichannel statistical classifier," *Med Phys*, vol. 30, no. 9, pp. 2390–2398, Sep 2003.

[121] A. Chung, F. Khalvati, M. Shafiee, M. Haider, and A. Wong, "Prostate cancer detection via a quantitative radiomics-driven conditional random field framework," *IEEE Access*, vol. 3, pp. 2531–2541, 2015.

[122] V. Giannini, A. Vignati, S. Mazzetti, M. De Luca, C. Bracco, M. Stasi, F. Russo, E. Armando, and D. Regge, "A prostate CAD system based on multiparametric analysis of DCE T1-w, and DW automatically registered images," in *Proc. SPIE 8670, Medical Imaging 2013: Computer-Aided Diagnosis*, 2013, pp. 86 703E–86 703E–6.

[123] V. Giannini, S. Mazzetti, A. Vignati, F. Russo, E. Bollito, F. Porpiglia, M. Stasi, and D. Regge, "A fully automatic computer aided diagnosis system for peripheral zone prostate cancer detection using multi-parametric magnetic resonance imaging," *Computerized Medical Imaging and Graphics*, vol. 46, pp. 219–226, 2015.

[124] B. M. Kelm, B. H. Menze, C. M. Zechmann, K. T. Baudendistel, and F. A. Hamprecht, "Automated estimation of tumor probability in prostate magnetic resonance spectroscopic imaging: pattern recognition vs quantification," *Magn Reson Med*, vol. 57, no. 1, pp. 150–159, Jan 2007.

[125] F. Khalvati, A. Wong, and M. A. Haider, "Automated prostate cancer detection via comprehensive multi-parametric magnetic resonance imaging texture feature models," *BMC medical imaging*, vol. 15, no. 1, p. 27, 2015.

[126] D. L. Langer, T. H. van der Kwast, A. J. Evans, J. Trachtenberg, B. C. Wilson, and M. A. Haider, "Prostate cancer detection with multi-parametric MRI: logistic regression analysis of quantitative T2, diffusion-weighted imaging, and dynamic contrast-enhanced MRI," *J Magn Reson Imaging*, vol. 30, no. 2, pp. 327–334, Aug 2009.

[127] J. Lehaire, R. Flamary, O. Rouvière, and C. Lartizien, "Computer-aided diagnostic system for prostate cancer detection and characterization combining learned dictionaries and supervised classification," in *2014 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2014, pp. 2251–2255.

[128] G. Lemaitre, "Computer-Aided Diagnosis for Prostate Cancer using Multi-Parametric Magnetic Resonance Imaging," Ph.D. dissertation, Universitat de Girona and Université de Bourgogne, 2016.

[129] G. J. S. Litjens, P. C. Vos, J. O. Barentsz, N. Karssemeijer, and H. J. Huisman, "Automatic computer aided detection of abnormalities in multi-parametric prostate MRI," in *Proc. SPIE 7963, Medical Imaging 2011: Computer-Aided Diagnosis*, 2011, pp. 79 630T–79 630T–7.

[130] G. J. S. Litjens, J. O. Barentsz, N. Karssemeijer, and H. J. Huisman, "Automated computer-aided detection of prostate cancer in MR images: from a whole-organ to a zone-based approach," in *Proc. SPIE 8315, Medical Imaging 2012: Computer-Aided Diagnosis*, 2012, pp. 83 150G–83 150G–6.

[131] G. Litjens, O. Debats, J. Barentsz, N. Karssemeijer, and H. Huisman, "Computer-iided detection of prostate cancer in MRI," *Medical Imaging, IEEE Transactions on*, vol. 33, no. 5, pp. 1083–1092, May 2014.

[132] X. Liu, D. L. Langer, M. A. Haider, Y. Yang, M. N. Wernick, and I. S. Yetik, "Prostate cancer segmentation with simultaneous estimation of Markov random field parameters and class," *IEEE Trans Med Imaging*, vol. 28, no. 6, pp. 906–915, Jun 2009.

[133] P. Liu, S. Wang, B. Turkbey, K. Grant, P. Pinto, P.and Choyke, B. J. Wood, and R. M. Summers, "A prostate cancer computer-aided diagnosis system using multimodal magnetic resonance imaging and targeted biopsy labels," in *Proc. SPIE 8670, Medical Imaging 2013: Computer-Aided Diagnosis*, 2013, pp. 86 701G–86 701G–6.

[134] R. Lopes, A. Ayache, N. Makni, P. Puech, A. Villers, S. Mordon, and N. Betrouni, "Prostate cancer characterization on MR images using fractal features," *Med Phys*, vol. 38, no. 1, pp. 83–95, Jan 2011.

[135] D. Lv, X. Guo, X. Wang, J. Zhang, and J. Fang, "Computerized characterization of prostate cancer by fractal analysis in MR images," *J Magn Reson Imaging*, vol. 30, no. 1, pp. 161–168, Jul 2009.

[136] L. Matulewicz, J. F. Jansen, L. Bokacheva, H. A. Vargas, O. Akin, S. W. Fine, A. Shukla-Dave, J. A. Eastham, H. Hricak, J. A. Koutcher, and K. L. Zakian, "Anatomic segmentation improves prostate cancer detection with artificial neural networks analysis of 1H magnetic resonance spectroscopic imaging," *Journal of Magnetic Resonance Imaging*, pp. n/a–n/a, 2013.

[137] S. Mazzetti, M. De Luca, C. Bracco, A. Vignati, V. Giannini, M. Stasi, F. Russo, E. Armando, S. Agliozzo, and D. Regge, "A CAD system based on multi-parametric analysis for cancer prostate detection on DCE-MRI," in *Proc. SPIE 7963, Medical Imaging 2011: Computer-Aided Diagnosis*, 2011, pp. 79 633Q–79 633Q–7.

[138] E. Niaf, O. Rouvire, and C. Lartizien, "Computer-aided diagnosis for prostate cancer detection in the peripheral zone via multisequence MRI," in *Proc. SPIE 7963, Medical Imaging 2011: Computer-Aided Diagnosis*, 2011.

[139] E. Niaf, O. Rouviere, F. Mege-Lechevallier, F. Bratan, and C. Lartizien, "Computer-aided diagnosis of prostate cancer in the peripheral zone using multiparametric MRI," *Phys Med Biol*, vol. 57, no. 12, pp. 3833–3851, Jun 2012.

[140] S. Ozer, M. Haider, D. L. Langer, T. H. Van der Kwast, A. Evans, M. Wernick, J. Trachtenberg, and I. Yetik, "Prostate cancer localization with multispectral MRI based on Relevance Vector Machines," in *Biomedical Imaging: From Nano to Macro, 2009. ISBI '09. IEEE International Symposium on*, 2009, pp. 73–76.

[141] S. Ozer, D. L. Langer, X. Liu, M. A. Haider, T. H. van der Kwast, A. J. Evans, Y. Yang, M. N. Wernick, and I. S. Yetik, "Supervised and unsupervised methods for prostate cancer segmentation with multispectral MRI," *Med Phys*, vol. 37, no. 4, pp. 1873–1883, Apr 2010.

[142] S. Parfait, P. Walker, G. Crhange, X. Tizon, and J. Mitran, "Classification of prostate magnetic resonance spectra using Support Vector Machine ," *Biomedical Signal Processing and Control*, vol. 7, no. 5, pp. 499 – 508, 2012.

[143] P. Puech, N. Betrouni, N. Makni, A. S. Dewalle, A. Villers, and L. Lemaitre, "Computer-assisted diagnosis of prostate cancer using DCE-MRI data: design, implementation and preliminary results," *Int J Comput Assist Radiol Surg*, vol. 4, no. 1, pp. 1–10, Jan 2009.

[144] A. Rampun, L. Zheng, P. Malcolm, and R. Zwiggelaar, "Classifying benign and malignant tissues within the prostate peripheral zone using textons," in *Medical Image Understanding and Analysis*

(MIUA 2015), 2015.

[145] A. Rampun, B. Tiddeman, R. Zwiggelaar, and P. Malcolm, "Computer aided diagnosis of prostate cancer: A texton based approach," *Medical Physics*, vol. 43, no. 10, pp. 5412–5425, 2016.

[146] A. Rampun, L. Zheng, P. Malcolm, B. Tiddeman, and R. Zwiggelaar, "Computer-aided detection of prostate cancer in t2-weighted mri within the peripheral zone." *Physics in medicine and biology*, vol. 61, no. 13, pp. 4796–4825, 2016.

[147] A. Rampun, L. Wang, P. Malcolm, and R. Zwiggelaar, "A quantitative study of texture features across different window sizes in prostate t2-weighted mri," *Procedia Computer Science*, vol. 90, pp. 74–79, 2016.

[148] G. Samarasinghe, A. Sowmya, and D. A. Moses, "Semi-quantitative analysis of prostate perfusion mri by clustering of pre and post contrast enhancement phases," in *Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on*. IEEE, 2016, pp. 943–947.

[149] Y. S. Sung, H. J. Kwon, B. W. Park, G. Cho, C. K. Lee, K. S. Cho, and J. K. Kim, "Prostate cancer detection on dynamic contrast-enhanced MRI: computer-aided diagnosis versus single perfusion parameter maps," *AJR Am J Roentgenol*, vol. 197, no. 5, pp. 1122–1129, Nov 2011.

[150] P. Tiwari, A. Madabhushi, and M. Rosen, "A hierarchical unsupervised spectral clustering scheme for detection of prostate cancer from magnetic resonance spectroscopy (MRS)," *Med Image Comput Comput Assist Interv*, vol. 10, no. Pt 2, pp. 278–286, 2007.

[151] P. Tiwari, M. Rosen, and A. Madabhushi, "Consensus-locally linear embedding (C-LLE): application to prostate cancer detection on magnetic resonance spectroscopy," *Med Image Comput Comput Assist Interv*, vol. 11, no. Pt 2, pp. 330–338, 2008.

[152] ——, "A hierarchical spectral clustering and nonlinear dimensionality reduction scheme for detection of prostate cancer from magnetic resonance spectroscopy (MRS)," *Med Phys*, vol. 36, no. 9, pp. 3927–3939, Sep 2009.

[153] P. Tiwari, M. Rosen, G. Reed, J. Kurhanewicz, and A. Madabhushi, "Spectral embedding based probabilistic boosting tree (ScEPTre): classifying high dimensional heterogeneous biomedical data," *Med Image Comput Comput Assist Interv*, vol. 12, no. Pt 2, pp. 844–851, 2009.

[154] P. Tiwari, J. Kurhanewicz, M. Rosen, and A. Madabhushi, "Semi supervised multi kernel (SeSMiK) graph embedding: identifying aggressive prostate cancer via magnetic resonance imaging and spectroscopy," *Med Image Comput Comput Assist Interv*, vol. 13, no. Pt 3, pp. 666–673, 2010.

[155] P. Tiwari, S. Viswanath, J. Kurhanewicz, A. Sridhar, and A. Madabhushi, "Multimodal wavelet embedding representation for data combination (MaWERiC): integrating magnetic resonance imaging and spectroscopy for prostate cancer detection," *NMR Biomed*, vol. 25, no. 4, pp. 607–619, Apr 2012.

[156] P. Tiwari, J. Kurhanewicz, and A. Madabhushi, "Multi-kernel graph embedding for detection, Gleason grading of prostate cancer via MRI/MRS," *Med Image Anal*, vol. 17, no. 2, pp. 219–235, Feb 2013.

[157] R. Trigui, J. Miteran, L. Sellami, P. Walker, and A. B. 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*. IEEE, 2016, pp. 113–118.

[158] R. Trigui, J. Mitéran, P. Walker, L. Sellami, and A. B. Hamida, "Automatic classification and localization of prostate cancer using multi-parametric mri/mrs," *Biomedical Signal Processing and Control*, vol. 31, pp. 189–198, 2017.

[159] S. Viswanath, P. Tiwari, M. Rosen, and A. Madabhushi, "A meta-classifier for detecting prostate cancer by quantitative integration of *In Vivo* magnetic resonance spectroscopy and magnetic resonance imaging," in *Medical Imaging 2008: Computer-Aided Diagnosis*, vol. 6915. SPIE, 2008.

[160] S. Viswanath, B. N. Bloch, E. Genega, N. Rofsky, R. Lenkinski, J. Chappelow, R. Toth, and A. Madabhushi, "A comprehensive segmentation, registration, and cancer detection scheme on 3 Tesla in vivo prostate DCE-MRI," *Med Image Comput Comput Assist Interv*, vol. 11, no. Pt 1, pp.

662–669, 2008.

[161] S. Viswanath, B. N. Bloch, M. Rosen, J. Chappelow, R. Toth, N. Rofsky, R. Lenkinski, E. Genega, A. Kalyanpur, and A. Madabhushi, "Integrating structural and functional imaging for computer assisted detection of prostate cancer on multi-protocol in vivo 3 Tesla MRI," in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, ser. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, vol. 7260, Feb. 2009.

[162] S. Viswanath, B. N. Bloch, J. Chappelow, P. Patel, N. Rofsky, R. Lenkinski, E. Genega, and A. Madabhushi, "Enhanced multi-protocol analysis via intelligent supervised embedding (EMPrAvISE): detecting prostate cancer on multi-parametric MRI," in *Proc. SPIE 7963, Medical Imaging 2011: Computer-Aided Diagnosis*, 2011.

[163] S. E. Viswanath, N. B. Bloch, J. C. Chappelow, R. Toth, N. M. Rofsky, E. M. Genega, R. E. Lenkinski, and A. Madabhushi, "Central gland and peripheral zone prostate tumors have significantly different quantitative imaging signatures on 3 Tesla endorectal, in vivo T2-weighted MR imagery," *J Magn Reson Imaging*, vol. 36, no. 1, pp. 213–224, Jul 2012.

[164] P. C. Vos, T. Hambrock, J. O. Barenstz, and H. J. Huisman, "Combining T2-weighted with dynamic MR images for computerized classification of prostate lesions," in *Medical Imaging 2008: Computer-Aided Diagnosis*, vol. 6915. SPIE, 2008.

[165] P. C. Vos, T. Hambrock, C. A. Hulsbergen-van de Kaa, J. J. Futterer, J. O. Barentsz, and H. J. Huisman, "Computerized analysis of prostate lesions in the peripheral zone using dynamic contrast enhanced MRI," *Med Phys*, vol. 35, no. 3, pp. 888–899, Mar 2008.

[166] P. C. Vos, T. Hambrock, J. O. Barenstz, and H. J. Huisman, "Computer-assisted analysis of peripheral zone prostate lesions using T2-weighted and dynamic contrast enhanced T1-weighted MRI," *Phys Med Biol*, vol. 55, no. 6, pp. 1719–1734, Mar 2010.

[167] P. C. Vos, J. O. Barentsz, N. Karssemeijer, and H. J. Huisman, "Automatic computer-aided detection of prostate cancer based on multiparametric magnetic resonance image analysis," *Phys Med Biol*, vol. 57, no. 6, pp. 1527–1542, Mar 2012.

[168] R. Nowak, "Wavelet-based Rician noise removal for magnetic resonance imaging," *Image Processing, IEEE Transactions on*, vol. 8, no. 10, pp. 1408–1419, 1999.

[169] J. V. Manjon, J. Carbonell-Caballero, J. J. Lull, G. Garcia-Marti, L. Marti-Bonmati, and M. Robles, "MRI denoising using non-local means," *Med Image Anal*, vol. 12, no. 4, pp. 514–523, Aug 2008.

[170] A. Buades, B. Coll, and J. Morel, "A review of image denoising algorithms, with a new one," *Simul*, vol. 4, pp. 490–530, 2005.

[171] J. Mohan, V. Krishnaveni, and Y. Guo, "A survey on the magnetic resonance image denoising methods," *Biomedical Signal Processing and Control*, vol. 9, no. 0, pp. 56 – 69, 2014.

[172] H. Ling and A. C. Bovik, "Smoothing low-snr molecular images via anisotropic median-diffusion," *IEEE Transactions on Medical Imaging*, vol. 21, no. 4, pp. 377–384, 2002.

[173] D. L. Donoho and J. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, 1994.

[174] A. Pizurica, "Image denoising using wavelets and spatial context modeling," Ph.D. dissertation, Universiteit Gent, 2002.

[175] S. Mallat, *A wavelet tour of signal processing, Third Edition: The sparse way*, 3rd ed. Academic Press, 2008.

[176] A. Pizurica, W. Philips, I. Lemahieu, and M. Acheroy, "A versatile wavelet domain noise filtration technique for medical imaging," *IEEE Trans Med Imaging*, vol. 22, no. 3, pp. 323–331, Mar 2003.

[177] M. Styner, C. Brechbuhler, G. Szckely, and G. Gerig, "Parametric estimate of intensity inhomogeneities applied to MRI," *Medical Imaging, IEEE Transactions on*, vol. 19, no. 3, pp. 153–165, 2000.

[178] M. Jungke, W. Von Seelen, G. Bielke, S. Meindl, M. Grigat, and P. Pfannenstiel, "A system for the diagnostic use of tissue characterizing parameters in NMR-tomography," in *Proc. of Information Processing in Medical Imaging*, vol. 87, 1987, pp. 471–481.

[179] U. Vovk, F. Pernus, and B. Likar, "A review of methods for correction of intensity inhomogeneity in MRI," *Medical Imaging, IEEE Transactions on*, vol. 26, no. 3, pp. 405–421, 2007.

[180] S. Z. Li, "Robustizing robust M-estimation using deterministic annealing," *Pattern Recognition*, vol. 29, pp. 159–166, 1996.

[181] M. Styner and G. Gerig, "Evaluation of 2D/3D bias correction with 1+1ES-optimization," ETH Zürich, Tech. Rep., 1997.

[182] J. G. Sled, A. P. Zijdenbos, and A. C. Evans, "A nonparametric method for automatic correction of intensity nonuniformity in MRI data," *IEEE Trans Med Imaging*, vol. 17, no. 1, pp. 87–97, Feb 1998.

[183] A. Madabhushi, J. Udupa, and A. Souza, "Generalized scale: Theory, algorithms, and application to image inhomogeneity correction ," *Computer Vision and Image Understanding*, vol. 101, no. 2, pp. 100 – 121, 2006.

[184] L. G. Shapiro and G. C. Stockman, *Computer vision*. Upper Saddle River, NJ: Prentice Hall, 2001.

[185] L. G. Nyul and J. K. Udupa, "On standardizing the MR image intensity scale," *Magn Reson Med*, vol. 42, no. 6, pp. 1072–1081, Dec 1999.

[186] A. Rampun, L. Zheng, P. Malcolm, and R. Zwiggelaar, "Computer aided diagnosis of prostate cancer within the peripheral zone in t2-weighted mri," in *Medical Image Understanding and Analysis (MIUA 2015)*, 2015.

[187] G. Lemaitre, M. R. Dastjerdi, J. Massich, J. C. Vilanova, P. M. Walker, J. Freixenet, A. Meyer-Baese, F. Mériaudeau, and R. Marti, "Normalization of t2w-mri prostate images using rician a priori," in *SPIE Medical Imaging*. International Society for Optics and Photonics, 2016, pp. 978 529–978 529.

[188] L. G. Nyul, J. K. Udupa, and X. Zhang, "New variants of a method of MRI scale standardization," *IEEE Trans Med Imaging*, vol. 19, no. 2, pp. 143–150, Feb 2000.

[189] A. Srivastava, E. Klassen, S. Joshi, and I. Jermyn, "Shape analysis of elastic curves in euclidean spaces," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 33, no. 7, pp. 1415–1428, July 2011.

[190] A. Madabhushi and J. K. Udupa, "New methods of MR image intensity standardization via generalized scale," *Med Phys*, vol. 33, no. 9, pp. 3426–3434, Sep 2006.

[191] M. Wiart, L. Curiel, A. Gelet, D. Lyonnet, J. Y. Chapelon, and O. Rouviere, "Influence of perfusion on high-intensity focused ultrasound prostate ablation: a first-pass MRI study," *Magn Reson Med*, vol. 58, no. 1, pp. 119–127, Jul 2007.

[192] L. Chen, Z. Weng, L. Goh, and M. Garland, "An efficient algorithm for automatic phase correction of {NMR} spectra based on entropy minimization ," *Journal of Magnetic Resonance*, vol. 158, no. 12, pp. 164 – 168, 2002.

[193] M. Osorio-Garcia, A. Croitor Sava, D. M. Sima, F. Nielsen, U. Himmelreich, and S. Van Huffel, *Magnetic Resonance Spectroscopy*. InTech, March 2012, ch. Quantification improvements of 1H MRS Signals, pp. 1–27.

[194] J. A. Nelder and R. Mead, "A simplex method for function minimization," *The Computer Journal*, vol. 7, no. 4, pp. 308–313, 1965.

[195] H. Zhu, R. Ouwerkerk, and P. B. Barker, "Dual-band water and lipid suppression for MR spectroscopic imaging at 3 Tesla," *Magn Reson Med*, vol. 63, no. 6, pp. 1486–1492, Jun 2010.

[196] W. Pijnappel, A. van den Boogaart, R. de Beer, and D. van Ormondt, "SVD-based quantification of magnetic resonance signals ," *Journal of Magnetic Resonance (1969)*, vol. 97, no. 1, pp. 122 – 134, 1992.

[197] T. Laudadio, N. Mastronardi, L. Vanhamme, P. V. Hecke, and S. V. Huffel, "Improved Lanczos algorithms for blackbox {MRS} data quantitation," *Journal of Magnetic Resonance*, vol. 157, no. 2, pp. 292 – 297, 2002.

[198] C. A. Lieber and A. Mahadevan-Jansen, "Automated method for subtraction of fluorescence from biological Raman spectra," *Appl Spectrosc*, vol. 57, no. 11, pp. 1363–1367, Nov 2003.

[199] A. Devos, L. Lukas, J. A. Suykens, L. Vanhamme, A. R. Tate, F. A. Howe, C. Majos, A. Moreno-Torres, M. van der Graaf, C. Arus, and S. Van Huffel, "Classification of brain tumours using short echo time 1H MR spectra," *J. Magn. Reson.*, vol. 170, no. 1, pp. 164–175, Sep 2004.

[200] Y. Xi and D. M. Rocke, "Baseline correction for nmr spectroscopic metabolomics data analysis," *BMC bioinformatics*, vol. 9, no. 1, p. 1, 2008.

[201] S. Ghose, A. Oliver, R. Marti, X. Llado, J. C. Vilanova, J. Freixenet, J. Mitra, D. Sidibe, and F. Meriaudeau, "A survey of prostate segmentation methodologies in ultrasound, magnetic resonance and computed tomography images," *Comput Methods Programs Biomed*, vol. 108, no. 1, pp. 262–287, Oct 2012.

[202] G. Litjens, O. Debats, W. van de Ven, N. Karssemeijer, and H. Huisman, "A pattern recognition approach to zonal segmentation of the prostate on MRI," *Med Image Comput Comput Assist Interv*, vol. 15, no. Pt 2, pp. 413–420, 2012.

[203] S. K. Warfield, K. H. Zou, and W. M. Wells, "Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation," *IEEE Trans Med Imaging*, vol. 23, no. 7, pp. 903–921, Jul 2004.

[204] M. Amadasun and R. King, "Textural features corresponding to textural properties," *Systems, Man and Cybernetics, IEEE Transactions on*, vol. 19, no. 5, pp. 1264–1274, 1989.

[205] H. Li, M. L. Giger, O. I. Olopade, A. Margolis, L. Lan, and M. R. Chinander, "Computerized texture analysis of mammographic parenchymal patterns of digitized mammograms," *Acad Radiol*, vol. 12, no. 7, pp. 863–873, Jul 2005.

[206] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognition*, vol. 29, no. 1, pp. 51–59, Jan. 1996.

[207] F. L. Bookstein, "Principal warps: thin-plate splines and the decomposition of deformations," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 11, no. 6, pp. 567–585, 1989.

[208] G. Litjens, R. Toth, W. van de Ven, C. Hoeks, S. Kerkstra, B. van Ginneken, G. Vincent, G. Guillard, N. Birbeck, J. Zhang, R. Strand, F. Malmberg, Y. Ou, C. Davatzikos, M. Kirschner, F. Jung, J. Yuan, W. Qiu, Q. Gao, P. E. Edwards, B. Maan, F. van der Heijden, S. Ghose, J. Mitra, J. Dowling, D. Barratt, H. Huisman, and A. Madabhushi, "Evaluation of prostate segmentation algorithms for MRI: the PROMISE12 challenge," *Med Image Anal*, vol. 18, no. 2, pp. 359–373, Feb 2014.

[209] T. R. Langerak, U. A. van der Heide, A. N. Kotte, M. A. Viergever, M. van Vulpen, and J. P. Pluim, "Label fusion in atlas-based segmentation using a selective and iterative method for performance level estimation (simple)," *IEEE Transactions on Medical Imaging*, vol. 29, no. 12, pp. 2000–2008, 2010.

[210] A. Rampun, P. Malcolm, and R. Zwiggelaar, "Detection and localisation of prostate cancer within the peripheral zone using scoring algorithm," in *Irish Machine Vision and Image Processing*, 2014.

[211] R. Toth, J. Chappelow, M. Rosen, S. Pungavkar, A. Kalyanpur, and A. Madabhushi, "Multi-attribute non-initializing texture reconstruction based active shape model (MANTRA)," *Med Image Comput Comput Assist Interv*, vol. 11, no. Pt 1, pp. 653–661, 2008.

[212] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models&Mdash;Their training and application," *Comput. Vis. Image Underst.*, vol. 61, no. 1, pp. 38–59, Jan. 1995.

[213] K. Pearson, "On lines and planes of closest fit to systems of points in space," *Philosophical Magazine*, vol. 2, no. 6, pp. 559–572, 1901.

[214] R. Toth, S. Doyle, S. Pungavkar, A. Kalyanpur, and A. Madabhushi, "A boosted ensemble scheme for accurate landmark detection for active shape models," in *SPIE Medical Imaging*, vol. 7260, Orlando, FL, 2009.

[215] H. Huisman, P. Vos, G. Litjens, T. Hambrock, and J. Barentsz, "Computer aided detection of prostate cancer using T2, DWI and DCE MRI: methods and clinical applications," in *Proceedings of the 2010 international conference on Prostate cancer imaging: computer-aided diagnosis, prognosis, and intervention*, ser. MICCAI'10. Berlin, Heidelberg: Springer-Verlag, 2010, pp. 4–14.

[216] N. Otsu, "A threshold selection method from gray-level histograms," *Automatica*, vol. 11, no. 285-296, pp. 23–27, 1975.

[217] J. Shi and J. Malik, "Normalized cuts and image segmentation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, no. 8, pp. 888–905, 2000.

[218] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," in *Advances in Neural Information Processing Systems 14*. MIT Press, 2001, pp. 585–591.

[219] J. B. Maintz and M. A. Viergever, "A survey of medical image registration," *Med Image Anal*, vol. 2, no. 1, pp. 1–36, Mar 1998.

[220] B. Zitová and J. Flusser, "Image registration methods: a survey," *Image and Vision Computing*, vol. 21, no. 11, pp. 977 – 1000, 2003.

[221] J. Mitra, R. Marti, A. Oliver, X. Llado, J. C. Vilanova, and F. Meriaudeau, "A comparison of thin-plate splines with automatic correspondences and B-splines with uniform grids for multimodal prostate registration," in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, ser. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, vol. 7964, Mar. 2011.

[222] J. Mitra, Z. Kato, R. Marti, A. Oliver, X. Llado, D. Sidibe, S. Ghose, J. C. Vilanova, J. Comet, and F. Meriaudeau, "A spline-based non-linear diffeomorphism for multimodal prostate registration," *Med Image Anal*, vol. 16, no. 6, pp. 1259–1279, Aug 2012.

[223] J. Pluim, J. Maintz, and M. Viergever, "Mutual-information-based registration of medical images: a survey," *IEEE Transcations on Medical Imaging*, vol. 22, no. 8, pp. 986–1004, 2003.

[224] J. Chappelow, B. N. Bloch, N. Rofsky, E. Genega, R. Lenkinski, W. DeWolf, and A. Madabhushi, "Elastic registration of multimodal prostate MRI and histology via multiattribute combined mutual information," *Med Phys*, vol. 38, no. 4, pp. 2005–2018, Apr 2011.

[225] M. Staring, U. A. van der Heide, S. Klein, M. A. Viergever, and J. P. Pluim, "Registration of cervical MRI using multifeature mutual information," *IEEE Trans Med Imaging*, vol. 28, no. 9, pp. 1412–1421, Sep 2009.

[226] A. Hero, B. Ma, O. Michel, and J. Gorman, "Applications of entropic spanning graphs," *Signal Processing Magazine, IEEE*, vol. 19, no. 5, pp. 85–95, 2002.

[227] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu, "A limited memory algorithm for bound constrained optimization," *SIAM J. Sci. Comput.*, vol. 16, no. 5, pp. 1190–1208, Sep. 1995.

[228] P. Viola and W. M. Wells, III, "Alignment by maximization of mutual information," *Int. J. Comput. Vision*, vol. 24, no. 2, pp. 137–154, Sep. 1997.

[229] J. Mitra, "Multimodal image registration applied to magnetic resonance and ultrasound prostatic images," Ph.D. dissertation, Universitat de Girona and Université de Bourgogne, 2012.

[230] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. Hill, M. O. Leach, and D. J. Hawkes, "Nonrigid registration using free-form deformations: application to breast MR images," *IEEE Trans Med Imaging*, vol. 18, no. 8, pp. 712–721, Aug 1999.

[231] Q. Li, S. Sone, and K. Doi, "Selective enhancement filters for nodules, vessels, and airway walls in two- and three-dimensional CT scans," *Med Phys*, vol. 30, no. 8, pp. 2040–2051, Aug 2003.

[232] J. Prewitt, *Picture processing and psychohistories*. Academic Press, 1970, ch. Object enhancement and extraction.

[233] I. Sobel, "Camera models and machine perception," DTIC Document, Tech. Rep., 1970.

[234] R. Kirsch, "Computer determination of the constituent structure of biological images ," *Computers and Biomedical Research*, vol. 4, no. 3, pp. 315 – 328, 1971.

[235] D. Gabor, "Theory of communication. Part 1: The analysis of information," *Electrical Engineers - Part III: Radio and Communication Engineering, Journal of the Institution of*, vol. 93, no. 26, pp. 429–441, 1946.

[236] J. G. Daugman, "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters," *J Opt Soc Am A*, vol. 2, no. 7, pp. 1160–

1169, Jul 1985.

[237] Y. Wang and C.-S. Chua, "Face recognition from 2d and 3d images using 3d gabor filters," *Image and vision computing*, vol. 23, no. 11, pp. 1018–1028, 2005.

[238] P. Kovesi, "Image features from phase congruency," *Videre: Journal of computer vision research*, vol. 1, no. 3, pp. 1–26, 1999.

[239] R. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *Systems, Man and Cybernetics, IEEE Transactions on*, vol. SMC-3, no. 6, pp. 610–621, 1973.

[240] H. Tamura, S. Mori, and T. Yamawaki, "Textural features corresponding to visual perception," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 8, no. 6, pp. 460–473, 1978.

[241] A. Benassi, S. Cohen, and J. Istas, "Identifying the multifractional function of a Gaussian process," *Statistics & Probability Letters*, vol. 39, no. 4, pp. 337 – 345, 1998.

[242] N. Ahmed, T. Natarajan, and K. Rao, "Discrete cosine transform," *Computers, IEEE Transactions on*, vol. C-23, no. 1, pp. 90–93, 1974.

[243] T. Leung and J. Malik, "Representing and recognizing the visual appearance of materials using three-dimensional textons," *International journal of computer vision*, vol. 43, no. 1, pp. 29–44, 2001.

[244] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 1, 2005, pp. 886–893 vol. 1.

[245] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 4, pp. 509–522, 2002.

[246] G. Zhao, T. Ahonen, J. Matas, and M. Pietikainen, "Rotation-Invariant Image and Video Description With Local Binary Pattern Features," *Image Processing, IEEE Transactions on*, vol. 21, no. 4, pp. 1465–1477, 2012.

[247] P. Castorina, P. P. Delsanto, and C. Guiot, "Classification scheme for phenomenological universalities in growth problems in physics and other sciences," *Phys. Rev. Lett.*, vol. 96, p. 188701, May 2006.

[248] H. Ratiney, M. Sdika, Y. Coenradie, S. Cavassila, D. van Ormondt, and D. Graveron-Demilly, "Time-domain semi-parametric estimation based on a metabolite basis set," *NMR Biomed*, vol. 18, no. 1, pp. 1–13, Feb 2005.

[249] L. Vanhamme, A. van den Boogaart, and S. Van Huffel, "Improved method for accurate and efficient quantification of MRS data with use of prior knowledge," *J. Magn. Reson.*, vol. 129, pp. 35–45, 1997.

[250] T. Coleman and Y. Li, "An interior trust region approach for nonlinear minimization subject to bounds," Cornell University, Tech. Rep., 1993.

[251] S. W. Provencher, "Estimation of metabolite concentrations from localized in vivo proton NMR spectra," *Magn Reson Med*, vol. 30, no. 6, pp. 672–679, Dec 1993.

[252] C. Gasparovic, T. Song, D. Devier, H. J. Bockholt, A. Caprihan, P. G. Mullins, S. Posse, R. E. Jung, and L. A. Morrison, "Use of tissue water as a concentration reference for proton spectroscopic imaging," *Magnetic resonance in medicine*, vol. 55, no. 6, pp. 1219–1226, 2006.

[253] R. Coifman and M. Wickerhauser, "Entropy-based algorithms for best basis selection," *Information Theory, IEEE Transactions on*, vol. 38, no. 2, pp. 713–718, 1992.

[254] G. Lemaître, F. Nogueira, and C. K. Aridas, "Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning," *Journal of Machine Learning Research*, vol. 18, no. 17, pp. 1–5, 2017. [Online]. Available: http://jmlr.org/papers/v18/16-365.html

[255] 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.

[256] M. R. Smith, T. Martinez, and C. Giraud-Carrier, "An instance level analysis of data complexity," *Machine learning*, vol. 95, no. 2, pp. 225–256, 2014.

[257] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: synthetic minority over-

sampling technique," *Journal of artificial intelligence research*, pp. 321–357, 2002.
[258] 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*. Springer, 2005, pp. 878–887.
[259] Y. Saeys, I. Inza, and P. Larranaga, "A review of feature selection techniques in bioinformatics," *Bioinformatics*, vol. 23, no. 19, pp. 2507–2517, 2007.
[260] S. N. Goodman, "Toward evidence-based medical statistics. 1: The P value fallacy," *Ann. Intern. Med.*, vol. 130, no. 12, pp. 995–1004, Jun 1999.
[261] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 27, no. 8, pp. 1226–1238, 2005.
[262] I. Fodor, "A survey of dimension reduction techniques," 2002.
[263] I. T. Jolliffe, *Principal Component Analysis*, 2nd ed. Springer, Oct. 2002.
[264] H. Zou, T. Hastie, and R. Tibshirani, "Sparse principal component analysis," *Journal of computational and graphical statistics*, vol. 15, no. 2, pp. 265–286, 2006.
[265] P. Comon, "Independent component analysis, a new concept?" *Signal processing*, vol. 36, no. 3, pp. 287–314, 1994.
[266] K. P. Murphy, *Machine learning: a probabilistic perspective*. MIT press, 2012.
[267] M. A. Aizerman, E. A. Braverman, and L. Rozonoer, "Theoretical foundations of the potential function method in pattern recognition learning," in *Automation and Remote Control,*, no. 25, 1964, pp. 821–837.
[268] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.
[269] R. Rubinstein, M. Zibulevsky, and M. Elad, "Efficient implementation of the k-svd algorithm using batch orthogonal matching pursuit," *CS Technion*, vol. 40, no. 8, pp. 1–15, 2008.
[270] M. Elad, *Sparse and redundant representations: from theory to applications in signal and image processing*, 1st ed. Springer Publishing Company, Incorporated, 2010.
[271] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *Signal Processing, IEEE Transactions on*, vol. 41, no. 12, pp. 3397–3415, 1993.
[272] Y. C. Pati, R. Rezaiifar, and P. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Signals, Systems and Computers, 1993. 1993 Conference Record of The Twenty-Seventh Asilomar Conference on*. IEEE, 1993, pp. 40–44.
[273] G. Davis, S. Mallat, and M. Avellaneda, "Adaptive greedy approximations," *Constructive approximation*, vol. 13, no. 1, pp. 57–98, 1997.
[274] J. Sivic and A. Zisserman, "Video google: a text retrieva approach to object matching in videos," in *IEEE ICCV*, 2003, pp. 1470–1477.
[275] C. M. Bishop, *Pattern recognition and machine learning*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006.
[276] A. Fred and A. Jain, "Combining multiple clusterings using evidence accumulation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 27, no. 6, pp. 835–850, 2005.
[277] J. Friedman, "Regularized discriminant analysis," *Journal of the American Statistical Association*, vol. 84, no. 405, pp. pp. 165–175, 1989.
[278] I. Rish, "An empirical study of the naive Bayes classifier," in *IJCAI 2001 workshop on empirical methods in artificial intelligence*, vol. 3, no. 22, 2001, pp. 41–46.
[279] Y. Freund and R. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of Computer and System Sciences*, vol. 55, no. 1, pp. 119 – 139, 1997.
[280] J. Quinlan, "Induction of decision trees," *Machine Learning*, vol. 1, no. 1, pp. 81–106, 1986.
[281] ——, *C4.5: Programs for machine learning*. San Francisco, CA, USA: Morgan Kaufmann

Publishers Inc., 1993.
[282] L. Breiman, J. Friedman, R. Olshen, and C. Stone, *Classification and regression trees*. Monterey, CA: Wadsworth and Brooks, 1984.
[283] J. Friedman, T. Hastie, and R. Tibshirani, "Additive logistic regression: a statistical view of boosting," *Annals of Statistics*, vol. 28, p. 2000, 1998.
[284] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," *Annals of statistics*, pp. 1189–1232, 2001.
[285] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
[286] B. Efron, "Bootstrap methods: Another look at the jackknife," *The Annals of Statistics*, vol. 7, no. 1, pp. 1–26, 01 1979.
[287] Z. Tu, "Probabilistic boosting-tree: learning discriminative models for classification, recognition, and clustering," in *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, vol. 2, 2005, pp. 1589–1596 Vol. 2.
[288] C. Rasmussen and C. Williams, *Gaussian processes for machine learning*. The MIT Press, 2005.
[289] V. Vapnik and A. Lerner, "Pattern Recognition using Generalized Portrait Method," *Automation and Remote Control*, vol. 24, 1963.
[290] C. Cortes and V. Vapnik, "Support-Vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.
[291] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, ser. COLT '92. New York, NY, USA: ACM, 1992, pp. 144–152.
[292] M. Tipping, "Sparse Bayesian learning and the relevance vector machine," *Journal of Machine Learning Research*, vol. 1, pp. 211–244, 2001.
[293] J. Quinonero-Candela, A. Girard, and C. Rasmussen, "Prediction at an Uncertain Input for Gaussian processes and relevance vector machines application to Multiple-Step ahead time-series forecasting," DTU Informatics, Tech. Rep., 2002.
[294] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation," in *Neurocomputing: foundations of research*, J. A. Anderson and E. Rosenfeld, Eds. Cambridge, MA, USA: MIT Press, 1988, ch. Learning Internal Representations by Error Propagation, pp. 673–695.
[295] D. F. Specht, "Probabilistic neural networks for classification, mapping, or associative memory," in *Neural Networks, 1988., IEEE International Conference on*, 1988, pp. 525–532 vol.1.
[296] Z. Kato and T. Pong, "A Markov random field image segmentation model using combined color and texture features," in *Computer Analysis of Images and Patterns*, ser. Lecture Notes in Computer Science, W. Skarbek, Ed. Springer Berlin Heidelberg, 2001, vol. 2124, pp. 547–554.
[297] Z. Kato and J. Zerubia, *Markov Random Fields in Image Segmentation. Collection Foundation and Trends in Signal Processing*. Now Editor, World Scientific, Sep. 2012.
[298] B. Efron, "Estimating the error rate of a prediction rule: Improvement on cross-validation," *Journal of the American Statistical Association*, vol. 78, no. 382, pp. pp. 316–331, 1983.
[299] C. E. Metz, "Receiver operating characteristic analysis: a tool for the quantitative evaluation of observer performance and imaging systems," *J Am Coll Radiol*, vol. 3, no. 6, pp. 413–422, Jun 2006.
[300] H. He, E. Garcia *et al.*, "Learning from imbalanced data," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 21, no. 9, pp. 1263–1284, 2009.

PLACE PHOTO HERE

Guillaume Lemaitre G. Lemaitre received B.Sc. degree (Hons) in electrical, signal, and image engineering from the Universté de Bourgogne and the M.Sc. Erasmus Mundus Master of Excellence (Hons) in Vision and Robotics co-jointly delivered by the Université de Bourgogne, Universitat de Girona, and Heriot-Watt University. Additionally, he received a M.Sc. in Business Innovation and Technology Management by the Universitat de Girona. He got a joint Ph.D. from the Université de Bourgogne and the Universitat de Girona. He is currently working for the Center for Data Sciences Paris-Saclay and for the PARIETAL team of the Inria Saclay. His research interests include machine learning applied to pluridisciplinary fields as medical imaging.

PLACE PHOTO HERE

Fabrice Mériaudeau F. Mériaudeau was born in Villeurbanne, France, on March 18, 1971. He received both the master degree in physics at Dijon University, France as well as an Engineering Degree (FIRST) in material sciences in 1994. He also obtained a Ph.D. in image processing at the same University in 1997. He was a postdoc for a year at The Oak Ridge National Laboratory. He is currently "Professeur des Universite" and was Director of the Le2i (UMR CNRS), which has more than 200 staff members, from 2011 to 2016. His research interests were focused on image processing for non-conventional imaging systems (UV, IR, polarization) and more recently on medical/biomedical imaging. He has coordinated an Erasmus Mundus Master in the field of Computer Vision and Robotics from 2006 to 2010 and he was the Vice President for International Affairs for the University of Burgundy from 2010 to 2012. He has authored and co-authored more than 150 international publications and holds three patents. Since 2016, he has joined Universiti Teknologi PETRONAS as a professor.

PLACE PHOTO HERE

Robert Martí R. Martí is an associate professor at Computer Vision and Robotics Institute of the University Girona, Spain. He received his BS and MS degrees in Computer Science from the University of Girona in 1997 and 1999, respectively, and his PhD degree from the School of Information Systems at the University of East Anglia in 2003. He is the author of more than 30 international peer reviewed journal and 70 conference papers. His current research interests include medical image analysis, computer aided diagnosis, and breast and prostate imaging.