Computer-Aided Detection and Diagnosis for prostate cancer based on mono and multi-parametric MRI: A review

Guillaume Lemaître[a,c,*], Robert Martí[c], Jordi Freixenet[c], Joan C. Vilanova[d], Paul M. Walker[b], Fabrice Meriaudeau[a]

[a]*LE2I-UMR CNRS 6306, Université de Bourgogne, 12 rue de la Fonderie, 71200 Le Creusot, France*
[b]*LE2I-UMR CNRS 6306, Université de Bourgogne, Avenue Alain Savary, 21000 Dijon, France*
[c]*ViCOROB, Universitat de Girona, Campus Montilivi, Edifici P4, 17071 Girona, Spain*
[d]*Department of Magnetic Resonance, Clínica Girona, Lorenzana 36, 17002 Girona, Spain*

Abstract

Prostate cancer is reported to be the second most diagnosed cancer of men all over the world. In the last decades, new imaging techniques based on MRI have been developed improving the diagnosis task of radiologists. In practise, diagnosis can be affected by multiple factors reducing the chance to detect potential lesions. Computer-aided detection and computer-aided diagnosis have been designed to answer to these needs and provide help to radiologists in their daily duties. Research on computer-aided systems specifically focused for prostate cancer is a young technology and part of a dynamic field for the last ten years. This survey aimed to provide an overview of the researches carried out in this lapse of time and more precisely a comprehensive review of all the different stages composing the work-flow of computer-aided system. We also provide a comparison between these studies and potential avenues for future research are also discussed.

Keywords: computer-aided detection, computer-aided diagnosis, CAD, magnetic resonance imaging, magnetic resonance spectroscopy imaging, computer vision

1. Introduction

During the last century, physicists focused on constantly innovating in terms of imaging techniques assisting radiologists to overcome different tasks as cancer detection and diagnosis. However, human diagnosis still suffers from low repeatability, synonymous with erroneous detection or interpretations of abnormalities throughout clinical decisions (Giger et al. (2008), Hambrock et al. (2013)). These errors are driven by two majors causes (Giger et al. (2008)). On the one hand, observer limitations (e.g., constrained human visual perception, fatigue or distraction) are the principal human issues. On the other hand, the second reason is linked to the clinical cases themselves, for instance due to unbalanced data (number of healthy cases more abundant than malignant cases) or overlapping structures resulting from limitations of imaging techniques.

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 (Giger et al. (2008)). 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. (1999) reported a significant 4 % improvement in breast cancer detection, in accordance with later studies (Dean and Ilvento (2006)). Similar conclusions were drawn in the case of lung nodule detection (Li et al. (2004)), colon cancer (Petrick et al. (2008)) and prostate cancer (CaP) as well (Hambrock et al. (2013)). Chan et al. (1999) also hypothesized that CAD systems will be even more efficient assisting inexperienced radiologists to senior radiologists. That hypothesis was tested by Hambrock et al. (2013) and was confirmed in case of CaP detection. In this particular study, inexperienced radiologists obtained equivalent performance to senior radiologists, both with the help of a CAD system whereas the accuracy of their diagnosis was significantly poorer without this assistance.

In the late eighties, the first CAD systems were developed to detect anomalies on chest radiographies and mammograms (Chan et al. (1987), Doi et al. (1990), Giger et al. (1988)). In the past twenty years, extensive investigations were conducted in the advancement of CAD systems, migrating from intensive time consuming algorithms performed on reduced number of cases to "fast" processing on a large medical dataset. These works were focused on diverse organ cancer diagnosis making use of numerous imaging modalities: micro-calcification detection in breast mammography (Elter and Horsch (2009), Rangayyan et al. (2007)) and ultrasound (US) imaging (Cheng et al. (2010)), lung nodules detection based on computer tomography (CT) (Chan et al. (2008), Suzuki (2012)), colon tumours detection (Suzuki (2012)) and melanoma detection using dermoscopy imaging (Korotkov and Garcia (2012)). Noting the

---

*Corresponding author.
Email addresses: guillaume.lemaitre@udg.edu (Guillaume Lemaître), marly@eia.udg.edu (Robert Martí), jordif@eia.udg.edu (Jordi Freixenet), pwalker@u-bourgogne.fr (Paul M. Walker), fabrice.meriaudeau@u-bourgogne.fr (Fabrice Meriaudeau)

abundance of diverse CAD systems, these fields achieved a certain maturity which can be explained by the imaging techniques employed. Indeed, x-rays, US as well as CT are medical imaging techniques developed all before the 1970s and were subject to intensive research.

In contradiction with the aforementioned statement, CaP detection using CAD is a young technology due to the fact that magnetic resonance imaging (MRI) is the keystone medical imaging technique (Hegde et al. (2013)). Four distinct MRI modalities are employed in CAD for CaP and were mainly developed after the mid-1990s: (i) T_2 Weighted (T_2-W) MRI (Hricak et al. (1983)), (ii) dynamic contrast-enhanced (DCE) MRI (Huch Boni et al. (1995)), (iii) magnetic resonance spectroscopy imaging (MRSI) (Kurhanewicz et al. (1996)) and (iv) diffusion weighted (DW) MRI (Scheidler et al. (1999b)). It can be noted that these techniques came into existence relatively recently mainly due to technological progress. In addition, the increase of magnetic field strength and the development of endorectal coil, both improved image spatial resolution (Swanson et al. (2001)) needed to perform more accurate diagnosis. It is for this matter that development of CAD for CaP is lagging behind the other fields stated above.

The first study using MRI as inputs of CAD system was published ten years ago by Chan et al. (2003). Despite this, no less than fifty studies have been reviewed for this survey since that seminal work. To the best of our knowledge, there is no review in the literature regarding the advancement of CAD systems devoted specifically to CaP detection and diagnosis. Thus, our aim with this survey is threefold: (i) provide an overview of developed CAD systems for CaP detection and diagnosis based on MRI modalities (ii) assess the different work and (iii) pointing out avenues for future work.

As discussed further in Sect. 2.3.3, CAD systems share a common framework. Stages involved in CAD work-flow can be categorized into six distinctive processes: (i) pre-processing, (ii) segmentation, (iii) registration, (iv) feature detection, (v) feature selection and extraction and (vi) classification. The first three stages are used to enhance data as well as to extract regions of interest and, in the case of multi-modal sources, to merge information of those heterogeneous sources in a joint reference system. The last three categories deal with pattern recognition, machine learning and data mining notions and more precisely with the data classification problem. First, information is detected from the different data sources and a subset of relevant features is selected and/or extracted. Then, this meaningful data will then be classified in order to provide the probability of malignancy of the area of interest and will assist radiologists in their diagnosis decisions (see Fig. 2).

This paper is organized as follows: Sect. 2 deals with general information about human prostate and background about CaP. Methods regarding CaP screening and imaging techniques used are also presented as well as an introduction on the CAD framework. Sections 3 - 4 review techniques used in different steps involved in a CAD work-flow which will be our main contribution. Image regularization framework including pre-processing (Sect. 3.1), segmentation (Sect. 3.2) and registration (Sect. 3.3) will be covered as well as the image classification framework

comprising of feature detection (Sect.4.2), feature selection and extraction (see Sect. 4.3) and feature classification (Sect. 4.4). Results and discussion are reported in Sect. 5 followed by a concluding section.

2. Background

This section provides an overview of CaP and its detection and diagnosis. We start with a summary of the prostate anatomy and a brief overview of different CaPs. Subsequently a dicussion of the current screening strategy for CaP and its drawbacks is presented. MRI plays an important role in improving the current strategy and a more detailed description of MRI modalities is given. These different techniques are used as inputs to the CAD system which is finally discussed.

2.1. The human prostate

The prostate is an exocrine gland of the male reproductive system having an inverted pyramidal shape. It measures approximately three centimetres in height by two and half centimetres in depth and its weight is estimated to be between seven and sixteen grams for an adult (Leissner and Tisell (1979)). The prostate size increases at two distinct stages during physical development: initially at puberty to reach its normal size, then again after sixty years of age leading to benign prostatic hyperplasia (BPH) (Parfait (2010)).

A zonal classification of the prostate, depicted in Fig. 1, was suggested by McNeal (McNeal (1981)). Subsequently, this categorization was widely accepted in the literature (cf., Coakley and Hricak (2000), Hricak et al. (1987), Parfait (2010), Villers et al. (1991)) and is used in all medical examinations (e.g., biopsy, MRI screening). The classification is based on dividing the gland into three distinct regions: (i) central zone (CZ) accounting for 20-25% of the whole prostate gland, (ii) transitional zone (TZ) standing for 5% and (iii) 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 can be divided in three longitudinal portions depicted in Fig. 1(b): (i) base, (ii) median gland and (iii) apex.

2.2. Prostate carcinoma

CaP has been reported on a worldwide scale to be the second most frequently diagnosed cancer of men accounting for 13.6% (Ferlay et al. (2010)). Statistically, in 2008, the number of new diagnosed cases was estimated to be 899,000 with no less than 258,100 deaths (Ferlay et al. (2010)). In United States, aside from skin cancer, CaP was declared to be the most commonly diagnosed cancer among men, implying that approximately one in six men will be diagnosed with CaP during their lifetime and one in thirty-six will die from this disease causing CaP to be the second most common cause of cancer death among men (Siegel et al. (2013), American Cancer Society (2013)).
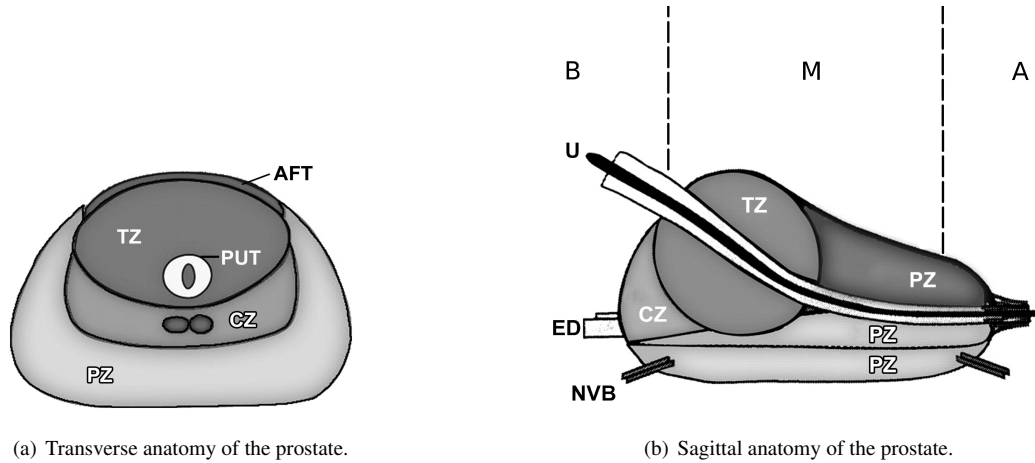
(a) Transverse anatomy of the prostate. (b) Sagittal anatomy of the prostate.

Figure 1: Prostate anatomy with division in different zones. *AFT:* anterior fibromuscular tissue, *CZ:* central zone, *ED:* ejaculatory duct, *NVB:* neurovascular bundle, *PUT:* tissue, *PZ:* peripheral zone, *U:* urethra, *TZ:* transitional zone, *B:* base, *M:* median, *A:* apex (copyright by Choi et al. (2007)).

Despite active research to determine the causes of prostate cancer, a fuzzy list of risk factors has arisen (American Cancer Society (2010)). The etiology was linked to the following factors (American Cancer Society (2010)): (i) family history (Giovannucci et al. (2007), Steinberg et al. (1990)), (ii) genetic factors (Agalliu et al. (2009), Amundadottir et al. (2006), Freedman et al. (2006)), (iii) race-ethnicity (Giovannucci et al. (2007), Hoffman et al. (2001)), (iv) diet (Alexander et al. (2010), Giovannucci et al. (2007), Ma and Chapman (2009)), (v) obesity (Giovannucci et al. (2007), Rodriguez et al. (2007)). 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 (Strum and Pogliano (2005)): slow-growing tumours, accounting for up to 85 % of all CaPs (Lu-Yao et al. (2009)), 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 others organs, primarily the bones (Oster et al. (2013)). Bone metastases, being an incurable disease, significantly affects the morbidity and mortality rate (Ye et al. (2007)). Hence, the results of the surveillance have to be trustworthy in order 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 70-80 % of CaPs originate in PZ whereas 10-20 % in TZ (Carrol et al. (1987), McNeal et al. (1988), Stamey et al. (1998)). Only about 5 % of CaPs occur in CZ (Cohen et al. (2008), McNeal et al. (1988)). However, those cancers appear to be more aggressive and more likely to invade other organs due to their location (Cohen et al. (2008)).

2.3. CaP screening and imaging techniques
2.3.1. Current CaP screening

Current CaP screening consists of three different stages. First, prostate-specific antigen (PSA) control is performed to distinguish between low and high risk CaP. Then, for confirmation, samples are taken during prostate biopsy and finally

analysed 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 (Etzioni et al. (2002)). A higher-than-normal level of PSA can indicate an abnormality of the prostate either as a BPH or a cancer (Hoeks et al. (2011)). However, other factors can lead to an increased PSA level such as prostate infections, irritations, a recent ejaculation or a recent rectal examination (Parfait (2010)). PSA can be found in the bloodstream in two different forms: free PSA (about 10%), and linked to another protein (about 90%). A level of PSA higher than 10 ng.mL^{-1} is considered to be at risk (Parfait (2010)). If the PSA level is between 10 ng.mL^{-1} and 4 ng.mL^{-1}, the patient is considered as suspicious (Barentsz et al. (2012)). 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 (Parfait (2010)).

A transrectal ultrasound (TRUS) biopsy is carried out for cases which are considered as pathological. At least six different samples are taken randomly from the right and left parts of three three different zones: apex, median and base. These samples are further evaluated using the Gleason grading system (Gleason (1977)). The scoring scheme to characterize the biopsy sample is composed of five different patterns which correspond to grades ranging from 1 to 5. Higher grades are associated with poor prognosis (Epstein et al. (2005)). Then, in the Gleason system, two 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 (Epstein et al. (2005)). A higher Gleason score (GS) indicates a more aggressive tumour (Epstein et al. (2005)). Also, it should be noted that biopsy is an invasive procedure which can result in serious infection or urine retention (Chou et al. (2011), Hara et al. (2005)).

Although PSA screening has been shown to improve early detection of CaP (Chou et al. (2011)), its lack of reliability

3

motivates further investigations using MRI-CAD. Two reliable studies, carried out in the United States (Andriole et al. (2009)) and in Europe (Hugosson et al. (2010), Schröder et al. (2012)), have attempted to assess the impact of early detection of CaP, with diverging outcomes (Chou et al. (2011), Heidenreich et al. (2013)). The study carried out in Europe[1] concluded that PSA screening reduces CaP-related mortality by 21-44% (Hugosson et al. (2010), Schröder et al. (2012)), while the American[2] trial found no such effect (Andriole et al. (2009)). However, both studies agree that PSA screening suffers from low specificity, with an estimated rate of 36 % (Schroder et al. (2008)). 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 (Delpierre et al. (2013)).

Hence, new screening methods should be developed with improved specificity of detection as well as more accurate risk assessment (aggressiveness and progression). Current research is focused on identifying new biological markers to replace PSA-based screening (Bourdoumis et al. (2010), Brenner et al. (2013), Morgan et al. (2011)). Until such research comes to fruition, these needs can be met through active-surveillance strategy using multi-parametric MRI techniques (Hoeks et al. (2011), Moore et al. (2013)). An MRI-CAD system, which is an area of active research and forms the focus of this paper, can be incorporated into this screening strategy allowing a more systematic and rigorous follow-up.

Another weakness of the current screening strategy lies in the fact that TRUS biopsy does not provide trustworthy results. Due to its "blind" nature, there is a chance of missing aggressive tumours or detecting microfocal "cancers", which influences the aggressiveness-assessment procedure (Noguchi et al. (2001)). As a consequence, over-diagnosis is estimated at up to 30 % (Haas et al. (2007)), while missing clinically significant CaP is estimated at up to 35 % (Taira et al. (2010)). In an effort to solve both issues, alternative biopsy approaches have been explored. MRI/US-guided biopsy has been shown to outperform standard TRUS biopsy (Delongchamps et al. (2013)). There, multimodal MRI images are fused with US 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 be used to improve repeatability of examination.

CaP detection and diagnosis benefit from the use of CAD and MRI techniques. In the following sections, these techniques will be presented in addition to an overview of CAD for CaP.

2.3.2. MRI imaging techniques

MRI provides promising imaging techniques to overcome the previous mentioned drawbacks. Unlike TRUS biopsy, MRI ex-

amination is a non-invasive protocol and has been shown to be the most acute and harmless technique available currently (Turkbey and Choyke (2012)). In this section, we review different MRI 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. 4.2 for more details on automatic feature detection methods since they are part and parcel of the CAD framework. Table 1 provides an overview of the following discussion.

– **T_2-W MRI:** T_2-W MRI was the first MRI-modality used to perform CaP diagnosis using MRI (Hricak et al. (1983)). 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 (Barentsz et al. (2012)).

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-SI while the latter is depicted by a low-SI (Hricak et al. (1987)). An example of a healthy prostate is shown in Fig. 3(a).

In PZ, round or ill-defined low-SI masses are synonymous with CaPs (Hricak et al. (1983)) as shown in Fig. 3(b). Detecting CaP in CG is more challenging. In fact both normal CG tissue and malignant tissue, have a low-SI in T_2-W MRI reinforcing difficulties to distinguish between them. However, CaPs in CG appear often as homogeneous mass possessing ill-defined edges with lenticular or "water-drop" shapes (Akin et al. (2006), Barentsz et al. (2012)) as depicted in Fig. 3(c).

CaP aggressiveness was 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 (Wang et al. (2008)).

In spite of the availability of these useful and encouraging features, the T_2-W modality lacks reliability (Hoeks et al. (2011), Kirkham et al. (2006)). Sensitivity is affected by the difficulties in detecting cancers in CG (Kirkham et al. (2006)) while specificity rate is highly affected by outliers (Barentsz et al. (2012)). In fact, various conditions emulate patterns of CaP such as BPH, post-biopsy haemorrhage, atrophy, scars and post-treatment (Barentsz et al. (2012), Cruz et al. (2002), Hricak et al. (1987), Quint et al. (1991), Scheidler et al. (1999b)). These issues can be partly addressed using more innovative and advanced modalities.

– **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 (Hegde et al. (2013)). However, T_2 values alone have been shown to be more discriminative (Liu et al. (2011)) and highly correlated with citrate concentration, a biological marker in CaP (Liney et al.

---

[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.

Table 1: Overview of the features associated with each MRI modality. 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	intermediate to high-SI	+
T_2 map	SI	low-SI	intermediate to high-SI	+
DCE MRI	Semi-quantitative features: − wash-in − wash-out − integral under the curve − maximum signal intensity − time-to-peak enhancement	faster faster higher higher faster	slower slower lower lower slower	0 0 0 0 0
	Quantitative features (Tofts' parameters): − k_{ep} − K^{trans}	higher higher	lower lower	0 0
DW MRI	SI	higher-SI	lower-SI	+
ADC map	SI	low-SI	high-SI	+
MRSI	Metabolites: Citrate (2.64 ppm) Choline (3.21 ppm) Spermine (3.11 ppm)	lower concentration higher concentration lower concentration	higher concentration lower concentration higher concentration	+ 0 +

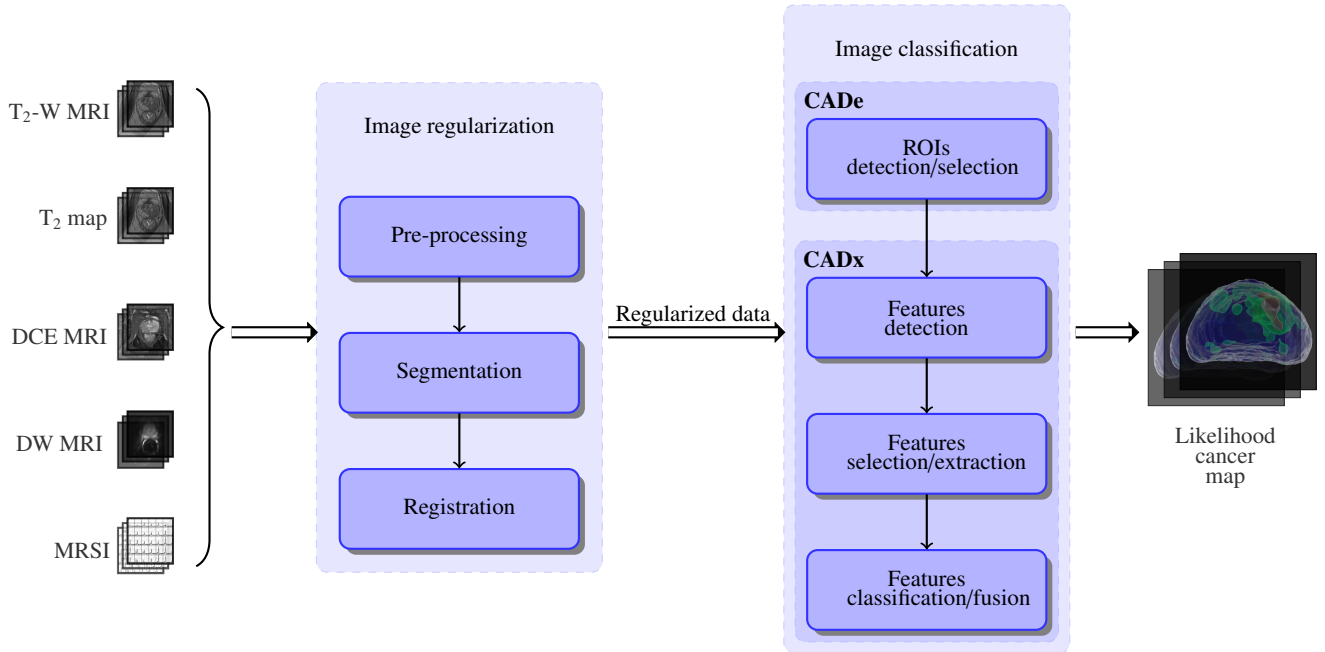Notes:

+ = significantly correlated.

0 = no correlation.

Figure 2: CAD framework using MRI images. Multiparametric MRI images are provided as inputs. These data arise from heterogeneous sources and need to be regularized. Some studies do not consider this stage as mandatory and do not implement or only partly those processes (see Tab. 2). A pre-processing stage is usually applied to standardize the intensity of images, reduce noise and artefacts. Then, in the image set, the prostate organ has to be segmented to focus the next processing stages only on that particular ROI. Moreover, prostate location can vary depending of the modality chosen. Therefore, the images are registered so that all segmented images will be in the same reference frame. Once the image regularisation performed, image classification can be carried out. First, a strategy defining ROIs to focus on is decided. Then, distinctive features are extracted before to be post-processed to select the most salient features. Finally, these salient features will feed a classifier previously trained which will provide a likelihood cancer map associated with either CaP detection or diagnosis.
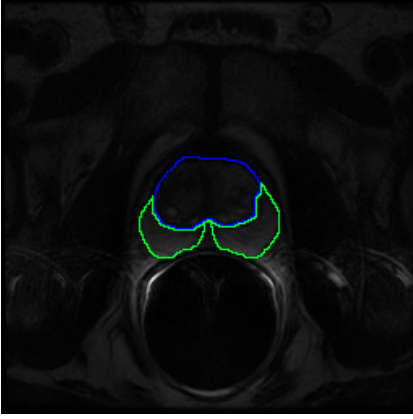
(a) T_2-W-acmri slice of an healthy prostate acquire with a 1.5 Tesla MRI. The blue contour represents the CG while the PZ corresponds to the green contour.
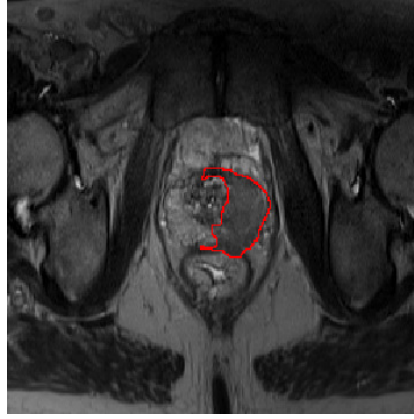
(b) T_2-W-acmri slice of a prostate with a CaP highlighted in the PZ using a 3.0 Tesla MRI scanner.

(c) T_2-W-acmri slice of a prostate with a CaP highlighted in the CG using a 3.0 Tesla MRI scanner.

Figure 3: Rendering of T_2-W-MRI prostate image with both 1.5 and 3.0 Tesla MRI scanner.

(1996b, 1997)).

T_2 values are computed using the characteristics of transverse relaxation. Transverse relaxation is formalized as:

$$M_{x,y}(t) = M_{x,y}(0) \exp\left(-\frac{t}{T_2}\right), \qquad (1)$$

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

By rearranging Eq. 1, T_2 map is computed performing a linear fitting on the model in Eq. 2 using several TE, $t = \{TE_1, TE_2, \ldots, TE_m\}$.

$$\ln\left[\frac{M_{x,y}(t)}{M_{x,y}(0)}\right] = -\frac{t}{T_2}. \qquad (2)$$

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 (Liney et al. (1996a)).

Similar to T_2-W MRI, T_2 values associated with CaP are significantly lower than those of healthy tissues (Gibbs et al. (2001), Liney et al. (1996b)).

– **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 the released back into the vasculature before being eliminated by the kidneys (Gribbestad et al. (2005)). Furthermore, the diffusion speed of the contrast agent may vary due to several parameters: (i) the permeability of the micro-vessels, (ii) their surface area and (iii) the blood flow (Padhani (2002)).

Healthy PZ is mainly made up of glandular tissue, around 70 % (Choi et al. (2007)), which implies a reduced interstitial space restricting exchanges between vessels and EES (Buckley et al. (2004), van Niekerk et al. (2009)). Normal CG has a more disorganised structure, composed of mainly

fibrous tissue (Choi et al. (2007), Hoeks et al. (2011)), which facilitates the arrival of the contrast agent in EES (van Niekerk et al. (2013)). 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 (Carmeliet and Jain (2000)). In order to ensure growth, malignant tumours produce and release angiogenic promoter substances (Carmeliet and Jain (2000)). These molecules stimulate the creation of new vessels towards the tumour (Carmeliet and Jain (2000)). However, the new vessel networks in tumours differ from those present in healthy tissue (Gribbestad et al. (2005)). They are more porous due to the fact that their capillary walls have a large number of "openings" (Choi et al. (2007), Gribbestad et al. (2005)). In contrast to healthy cases, this increased vascular permeability results in increased contrast agent exchanges between vessels and EES (Verma et al. (2012)).

By making use of the previous aspects, DCE MRI is based on an acquisition of a set of 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-analyse the pharmacokinetic behaviour of the contrast media concentration in prostate tissues (Verma et al. (2012)). 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. 4(b) (Tofts (2010)).

By taking the previous remarks regarding medical aspects and signal theory into account, CaPs are characterized by a signal having an earlier and faster enhancement and an earlier wash-out (cf., the rate of the contrast agent flowing out of the tissue) (see Fig. 4(b)) (Verma et al. (2012)). Three different approaches exist to analyse these signals with the aim of tagging them as corresponding to either normal or
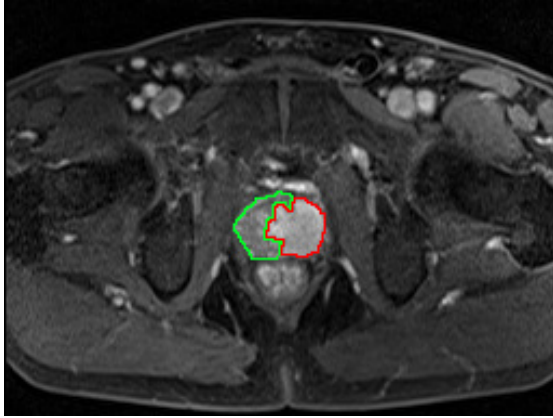
(a) T_1-W-MRI image where the cancer is delimited by the red contour. The green area was still not invaded by the CaP
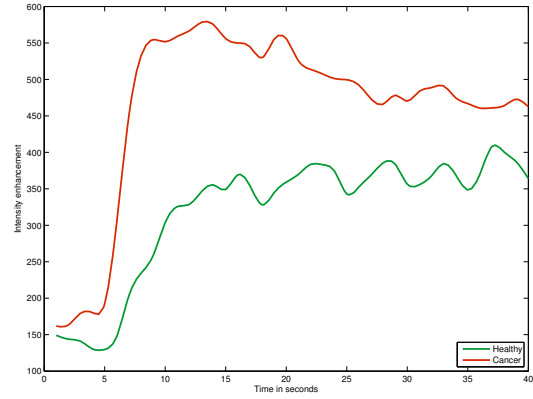
(b) Enhancement curve computed during the DCE-MRI analysis. The red curve is typical from CaP cancer while the green curve is characteristic of healthy tissue.

Figure 4: Illustration of typical enhancement signal observed in DCE-MRI analysis collected with a 3.0 Tesla MRI scanner.

malignant tissues. Qualitative analysis is based on assessment of the signal shape (Hoeks et al. (2011)). Quantitative approaches consist of inferring pharmocokinetic parameter values (Tofts (2010)). Those parameters are part of mathematical-pharmacokinetic models which are directly based on physiological exchanges between vessels and EES. Several pharmacokinetic models were proposed such as the Kety model (Kety (1951)), the Tofts model (Tofts (1997)) and mixed models (Larsson et al. (1996), St Lawrence and Lee (1998)). 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. These parameters will be discussed in a later section (see Fig. 17) (Hoeks et al. (2011), Verma et al. (2012)). It was shown that semi-quantitative and quantitative methods improve localization of CaP when compared with qualitative methods (Rosenkrantz et al. (2013)). Section 4.2.2 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 (Jager et al. (1997), Kim et al. (2005), Schlemmer et al. (2004), Zelhof et al. (2009)). Despite this fact, DCE MRI possesses some drawbacks. Due to its "dynamic" nature, patient motions during the image acquisition lead to spatial misregistration of the image set (Verma et al. (2012)). Furthermore, it has been suggested that malignant tumours are difficult to distinguish from prostatitis located in PZ and BPH located in CG (Hoeks et al. (2011), Verma et al. (2012)). 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 illus-

trates the challenge of CaP detection in CG (van Niekerk et al. (2013)).

- **DW MRI:** As previously mentioned in the introduction, DW MRI is the most recent MRI imaging technique aiming at CaP detection and diagnosis (Scheidler et al. (1999b)). This modality exploits the variations in the motion of water molecules in different tissues (Koh and Collins (2007), Le Bihan et al. (1988)).

From a physiological point of view, the following facts can be claimed. On the one hand, PZ, as previously mentioned, is mainly glandular and tubular in structure allowing water molecules to move freely (Choi et al. (2007), Hoeks et al. (2011)). 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 (Hoeks et al. (2011)). Then, CaP growth leads to the destruction of normal glandular structure and is associated with an increase in cellular density (Hoeks et al. (2011), Koh and Collins (2007), Somford et al. (2008)). Furthermore, these factors both have been shown to be inversely correlated with water diffusion (Koh and Collins (2007), Somford et al. (2008)): higher cellular density implies a restricted water diffusion. Thus, water diffusion in CaP will be more restricted than both healthy PZ and CG (Hoeks et al. (2011), Koh and Collins (2007)).

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 (Huisman (2003)). 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
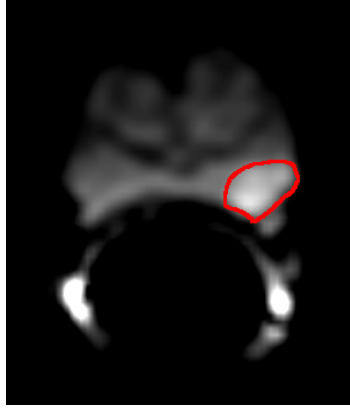
(a) DW-MRI image acquired with a 1.5 Tesla MRI scanner. The cancer corresponds to the high SI region highlighted in red.
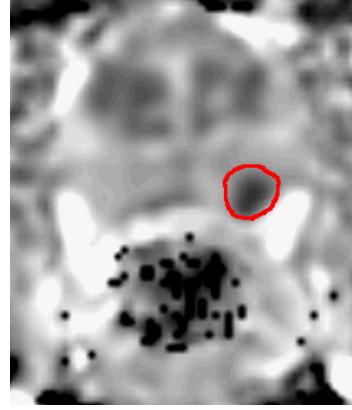
(b) ADC map computer after acquisition of DW-MRI iages with a 1.5 Tesla MRI scanner. The cancer corresponds to the low SI region highlighted in red.

Figure 5: Illustration of of DW-MRI and ADC map. The signal intensity corresponding to cancer are inversely correlated on these two types of imaging techniques.

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 (Huisman (2003)). Under these conditions, the MRI signal is measured as:

$$M_{x,y}(t, b) = M_{x,y}(0) \exp\left(-\frac{t}{T_2}\right) S_{\text{ADC}}(b) , \qquad (3)$$

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

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 gradient pulses parameters: (i) gradient intensity and (ii) gradient duration (Le Bihan et al. (1986)).

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

To summarize, in DW MRI images, CaPs are characterized by high-SI compared to normal tissues in PZ and CG as shown in Fig. 5(a) (Barentsz et al. (2012)). However, some tissues in CG can look similar to CaP with higher SI (Barentsz et al. (2012)).

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 (Choi et al. (2007), Padhani (2011), Shimofusa et al. (2005)). As draw-

backs, this modality suffers from poor spatial resolution and specificity due to false positive detection (Choi et al. (2007)).

With a view to eliminate these drawbacks, radiologists are extracting quantitative maps from DW MRI. This imaging technique is presented next.

− *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 (Le Bihan et al. (1986)). The ADC is considered as a "pure" diffusion coefficient and can be extracted to build a quantitative map.

From Eq. 3, it is clear that performing multiple acquisitions only varying b will not have any effect on the term $M_{x,y}(0) \exp\left(-\frac{t}{T_2}\right)$. Thus, Eq. 3 can be rewritten as:

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

To compute the ADC map, a minimum of two acquisitions are necessary: (i) for $b_0 = 0$ s.mm^{-2} where the measured signal is equal to S_0, and (ii) $b_1 > 0$ s.mm^{-2} (typically 1000 s.mm^{-2}). Then, the ADC map can be computed as:

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

More accurate computation of the ADC map can be obtained by performing several acquisitions with different values for the parameter b and performing a semi-logarithmic linear fitting using the model presented in Eq. (5).

Regarding the appearance of the ADC maps, it was previously stated that by increasing the value of b, the signal of CaP tissue increases significantly. From Eq. (6), it can be shown that tissue appearance in the ADC map will be the inverse of DW MRI images. Then, CaP tissue is associated with low-SI whereas healthy tissue appears brighter as depicted in Fig. 5(b) (Barentsz et al. (2012)).

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 (Choi et al. (2007), Doo et al. (2012)). Moreover, it has been shown that ADC is correlated with GS (Hambrock et al. (2011), Itou et al. (2011), Peng et al. (2013)).

However, some tissues of the CG zone mimic CaP with low-SI (Kirkham et al. (2006)) and image distortion can arise due to haemorrhage (Choi et al. (2007)). 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 (Choi et al. (2007)).

– **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 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 (Awwad et al. (2012), Costello and Franklin (2006), Giskeodegard et al. (2013)).

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 (Costello and Franklin (2006)). However, the metabolism allowing the accumulation of citrate requires a large amount of energy (Costello and Franklin (2006)). In contrast, malignant cells do not have high zinc levels leading to lower citrate levels due to citrate oxydation (Costello and Franklin (2006)). Furthermore, this change results in a more energy-efficient metabolism enabling malignant cells to grow and spread (Costello and Franklin (2006)).

An increased concentration of choline is related to CaP (Awwad et al. (2012)). Malignant cell development requires epigenetic mechanisms resulting in metabolic changes and relies on two mechanisms: DNA methylation and phospholid metabolism which both result in choline uptake, explaining its increased level in CaP tissue (Awwad et al. (2012)).

Spermine is also considered as a biological marker in CaP (Giskeodegard et al. (2013), van der Graaf et al. (2000)). In CaP, reduction of the ductal volume due to shifts in polyamine homeostasis might lead to a reduced spermine concentration (van der Graaf et al. (2000)).

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 (Haacke et al. (1999)). 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 (Haacke et al. (1999)), a phenomenon known as the chemical shift effect (CSE) (Parfait (2010)). Given this property, metabolites can be identified and their concentrations can be determined. In this regard, the Fourier transform is used to obtain the frequency spectrum of the NMR signal (Haacke et al. (1999), Parfait (2010)). 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 (see Fig. 6) (Parfait (2010)).

Hence, frequencies of interest in regard to CaP detection and diagnosis should correspond to the earlier mentioned metabolites. Choline and spermine are represented by a single peak at respectively 3.21 ppm and 3.11 ppm (Verma et al. (2010)). Due to the coupling effect, citrate is represented by three or four peaks depending on the magnetic field strength. Citrate ranges from 2.47 ppm to 2.81 ppm with a central frequency at 2.64 ppm (Verma et al. (2010)). Then, relative concentrations of these metabolites are obtained by computing the area under the curve of the spectrum between the lower and upper frequency limits of each peak (see Fig. 6).

Two different quantitative approaches are used to decide or whether not the spectra of a ROI is associated with CaP classified either as relative quantification or absolute quantification (Lemaître (2011)). In relative quantification, the ratio of choline-polyamines-creatine to citrate is computed. The integral of the signal is computed from choline (cf., 3.21 ppm) to creatine (cf., 3.02 ppm) because the peaks in this region can be merged at clinical magnetic field strengths (see Fig. 6) (van der Graaf et al. (2000), Hoeks et al. (2011)). Considering the previous assumption 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 (Lemaître (2011)). 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 (Choi et al. (2007)). Furthermore, it has been shown that combining MRSI with MRI improves detection and diagnosis performance (Kaji et al. (1998), Scheidler et al. (1999a), Vilanova et al. (2009)). Citrate and spermine concentrations are inversely correlated with the GS allowing us to distinguish low from high grade CaPs (Giskeodegard et al. (2013)). However, choline concentration does not provide the same properties (Giskeodegard et al. (2013)).

Unfortunately, MRSI also presents several drawbacks. First, MRSI acquisition is time consuming which prevents this modality from being used in daily clinical practise (Barentsz et al. (2012)). 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 ad-

Table 2: Overview of the different studies reviewed with their main characteristics. Acronyms: number (#) - image regularization (Img. Reg.).

| Index	Study	# patients	MRI-modality				Strength of field		Studied zones		CAD stages		
			T₂-W MRI	DCE MRI	DW MRI	MRSI	1.5 T	3.0 T	PZ	CG	Img. Reg.	CADe	CADx
[1]	Ampeliotis et al. (2007)	25	✓	✓	✗	✗	✓	✗	✓	✗	✓!	✗	✓
[2]	Ampeliotis et al. (2008)	25	✓	✓	✗	✗	✓	✗	✓	✗	✓!	✗	✓
[3]	Antic et al. (2013)	53	✓	✗	✓	✗	✓	✗	✓	✓	✗	✗	✓
[4]	Artan et al. (2009)	10	✓	✓	✓	✗	✓	✗	✓	✗	✗	✓	✓
[5]	Artan et al. (2010)	21	✓	✓	✓	✗	✓	✗	✓	✗	✓!	✓	✓
[6]	Chan et al. (2003)	15	✓	✗	✓	✗	✓	✗	✓	✗	✗	✗	✓
[7]	Giannini et al. (2013)	10	✓	✓	✓	✗	✓	✗	✓	✗	✓	✓	✓
[8]	Kelm et al. (2007)	24	✗	✗	✗	✓	✓	✗	✓	✓	✓!	✓	✓
[9]	Langer et al. (2009)	25	✓	✓	✓	✗	✓	✗	✓	✗	✓!	✗	✓
[10]	Litjens et al. (2011)	188	✓	✓	✓	✗	✗	✓	✓	✗	✓!	✓	✓
[11]	Litjens et al. (2012b)	288	✓	✓	✓	✗	✗	✓	✓	✓	✓!	✓	✓
[12]	Liu et al. (2009)	11	✓	✓	✓	✗	✓	✗	✓	✗	✓!	✓	✓
[13]	Liu et al. (2013)	54	✓	✓	✓	✗	✗	✓	✓	✓	✓!	✗	✓
[14]	Lopes et al. (2011)	27	✓	✗	✗	✗	✓	✗	✓	✗	✓!	✓	✓
[15]	Lv et al. (2009)	55	✓	✗	✗	✗	✓	✗	✓	✗	✓!	✗	✓
[16]	Matulewicz et al. (2013)	18	✗	✗	✗	✓	✗	✓	✓	✓	✗	✓	✓
[17]	Mazzetti et al. (2011)	10	✗	✓	✗	✗	✓	✗	✓	✗	✓!	✓	✓
[18]	Niaf et al. (2011)	23	✓	✓	✓	✗	✓	✗	✓	✗	✓!	✗	✓
[19]	Niaf et al. (2012)	30	✓	✓	✓	✗	✓	✗	✓	✗	✓!	✗	✓
[20]	Ozer et al. (2009)	20	✓	✓	✓	✗	✓	✗	✓	✗	✓!	✓	✓
[21]	Ozer et al. (2010)	20	✓	✓	✓	✗	✓	✗	✓	✗	✓!	✓	✓
[22]	Parfait et al. (2012)	22	✗	✗	✗	✓	✗	✓	✓	✓	✓!	✓	✓
[23]	Peng et al. (2013)	48	✓	✓	✓	✗	✗	✓	✓	✓	✗	✗	✓
[24]	Puech et al. (2009)	100	✗	✓	✗	✗	✓	✗	✓	✓	✗	✗	✓
[25]	Sung et al. (2011)	42	✗	✓	✗	✗	✗	✓	✓	✗	✗	✓	✓
[26]	Tiwari et al. (2007)	14	✗	✗	✗	✓	✓	✗	✓	✓	✓!	✓	✓
[27]	Tiwari et al. (2008)	18	✗	✗	✗	✓	✓	✗	✓	✓	✓!	✓	✓
[28]	Tiwari et al. (2009a)	18	✗	✗	✗	✓	✓	✗	✓	✗	✓!	✓	✓
[29]	Tiwari et al. (2009b)	15	✓	✗	✗	✓	✓	✗	✓	✗	✓!	✓	✓
[30]	Tiwari et al. (2010)	19	✓	✗	✗	✓	✓	✗	✓	✗	✓!	✓	✓
[31]	Tiwari et al. (2012)	36	✓	✗	✗	✓	✓	✗	✓	✓	✗	✓	✓
[32]	Tiwari et al. (2013)	29	✓	✗	✗	✓	✓	✗	✓	✓	✓!	✓	✓
[33]	Viswanath et al. (2008b)	16	✓	✗	✗	✓	✓	✗	✓	✓	✗	✓	✓
[34]	Viswanath et al. (2008a)	6	✓	✓	✗	✗	✗	✓	✓	✓	✓!	✓	✓
[35]	Viswanath et al. (2009)	6	✓	✓	✗	✗	✗	✓	✓	✓	✓	✓	✓
[36]	Viswanath et al. (2011)	12	✓	✓	✓	✗	✗	✓	✓	✓	✓!	✓	✓
[37]	Viswanath et al. (2012)	22	✓	✗	✗	✗	✗	✓	✓	✓	✓	✓	✓
[38]	Vos et al. (2008a)	29	✓	✓	✗	✗	✓	✗	✓	✗	✓!	✗	✓
[39]	Vos et al. (2008b)	29	✗	✓	✗	✗	✓	✗	✓	✗	✓!	✗	✓
[40]	Vos et al. (2010)	29	✓	✓	✗	✗	✓	✗	✓	✗	✓!	✗	✓
[41]	Vos et al. (2012)	NA	✓	✓	✓	✗	✗	✓	✓	✗	✓!	✓	✓

Notes:

✗: not used or not implemented.

✓!: partially implemented.

✓ : used or implemented.

10
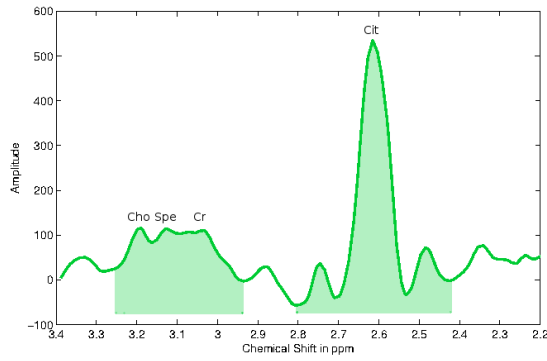
(a) Illustration of an MRSI spectrum of an healthy voxel acquired with a 3.0 Tesla MRI.
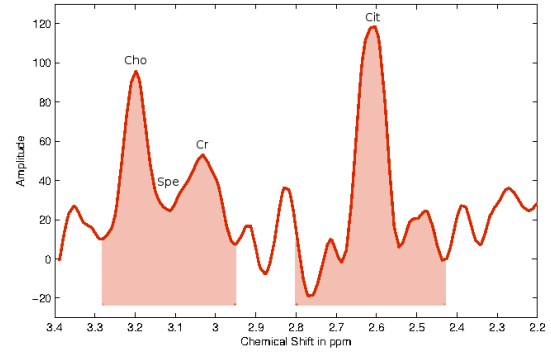
(b) Illustration of an MRSI spectrum of a cancerous voxel acquired with a 3.0 Tesla MRI.

Figure 6: Illustration of an MRSI spectrum both healthy and cancerous voxel with a 3.0 Tesla MRI. The highlighted areas corresponds 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).

dressed by developing new scanners with higher magnetic field strengths such as 7.5 T (Giskeodegard et al. (2013)). Finally, a high variability of the relative concentrations between patients was observed (Choi et al. (2007)). The same observation was made depending on the zones studied (cf., PZ, CG, base, mid-gland, apex) (Lemaître (2011), Walker et al. (2010)). Due to this variability, it is difficult to use a fixed thresholds in order to differentiate CaP from healthy tissue.

2.3.3. Computer-aided systems for CaP: CADe - CADx

As previously mentioned in the introduction (see Sect. 1), CADs are developed to advise and backup radiologists in their tasks of CaP detection and diagnosis; CADs are not aimed to provide fully automatic decisions (Giger et al. (2008)). 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 (Giger et al. (2008)). 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. The CAD work-flow is presented in Fig. 2.

MRI modalities mentioned in Sect. 2.3.2 are used as inputs of CAD for CaP. It can be noted that ADC map is not considered as an input since it is a feature derived from the DW MRI images. The images acquired from the different modalities show a large variability between patients: the prostate organ can be located at different positions in images (e.g., patient motion, variation of acquisition plan), and the SI can be corrupted with noise or artefacts during the acquisition process (eg., magnetic field inhomogeneity, use of endorectal coil). To address these issues, the first stage of CAD is to pre-process multiparametric MRI images to reduce noise, remove artefacts and standardize the SI. Then, it is important to mention that most of the later processes would be only focused on the prostate. Thus, it is

necessary to segment the prostate in each MRI-modality to define it as a ROI. However, data suffers of misalignment due to patient motions or different acquisition plan. Therefore, a registration step is performed so that all the previously segmented MRI images will be in the same reference frame.

Some studies do not fully apply the methodology depicted in Fig. 2. Details about those can be found in Tab. 2. Some studies preferred to work directly with raw data in order to demonstrate the robustness of their approaches to noise or artefacts. In some cases, prostate segmentation is performed manually as well as registration. It is also sometimes assumed that no patient motions occur during the acquisition procedure, removing the need of registering the multiparametric MRI images.

Once the data are regularized, it becomes possible to extract features and classify the data to obtain the probabilistic maps. We refereed this stage to image classification where CADe and CADx are the main components.

In a CADe framework, possible lesions will be segmented automatically and further used as input of CADx. We also included in CADe studies, the methods using voxel-based delineation in which the final results will highlight the boundaries of the lesions. On the other hand, manual lesions segmentation are not considered to be part of a CADe. The output of the CADe is used as input of the CADx.

CADx is composed of the processes allowing to distinguish malignant from non-malignant tumours. We divided CADx into three different stages. First, salient features are extracted, in an pixel-based or region-based manner, from MRI images to characterize the lesion. Of course, more discriminative features will be associated with a robust and accurate likelihood cancer map. Frequently, the number of features extracted can be large resulting in redundant or insufficient discriminative features which will negatively affect the performances of the further classification. Therefore, a step consists of selecting the best features or/and reducing the number of dimensions is commonly used. Then, this modified feature vector is finally classified using different pattern recognition approaches.

As pointed out in the introduction, performance of CaP de-

Table 3: Overview of the pre-processing methods used in CAD systems.

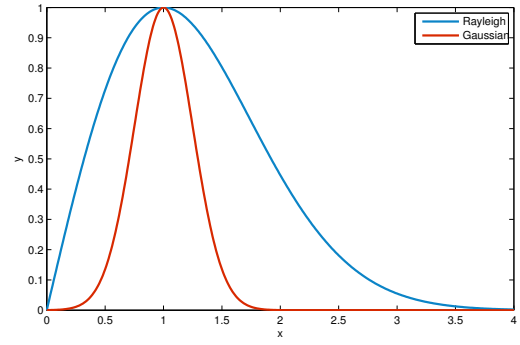| Pre-processing operations	References
MRI pre-processing:	
Noise filtering:	
Median filtering	[20-21]
Wavelet-based filtering	[1-2,14]
Bias correction:	
Parametric methods	[15-35]
Non-parametric methods	[36]
Standardization:	
Statistical-based normalization:	[3-4,15,20-21,35,37]
Organ SI-based normalization	[18-19]
MRSI pre-processing:	
Phase correction	[22]
Water and lipid residuals filtering	[8]
Baseline correction	[22,31]
Frequency alignment	[31]
Normalization	[22]

Figure 7: Illustration of a Gaussian distribution ($\mu = 1, \sigma = 0.25$) and a Rayleigh distribution ($\sigma = 2$). It can be seen that the Rayleigh distribution is suffering of a bias term when compared with the Gaussian distribution.

tection and diagnosis are affected by observer interpretation and limitations (Giger et al. (2008), Hambrock et al. (2013)). CAD offers a possible solution in order to reduce this variability. As mentioned in the introduction, the effects of CAD on the observer performance has been studied (Hambrock et al. (2013)), with results showing that CADs benefit to less-experienced radiologist to perform similarly as experienced radiologist in their tasks (Hambrock et al. (2013)).

2.4. Literature classification

The CAD review is organized using the methodology presented in Fig. 2. Methods embedded in the image regularization framework are presented before to focus on the image classification framework, the later being divided into CADe and CADx. Table 2 summarizes the different CAD studies reviewed in this paper. Characteristics related to MRI acquisition as well as CAD strategies are reported. Only methods used in CAD system are discussed.

3. Image regularization framework

This section provides a review of the methods used in CADs in order to regularize input images. We start with pre-processing methods presented in Sect. 3.1, focusing mainly on the reduction of noise level and artefacts as well as standardization of SI. Sections 3.2 and 3.3 will be dedicated to segmentation methods, so that later methods only operate on the segmented prostate, and registration to align segmented images from different MRI-modalities in the same reference frame.

3.1. Pre-processing
3.1.1. MRI images pre-processing

Three different groups of pre-processing methods are commonly applied to images as initial stage in CAD.

– *Noise filtering:* The NMR signal measured and recorded in the k-space during an MRI acquisition is affected by noise. This noise obeys a complex Gaussian white noise mainly due to thermal noises in the patient area (Nowak (1999)). 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, being a linear and orthogonal transform, does not affect the Gaussian noise characteristics (Nowak (1999)). However, the function involved in the magnitude computation 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 harder. Briefly, a Rician distribution can be 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 (see Fig. 7) (Manjon et al. (2008)). Reviews of all denoising methods can be found in the work of Buades et al. (2005) and Mohan et al. (2014).

Median filtering is the simplest approach used to address the denoising issue in MRI images (Ozer et al. (2009, 2010)). In both studies, Ozer et al. used a square kernel of size 5×5 pixels with the image resolutions ranging from 320×256 (cf., T_2-W MRI) to 256×128 (cf., T_2 map, DCE and DW MRI) and a field of view (FOV) ranging from 14 cm (cf, T_2-W and DW MRI) to 20 cm (cf, T_2 map and DCE MRI). However, from a theoretical point of view, this simple filtering method is not well formalized to address the noise distribution in MRI images.

More complex approaches were proposed to overcome this problem. A common method used to denoise MRI images is based on wavelet-based filtering. 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 (Donoho and Johnstone (1994)). Therefore, denoising is performed by threshold-
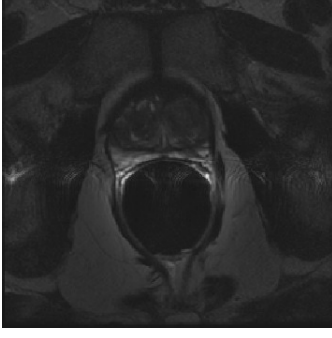
Figure 8: Example of artefacts with high SI due to perturbation from the endorectal coil which create inhomogeneity.

ing/attenuating the insignificant wavelet coefficients to enforce the sparsity in the wavelet-domain. Investigations focus on the strategies to perform the most adequate coefficient shrinkage method (e.g., using thresholding, singularity property or Bayesian framework) (Pizurica (2002)).

Ampeliotis et al. (2007, 2008) performed wavelet shrinkage to denoise magnitude MRI images (cf., T_2-W-MRI and DCE-MRI) using thresholding techniques (Mallat (2008)). 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 (Nowak (1999)).

Lopes et al. (2011) used a technique proposed by Pizurica et al. (2003) to denoise T_2-W-MRI. Pizurica et al. (2003) proposed a filtering technique based on joint detection and estimation theory (Middleton and Esposito (1968)). The wavelet coefficients "free-of-noise" are estimated from the noisy wavelet coefficients using a maximum *a posteriori* (MAP) estimate. Furthermore, the estimator designed 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 used instead of the magnitude MRI image as proposed by Nowak (1999). This involves changing the Rician distribution to a scaled non-central Chi-square 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,

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

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 (Styner et al. (2000)). This bias field results in a smooth variation of the SI through the image. When an endorectal coil is used, an artefact resulting of an hyper-intense signal can be observed around the coil on the images (see Fig. 8). As a consequence, the SI of identical tissues varies depending on their spatial location in the image making further processes such as segmentation or registration harder (Jungke et al. (1987), Vovk et al. (2007)). A review of bias correction methods can be found in Vovk et al. (2007).

The model of image formation is usually formalized such that:

$$ss(\mathbf{x}) = o(\mathbf{x})b(\mathbf{x}) + \eta(\mathbf{x}) \, , \qquad (8)$$

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. (2009) performed bias correction on T_2-W-MRI using a parametric Legendre polynomial model proposed by Styner et al. (2000) and available in the Insight Segmentation and Registration Toolkit (ITK) library[3]. Styner et al. (2000) chose to model the bias field by using a linear combination of Legendre polynomials as:

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

where $\hat{b}$ 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 allows us to model the bias as a smooth inhomogeneity 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 μ_k being the mean SI and a variance σ_k^2 of 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:

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

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

where $\rho_k(\cdot)$ is a M-estimator allowing estimations to be less sensitive to outliers than usual square distance (Li (1996)).

Finally, estimation of the parameters $\mathbf{p}$ results in finding the minimum of the cost function $C(\mathbf{p})$. This optimization

---

[3]The ITK library is available at: http://www.itk.org/

was performed using the non-linear $(1 + 1)$ Evolution Strategy (ES) optimizer (Styner and Gerig (1997)).

In a later publication, Viswanath et al. (2012) make use of the well known N3 algorithm[4] to correct T_2-W-MRI developed by Sled et al. (1998). To estimate the bias function, Sled et al. (1998) proposed to estimate the probability density functions (PDFs) of the signal and bias.

Recalling Eq. (8) and taking advantage of logarithm property, it implies that this model becomes additive such that:

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

where $\hat{o}(\mathbf{x})$ is the signal only degraded by noise. Sled et al. (1998) shows that Eq. (12) can be related to PDFs such that:

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

where S, B and O are respectively the probability densities of s, b and o.

Restoring the corrupted signal s is carried out by finding the multiplicative field b which maximizes the frequency content of the distribution O. Sled et al. (1998) argue that a search through all possible fields b and selection of the one which maximizes the high frequency content of O could be carried out but results in an exhaustive search. However, they show that the bias field distribution can be assimilated to a near Gaussian distribution. Using this fact as *a priori*, it is then possible to infer the distribution O using Wiener deconvolution given B and S and later estimate the corresponding smooth field b.

Lv et al. (2009) corrected the inhomogeneity in T_2-W-MRI images by using the method proposed by Madabhushi et al. (2006). In this method, the MRI images are corrected iteratively by successively detecting the image foreground via generalized scale (g-scale) and estimating a bias field function based on a second-order polynomial model. First the background of the MRI image is eliminated by thresholding. The threshold value is commonly equal to the mean SI of the considered image. Then, in the seeded region growing algorithm is applied considering every thresholded pixel as a potential seed. However, pixels already assigned to a region will not be considered any more as seed. As in seeded region growing algorithm (Shapiro and Stockman (2001)), two criteria are taken into account to expand the region. First, the region will grow 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 the SIs of two pixels. Depending on the membership value (cf., a threshold has 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 second order polynomial $\hat{b}_\Theta(\mathbf{x})$ is fitted in a least-squares sense (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 $\hat{b}_\Theta(\mathbf{x})$. This process is repeated until 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 performed by first learning from a training set of patients. Hence, one can emphasize the desire to perform MRI examinations with a high repeatability or in other words, one would ensure to obtain similar MRI images (cf., similar SIs) for patients of the same group (cf., healthy patients *vs.* patients with CaP), for a similar sequence.

However, it is a known fact that variability between patients occurs during the MRI examinations even using the same scanner, protocol or sequence parameters (Nyul and Udupa (1999)). 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.

Approaches used to standardize MRI images can be either categorized as statistical-based standardization or organ SI-based standardization.

Artan et al. (2010, 2009) as well as Ozer et al. (2009, 2010) standardized T_2-W, DCE 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, mu_{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. (2013) normalized T_2-W-MRI by making use of the median and interquartile range for all the pixels.

$$
I_s(\mathbf{x}) = \frac{I_r(\mathbf{x}) - Q_2}{Q_3 - Q_1}, \ \forall \mathbf{x}, \tag{17}
$$

where $I_s(\mathbf{x})$ is the standardized SI with the images coordinates $\mathbf{x} = \{x, y\}$, $I_r(\mathbf{x})$ is the raw SI and with Q_1, Q_2 and Q_3 being respectively the first quartile, the median and the third quartile respectively.

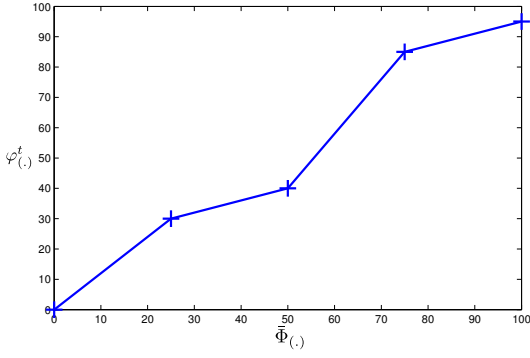[4]The N3 algorithm implementation is available at: `http://www.bic.mni.mcgill.ca/software/N3/`

Figure 9: Example of linear mapping by parts as proposed by Nyul et al. (2000).

(a) Illustration and location of the bladder on a T_2-W-MRI image acquired with a 3.0 Tesla MRI scanner

(b) Illustration and location of the femoral arteries on a T_1-W-MRI image acquired with a 3.0 Tesla MRI scanner

Figure 10: Illustration of the two organs used by Niaf et al. (2012, 2011) to normalize T_2-W and T_1-W MRI images.

Lv et al. (2009) scaled the SI of T_2-W-MRI images using the method proposed by Nyul et al. (2000) based on PDF matching. 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 median and different quantiles. Using a training set, these statistical landmarks are extracted for N training images as for instance for the minimum, the 25th quantile, the median, the 75th quantile and the maximum:
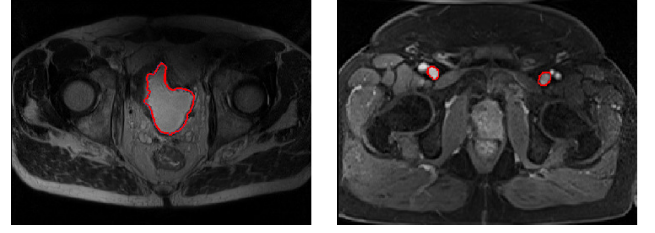
$$
\begin{aligned}
\Phi_0 &= \{\phi_0^1, \phi_0^2, \cdots, \phi_0^N\}, \\
\Phi_{25} &= \{\phi_{25}^1, \phi_{25}^2, \cdots, \phi_{25}^N\}, \\
\Phi_{50} &= \{\phi_{50}^1, \phi_{50}^2, \cdots, \phi_{50}^N\}, \\
\Phi_{75} &= \{\phi_{75}^1, \phi_{75}^2, \cdots, \phi_{75}^N\}, \\
\Phi_{100} &= \{\phi_{100}^1, \phi_{100}^2, \cdots, \phi_{100}^N\},
\end{aligned}
\tag{18}
$$

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

Then, the mean of each quantile $\{\bar{\Phi}_0, \bar{\Phi}_{25}, \bar{\Phi}_{50}, \bar{\Phi}_{75}, \bar{\Phi}_{100}\}$ is also calculated. Once this training stage is performed, a linear transformation by parts $\mathcal{T}(\cdot)$ can be computed (Eq. (19)) for each test image t by mapping each statistical landmark $\varphi_{(cdot)}^t$ of this image with the pre-learned statistical landmarks $\bar{\Phi}_{(\cdot)}$. This linear mapping is also depicted in Fig. 9.

$$
\mathcal{T}(s(\mathbf{x})) = \begin{cases}
\lceil \bar{\Phi}_0 + (s(\mathbf{x}) - \varphi_0^t)\left(\frac{\Phi_{25}-\bar{\Phi}_0}{\varphi_{25}^t-\varphi_0^t}\right)\rceil, & \text{if } \varphi_0^t \leq s(\mathbf{x}) < \varphi_{25}^t, \\
\lceil \bar{\Phi}_{25} + (s(\mathbf{x}) - \varphi_{25}^t)\left(\frac{\bar{\Phi}_{50}-\bar{\Phi}_{25}}{\varphi_{50}^t-\varphi_{25}^t}\right)\rceil, & \text{if } \varphi_{25}^t \leq s(\mathbf{x}) < \varphi_{50}^t, \\
\lceil \bar{\Phi}_{50} + (s(\mathbf{x}) - \varphi_{50}^t)\left(\frac{\bar{\Phi}_{75}-\bar{\Phi}_{50}}{\varphi_{75}^t-\varphi_{50}^t}\right)\rceil, & \text{if } \varphi_{50}^t \leq s(\mathbf{x}) < \varphi_{75}^t, \\
\lceil \bar{\Phi}_{75} + (s(\mathbf{x}) - \varphi_{75}^t)\left(\frac{\bar{\Phi}_{100}-\bar{\Phi}_{75}}{\varphi_{100}^t-\varphi_{75}^t}\right)\rceil, & \text{if } \varphi_{75}^t \leq s(\mathbf{x}) \leq \varphi_{100}^t,
\end{cases}
\tag{19}
$$

Viswanath et al. (2011, 2009, 2012) use a variant of this previous approach presented in the work of Madabhushi and Udupa (2006) aiming to standardize the T_2-W-MRI images. Instead of computing the PDF of an entire image, a presegmentation of the foreground is carried out via g-scale which was discussed in the bias correction section. Once the foreground is detected, the largest region is extracted and the

same process than previously mentioned (see Eq. (19)) takes place in order to align PDFs of the foreground of the MRI images.

The methods described above were statistical-based methods. However, 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. (2012, 2011) normalized T_2-W-MRI images by dividing the original SI of the images by the mean SI of the bladder (see Fig. 10(a)). Likewise, Niaf et al. (2011) standardized the T_1-W-MRI images using the arterial input function (AIF). They computed the AIF by taking the mean of the SI in the most enhanced part of the common femoral arteries (see Fig. 10(b)) as proposed by Wiart et al. (2007).

3.1.2. MRSI spectra

Presented in Sect. 2.3.2, 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 processing methods.

– **Phase correction:** MRSI data acquired suffer from zero-order and first-order phase misalignments as shown in Fig. 11 (Chen et al. (2002), Osorio-Garcia et al. (2012)).

Parfait et al. (2012) used a method proposed by Chen et al. (2002) where the phase of MRSI signal is corrected based on entropy minimization in the frequency domain. The corrected MRSI signal $o(\xi)$ can be expressed as:

$$
\begin{aligned}
\Re(o(\xi)) &= \Re(s(\xi))\cos(\Phi(\xi)) - \Im(\xi)\sin(\Phi(\xi)), \\
\Im(o(\xi)) &= \Im(s(\xi))\cos(\Phi(\xi)) + \Re(\xi)\sin(\Phi(\xi)), \\
\Phi(\xi) &= \phi_0 + \phi_1\frac{\xi}{N},
\end{aligned}
\tag{20}
$$

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. (2002) tackled this problem using an optimization framework where ϕ_0 and ϕ_1 had to be inferred. Hence,
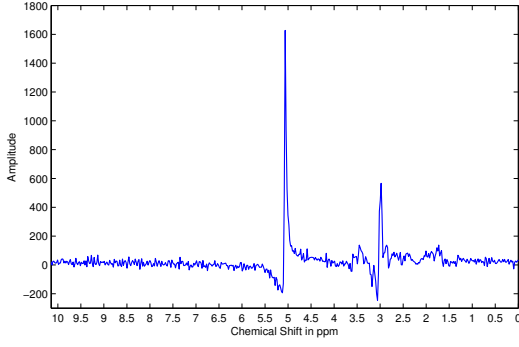
15

Figure 11: Illustration of phase misalignment in an MRSI spectra acquire with a 3.0 Tesla MRSI scanner. Note the distortion of the signal specially visible for the water and citrate peaks.
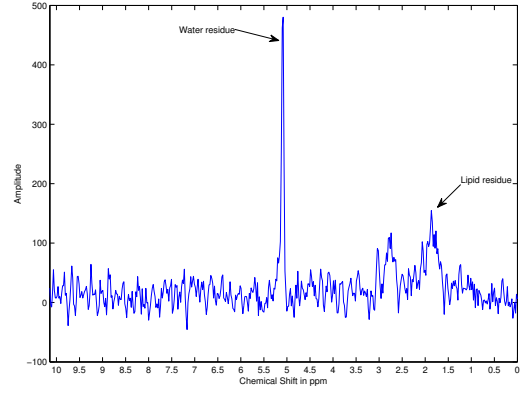

Figure 12: Illustration of the residues of water and fat even after their suppression during the acquisition protocol. The acquisition was carried out with a 3.0 Tesla MRI.

the simplex Nelder-Mead optimization method was used to minimize the following cost function based on the Shannon entropy formulation:

$$
\hat{\Phi} = \arg \min_{\Phi} \left[-\sum \mathfrak{R}(s'(\xi)) \ln \mathfrak{R}(s'(\xi)) + \lambda \|\mathfrak{R}(s(\xi))\|_2 \right] ,
$$
(21)

where $s'(\xi)$ is the first derivative of the corrupted signal $s(\xi)$ and λ is a regularization parameter.

Once the best parameter Φ is obtained, the MRSI signal is corrected using Eq. (20).

– **_Water and lipid residuals filtering:_** The water and lipid metabolites occur in much higher concentrations the metabolites of interests (cf., choline, creatine and citrate) (Osorio-Garcia et al. (2012), Zhu et al. (2010)). Fortunately, specific MRSI sequences were developed in order to suppress water and lipid metabolites using pre-saturation techniques (Zhu et al. (2010)). However, these techniques do not perfectly remove water and lipids peaks and some residuals are still present in the MRSI spectra as shown in Fig. 12. Therefore, different post-processing methods have been proposed to enhance the quality of the MRSI spectra by removing these residuals.

Kelm et al. (2007) used the well known HSVD algorithm proposed by Pijnappel et al. (1992). In the time domain, a MRSI signal $s(t)$ is modelled by a sum of K exponentially damped sinusoids such that:

$$
s(t) = \sum_{k=1}^{K} a_k \exp(i\phi_k) \exp(-d_k + i2\pi f_k)t + \eta(t) ,
$$
(22)

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.

Pijnappel et al. (1992) showed that the "noise-free signal" can be found using the singular value decomposition (SVD) decomposition. First the noisy signal is reorganized inside a Hankel matrix H. It can be shown that if the signal considered would be a "noise-free signal", the rank of

H would be 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 can be truncated to K using its SVD decomposition. Hence, knowing the cut off frequencies of water (cf., 4.7 ppm) and lipid (cf., 2.2 ppm) metabolites, their corresponding peaks can be reconstructed and subtracted from the original signal (Laudadio et al. (2002)).

– **_Baseline correction:_** Sometimes, the problem discussed in the above section regarding the lipid molecules is not addressed simultaneously with water residuals suppression. Lipids and macromolecules are known to affect the baseline of the MRSI spectra. They could cause errors during further fitting processes aiming to quantify the metabolites, especially regarding the citrate metabolite.

Parfait et al. (2012) made the comparison of two different methods to detect the baseline and correct the MRSI spectra which are based on the work of Lieber and Mahadevan-Jansen (2003) and Devos et al. (2004). Lieber and Mahadevan-Jansen (2003) addressed the problem of baseline detection in the frequency domain by fitting a polynomial of low degree $p(x)$ (e.g., second or third degree) to the MRSI signal $s(x)$ in a least-squares sense. Then, the values of the fitted polynomial are re-assigned as:

$$
p_f(x) = \begin{cases} p(x) , & \text{if } p(x) \le s(x) , \\ s(x) , & \text{if } p(x) > s(x) . \end{cases}
$$
(23)

Finally, this procedure of fitting and re-assignment is iteratively repeated on $p_f(x)$ until a stopping criterion is reached. The final polynomial function can be subtracted from the original signal s(x) to correct it.

Parfait et al. (2012) 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 (2003), Devos et al. (2004) proposed to correct the baseline in the time domain
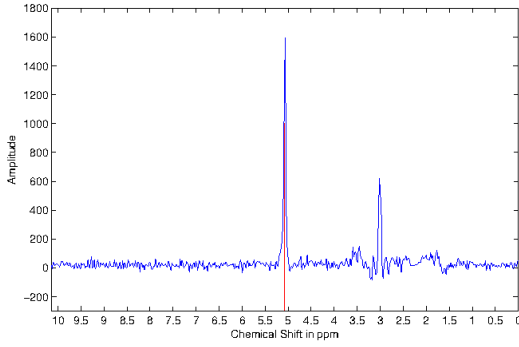
16

Figure 13: Illustration of frequency misalignment in an MRSI spectra acquired with a 3.0 Tesla MRSI scanner. The water peak is known to be aligned at 4.65 ppm. However, it can be seen that the peak on this spectra is aligned at around 5.1 ppm.

by multiplying the MRSI signal by a decreasing exponential function as:

$$c(t) = \exp(-\beta t) \, , \qquad (24)$$

Having a typical value for β of 0.15.

However, Parfait et al. (2012) concluded that the method proposed by Lieber and Mahadevan-Jansen (2003) outperforms the one of Devos et al. (2004). In the contemporary work of Tiwari et al. (2012), the authors detected the baseline using a local non-linear fitting method avoiding regions with significant peaks which were detected using a experimentally parametrised signal-to-noise ratio (i.e. a value larger than 5 dB).

− **Frequency alignment:** Due to variations of the experimental conditions, a frequency shift can be observed in the MRSI spectra as shown in Fig. 13 (Chen et al. (2002), Osorio-Garcia et al. (2012)).

Tiwari et al. (2012) corrected the frequency shift by first detecting known metabolite peaks such as choline, creatine and citrate. The frequency shift is corrected by minimizing the frequency error between the experimental and theoretical values of each of these peaks.

− **Normalization:** Due to variations of the experimental conditions, the MRSI signal may also vary between patients.

Parfait et al. (2012) as Devos et al. (2004) compared two methods to normalize MRSI signal. In each method, the original MRSI spectra is divided by a normalization factor, similar to the intensity normalization described earlier.

The first approach to obtain the normalization factor is based on an estimation of the water concentration. It is required to have an additional MRSI sequence where the water metabolites are unsuppressed. Using this sequence, an estimation of the water concentration can be performed using the previously reported HSVD algorithm. The second approach to normalization is based on using the L_2 norm of the MRSI spectra $\|s(\xi)\|_2$. It should be noted that both Parfait et al. (2012) and Devos et al. (2004) concluded that the L_2 normalization was more efficient in their framework.

Table 4: Overview of the segmentation methods used in CAD systems.

Segmentation methods	References
MRI-based segmentation:	
Manual segmentation	[4-5,16,18-21,24,38-40]
Region-based segmentation	[11]
Model-based segmentation	[10,34-36,41]
MRSI-based segmentation:	
Clustering	[28]

3.2. Segmentation

The segmentation task consists of delineating the prostate boundaries in the MRI. This procedure is of particular importance for focusing the posterior processing on the organ of interest (Ghose et al. (2012)). In this section, only the segmentation methods used in CAD systems are presented and summarized in Tab. 4. An exhaustive review of prostate segmentation methods in MRI can be found in Ghose et al. (2012).

3.2.1. MRI-based segmentation

− **Manual segmentation:** To highlight the importance of prostate segmentation task in CAD systems, it is interesting to note the large number of studies which segment manually the prostate organs (Artan et al. (2010, 2009), Matulewicz et al. (2013), Niaf et al. (2012, 2011), Ozer et al. (2009, 2010), Puech et al. (2009), Vos et al. (2008a,b)). In all the cases, the boundaries of the prostate gland is defined in order to limit the further processing to only this area. This approach ensure the right delineation of the organ nevertheless this procedure is highly time consuming and should be perform by a radiologist.

− **Region-based segmentation:** Litjens et al. (2012b) used a multi-atlas-based segmentation using multi-modal images (e.g., T_2-W-MRI and ADC map) to segment the prostate with an additional pattern recognition method to differentiate CG and PZ as proposed in Litjens et al. (2012a). This method consists of 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 segmented voxels in either CG or PZ.

The registration between each atlas and the MRI images is performed using two successive registrations: the first registration is a rigid registration to roughly align the atlases and the MRI images and the second is an elastic registration using B-spline transformation. The objective function to perform the registration is defined as the weighted sum of the metric of both T_2-W-MRI and ADC map. The metric is based on mutual information (MI). We refer to the next section for more details in regard to registration. Two strategies of atlas selection were performed by using either a majority voting approach or the simultaneous truth and performance level estimation (STAPLE) approach (Warfield et al.

17

(2004)).

Subsequently, CG and PZ segmentation within the prostate region is achieved by classifying each voxel using a linear discriminant analysis (LDA) classifier. Three types of features were considered: (i) anatomy, (ii) intensity and (iii) texture. Regarding the anatomy, relative position and relative distance from the pixel to the border of the prostate were used. The intensity features consist of the intensity of the voxel in the ADC coefficient and the T_2 map. The texture features were composed of five different features: homogeneity, correlation (Amadasun and King (1989)), entropy, texture strength (Li et al. (2005)) and local binary pattern (LBP) (Ojala et al. (1996)). Finally, morphological operations were applied to remove artefacts and the contours between the zones were smoothed using thin plate spline (TPS) (Bookstein (1989)).

– **_Model-based segmentation:_** Viswanath et al. (2008a, 2009) used the multi-attribute non-initializing texture reconstruction based active shape model (MANTRA) method as proposed by Toth et al. (2008). MANTRA is closely related to the active shape model (ASM) from Cootes et al. (1995). This algorithm consists of two stages: (i) a training stage where a shape and appearance model is generated and (ii) the actual segmentation performed 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 (Cootes et al. (1995)). 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 are computed. For a given landmark l with its given neighbourhood $\mathcal{N}(l)$, its feature matrix extracted can be expressed as:

$$f_l = \{I_1(\mathcal{N}(l)), I_2(\mathcal{N}(l)), \cdots, I_k(\mathcal{N}(l))\} . \qquad (25)$$

where $I_k(\mathcal{N}(l))$ represents a feature vector obtained by sampling the kth texture map using the neighbourhood $\mathcal{N}(l)$.

By generating multiple landmarks in the same fashion as ASM, principal components analysis (PCA) (Pearson (1901)) is applied to learn the appearance variations.

For the segmentation stage, the mean shape learned previously is initialised 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 Cootes et al. (1995).

Subsequently, the same authors (Viswanath et al. (2012)) used the weighted ensemble of regional image textures for active shape model segmentation (WERITAS) method also proposed by Toth et al. (2009). As with the MANTRA method, WERITAS is based on the ASM formulation. In fact it is very close to the MANTRA itself. The same texture features are used to construct the appearance models,

but instead of using MI between the landmarks and neighbour patches for adapting the landmark positions, it defines a metric based on the Mahalanobis distance. 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 maximises the correlation with the Euclidean distance between the patches and the true landmarks. In the segmentation step, this metric is then computed between the initialised landmarks and neighbouring patches in order to update landmark positions, in a similar fashion to other active contour model (ACM) models.

Litjens et al. (2011) and Vos et al. (2012) used an approach proposed by Huisman et al. (2010) in which the bladder, prostate and rectum are segmented tackling the segmentation task as an optimization problem. A probabilistic model is first trained by embedding the three following aspects: (i) the shape by defining each organ as an ellipse, (ii) the position by defining the distance and the angle between each organ center and (iii) the appearance using the PDFs of SI of each organ. Litjens et al. (2011) used only ADC map to encode the appearance whereas Vos et al. (2012) used both ADC and T_2 maps. Then, during the optimization using a quasi-Newton optimizer, an objective function is minimized. This function is defined as the sum of the deviations from the above model learnt. This rough segmentation is then used inside a Bayesian framework to refine the segmentation.

3.2.2. MRSI-based segmentation

Tiwari et al. (2009a) localized the voxels corresponding to the prostate organ using a hierarchical spectral clustering. First, each MRSI spectrum is projected into a lower dimension space using graph embedding (Shi and Malik (2000)). To proceed, a similarity matrix W is computed using a Gaussian similarity measure from the Euclidean distance (Belkin and Niyogi (2001)) such that:

$$W(\mathbf{x}, \mathbf{y}) = \begin{cases} \exp\left(\frac{\|s(\mathbf{x}) - s(\mathbf{y})\|_2^2}{\sigma^2}\right) , & \text{if } \|\mathbf{x} - \mathbf{y}\|_2 < \epsilon , \\ 0 , & \text{if } \|\mathbf{x} - \mathbf{y}\|_2 > \epsilon . \end{cases} \qquad (26)$$

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 a parameter used to define an ϵ-neighbourhood.

The MRSI spectra projection into the lower dimension space is approached as a generalized eigenvector problem. Subsequently, a replicate k-means clustering method is run defining two clusters. The data corresponding to larger cluster is assumed to belong to the non-prostate voxels and these voxels will be eliminated from the processing. The full procedure is repeated until the total number of voxels left is inferior to a given threshold set experimentally.

3.3. Registration

The role of image registration is vital in CAD systems using multi-parametric MRI images. As it will be discussed in Sect.

Table 5: Classification of the different registration methods used in the CAD systems reviewed. Acronyms: gradient descent (GD), Nelder-Mead (NM).

Study index	Modality registered	Type	Geometric transformation			Similarity measure			Optimizer		
			Rigid	Affine	Elastic	MSE	MI	CMI	GD	L-BFGS-B	NM simplex
[1-2]	T_2-W - DCE	2D	–	✓	–	✓	–	–	–	–	–
[7]	T_2-W - DW	2D	–	✓	✓	–	–	–	–	–	–
[7]	T_2-W - DCE	2D	–	✓	✓	–	✓	–	✓	–	–
[34-35]	T_2-W - DCE	2D	–	✓	–	–	✓	–	–	–	–
[36]	T_2-W - DCE - DW	3D	–	✓	–	–	–	✓	✓	–	–
[38]	T_2-W - DCE	3D	–	✓	–	–	✓	–	–	–	–
[40]	T_2-W - DCE	3D	–	✓	✓	–	✓	–	–	✓	–

Notes:

−: not used or not mentioned.
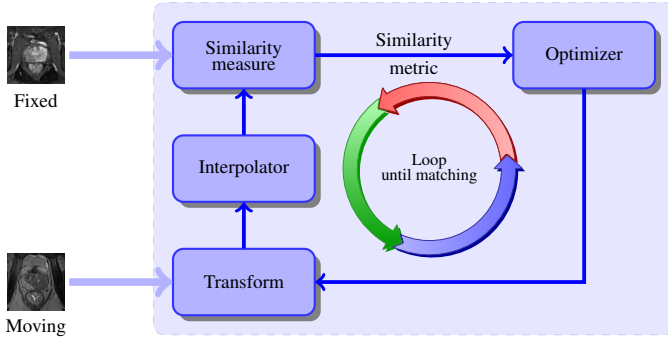
✓ : used or implemented.

Figure 14: Typical framework involved to solve the registration problem.

4, for the sake of an optimal classification, the features detected in each modality will be grouped depending of their spatial locations. Hence, one has to ensure the perfect alignment of the multi-modal MRI images ahead of performing any classification scheme.

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 presented in Fig. 14. 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.

From Sect. 3.3.1 to 3.3.4, we individually review the different components of a typical registration framework (Fig 14). Section 3.3.5 will summarize the combinations of these components especially for the frameworks used in CAD systems. Exhaustive reviews covering all registration methods in computer science and medical fields can be found in Maintz and Viergever (1998) and Zitová and Flusser (2003).

3.3.1. Geometric transformation models

As previously mentioned, the registration problem is to align two images or volumes by finding the geometric transforma-

tion. Regarding the transformation, from all CAD systems reviewed, only parametric methods have been implemented.

Three different groups of parametric transformation models have been used, each of them are characterized by the degree of freedom that they offer.

The first type of transformation is usually referred to as rigid transformation. These transformations are only composed of rotation and translation transforms. Hence, for a 2D space where $\mathbf{x} = (x, y) \in \mathbb{R}^2$ a rigid transformation $\mathcal{T}_R$ is formalized as:

$$
\begin{aligned}
\mathcal{T}_R(\mathbf{x}) &= \begin{bmatrix} R & \mathbf{t} \\ \mathbf{0}^{\mathbf{T}} & 1 \end{bmatrix} \mathbf{x}, \\
&= \begin{bmatrix} \cos\theta & -\sin\theta & t_x \\ \sin\theta & \cos\theta & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix},
\end{aligned} \tag{27}
$$

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 has to be taken into account 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 three elements. Hence, the geometric transformation $\mathcal{T}_R(\cdot)$ is embedded into a matrix of size 4×4.

Affine transformations provide additional degrees of freedom managing rotations and translation as with the rigid transformations but 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:

$$
\begin{aligned}
\mathcal{T}_A(\mathbf{x}) &= \begin{bmatrix} A & \mathbf{t} \\ \mathbf{0}^{\mathbf{T}} & 1 \end{bmatrix} \mathbf{x}, \\
&= \begin{bmatrix} a_{11} & a_{12} & t_x \\ a_{21} & a_{22} & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix},
\end{aligned} \tag{28}
$$

Hence the four parameters $\{a_{11}, a_{12}, a_{21}, a_{22}\}$ of the affine matrix and $\{t_x, t_y\}$ of the translation encode an affine transformation.

Regarding volume registered, the same remark as previously mentioned can be applied in this case. Thus the geometric transformation $\mathcal{T}_A(\cdot)$ is of size 4×4 with nine parameters involved.

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

$$\mathcal{T}_E(\mathbf{x}) = \frac{a_{11}x - a_{12}y + t_x + \sum_i c_i g(\|\mathbf{x} - p_i\|)}{a_{21}x + a_{22}y + t_y + \sum_i c_i g(\|\mathbf{x} - p_i\|)} . \tag{29}$$

where $\mathbf{x}$ are the control points in both images and $g(\cdot)$ 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 where as with TPS, the control points correspond to detected or selected key points. By using TPS, Mitra et al. (2011) obtained more accurate and time efficient results than with the B-splines strategy (Mitra et al. (2012)).

It is reasonable to point out that usually only rigid or affine registrations are used to register multi-parametric images from a same protocol. Elastic registration methods are more commonly used to register multi-protocol images (e.g., histopathology with MRI images) (Toth et al. (2008, 2009)).

3.3.2. Similarity measure

During the registration procedure, a similarity criterion is computed in order to evaluate the quality of the alignment performed. Roughly speaking, this criterion will give the direction to take to the optimizer, in order to assign the most optimal values to the geometric transformation parameters.

The most naive similarity measure 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:

$$\text{MSE} = \frac{1}{N} \sum_x \sum_y (I(x,y) - J(x,y))^2 , \tag{30}$$

where N is the total number of pixels.

However, this metric is not well suited when multi-parametric images are involved due to the tissue appearance variations between the different modalities.

In that regard, MI was introduced as a registration measure in the late 1990's by Pluim et al. (2003). 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 can be achieved by maximizing this common information. Hence, MI of two images A and B is defined as:

$$MI(A; B) = S(A) + S(B) - S(A, B) , \tag{31}$$

where $S(A)$ and $S(B)$ are the marginal entropies and $S(A, B)$ is the joint entropy.

Then, maximizing the MI is equivalent to minimizing the joint entropy. The joint entropy measure is related with the degree of uncertainty or dispersion of the data in the joint histogram of the images A and B. As shown in Fig. 15, the data in the joint histogram will be concentrated in the case of aligned images while will be more randomly distributed in the case of misaligned images. Regarding the computation of the entropies, an estimation of the PDFs have to be carried out. Histogram or Parzen window methods are a common way to estimate these PDFs.

A generalized form of MI, CMI, was proposed by Chappelow et al. (2011). CMI encompasses interdependent information such as texture and gradient 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 can be formulated such as:

$$CMI(\epsilon_n^A; \epsilon_m^B) = S(\epsilon_n^A) + S(\epsilon_m^B) - S(\epsilon_n^A, \epsilon_m^B) . \tag{32}$$

From Eq. (32), it can be seen that CMI is estimated using high dimensional data and that histogram-based methods to estimate the PDFs are not suitable any more (Chappelow et al. (2011)). However, other approaches can be used such as the one employed by Staring et al. (2009) to compute the α-MI (Hero et al. (2002)) which is based on the construction of entropic graphs using k-neareast neighbour (k-NN) inside the high dimensional feature space later used to estimate the MI.

3.3.3. Optimization methods

Registration is usually regarded as an optimization problem where the parameters of the geometric transformation model have to be inferred by minimizing the similarity measure. Iterative estimation methods are commonly used: Nelder-Mead simplex method (Nelder and Mead (1965)), L-BFGS-B quasi-Newton method (Byrd et al. (1995)) and gradient descent (Viola and Wells (1997)). During our review, we noticed that authors do not usually linger over optimizer choice.

3.3.4. 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, commonly used methods are bilinear, nearest-neighbour, bi-cubic, spline and inverse-distance weighting method (Mitra (2012)).

3.3.5. Review of the methods used in CAD system

Studies presenting CAD pipeline incorporating an automatic registration procedure are summarized in Tab. 5.

Ampeliotis et al. (2007, 2008) did not use the framework as presented in Fig. 14 to register 2D T_2-W and DCE images. By using image symmetries and the MSE metric, they find the parameters of an affine transformation but not using a common

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

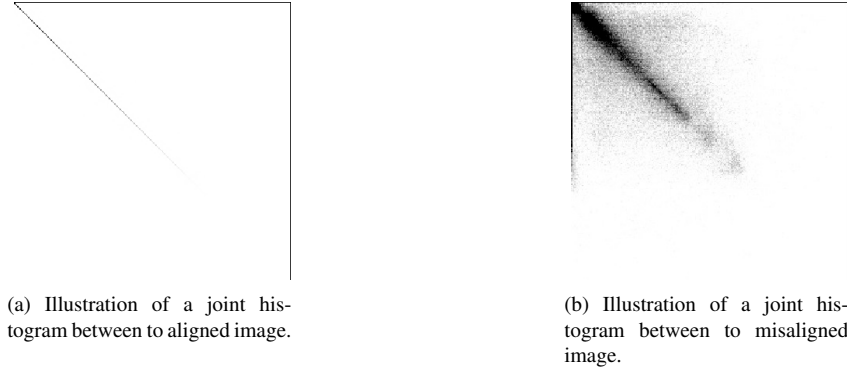(b) Illustration of a joint histogram between to misaligned image.

Figure 15: 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.

objective function. They are finding independently and consecutively the scale factor, the rotation and finally the translation.

Giannini et al. (2013) used also a in-house development registration method to register 2D T_2-W and DW images. To register both images using an affine transformation model, the bladder is first segmented in both modalities in order to obtain its contours.

Giannini et al. (2013) and also Vos et al. (2010) used the same framework (Rueckert et al. (1999)) which is based on finding an affine transformation to register the T_2-W and DCE images using MI. 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 approaches differ regarding the choice of the optimizer since Giannini et al. (2013) used a gradient descent and Vos et al. (2010) tackle the optimization via a quasi-Newton method. Moreover, Giannini et al. (2013) performed a 2D registration whereas Vos et al. (2010) registered 3D volumes.

Viswanath et al. (2008a, 2009) as well as Vos et al. (2008a) performed an affine registration using the MI as similarity measure to correct the misalignment between T_2-W and DCE images. The choice of the optimizer was not specified. Viswanath et al. (2008a, 2009) focused on 2D registration while Vos et al. (2008a) performed 3D registration.

Finally, Viswanath et al. (2011) performed a 3D registration with the three modalities, T_2-W and DCE and DW MRI, by using an affine transformation model combined with the CMI similarity measure as presented in Chappelow et al. (2011). Moreover, Chappelow et al. (2011) employed gradient descent to solve this problem but suggested Nelder-Mead simplex and quasi-Newton method as other solutions.

4. Image classification framework

4.1. CADe: ROIs detection/selection

As discussed in the introduction and shown in Fig. 2, the image classification framework is composed of eventually a CADe and a CADx. In this section, we will focus on studies embedding a CADe in their framework. Two approaches are considered to define a CADe (see Tab. 6): (i) voxel-based delineation and (ii) lesion segmentation.

Table 6: Overview of the CADe strategies employed in CAD systems.

CADe: ROIs selection strategy	References
All voxels-based approach	[4-5,7-8,12,14,16-17,20-22,25-37]
Lesions candidate detection	[10-11,41]

The first strategy, which concerns the majority of the studies reviewed (see Tab. 6), is in fact linked to the nature of the classification framework. All voxels are considered as a possible threat and the output of the framework will be the lesions classified as cancerous.

The secondary 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 was borrowed from methodologies applied successfully in CADe for breast cancers.

Vos et al. (2012) highlighted lesion candidates by detecting blobs in the ADC map. These candidates were filter using some *a priori* criteria such as SI or diameter. As mentioned in Sect. 2.3.2 (see also Tab. 1), CaP can be interpreted as region of lower SI in ADC map. 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 as proposed by Li et al. (2003). The probability p of a voxel $\mathbf{x}$ being a part of a blob at the scale σ is given by:

$$P(\mathbf{x}, \sigma) = \begin{cases} \frac{\|\lambda_{\sigma,3}(\mathbf{x})\|^2}{\|\lambda_{\sigma,1}(\mathbf{x})\|} , & \text{if } \lambda_{\sigma,k}(\mathbf{x}) > 0 \text{ with } k = \{1, 2, 3\} , \\ 0 , & \text{otherwise} . \end{cases} \quad (33)$$

The fusion of the different scales is computed as:

$$L(\mathbf{x}) = \max P(\mathbf{x}, \sigma), \forall \sigma . \quad (34)$$

The resulting map $L(\mathbf{x})$ is then filtered depending on its appearance (cf. maximum of the likelihood of the region, diam-

eter of the lesion) and their SI in ADC and T_2-W images. The detected regions are then used as inputs for the CADx.

Litjens et al. (2011) used a pattern recognition approach in order to delineate the ROIs. A blobness map was calculated in the same manner as previously mentioned using the multi-resolution Hessian blob detector on the ADC map, T_2-W and pharmacokinetic parameters maps (see Sect. 4.2 for details about those parameters). Additionally, the position of the voxel $\mathbf{x} = \{x, y, z\}$ was used as a feature as well as the Euclidean distance of the voxel to the prostate center. Hence, the feature vectors were composed of eight features and a support vector machines (SVM) classifier was trained using a radial basis function (RBF) kernel (see Sect. 4.4 for more details).

Subsequently, Litjens et al. (2012b) 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. Two new maps were introduced based on texture. Instead of a SVM classifier, a k-NN classifier was used. The candidate regions were then extracted by performing a local maxima detection followed by post-processing region-growing and morphological operations.

4.2. Feature detection

Discriminative features which can be used to recognize CaP from healthy tissue have to be first detected. This section will summarize the different strategies employed for this task. The feature type used is summarized in Tab. 7 while Tab. 8 sums up which strategies were used by the different studies reviewed.

4.2.1. Image-based features

This section will focus on image-based features detection. Two main strategies to detect features have been identified and used for the purpose of our classification: (i) voxel-wise detection and (ii) region-wise detection.

– **Voxel-wise detection:** This strategy refers to the fact that a feature is extracted at each voxel location of an image.

CaP as previously discussed (see Tab. 1) can be discerned due to SI changes. Hence, intensity-based features are one of the most common features used to build the feature vector which has to be classified (Ampeliotis et al. (2007, 2008), Artan et al. (2010, 2009), Chan et al. (2003), Langer et al. (2009), Liu et al. (2009), Niaf et al. (2012, 2011), Viswanath et al. (2011, 2008a)). This type of feature consists simply of the SI of each voxel of the different MRI modalities.

SI changes can be viewed as heterogeneous regions and edge-based features are used in that regard. Each feature is computed by convolving the original image with an edge operator. Three of these operators are used in CAD systems: (i) Prewitt operator (Prewitt (1970)), (ii) Sobel operator (Sobel (1970)) and (iii) Kirsch operator (Kirsch (1971)). Results obtained with these operators vary, due to their different kernels.

These features are commonly incorporated in the feature vector for further classification in the CAD systems reviewed (Niaf et al. (2012, 2011), Tiwari et al. (2013, 2010, 2009b), Viswanath et al. (2011, 2008b)).

Table 7: Overview of the feature detection methods used in CAD systems.

Feature detection methods	Indexes
MRI image:	
Voxel-wise detection	
Intensity-based	A
Edge-based	
Prewitt operator	B_1
Sobel operator	B_2
Kirsch operator	B_3
Gabor filtering	B_4
Texture-based	
Haralick features	C_1
Fractal analysis	C_2
Discrete cosine transform (DCT)	C_3
Wavelet-based features	C_4
Position-based	D
Region-wise detection	
Statistical-based	
Percentiles	E_1
Statistical-moments	E_2
Histogram-based	
PDF	F_1
Histogram of oriented gradient (HOG)	F_2
Shape context	F_3
LBP	F_4
Anatomical-based	G
DCE signal:	
Whole spectra approach	H
Semi-quantitative approach	I
Quantitative approach	
Toft model	J_1
Brix model	J_2
Weibull function	J_3
Phenomenological universalities model	J_4
MRSI signal:	
Whole spectra approach	K
Quantification approach	L
Wavelet-based approach	M

(a) $\theta = 0°$.   (b) $\theta = 60°$.   (c) $\theta = 120°$.   (d) $\theta = 180°$.
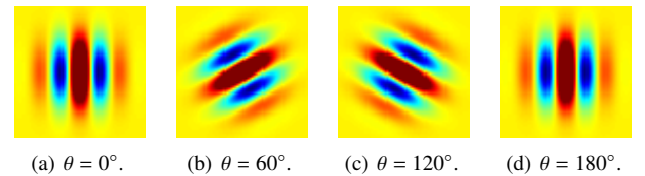
Figure 16: Illustration of four different Gabor filters varying their orientations θ.

Table 8: Overview of the different features extracted by studies reviewed. The indexes of the features are presented in Tab. 7.

	A	B₁	B₂	B₃	B₄	C₁	C₂	C₃	C₄	D	E₁	E₂	F₁	F₂	F₃	F₄	G	H	I	J₁	J₂	J₃	J₄	K	L	M
[1]	△––											△––						✓								
[2]	△––											△––						✓								
[3]						△––					––◦	––◦														
[4]	△–◦																				✓!					
[5]	△–◦																				✓!					
[6]	△–◦							△–◦		✓																
[7]	––◦																			✓		✓	✓			
[8]																								✓	✓	
[9]	△–◦																	✓								
[10]	△–◦									✓	△•◦	△•◦						✓								
[11]	△–◦					△•◦				✓	△•◦	△•◦					✓	✓								
[12]	△–◦																			✓						
[13]													△•◦	△•◦	△•◦	△•◦				✓!						
[14]							△––																			
[15]							△––																			
[16]																	✓							✓		
[17]																			✓	✓		✓	✓			
[18]	△•◦		△•◦	△•◦		△•◦					△•◦	△•◦						✓	✓							
[19]	△•◦		△•◦	△•◦		△•◦					△•◦	△•◦						✓	✓							
[20]	△–◦																			✓!						
[21]	△–◦																			✓!						
[22]																									✓	✓
[23]											––◦	△–◦								✓!						
[24]																		✓!								
[25]																		✓		✓						
[26]																								✓		
[27]																								✓		
[28]																								✓		
[29]		△––	△––	△––		△––						△––												✓		
[30]		△––	△––	△––		△––						△––												✓		
[31]					△––																					✓
[32]		△––	△––	△––		△––						△––												✓		
[33]		△––	△––	△––	△––	△––						△––												✓		
[34]																		✓								
[35]			△––	△––		△––						△––						✓								
[36]			△––	△––		△•–						△•–						✓								
[37]			△––	△––	△––	△––			△––			△––														
[38]	△––																			✓						
[39]											–•–									✓						
[40]											△•–									✓						
[41]											△•◦									✓						

Notes:
△: implemented for T₂-W-MRI images.
•: implemented for DCE-MRI images.
◦: implemented for DW-MRI images.
–: not used or not implemented.
✓: used or implemented.
✓!: partially implemented.

23

Gabor filters offer another approach to extract information related to edges and texture and were found to perform similar processing to that in the human visual system (Daugman (1985), Gabor (1946)). A Gabor filter is defined by the modulation of a Gaussian function with a sinusoid which can be further rotated. Hence, a Gabor filter g can be formalized as:

$$g(x, y; \theta, \psi, \sigma, \gamma) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \times \cos\left(2\pi \frac{x'}{\lambda} + \phi\right), \quad (35)$$

with

$$x' = s(x\cos\theta + y\sin\theta),$$
$$y' = s(-x\sin\theta + y\cos\theta),$$

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.

To perform Gabor analysis to extract features for a classification scheme, a bank of Gabor filters is usually created with different angles, scale and dilatations (see Fig. 16) and then convolved with the image. Viswanath et al. (2008a, 2012) and Tiwari et al. (2012) integrated Gabor analysis in their feature vector.

Texture-based features provide other characteristics discerning CaP from healthy tissue. The most common texture analysis for image classification was proposed by Haralick et al. (1973) and is commonly used in CAD systems (Antic et al. (2013), Niaf et al. (2012, 2011), Tiwari et al. (2013, 2010, 2009b), Viswanath et al. (2011, 2008a,b, 2012)). At each voxel, a neighbourhood is defined around this center and a gray-level co-occurrence matrix is built by selecting a pair of voxels based on a defined distance and angle. Then, using this co-occurrence matrix, a set of features can be computed based on the statistics describing the texture around each voxel. Computation of these features is presented in Tab. 9.

Fractal analysis and more precisely a local estimation of the fractal dimension (Benassi et al. (1998)) describing the texture roughness at a specific location was used in Lopes et al. (2011). A wavelet-based method in a multi-resolution framework was used to estimate the fractal dimension. Cancerous tissue were characterized to have a higher fractal dimension than healthy tissue.

Chan et al. (2003) aimed to describe the texture using the frequency signature via the discrete cosine transform (DCT) (Ahmed et al. (1974)) defining a neighbourhood of 7×7 pixels for each of the modalities that they used.. The DCT allows to decompose a portion of image into a coefficients space where few of these coefficients encoded the visually significant information. The DCT coefficients are computed such as:

Table 9: The fourteen statistical features for texture analysis commonly computed from the gray level co-occurrence matrix p as presented by Haralick et al. (1973).

Statistical features	Formula
Angular second moment	$\sum_i \sum_j p(i, j)^2$.
Contrast	$\sum_{n=0}^{N_g-1} n^2 \{\sum_{i=1}^{N_g-1} \sum_{j=1}^{N_g-1} p(i, j)\}, |i - j| = n$.
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	$\sqrt{(1 - \exp[-2(H_2(X;Y) - H(X;Y))])}$.
Max. corr. coeff.	$\sqrt{\lambda_2}$, of $Q(i, j) = \sum_k \frac{p(i,k)p(j,k)}{p_x(i)p_y(k)}$.

$$C_{k_1,k_2} = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} p_{m,n} \cos\left[\frac{\pi}{M}\left(m + \frac{1}{2}\right)k_1\right] \cos\left[\frac{\pi}{N}\left(n + \frac{1}{2}\right)k_2\right], \quad (36)$$

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 $p_{m,n}$.

In the same spirit, Viswanath et al. (2012) projected T$_2$-W images into the wavelet space and used the coefficients obtained from the decomposition as features. The wavelet family used for the decomposition was the Haar wavelet.

The position of a voxel within the prostate was also considered a possible feature. Litjens et al. (2011) computed the Euclidean distance from each voxel to the prostate center as well as the individual distance in the three directions x, y and z. Chan et al. (2003) embedded the same information but this time using cylindrical coordinate r, θ and z corresponding to the radius, azimuth and elevation respectively.

– Region-wise detection:
Unlike the previous section, another strategy is to study an entire region and extract characteristic features corresponding to this region.

The most common approach reviewed can be classified as statical methods. Full map corresponding to feature types presented in the previous section are computed. Then, ROIs are defined and statistics are extracted from each of these regions. The first type of statistic is based on percentiles and is widely used (Antic et al. (2013), Litjens et al. (2012b,

2011), Peng et al. (2013), Tiwari et al. (2013, 2010, 2009b), Viswanath et al. (2011, 2008a,b, 2012), Vos et al. (2012, 2008a, 2010, 2008b)). In fact, once that a ROI is defined, the features corresponding the nth percentile is used as feature. This n can take any value between 0 and 100. This threshold is usually manually determined observing the distribution and corresponds to the best discriminant value differentiating malignant and healthy tissue. in addition, statistic-moments such as mean, standard deviation, kurtosis and skewness are also used (Ampeliotis et al. (2007, 2008), Antic et al. (2013), Niaf et al. (2012, 2011), Peng et al. (2013)).

Another subset of features are anatomic which were also used by Litjens et al. (2012b) and Matulewicz et al. (2013). Litjens et al. (2012b) computed the volume, compactness and sphericity related to the region to integrate it in their feature vector to later classify. Matulewicz et al. (2013) introduced four features corresponding to the percentage of tissue belonging to the regionsPZ, CG, periurethral region or outside prostate region for the considered ROI.

In contrast to anatomical are histogram-based features. For instance, Liu et al. (2013) introduced four different types of histogram-based features. The first type corresponds to the histogram of the SI of the image. The second type is the histogram of oriented gradient (HOG) (Dalal and Triggs (2005)). HOG descriptor describes the local shape of the object of interest by using gradient directions distribution. 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 will vote for a bin of the histogram and this vote is weighted by the magnitude of the same gradient. Finally, The cells are grouped into block and each block is normalized. The third histogram-based type used by Liu et al. (2013) was shape context (Belongie et al. (2002)). 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 of Zhao et al. (2012) which is using the Fourier transform of the histogram created via local binary pattern (LBP) (Ojala et al. (1996)). LBP is generated by comparing the value of the central pixel with its 8-connected neighbours. Then, in the ROI, the histogram of the LBP distribution is computed. The discrete fourier transform (DFT) of the LBP histogram is ised to make the feature invariant to rotation.

The last group of region-based feature is based on fractal analysis. The features proposed are based on estimating the fractal dimension which is a statistical index representing the complexity of what is analysed. Lv et al. (2009) proposed two features based on fractal dimension: (i) texture fractal dimension and (ii) histogram fractal dimension. The first feature is based on estimating the fractal dimension on the SI of each image. Hence, this feature is a statistical characteristic of the image roughness. The second fractal di-
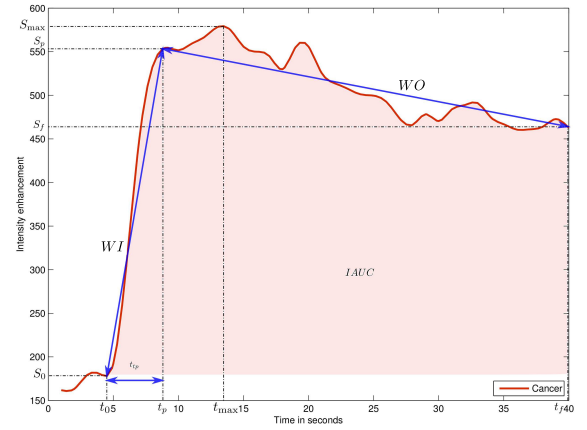

Figure 17: Graphical representation of the different semi-quantitative features used for DCE-MRI analysis.

mension is estimated in the PDF of each image and is characteristic of the complexity of the PDF. Lopes et al. (2011) proposed a 3D version to estimate the fractal dimension of a volume using wavelet decomposition.

4.2.2. DCE-based features

DCE-MRI is more commonly based on a SI analysis over time as presented in Sect. 2.3.2. The parameters extracted used in CAD system during the DCE-MRI analysis are presented

– **Whole-spectra approach:** Some studies are using the whole DCE time series as feature vector such as Ampeliotis et al. (2007, 2008), Tiwari et al. (2012) and Viswanath et al. (2008a,b). In some cases, the high-dimensional feature space is reduced using dimension reduction methods as it will be presented in the next section (see Sect. 4.3).

– **Semi-quantitative approach:** Semi-quantitative approaches are based on modelling mathematically the DCE time series. The parameters modelling the signal are commonly used mainly due to the simplicity of their computation. Parameters included in semi-quantitative analysis are summarized in Tab. 10 and also graphically depicted in Fig. 17. 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. 2.3.2, quantitative approaches correspond to mathematical-pharmacokinetic models based on physiological exchanges. Four different models have been used in CAD systems.

The most common model reviewed was the Brix model (Artan et al. (2010, 2009), Liu et al. (2009), Ozer et al. (2009, 2010), Sung et al. (2011)). This model is formalized such as:

$$\frac{S(t)}{S(0)} = 1 + Ak_{ep}\left(\frac{\exp(-k_{ep}t) - \exp(-k_{el}t)}{k_{el} - k_{ep}}\right), \qquad (37)$$

25

Table 10: 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

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 and the parameters k_{ep}, k_{el} and A are computed from the MRI data and used as features.

The Tofts model (Tofts (1997)) was used by Giannini et al. (2013), Langer et al. (2009), Mazzetti et al. (2011), Niaf et al. (2012, 2011). The DCE signal relative to the concentration is modelled as:

$$C_t(t) = v_p C_p(t) + K_{trans} \int_0^t C_p(\tau) \exp(-k_{ep}(t - \tau))\, d\tau , \quad (38)$$

where $C_t(\cdot)$ is the concentration of the medium, $C_p(\cdot)$ is the AIF which have 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 $v_e = 1 - v_p$. Thus, the parameters K_{trans}, k_{ep} and v_e are computed and used as features in this model.

Mazzetti et al. (2011) and Giannini et al. (2013) used the Weibull function empirically formalized as:

$$S(t) = At \exp(-t^B) , \quad (39)$$

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 model (Castorina et al. (2006)) formalized as:

$$S(t) = \exp\left[rt + \frac{1}{\beta}a_0 - r\left(\exp(\beta t) - 1\right) \right] , \quad (40)$$

where the parameters β, a_0 and r are inferred.

For all these models, the parameters are inferred using an optimization curve fitting approach.

4.2.3. MRSI-based features

– **Whole spectra approach:** As in the case of DCE analysis, one common approach is to incorporate the whole MRSI spectra in the feature vector for classification (Kelm et al. (2007), Matulewicz et al. (2013), Parfait et al. (2012), Tiwari et al. (2013, 2010, 2007, 2009a,b), Viswanath et al. (2008a)). Sometimes post-processing involving dimension reduction methods is performed to reduce the complexity during the classification as it will be presented in Sect. 4.3.

– **Quantification approach:** We can reiterate that in MRSI only few biological markers (cf., choline, creatine and citrate metabolites mainly) are known to be useful to discriminate CaP and healthy tissue. Then, concentrations of these metabolites can be considered as a feature used for classification. In order to perform this quantification, four different approaches have been used. The QUEST (Ratiney et al. (2005)), AMARES (Vanhamme et al. (1997)) and VARPRO (Coleman and Li (1993)) models were used by Kelm et al. (2007). 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. (2012) used the LcModel approach (Provencher (1993)) which solves the optimization problem in the frequency domain.

Once the different concentrations are computed, Kelm et al. (2007) calculated the relative concentrations (Eq. (41) and (42)) and used them as features. However, Parfait et al. (2012) used each metabolite concentrations individually.

$$R_1 = \frac{[\text{Cho}] + [\text{Cr}]}{[\text{Cit}]} . \quad (41)$$

$$R_2 = \frac{[\text{Cit}]}{[\text{Cho}] + [\text{Cr}] + [\text{Cit}]} , \quad (42)$$

26

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

- **Wavelet decomposition approach:** Tiwari et al. (2012) performed a wavelet packet decomposition (Coifman and Wickerhauser (1992)) with the Haar wavelet basis function and used the coefficients of this decomposition as features for further classification.

4.3. Feature selection and feature extraction

As presented in the previous section, a wide variety of features can be computed (see Tab. 7). This often leads from multi-parametric MRI data to a high complexity feature space which might mislead or corrupt the classifier which used for training. Thus, 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. Those methods used in CAD system are summarized in Tab. 11.

4.3.1. 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 the studies reviewed will be briefly presented. More extensive reviews specific to feature selection can be found in Saeys et al. (2007).

Niaf et al. (2012, 2011) make use of the p-value by using the independent two-sample t-test with equal mean for each feature dimension. In this statistical test, it is that there are two classes: CaP and healthy tissue. Hence, for each particular feature, the distribution of each class can be characterized by their means $\bar{X}_1$ and $\bar{X}_2$ and standard deviation s_{X_1} and s_{X_2}, respectively. Hence, the null hypothesis test is based on the fact that these both distribution means are equal. Thus, the t-statistic used to verify the null hypothesis is then formalized such that:

$$
\begin{aligned}
t &= \frac{\bar{X}_1 - \bar{X}_2}{s_{X_1 X_2} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} , \\
s_{X_1 X_2} &= \sqrt{\frac{(n_1 - 1)s_{X_1}^2 + (n_2 - 1)s_{X_2}^2}{n_1 + n_2 - 2}} ,
\end{aligned} \tag{43}
$$

where n_1 and n_2 are the number of samples in each class.

From Eq. (43), it can be seen that 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 can be deduced from the t-test and corresponds to the probability of obtaining such an extreme test assuming that the null hypothesis is true (Goodman (1999)). Hence, smaller the p-value, the more likely we are to reject the null hypothesis and more relevant the feature is likely to be.

Finally, the features can be ranked and the most significant features can be selected. However, this technique suffers from

Table 11: Overview of the feature selection and extraction methods used in CAD systems.

Dimension reduction methods	References
Feature selection:	
Statistical test	[17-18,41]
MI-based methods	[18-19,37]
Feature extraction:	
Linear mapping	
PCA	[27-28,31]
Non-linear mapping	
Laplacian eigenmaps	[26,28-30,33,36]
LLE and LLE-based	[27-28,33-34]

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. (2012) 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:

$$
F_r = \frac{\bar{X}_1 - \bar{X}_2}{s_{X_1}^2 + s_{X_2}^2} . \tag{44}
$$

Hence, a feature dimension can be seen as more relevant when the interclass variance is maximum and the intraclass variance in minimum. Once the features are ordered, Vos et al. (2012) select the feature dimensions with the larger Fisher discriminant ratio.

MI can also be used to select a subset of feature dimensions. Definition of the MI was presented in Sect. 3.3.2 and formalized in Eq. (31) and as previously mentioned, the computation of the entropies involves the estimation of some PDFs and the data being usually continuous variables, it is then necessary to estimate the PDFs using a method such as Parzen windows.

Peng et al. (2005) introduced two main criteria to select the feature dimensions: (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 can be formalized:

$$
\arg\max Rel(\mathbf{x}, c) = \frac{1}{|\mathbf{x}|} \sum_{x_i \in \mathbf{x}} MI(x_i, c) , \tag{45}
$$

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 will imply independence between each feature dimension which is usually not true.

Minimal redundancy criterion will force selection of a new feature dimension which shares as little as possible MI previously selected feature dimension. It can be formalized as:

$$\arg\min Red(\mathbf{x}) = \frac{1}{|\mathbf{x}|^2} \sum_{x_i, x_j \in \mathbf{x}} MI(x_i, x_j) \,. \qquad (46)$$

Combination of these two criteria is known as minimum redundancy maximum relevance (mRMR)[5] (Peng et al. (2005)) and are computed as a difference or quotient of Eqs. (45) and (46).

Niaf et al. (2012, 2011) make use of maximal relevance criterion alone and also of both mRMR difference and quotient criterion. Viswanath et al. (2012) also reduced their feature vector via mRMR difference and quotient.

4.3.2. 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 created to maximize the separability between the classes. The mapping can be performed in a linear or a non-linear manner. Again, only methods employed in CAD system will be reviewed in this section. We refer the reader to the review of Fodor (2002) 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 will lie on the kth principal components (Jolliffe (2002)).

The principal components can then be computed by using the eigenvectors-eigenvalues decomposition of the covariance matrix. Let $\mathbf{x}$ denote the data matrix. Then the covariance matrix is defined as:

$$\Sigma = \mathbf{x}^{\mathrm{T}} \mathbf{x} \,. \qquad (47)$$

The eigenvectors-eigenvalues decomposition can be formalized as:

$$\mathbf{v}^{-1} \Sigma \mathbf{v} = \Lambda \,, \qquad (48)$$

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 selecting the largest eigenvalues. The total variation that is the sum of the principal eigenvalues of the covariance matrix (Fodor (2002)), usually corresponds to the 95% to 98% of the cumulative sum of the eigenvalues. Tiwari et al. (2008, 2009a, 2012) used PCA in order to reduce the dimensionality of their feature vector.

Non-linear mapping was also used for dimension reduction. It is mainly based on Laplacian eigenmaps and locally linear

embedding (LLE) methods. Laplacian eigenmaps[6], 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 (Belkin and Niyogi (2001), Shi and Malik (2000)). Thus, two adjacent data points in the high-dimensional space should also be close in the low-dimensional space. Similarly, two far away data points in the high-dimensional space also should be distant in the low-dimensional space. To compute this projection, an adjacency matrix is defined as:

$$W(i, j) = \exp \|\mathbf{x}_i - \mathbf{x}_j\|_2 \,, \qquad (49)$$

where $\mathbf{x}_i$ and $\mathbf{x}_j$ are the two samples considered.

Then, the low-dimensional space will be found by solving the generalized eigenvectors-eigenvalues problem:

$$(D - W)\mathbf{y} = \lambda D \mathbf{y} \,, \qquad (50)$$

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 (Belkin and Niyogi (2001)). Tiwari et al. (2007, 2009a,b), Viswanath et al. (2008b) 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. (2013, 2009a) used a slightly different approach by combining the Laplacian eigenmaps techniques with a prior multi-kernel learning strategy. First, multiple features were extracted for multiple modalities. The features of a single modality were then mapped to a higher dimensional space via the Kernel trick (Aizerman et al. (1964)) and more precisely using a Gaussian kernel. Then, each kernel associated with each modality was linearly combined to obtain a combined kernel K. Then, the computation of the adjacency matrix W took place and the same scheme as in Laplacian eigenmaps is applied. However, in order to use the combined kernel, Eq. (50) is rewritten as:

$$K(D - W)K^{\mathrm{T}}\mathbf{y} = \lambda KDK^{\mathrm{T}}\mathbf{y} \,. \qquad (51)$$

It can be solved as a generalized eigenvectors-eigenvalues problem as previously. Viswanath et al. (2011) used Laplacian eigenmaps inside a bagging framework in which multiple embeddings are generated by successively selecting feature dimensions.

LLE[7] is another common non-linear dimension reduction technique widely used, first proposed by Roweis and Saul (2000). LLE is based on the fact that a data point in the feature space can be characterized by its neighbours. Thus, it was proposed to represent each data point in the high-dimensional

---

[5]mRMR implementation can be found at: `http://penglab.janelia.org/proj/mRMR/`

[6]Laplacian eigenmap implementation is available at: `http://www.cse.ohio-state.edu/~mbelkin/algorithms/algorithms.html`

[7]LLE implementation is available at: `http://www.cs.nyu.edu/~roweis/lle/code.html`

space as the linear combination of its *k*-nearest neighbours. This can be expressed as:

$$\hat{\mathbf{x}}_i = \sum_j W(i, j)\mathbf{x}_j \,, \qquad (52)$$

where $\hat{\mathbf{x}}_i$ are the data point estimated using its neighbouring data points $\mathbf{x}_j$.

Hence, this problem which has to be solved at this stage is to estimate the weight matrix W. This problem can be tackled using a least square optimization scheme by optimizing the following objective function:

$$\hat{W} = \arg\min_W \sum_i |\mathbf{x}_i - \sum_j W(i, j)\mathbf{x}_j|^2 \,, \qquad (53)$$
$$\text{subject to } \sum_j W(i, j) = 1 \,,$$

Then, the essence of LLE is to project the data into a low dimension space, while retaining the data organization. Thus, the projection into the low dimension space can be seen as an optimization problem as:

$$\hat{\mathbf{y}} = \arg\min_{\mathbf{y}} \sum_i |\mathbf{y}_i - \sum_j W(i, j)\mathbf{y}_j|^2 \,. \qquad (54)$$

This optimization can be performed as an eigenvectors-eigenvalues problem by finding the kth eigenvectors corresponding to the kth smallest eigenvalues of the sparse matrix $(I - W)^{\text{T}}(I - W)$. Tiwari et al. (2008, 2009a), Viswanath et al. (2008a,b) used LLE as a dimension reduction technique to reduce the complexity of their feature vector.

Tiwari et al. (2008) used a modified version of the LLE algorithm in which they applied LLE in a bagging approach with multiple neighbourhood sizes. The different embeddings obtained are then fused using the maximum likelihood (ML) estimation.

4.4. Classification

4.4.1. Classifier

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. Classification methods used in CAD system to distinguish these two classes are summarized in Tab. 12. A full review of classification methods used in pattern recognition can be found in Bishop (2006).

– ***Rule-based method:***

Lv et al. (2009) make use of a decision stump classifier to distinguish CaP and healthy classes.

Puech et al. (2009) detect CaP by implementing a given set of rules using a score medical decision making approach. The feature values are compared with a pre-defined threshold. Then, at each comparison, the final score is incremented or not, depending on the threshold and the final decision is taken depending of the final score.

Table 12: Overview of the classifiers used in CAD systems.

Classifier	References
Rule-based method:	[15,24]
Clustering methods:	
k-means clustering	[26-28,33-34]
k-NN	[11,18-19]
Linear model classifiers:	
LDA	[3,6,18-19,41]
Logistic regression	[8-9]
Non-linear classifier:	
QDA	[37]
Probabilistic classifier:	
Naive Bayes	[7,17-19]
Ensemble learning classifiers:	
AdaBoost	[14]
Random forest	[8,31-32,35]
Probabilistic boosting tree	[29-31,36]
Kernel method:	
Gaussian processes	[8]
Sparse kernel methods:	
SVM	[4-6,8,10-11,13-14,18-23,25,31,38-40]
RVM	[20-21]
Neural network:	
Multiple layer perceptron	[16,22]
Probabilistic neural network	[1-2,36]
Graphical model classifiers:	
Markov random field	[12,21]
Conditional random field	[4-5]

– *Clustering methods:*

k-neareast 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 was one of the method used by Niaf et al. (2012, 2011) mainly to make a comparison with different machine learning techniques. Litjens et al. (2012b) used this method to roughly detect potential CaP voxels before performing a region-based classification.

k-means is an unsupervised clustering method in which the data have to be partitioned into *k* clusters. The discovery of the clusters is an iterative procedure. 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 data points 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. (2007, 2009a) used *k*-means in an iterative procedure. Three clusters were defined corresponding to CaP, healthy and non-prostate, respectively. *k*-means was applied iteratively and the voxels corresponding to the largest cluster were excluded under the assumption that it is assigned to "non-prostate" cluster. The iteration stop when the number of voxels composing all clusters remaining is smaller than a threshold.

Tiwari et al. (2008), Viswanath et al. (2008a,b) used *k*-means in a repetitive manner to be less sensitive to the centroids initialisation. Thus, *k* clusters were generated *T* times. The final assignment was performed by majority voting using a co-association matrix as proposed by Fred and Jain (2005).

– *Linear model classifiers:*

Linear discriminant analysis (LDA) can be used as a classification method in which the optimal linear separation between two classes is found by maximizing the interclass variance and minimizing the intraclass variance (Friedman (1989)). The linear discriminant function is defined as:

$$\delta_k(\mathbf{x}_i) = \mathbf{x}_i^{\mathrm{T}} \Sigma^{-1} \mu_k - \frac{1}{2} \mu_k^{\mathrm{T}} \Sigma^{-1} \mu_k + \log(\pi_k) , \qquad (55)$$

where $\mathbf{x}_i$ is an unlabelled feature vector, 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 $\mathbf{x}_i$ will be assigned to the class which maximizes the discriminant function:

$$C(\mathbf{x}_i) = \arg\max_k \delta_k(\mathbf{x}_i) . \qquad (56)$$

Antic et al. (2013), Chan et al. (2003), Niaf et al. (2012, 2011), Vos et al. (2012) used LDA to classify their feature vectors defining two classes CaP *versus* healthy.

Logistic regression can be used to perform binary classification and can provide the probability of an observation to belong to a class. The posterior probability of one of the class c_1 can be written as:

$$p(c_1|\mathbf{x}_i) = \frac{1}{1 + \exp(-\mathbf{w}^{\mathrm{T}}\mathbf{x}_i)} , \qquad (57)$$

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 $\mathbf{x}_i$ will be assigned to the class which maximizes the posterior probability:

$$C(\mathbf{x}_i) = \arg\max_k p(C = k|\mathbf{x}_i) . \qquad (58)$$

From Eq. (57), 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 in the training set. This vector of parameters $\mathbf{w}$ can be inferred by finding the maximum likelihood estimates. This step can be performed through an optimization scheme, using a quasi-Newton method (Byrd et al. (1995)), which iteratively seeks for the local minimum in the derivative of Eq. (57).

Kelm et al. (2007), Puech et al. (2009) used a logistic regression to create a linear probabilistic model in order to classify their feature vectors.

– *Non-linear model classifier:*

Viswanath et al. (2012) used the most general quadratic discriminant analysis (QDA) instead of LDA. Unlike in LDA in which one assumes that the class covariance matrix Σ is identical for all the classes, in QDA, a covariance matrix Σ_k specific to each class is computed. Thus, Eq. (55) becomes:

$$\delta_k(\mathbf{x}_i) = \mathbf{x}_i^{\mathrm{T}} \Sigma_k^{-1} \mu_k - \frac{1}{2} \mu_k^{\mathrm{T}} \Sigma_k^{-1} \mu_k + \log(\pi_k) . \qquad (59)$$

The classification scheme in the case of the QDA is identical to Eq. (56).

– *Probabilistic classifier:*

The most commonly used classifier is the naive Bayes classifier which is a probabilistic classifier assuming independence between each feature dimension (Rish (2001)). This classifier is based on Bayes' theorem:

$$p(C = k|\mathbf{x}) = \frac{p(C)p(\mathbf{x}|C)}{p(\mathbf{x})} , \qquad (60)$$

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 will be classified to the class which maximizes the posterior probability as:

$$C(\mathbf{x}_i) = \arg\max_k p(C = k | \mathbf{x}_i) , \qquad (61)$$

$$p(C = k | \mathbf{x}_i) = p(C = k) \prod_{j=1}^{n} p(x_{ij}, | C = k) , \qquad (62)$$

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 Normal distribution to characterize each class. Thus, each class will be characterized by two parameters: (i) the mean and (ii) the standard deviation. These parameters can be inferred from the training set by using the ML approach.

Giannini et al. (2013), Mazzetti et al. (2011), Niaf et al. (2012, 2011) used the naive Bayes classifier to classify their feature vectors as either malignant or healthy. The Normal distribution was used as the likelihood probability for that model.

– **Ensemble learning classifiers:**

AdaBoost is an adaptive method based on an ensemble learning method and was initially proposed by Freund and Schapire (1997). AdaBoost linearly combines several weak learners resulting into a final strong classifier. A weak learner is defined as a classification method performing slightly better than random classification. Popular choices regarding the weak learner classifiers are: decision stump, decision tree learners (cf., iterative dichotomiser 3 (ID3) (Quinlan (1986)), C4.5 (Quinlan (1993)), classification and regression tree (CART) (Breiman et al. (1984))).

AdaBoost 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:

$$\alpha_t = \frac{1}{2} \ln \frac{1 - \epsilon_t}{\epsilon_t} , \qquad (63)$$

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 samples misclassified by h_t will be increased and the weights of well classified samples will decrease as shown in Eq. (64).

$$D_{t+1}(i) = \frac{D_t(i) \exp\left(-\alpha_t y_i h_t(\mathbf{x}_i)\right)}{Z_t} , \qquad (64)$$

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 us to select a weak learner at the next iteration $t + 1$ which will classify in priority 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:

$$C(\mathbf{x}_i) = \text{sign}\left(\sum_{t=1}^{T} \alpha_t h_t(\mathbf{x}_i)\right) . \qquad (65)$$

Lopes et al. (2011) make use of the AdaBoost classifier to perform their classification.

Random forest[8] is a classification method which is based on creating an ensemble of decision trees and was introduced by Breiman (2001). In the learning stage, multiple decision tree learners (Breiman et al. (1984)) will be trained. However, each decision tree will be trained with 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 (Efron (1979)). 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 (see Sect. 3.3.2). 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 will assign a class to this sample. Finally, it is common to use a majority voting approach to choose the final class label.

Kelm et al. (2007), Tiwari et al. (2013, 2012), Viswanath et al. (2009) make use of the random forest classifier to classify their feature vector.

Probabilistic boosting-tree is another ensemble learning classifier which shares principles with AdaBoost but using them inside a decision tree (Tu (2005)). In the training stage, the probabilistic boosting-tree method grows a decision tree and at each node, a strong classifier is learnt in an almost comparable scheme to AdaBoost (see Eq. 65). Once the strong learner is trained, the training set will be split into two subsets which will be used to train the next strong classifiers in the next descending nodes. Thus, three cases are conceivable to decide 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.

---

[8]Random forest implementation can be found at: `http://www.stat.berkeley.edu/~breiman/RandomForests/cc_software.htm`
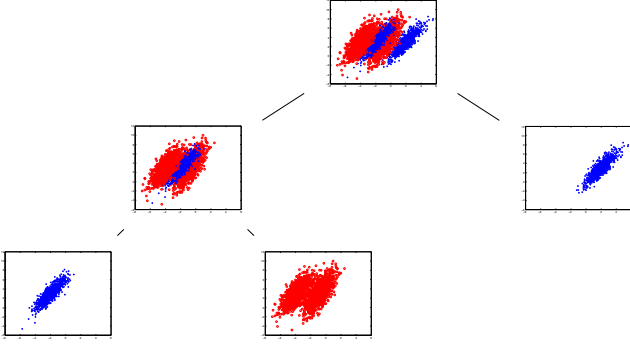
Figure 18: Representation of the capabilities of the probabilistic boosting tree algorithm to split at each node of the tree the positive and negative samples.

- 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.

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

$$
\begin{aligned}
q(+1, \mathbf{x}_i) &= \frac{\exp(2H(\mathbf{x}_i))}{1 + \exp(2H(\mathbf{x}_i))}, &(66)\\
q(-1, \mathbf{x}_i) &= \frac{\exp(-2H(\mathbf{x}_i))}{1 + \exp(-2H(\mathbf{x}_i))}. &(67)
\end{aligned}
$$

Employing such a scheme tends to divide the data in such a way that positive and negative samples are naturally split as shown in Fig. 18.

In the classification stage, the unlabelled sample $\mathbf{x}$ is propagated through the tree, where at each node, it will be classified by each strong classifier previously learned and where an estimation of the posterior distribution will be computed. The posterior distribution will correspond to the sum of the posterior distribution at each node of the decision tree.

Tiwari et al. (2010, 2009b, 2012), Viswanath et al. (2011) make use of the probabilistic boosting-tree classifier.

– *Kernel method:*

A Gaussian process[9] 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 (Rasmussen and Williams (2005)). In the case of linear logistic regression for classification, the posterior probability can be expressed as:

$$
\begin{aligned}
p(y_i|\mathbf{x}_i, \mathbf{w}) &= \sigma(y_i f(\mathbf{x}_i)), &(68)\\
f(\mathbf{x}_i) &= \mathbf{x}_i^{\mathrm{T}} \mathbf{w},
\end{aligned}
$$

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 will be characterized by a mean function $\bar{f}$ and covariance function K. Thus, 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 can be performed in two stages. First, for a new observation $\mathbf{x}_*$, the corresponding probability $p(f(\mathbf{x}_*)|f(\mathbf{x}))$ can be computed such that:

$$
\begin{aligned}
p(f(\mathbf{x}_*)|f(\mathbf{x})) &= \mathcal{N}(K_* K^{-1} \bar{f}, K_{**} - K_*(K')^{-1} K_*^{\mathrm{T}}),\\
K' &= K + W^{-1}, &(69)\\
W &= \nabla\nabla \log p(\mathbf{y}|f(\mathbf{x})),
\end{aligned}
$$

where K_{**} is the variance of the testing sample $\mathbf{x}_*$, K_* is the covariance 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 can be defined such that:

$$
C(\mathbf{x}_*) = \sigma\left(\frac{\bar{f}(\mathbf{x}_*)}{\sqrt{1 + var(f(\mathbf{x}_*))}}\right). \qquad (70)
$$

Only the work of Kelm et al. (2007) used Gaussian process for classification in order to distinguish CaP in MRSI data.

– *Sparse kernel methods:*

In a classification scheme using Gaussian processes, when a prediction has to be performed, the whole training data will be 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 (Bishop (2006)).

Support vector machines (SVM)[10] is a sparse kernel method aims at finding the best linear hyperplane (nonlinear separation is discussed further) which separates two classes such that the margin between the two classes is maximized (Vapnik and Lerner (1963)). The margin is in fact the region defined by two hyperplanes splitting the two classes, such that there are no points lying in between. The distance between these two hyperplanes is equal to $\frac{2}{\|\mathbf{w}\|}$ where $\mathbf{w}$ is the normal vector of the hyperplane 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:

$$
\begin{aligned}
&\underset{\mathbf{w}}{\arg\min} && \frac{1}{2}\|\mathbf{w}^2\|,\\
&\text{subject to} && y_i(\mathbf{w}.\mathbf{x}_i - b) \geq 1, \ i = \{1, \ldots, N\},
\end{aligned} \qquad (71)
$$

where $\mathbf{x}_i$ is a training sample with is corresponding class label y_i.

---

[9]Gaussian process implementation can be found at: `http://www.gaussianprocess.org/gpml/code/matlab/doc/index.html`

[10]SVM implementation can be found at: `http://www.csie.ntu.edu.tw/~cjlin/libsvm/`

From Eq. (71), it is important to notice that only few points from the set of N points have to be selected which will later define the hyperplane. This can be introduced in the optimization problem using Lagrange multipliers $\boldsymbol{\alpha}$. All points which are not lying on the margin will be assigned a corresponding $\alpha_i = 0$. This can be formalized as:

$$\arg \min_{\mathbf{w},b} \max_{\boldsymbol{\alpha} \geq 0} \left\{ \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^{n} \alpha_i [y_i(\mathbf{w} \cdot \mathbf{x_i} - b) - 1] \right\} . \quad (72)$$

The different parameters can be 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 to not find any hyperplane splitting the classes such as specified previously. Thus, a soft-margin optimization approach was proposed (Cortes and Vapnik (1995)), where points can lie on the margin but at the cost of a penalty ξ_i which will be minimized in the optimization process such that:

$$\arg \min_{\mathbf{w},\xi,b} \max_{\boldsymbol{\alpha},\boldsymbol{\beta}} \left\{ \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^{n} \xi_i \right.$$
$$\left. - \sum_{i=1}^{n} \alpha_i [y_i(\mathbf{w} \cdot \mathbf{x_i} - b) - 1 + \xi_i] - \sum_{i=1}^{n} \beta_i \xi_i \right\} . \quad (73)$$

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

$$C(\mathbf{x}_i) = \mathrm{sign} \left(\sum_{n=1}^{N} \alpha_n (\mathbf{x}_n . \mathbf{x}_i) + b_0 \right) , \quad (74)$$

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 (Boser et al. (1992)). The original data $\mathbf{x}$ can be projected to a high-dimension space in which it is assumed that a linear hyperplane will split the classes. Different kernels are popular such as the RBF kernel, polynomial kernels or Gaussian kernel.

In prostate CAD system, SVM is the most popular classification method and was used in a multitude of research works: Artan et al. (2010, 2009), Chan et al. (2003), Kelm et al. (2007), Litjens et al. (2012b, 2011), Liu et al. (2013), Lopes et al. (2011), Niaf et al. (2012, 2011), Ozer et al. (2009, 2010), Parfait et al. (2012), Peng et al. (2013), Sung et al. (2011), Tiwari et al. (2012), Vos et al. (2012, 2008a, 2010, 2008b).

Relevant vector machine (RVM) is a sparse version of Gaussian process previously presented and was proposed by Tipping (2001). RVM is identical to a Gaussian process with the following covariance function (Quinonero-Candela et al. (2002)):

$$K_{RVM}(\mathbf{x}_p, \mathbf{x}_q) = \sum_{j=1}^{M} \frac{1}{\alpha_j} \Phi_j(\mathbf{x}_p) \Phi_j(\mathbf{x}_q) , \quad (75)$$
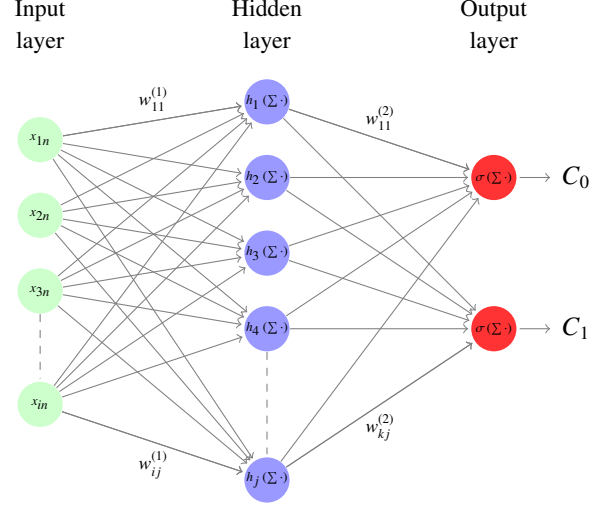

Figure 19: Representation of a neural network of the multilayer perceptron family.

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 Quinonero-Candela et al. (2002), the sparsity regarding the relevance vector arises if $j\alpha_j^{-1} = 0$. The set of weights $\boldsymbol{\alpha}$ is inferred using the expectation maximization algorithm. Ozer et al. (2009, 2010) make use of RVM and make a comparison with SVM for the task of CaP detection.

– **Neural network:**

Multilayer perceptron is a feed-forward neural networks considered as the most successful model of this kind in pattern recognition (Bishop (2006)). The most well known model used is based on a two layers where a prediction of an observation is computed as:

$$C(\mathbf{x}_n, w_{ij}^{(1)}, w_{kj}^{(2)}) = \sigma \left[\sum_{j=0}^{M} w_{kj}^{(2)} h \left(\sum_{i=0}^{D} w_{ij}^{(1)} x_{in} \right) \right] , \quad (76)$$

where $h(\cdot)$ and $\sigma(\cdot)$ are two 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, respectively.

A graphical representation of this network is presented in Fig. 19. Relating Fig. 19 with Eq. (76), 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
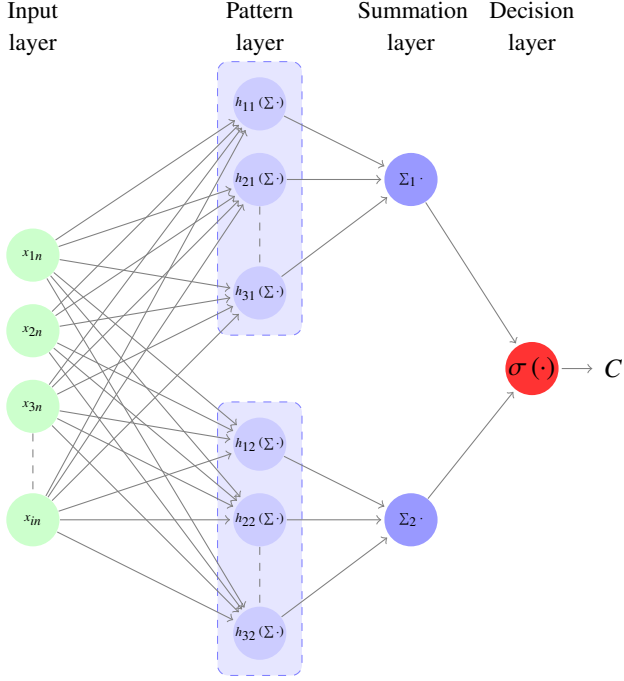
Figure 20: Representation of a neural network of the probabilistic neural network family.

best weights $w_{ij}^{(1)}$ and $w_{kj}^{(2)}$ which will model the best our data. The error of this model can be computed as:

$$E(w_{ij}^{(1)}, w_{kj}^{(2)}) = \frac{1}{2} \sum_{n=1}^{N} \left(C(\mathbf{x}_n, w_{ij}^{(1)}, w_{kj}^{(2)}) - y(\mathbf{x}_n) \right)^2 , \quad (77)$$

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

Thus the best set of weights can be inferred in an optimization framework where the error $E(\cdot)$ has to be minimized. This optimization can be performed using a gradient descent method where the derivative of Eq. (77) can be computed using the backpropagation algorithm proposed by Rumelhart et al. (1988). Matulewicz et al. (2013), Parfait et al. (2012) used this classifier to classify MRSI spectra.

Probabilistic neural networks are another type of feedforward networks which can be derived from the multilayer perceptron case and was proposed by Specht (1988). This classifier can be modelled by changing the activation function $h(\cdot)$ in Eq. (76) to an exponential function such that:

$$h(\mathbf{x}_n) = \exp\left(-\frac{(\mathbf{w}_j - \mathbf{x})^{\mathrm{T}}(\mathbf{w}_j - \mathbf{x})}{2\sigma^2} \right) , \quad (78)$$

where σ is a free parameter set by the user.

The other difference of the probabilistic neural networks when compared with the multilayer perceptron networks resides in the architecture as shown in Fig. 20. This network is formed by two hidden layers. The first hidden layer consists of the pattern layer, in which the mapping is done using Eq. (78). 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 was used by Ampeliotis et al. (2007, 2008), Viswanath et al. (2011) in order to perform the classification of their feature vector.

– *Graphical model classifiers:*

Markov random fields can also be used as a lesion segmentation method to detect CaP. First, we define s as a pixel which will belong to a certain class denoted by ω_s. The labelling process can be noted 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 (Kato and Pong (2001)). 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 can be formulated such that:

$$p(\omega|\mathcal{F}) = \arg\max_{\omega} \prod_{s \in I} p(f_s|\omega_s) p(\omega) . \quad (79)$$

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 will be characterized by their respective mean μ_λ and standard deviation σ_λ. Then, ω is a Markov random field, thus:

$$p(\omega) = \frac{1}{Z} \exp(-U(\omega)) , \quad (80)$$

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

Thus the segmentation problem can be 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 will 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 (Kato and Pong (2001)).

Liu et al. (2009), Ozer et al. (2010) used Markov random fields as an unsupervised method to segment lesions in multi-parametric MRI images.

Artan et al. (2010, 2009) used conditional random fields instead of Markov random fields to segment their MRI images. The difference between these two methods resides in the fact that conditional probabilities are defined such as:

$$p(\omega|\mathcal{F}) = \frac{1}{Z} \exp\left[-\sum_{s \in I} V_{C1}(\omega_s|\mathcal{F}) - \sum_{\{s,r\} \in C} V_{C2}(\omega_s, \omega_r|\mathcal{F}) \right] . \quad (81)$$

$V_{C1}(\cdot)$ is the state (or partition) feature function and $V_{C2}(\cdot)$ is the transition (or edge) feature function (Kato and Zerubia

Table 13: Overview of the model validation techniques used in CAD systems.

Model validation techniques	References
LOOCV	[1-8,11,17-21,23,25,32,36,38-40]
k-CV	[10,22,28-32,37,35,41]

Table 14: Overview of the evaluation metrics used in CAD systems.

Evaluation metrics	References
Accuracy	[4-5,12,25,31]
Sensitivity - Specificity	[4-5,7,12,14,17,20-23,25,27-28,33-34]
ROC - AUC	[2-3,6-9,13-19,23,29-32,35-40]
FROC	[10-11,41]
Dice's coefficient	[4-5,12,20]

(2012)).

4.4.2. Model validation

In pattern recognition, the use of model validation techniques to assess the performance of trained classifiers is quite important. Two techniques are broadly used in the development of CAD system and are summarized in Tab. 13.

The most popular technique used in CAD systems (see Tab. 13) 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 to train. This manipulation is repeated until each patient has been used for validation. This technique is popular when working with medical data due to the restricted number of patients included in datasets. Thus, it allows us to train on a fair number of patients even with a small dataset. However, this technique suffers from high variance and can be considered as an unreliable estimate (Efron (1983)).

The other very well known technique used for assessing classifiers is the k-fold cross-validation (k-CV) technique. This technique is based on splitting the dataset into k subsets where the samples are randomly selected. Then, one fold is kept for the validation and the remaining subsets for training. The classification is then repeated as in the LOOCV technique. In our review, the typical values used for k were set to three and five. This technique is more appropriate than the previous one since it does not suffer from large variance. However, the number of patients in the dataset needs to be large enough to apply such technique.

4.4.3. Evaluation measure

Several metrics can be used in order to assess the performance of the classifier trained when tested on the test data. The techniques used for evaluation of the CAD system for CaP detection are summarized in Tab. 14.

Using the classification approach previously presented, each voxel in the MRI image will be classified. Comparison with a ground-truth can give rise to a confusion matrix by counting true positive, true negative, false positive and false negative samples. From this analysis, different statistics can be 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 can be highly biased by a high number of true negative samples which will boost the accuracy score and does not represent the actual performance of the classifier.

That is why, the most common statistic computed are sensitivity and specificity which give a full overview of the performance of the classifier trained. Sensitivity is also called the true positive rate and is equal to the ratio of the true positive samples over the true positive added with the false negative samples as shown in Eq. (82). Specificity is also named the true negative rate and is equal to the ratio of the true negative samples over the true negative added with the false positive samples as shown in Eq. (83).

$$S EN = \frac{TP}{TP + FN} , \tag{82}$$

$$S PE = \frac{TN}{TN + FP} . \tag{83}$$

These statistics gave rise to the receiver operating characteristic (ROC) analysis. This analysis represents graphically the sensitivity as a function of (1 - 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 will be found but 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 were missed even if more false alarms have to be investigated. 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 previous method could have been classified on pixel-based evaluation method. However, a cancer can be also considered as a region. The free-response receiver operating characteristic (FROC) extends the ROC analysis but to a region-based level. The same confusion matrix can be computed were the sample are not pixels any more but refer to a lesion. 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!!!

Finally, Dice's coefficient is sometimes computed which corresponds to the similarity between a detected lesion and its ground-truth. This coefficient consists of the ratio of twice the number of pixels in common to 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. (84).
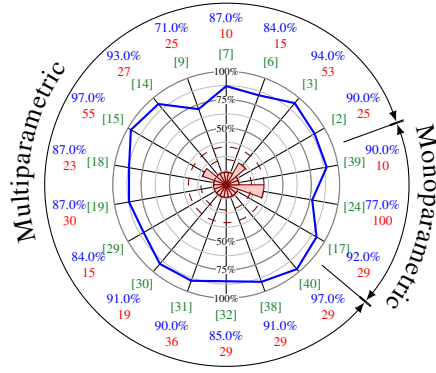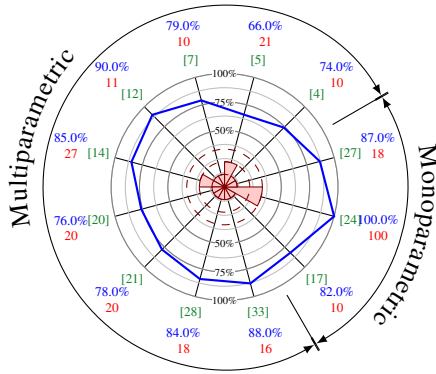
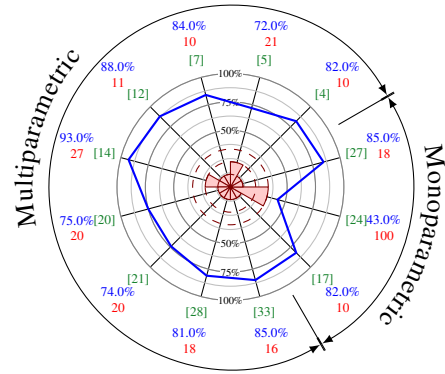(a) Comparison in terms of AUC-ROC of the methods using data from 1.5 Tesla MRI scanner.
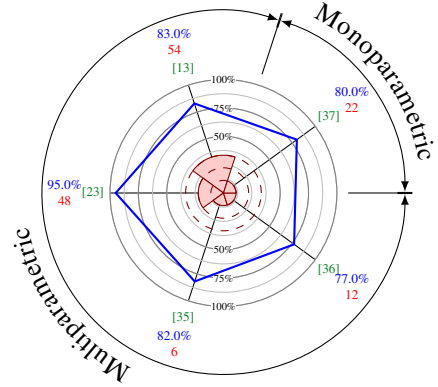
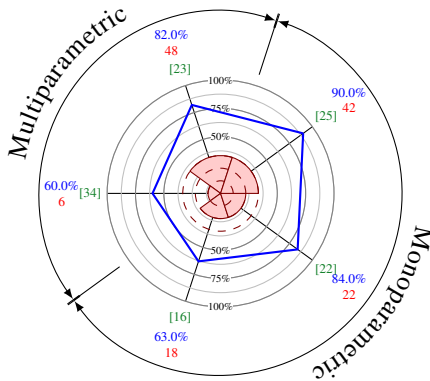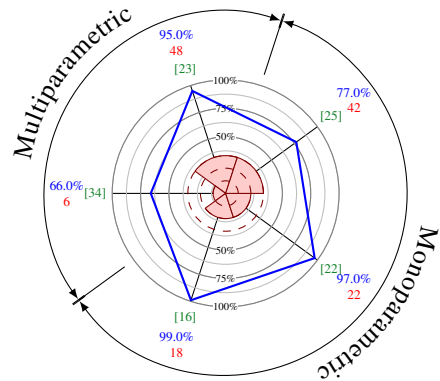(b) Comparison in terms of AUC-ROC of the methods using data from 3.0 Tesla MRI scanner.

Figure 21: Comparison of the results in terms of AUC for 1.5 and 3.0 Tesla MRI scanners. The blue value represents the metric and are graphically reported in the blue 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 green correspond to the reference as reported in Tab. 2.

(a) Comparison in terms of sensitivity of the methods using data from 1.5 Tesla MRI scanner.

(b) Comparison in terms of specificity of the methods using data from 1.5 Tesla MRI scanner.

(c) Comparison in terms of sensitivity of the methods using data from 3.0 Tesla MRI scanner.

(d) Comparison in terms of specificity of the methods using data from 3.0 Tesla MRI scanner.

Figure 22: Comparison of the results in terms of sensitivity (22(a),22(c)) and specificity (22(b),22(d)) for 1.5 and 3.0 Tesla MRI scanners. The blue value represents the metric and are graphically reported in the blue 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 green correspond to the reference as reported in Tab. 2.
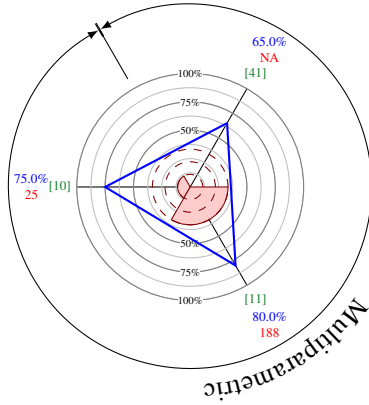
Figure 23: Comparison in terms of FROC of the methods using data from 3.0 Tesla MRI scanner. The blue value represent the metric and are graphically reported in the blue curve in the center of the figure. The red value correspond to the number of patients in the dataset and is also reported in the center of the figure. The numbers between brackets correspond to the reference as reported in Tab. 2.

$$Q_D = \frac{2|GT \cap S|}{|GT| + |S|}. \tag{84}$$

5. Discussion

5.1. Brief comparison

As discussed previously in Sect. 4.4.3, 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 (cf., 1.5 *versus* 3.0 Tesla). For each field, the experiment obtaining the best result was reported in these figures.

The results given in terms of AUC-ROC are depicted in Fig. 21. The results vary between 71% and 97% for some experiments with a 1.5 Tesla MRI scanner and 77% and 95% with a 3.0 Tesla MRI scanner.

The results in regard of sensitivity and specificity are reported in Fig. 22. In the case that the data were collected with a 1.5 Tesla MRI scanner, the sensitivity ranges from 74% to 100% and the specificity from 43% to 93%. For the experiments carried out with a 3.0 Tesla MRI scanner, the sensitivity varies from 60% to 90% and the specificity from 66% to 99%.

Three studies also use FROC analysis to report their results and are shown in Fig. 23 and the sensitivity reported ranges from 65% to 80%.

However, we would like to emphasize the fact that the results obtained from these different experiments cannot be fairly compared. Different datasets were used implying different complexity involved and different sets of input parameters during the data acquisition. To our mind, the only way to provide a real and fair comparison would be to provide a common working dataset where those algorithms could be tested.

5.2. General discussion

This review leads to some general discussions which could direct to future avenues for research. As previously mentioned, no open multi-parametric dataset is currently available. This fact leads to an impossibility to fairly compare the different algorithms designed over years. Also, the availability of a full multi-parametric MRI dataset, could lead to the development of algorithms which use all the different modalities currently available. Recalling Tab. 2, it can be noted that none of the current works provides a solution using at the same time the four different modalities. Also, all the algorithms are focused on one type of scanner only, either 1.5 Tesla and 3.0 Tesla. A dataset including both these types of imaging could allow development of more generic algorithms.

Analysing the different stages of the CAD work-flow, it is seen that the actual CAD systems do not include all the different pre-processing steps. It could be interesting to evaluate the improvement using these pre-processing steps on the final results obtained after the classification. 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. Moreover, other methods specific to segmentation and registration which are not actually used in CAD systems could also perform better than the ones currently used in CADs.

Regarding the classification framework, it seems that the current well-known pattern recognition methods have been widely studied. However, more investigations should be carried out regarding the feature detection stage. Lately, histogram-based features have shown good capabilities in the field of computer vision and could be further investigated. Only one study by Liu et al. (2013) used some of these features.

An important point allowing a fair comparison between methods resides in the fact that no universal evaluation model and metric has been defined by the research community allowing such comparison. Usually, it is quite common to choose an evaluation model which fits the dataset limitations, usually the size. Regarding the evaluation, the community should agree on a standard metric to measure the performance of the algorithms designed.

Finally, we would like to focus the attentions of the reader on the availability of a multi-parametric dataset from 1.5 and 3.0 Tesla MRI provided by the authors of this review. This dataset is available at the following website address: `http://visor.udg.edu/dataset`. The dataset is composed of the four modalities discussed in this review with their corresponding ground-truth images. In addition of the repository activity, this website will aim at providing comparison between algorithms developed by the research community.

6. Conclusion

This review presented an overview and gave a staging of the research related to CAD development for CaP using multi-parametric MRI data. We aimed at providing background information regarding multi-parametric MRI imaging techniques.

A work-flow describing the CAD stages were proposed. The methods used in the literature for each of this stage were reviewed. The results of the available CAD systems were briefly reviewed. Subsequently, an insight discussion were given to conclude this survey.

7. Acknowledgement

References

Agalliu, I., Gern, R., Leanza, S., Burk, R.D., 2009. Associations of high-grade prostate cancer with BRCA1 and BRCA2 founder mutations. Clin. Cancer Res. 15, 1112–1120.

Ahmed, N., Natarajan, T., Rao, K., 1974. Discrete cosine transform. Computers, IEEE Transactions on C-23, 90–93.

Aizerman, M.A., Braverman, E.A., Rozonoer, L., 1964. Theoretical foundations of the potential function method in pattern recognition learning, in: Automation and Remote Control,, pp. 821–837.

Akin, O., Sala, E., Moskowitz, C.S., Kuroiwa, K., Ishill, N.M., Pucar, D., Scardino, P.T., Hricak, H., 2006. Transition zone prostate cancers: features, detection, localization, and staging at endorectal MR imaging. Radiology 239, 784–792.

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

Amadasun, M., King, R., 1989. Textural features corresponding to textural properties. Systems, Man and Cybernetics, IEEE Transactions on 19, 1264–1274.

American Cancer Society, A.C., 2010. Cancer Facts and Figures 2010. http://www.cancer.org/research/cancerfactsfigures. URL: http://www.cancer.org/research/cancerfactsfigures. accessed: 2013-08-01.

American Cancer Society, A.C., 2013. Cancer Facts and Figures 2013. http://www.cancer.org/research/cancerfactsfigures. Accessed: 2013-08-01.

Ampeliotis, D., Anonakoudi, A., Berberidis, K., Psarakis, E.Z., 2007. 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, pp. 888–891.

Ampeliotis, D., Anonakoudi, A., Berberidis, K., Psarakis, E.Z., Kounoudes, A., 2008. 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.

Amundadottir, L.T., Sulem, P., Gudmundsson, J., Helgason, A., Baker, A., Agnarsson, B.A., Sigurdsson, A., Benediktsdottir, K.R., Cazier, J.B., Sainz, J., Jakobsdottir, M., Kostic, J., Magnusdottir, D.N., Ghosh, S., Agnarsson, K., Birgisdottir, B., Le Roux, L., Olafsdottir, A., Blondal, T., Andresdottir, M., Gretarsdottir, O.S., Bergthorsson, J.T., Gudbjartsson, D., Gylfason, A., Thorleifsson, G., Manolescu, A., Kristjansson, K., Geirsson, G., Isaksson, H., Douglas, J., Johansson, J.E., Balter, K., Wiklund, F., Montie, J.E., Yu, X., Suarez, B.K., Ober, C., Cooney, K.A., Gronberg, H., Catalona, W.J., Einarsson, G.V., Barkardottir, R.B., Gulcher, J.R., Kong, A., Thorsteinsdottir, U., Stefansson, K., 2006. A common variant associated with prostate cancer in European and African populations. Nat. Genet. 38, 652–658.

Andriole, G.L., Crawford, E.D., Grubb, R.L., Buys, S.S., Chia, D., Church, T.R., Fouad, M.N., Gelmann, E.P., Kvale, P.A., Reding, D.J., Weissfeld, J.L., Yokochi, L.A., O'Brien, B., Clapp, J.D., Rathmell, J.M., Riley, T.L.,

Hayes, R.B., Kramer, B.S., Izmirlian, G., Miller, A.B., Pinsky, P.F., Prorok, P.C., Gohagan, J.K., Berg, C.D., 2009. Mortality results from a randomized Prostate-cancer screening trial. New England Journal of Medicine 360, 1310–1319.

Antic, T., Peng, Y., Jiang, Y., Giger, M.L., Eggener, S., Oto, A., 2013. 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, pp. 86701H–86701H–6.

Artan, Y., Haider, M.A., Langer, D.L., van der Kwast, T.H., Evans, A.J., Yang, Y., Wernick, M.N., Trachtenberg, J., Yetik, I.S., 2010. Prostate cancer localization with multispectral MRI using cost-sensitive support vector machines and conditional random fields. IEEE Trans Image Process 19, 2444–2455.

Artan, Y., Langer, D., Haider, M., Van der Kwast, T.H., Evans, A., Wernick, M., Yetik, I., 2009. 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, pp. 278–281.

Awwad, H.M., Geisel, J., Obeid, R., 2012. The role of choline in prostate cancer. Clin. Biochem. 45, 1548–1553.

Barentsz, J.O., Richenberg, J., Clements, R., Choyke, P., Verma, S., Villeirs, G., Rouviere, O., Logager, V., Futterer, J.J., 2012. ESUR prostate MR guidelines 2012. Eur Radiol 22, 746–757.

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

Belongie, S., Malik, J., Puzicha, J., 2002. Shape matching and object recognition using shape contexts. Pattern Analysis and Machine Intelligence, IEEE Transactions on 24, 509–522.

Benassi, A., Cohen, S., Istas, J., 1998. Identifying the multifractional function of a Gaussian process. Statistics & Probability Letters 39, 337 – 345.

Bishop, C.M., 2006. Pattern recognition and machine learning. Springer-Verlag New York, Inc., Secaucus, NJ, USA.

Bookstein, F.L., 1989. Principal warps: thin-plate splines and the decomposition of deformations. Pattern Analysis and Machine Intelligence, IEEE Transactions on 11, 567–585.

Boser, B.E., Guyon, I.M., Vapnik, V.N., 1992. A training algorithm for optimal margin classifiers, in: Proceedings of the Fifth Annual Workshop on Computational Learning Theory, ACM, New York, NY, USA. pp. 144–152.

Bourdoumis, A., Papatsoris, A.G., Chrisofos, M., Efstathiou, E., Skolarikos, A., Deliveliotis, C., 2010. The novel prostate cancer antigen 3 (PCA3) biomarker. Int Braz J Urol 36, 665–668.

Breiman, L., 2001. Random forests. Machine Learning 45, 5–32.

Breiman, L., Friedman, J., Olshen, R., Stone, C., 1984. Classification and regression trees. Wadsworth and Brooks, Monterey, CA.

Brenner, J., Chinnaiyan, A., Tomlins, S., 2013. ETS fusion genes in prostate cancer, in: Tindall, D.J. (Ed.), Prostate Cancer. Springer New York. volume 16 of *Protein Reviews*, pp. 139–183.

Buades, A., Coll, B., Morel, J., 2005. A review of image denoising algorithms, with a new one. Simul 4, 490–530.

Buckley, D.L., Roberts, C., Parker, G.J., Logue, J.P., Hutchinson, C.E., 2004. Prostate cancer: evaluation of vascular characteristics with dynamic contrast-enhanced T1-weighted MR imaging–initial experience. Radiology 233, 709–715.

Byrd, R.H., Lu, P., Nocedal, J., Zhu, C., 1995. A limited memory algorithm for bound constrained optimization. SIAM J. Sci. Comput. 16, 1190–1208.

Carmeliet, P., Jain, R.K., 2000. Angiogenesis in cancer and other diseases. Nature 407, 249–257.

Carrol, C.L., Sommer, F.G., McNeal, J.E., Stamey, T.A., 1987. The abnormal prostate: MR imaging at 1.5 T with histopathologic correlation. Radiology 163, 521–525.

Castorina, P., Delsanto, P.P., Guiot, C., 2006. Classification scheme for phenomenological universalities in growth problems in physics and other sciences. Phys. Rev. Lett. 96, 188701.

Chan, H.P., Doi, K., Galhotra, S., Vyborny, C.J., MacMahon, H., Jokich, P.M., 1987. Image feature analysis and computer-aided diagnosis in digital radiography. I. Automated detection of microcalcifications in mammography. Med Phys 14, 538–548.

Chan, H.P., Hadjiiski, L., Zhou, C., Sahiner, B., 2008. Computer-aided diagnosis of lung cancer and pulmonary embolism in computed tomography-a review. Acad Radiol 15, 535–555.

Chan, H.P., Sahiner, B., Helvie, M.A., Petrick, N., Roubidoux, M.A., Wilson, T.E., Adler, D.D., Paramagul, C., Newman, J.S., Sanjay-Gopal, S., 1999. Improvement of radiologists' characterization of mammographic masses by using computer-aided diagnosis: an ROC study. Radiology 212, 817–827.

Chan, I., Wells, W., Mulkern, R.V., Haker, S., Zhang, J., Zou, K.H., Maier, S.E., Tempany, C.M., 2003. Detection of prostate cancer by integration of line-scan diffusion, T2-mapping and T2-weighted magnetic resonance imaging; a multichannel statistical classifier. Med Phys 30, 2390–2398.

Chappelow, J., Bloch, B.N., Rofsky, N., Genega, E., Lenkinski, R., DeWolf, W., Madabhushi, A., 2011. Elastic registration of multimodal prostate MRI and histology via multiattribute combined mutual information. Med Phys 38, 2005–2018.

Chen, L., Weng, Z., Goh, L., Garland, M., 2002. An efficient algorithm for automatic phase correction of {NMR} spectra based on entropy minimization . Journal of Magnetic Resonance 158, 164 – 168.

Cheng, H.D., Shan, J., Ju, W., Guo, Y., Zhang, L., 2010. Automated breast cancer detection and classification using ultrasound images: A survey. Pattern Recogn. 43, 299–317.

Choi, Y.J., Kim, J.K., Kim, N., Kim, K.W., Choi, E.K., Cho, K.S., 2007. Functional MR imaging of prostate cancer. Radiographics 27, 63–75.

Chou, R., Croswell, J.M., Dana, T., Bougatsos, C., Blazina, I., Fu, R., Gleismann, K., Koenig, H.C., Lam, C., Maltz, A., Rugge, J.B., Lin, K., 2011. Screening for prostate cancer: a review of the evidence for the U.S. Preventive Services Task Force. Ann. Intern. Med. 155, 762–771.

Coakley, F.V., Hricak, H., 2000. Radiologic anatomy of the prostate gland: a clinical approach. Radiol. Clin. North Am. 38, 15–30.

Cohen, R.J., Shannon, B.A., Phillips, M., Moorin, R.E., Wheeler, T.M., Garrett, K.L., 2008. Central zone carcinoma of the prostate gland: a distinct tumor type with poor prognostic features. J. Urol. 179, 1762–1767.

Coifman, R., Wickerhauser, M., 1992. Entropy-based algorithms for best basis selection. Information Theory, IEEE Transactions on 38, 713–718.

Coleman, T., Li, Y., 1993. An interior trust region approach for nonlinear minimization subject to bounds. Technical Report. Cornell University.

Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J., 1995. Active shape models&Mdash;Their training and application. Comput. Vis. Image Underst. 61, 38–59.

Cortes, C., Vapnik, V., 1995. Support-Vector networks. Machine Learning 20, 273–297.

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

Cruz, M., Tsuda, K., Narumi, Y., Kuroiwa, Y., Nose, T., Kojima, Y., Okuyama, A., Takahashi, S., Aozasa, K., Barentsz, J.O., Nakamura, H., 2002. Characterization of low-intensity lesions in the peripheral zone of prostate on pre-biopsy endorectal coil MR imaging. Eur Radiol 12, 357–365.

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

Daugman, J.G., 1985. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. J Opt Soc Am A 2, 1160–1169.

Dean, J.C., Ilvento, C.C., 2006. Improved cancer detection using computer-aided detection with diagnostic and screening mammography: prospective study of 104 cancers. AJR Am J Roentgenol 187, 20–28.

Delongchamps, N.B., Peyromaure, M., Schull, A., Beuvon, F., Bouazza, N., Flam, T., Zerbib, M., Muradyan, N., Legman, P., Cornud, F., 2013. Prebiopsy magnetic resonance imaging and prostate cancer detection: comparison of random and targeted biopsies. J. Urol. 189, 493–499.

Delpierre, C., Lamy, S., Kelly-Irving, M., Molinie, F., Velten, M., Tretarre, B., Woronoff, A.S., Buemi, A., Lapotre-Ledoux, B., Bara, S., Guizard, A.V., Colonna, M., Grosclaude, P., 2013. Life expectancy estimates as a key factor in over-treatment: the case of prostate cancer. Cancer Epidemiol 37, 462–468.

Devos, A., Lukas, L., Suykens, J.A., Vanhamme, L., Tate, A.R., Howe, F.A., Majos, C., Moreno-Torres, A., van der Graaf, M., Arus, C., Van Huffel, S., 2004. Classification of brain tumours using short echo time 1H MR spectra. J. Magn. Reson. 170, 164–175.

Doi, K., Chan, H.P., Giger, M., 1990. Method and system for enhancement and detection of abnormal anatomic regions in a digital image.

Donoho, D.L., Johnstone, J.M., 1994. Ideal spatial adaptation by wavelet shrinkage. Biometrika 81, 425–455.

Doo, K.W., Sung, D.J., Park, B.J., Kim, M.J., Cho, S.B., Oh, Y.W., Ko, Y.H., Yang, K.S., 2012. Detectability of low and intermediate or high risk prostate cancer with combined T2-weighted and diffusion-weighted MRI. Eur Radiol 22, 1812–1819.

Efron, B., 1979. Bootstrap methods: Another look at the jackknife. The Annals of Statistics 7, 1–26.

Efron, B., 1983. Estimating the error rate of a prediction rule: Improvement on cross-validation. Journal of the American Statistical Association 78, pp. 316–331.

Elter, M., Horsch, A., 2009. CADx of mammographic masses and clustered microcalcifications: a review. Med Phys 36, 2052–2068.

Epstein, J.I., Allsbrook, W.C., Amin, M.B., Egevad, L.L., 2005. The 2005 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma. Am. J. Surg. Pathol. 29, 1228–1242.

Etzioni, R., Penson, D.F., Legler, J.M., di Tommaso, D., Boer, R., Gann, P.H., Feuer, E.J., 2002. Overdiagnosis due to prostate-specific antigen screening: lessons from U.S. prostate cancer incidence trends. J. Natl. Cancer Inst. 94, 981–990.

Ferlay, J., Shin, H.R., Bray, F., Forman, D., Mathers, C., Parkin, D.M., 2010. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. Int. J. Cancer 127, 2893–2917.

Fodor, I., 2002. A survey of dimension reduction techniques.

Fred, A., Jain, A., 2005. Combining multiple clusterings using evidence accumulation. Pattern Analysis and Machine Intelligence, IEEE Transactions on 27, 835–850.

Freedman, M.L., Haiman, C.A., Patterson, N., McDonald, G.J., Tandon, A., Waliszewska, A., Penney, K., Steen, R.G., Ardlie, K., John, E.M., Oakley-Girvan, I., Whittemore, A.S., Cooney, K.A., Ingles, S.A., Altshuler, D., Henderson, B.E., Reich, D., 2006. Admixture mapping identifies 8q24 as a prostate cancer risk locus in African-American men. Proc. Natl. Acad. Sci. U.S.A. 103, 14068–14073.

Freund, Y., Schapire, R., 1997. A decision-theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences 55, 119 – 139.

Friedman, J., 1989. Regularized discriminant analysis. Journal of the American Statistical Association 84, pp. 165–175.

Gabor, D., 1946. Theory of communication. Part 1: The analysis of information. Electrical Engineers - Part III: Radio and Communication Engineering, Journal of the Institution of 93, 429–441.

Ghose, S., Oliver, A., Marti, R., Llado, X., Vilanova, J.C., Freixenet, J., Mitra, J., Sidibe, D., Meriaudeau, F., 2012. A survey of prostate segmentation methodologies in ultrasound, magnetic resonance and computed tomography images. Comput Methods Programs Biomed 108, 262–287.

Giannini, V., Vignati, A., Mazzetti, S., De Luca, M., Bracco, C., Stasi, M., Russo, F., Armando, E., Regge, D., 2013. 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, pp. 86703E–86703E–6.

Gibbs, P., Tozer, D.J., Liney, G.P., Turnbull, L.W., 2001. Comparison of quantitative T2 mapping and diffusion-weighted imaging in the normal and pathologic prostate. Magn Reson Med 46, 1054–1058.

Giger, M.L., Chan, H.P., Boone, J., 2008. Anniversary paper: History and status of CAD and quantitative image analysis: the role of Medical Physics and AAPM. Med Phys 35, 5799–5820.

Giger, M.L., Doi, K., MacMahon, H., 1988. Image feature analysis and computer-aided diagnosis in digital radiography. 3. Automated detection of nodules in peripheral lung fields. Med Phys 15, 158–166.

Giovannucci, E., Liu, Y., Platz, E.A., Stampfer, M.J., Willett, W.C., 2007. Risk factors for prostate cancer incidence and progression in the health professionals follow-up study. Int. J. Cancer 121, 1571–1578.

Giskeodegard, G.F., Bertilsson, H., Selnaes, K.M., Wright, A.J., Bathen, T.F., Viset, T., Halgunset, J., Angelsen, A., Gribbestad, I.S., Tessem, M.B., 2013. Spermine and citrate as metabolic biomarkers for assessing prostate cancer aggressiveness. PLoS ONE 8, e62375.

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

Goodman, S.N., 1999. Toward evidence-based medical statistics. 1: The P value fallacy. Ann. Intern. Med. 130, 995–1004.

van der Graaf, M., Schipper, R.G., Oosterhof, G.O., Schalken, J.A., Verhofstad,

39

A.A., Heerschap, A., 2000. Proton MR spectroscopy of prostatic tissue focused on the detection of spermine, a possible biomarker of malignant behavior in prostate cancer. MAGMA 10, 153–159.

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

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

Haas, G.P., Delongchamps, N.B., Jones, R.F., Chandan, V., Serio, A.M., Vickers, A.J., Jumbelic, M., Threatte, G., Korets, R., Lilja, H., de la Roza, G., 2007. Needle biopsies on autopsy prostates: sensitivity of cancer detection based on true prevalence. J. Natl. Cancer Inst. 99, 1484–1489.

Hambrock, T., Somford, D.M., Huisman, H.J., van Oort, I.M., Witjes, J.A., Hulsbergen-van de Kaa, C.A., Scheenen, T., Barentsz, J.O., 2011. Relationship between apparent diffusion coefficients at 3.0-T MR imaging and Gleason grade in peripheral zone prostate cancer. Radiology 259, 453–461.

Hambrock, T., Vos, P.C., Hulsbergen-van de Kaa, C.A., Barentsz, J.O., Huisman, H.J., 2013. Prostate cancer: computer-aided diagnosis with multiparametric 3-T MR imaging–effect on observer performance. Radiology 266, 521–530.

Hara, N., Okuizumi, M., Koike, H., Kawaguchi, M., Bilim, V., 2005. Dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) is a useful modality for the precise detection and staging of early prostate cancer. Prostate 62, 140–147.

Haralick, R., Shanmugam, K., Dinstein, I., 1973. Textural features for image classification. Systems, Man and Cybernetics, IEEE Transactions on SMC-3, 610–621.

Hegde, J.V., Mulkern, R.V., Panych, L.P., Fennessy, F.M., Fedorov, A., Maier, S.E., Tempany, C.M., 2013. Multiparametric MRI of prostate cancer: an update on state-of-the-art techniques and their performance in detecting and localizing prostate cancer. J Magn Reson Imaging 37, 1035–1054.

Heidenreich, A., Abrahamsson, P.A., Artibani, W., Catto, J., Montorsi, F., Van Poppel, H., Wirth, M., Mottet, N., 2013. Early detection of prostate cancer: European Association of Urology recommendation. Eur. Urol. 64, 347–354.

Hero, A., Ma, B., Michel, O., Gorman, J., 2002. Applications of entropic spanning graphs. Signal Processing Magazine, IEEE 19, 85–95.

Hoeks, C.M., Barentsz, J.O., Hambrock, T., Yakar, D., Somford, D.M., Heijmink, S.W., Scheenen, T.W., Vos, P.C., Huisman, H., van Oort, I.M., Witjes, J.A., Heerschap, A., Futterer, J.J., 2011. Prostate cancer: multiparametric MR imaging for detection, localization, and staging. Radiology 261, 46–66.

Hoffman, R.M., Gilliland, F.D., Eley, J.W., Harlan, L.C., Stephenson, R.A., Stanford, J.L., Albertson, P.C., Hamilton, A.S., Hunt, W.C., Potosky, A.L., 2001. Racial and ethnic differences in advanced-stage prostate cancer: the Prostate Cancer Outcomes Study. J. Natl. Cancer Inst. 93, 388–395.

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

Hricak, H., Williams, R.D., Spring, D.B., Moon, K.L., Hedgcock, M.W., Watson, R.A., Crooks, L.E., 1983. Anatomy and pathology of the male pelvis by magnetic resonance imaging. AJR Am J Roentgenol 141, 1101–1110.

Huch Boni, R.A., Boner, J.A., Lutolf, U.M., Trinkler, F., Pestalozzi, D.M., Krestin, G.P., 1995. Contrast-enhanced endorectal coil MRI in local staging of prostate carcinoma. J Comput Assist Tomogr 19, 232–237.

Hugosson, J., Carlsson, S., Aus, G., Bergdahl, S., Khatami, A., Lodding, P., Pihl, C.G., Stranne, J., Holmberg, E., Lilja, H., 2010. Mortality results from the Göteborg randomised population-based prostate-cancer screening trial. Lancet Oncol. 11, 725–732.

Huisman, H., Vos, P., Litjens, G., Hambrock, T., Barentsz, J., 2010. 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, Springer-Verlag, Berlin, Heidelberg. pp. 4–14.

Huisman, T.A., 2003. Diffusion-weighted imaging: basic concepts and application in cerebral stroke and head trauma. Eur Radiol 13, 2283–2297.

Itou, Y., Nakanishi, K., Narumi, Y., Nishizawa, Y., Tsukuma, H., 2011. 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 33, 167–172.

Jager, G.J., Ruijter, E.T., van de Kaa, C.A., de la Rosette, J.J., Oosterhof, G.O., Thornbury, J.R., Ruijs, S.H., Barentsz, J.O., 1997. Dynamic TurboFLASH subtraction technique for contrast-enhanced MR imaging of the prostate: correlation with histopathologic results. Radiology 203, 645–652.

Jolliffe, I.T., 2002. Principal Component Analysis. Second ed., Springer.

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

Kaji, Y., Kurhanewicz, J., Hricak, H., Sokolov, D.L., Huang, L.R., Nelson, S.J., Vigneron, D.B., 1998. Localizing prostate cancer in the presence of postbiopsy changes on MR images: role of proton MR spectroscopic imaging. Radiology 206, 785–790.

Kato, Z., Pong, T., 2001. A Markov random field image segmentation model using combined color and texture features, in: Skarbek, W. (Ed.), Computer Analysis of Images and Patterns. Springer Berlin Heidelberg. volume 2124 of *Lecture Notes in Computer Science*, pp. 547–554.

Kato, Z., Zerubia, J., 2012. Markov Random Fields in Image Segmentation. Collection Foundation and Trends in Signal Processing. Now Editor, World Scientific.

Kelm, B.M., Menze, B.H., Zechmann, C.M., Baudendistel, K.T., Hamprecht, F.A., 2007. Automated estimation of tumor probability in prostate magnetic resonance spectroscopic imaging: pattern recognition vs quantification. Magn Reson Med 57, 150–159.

Kety, S., 1951. The theory and applications of the exchange of inert gas at the lungs and tissues. Pharmacol. Rev. 3, 1–41.

Kim, J.K., Hong, S.S., Choi, Y.J., Park, S.H., Ahn, H., Kim, C.S., Cho, K.S., 2005. Wash-in rate on the basis of dynamic contrast-enhanced MRI: usefulness for prostate cancer detection and localization. J Magn Reson Imaging 22, 639–646.

Kirkham, A.P., Emberton, M., Allen, C., 2006. How good is MRI at detecting and characterising cancer within the prostate? Eur. Urol. 50, 1163–1174.

Kirsch, R., 1971. Computer determination of the constituent structure of biological images . Computers and Biomedical Research 4, 315 – 328.

Koh, D.M., Collins, D.J., 2007. Diffusion-weighted MRI in the body: applications and challenges in oncology. AJR Am J Roentgenol 188, 1622–1635.

Korotkov, K., Garcia, R., 2012. Computerized analysis of pigmented skin lesions: a review. Artif Intell Med 56, 69–90.

Kurhanewicz, J., Vigneron, D.B., Hricak, H., Narayan, P., Carroll, P., Nelson, S.J., 1996. Three-dimensional H-1 MR spectroscopic imaging of the in situ human prostate with high (0.24-0.7-cm3) spatial resolution. Radiology 198, 795–805.

Langer, D.L., van der Kwast, T.H., Evans, A.J., Trachtenberg, J., Wilson, B.C., Haider, M.A., 2009. 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 30, 327–334.

Larsson, H.B., Fritz-Hansen, T., Rostrup, E., Sondergaard, L., Ring, P., Henriksen, O., 1996. Myocardial perfusion modeling using MRI. Magn Reson Med 35, 716–726.

Laudadio, T., Mastronardi, N., Vanhamme, L., Hecke, P.V., Huffel, S.V., 2002. Improved Lanczos algorithms for blackbox {MRS} data quantitation. Journal of Magnetic Resonance 157, 292 – 297.

Le Bihan, D., Breton, E., Lallemand, D., Aubin, M.L., Vignaud, J., Laval-Jeantet, M., 1988. Separation of diffusion and perfusion in intravoxel incoherent motion MR imaging. Radiology 168, 497–505.

Le Bihan, D., Breton, E., Lallemand, D., Grenier, P., Cabanis, E., Laval-Jeantet, M., 1986. MR imaging of intravoxel incoherent motions: application to diffusion and perfusion in neurologic disorders. Radiology 161, 401–407.

Leissner, K.H., Tisell, L.E., 1979. The weight of the human prostate. Scand. J. Urol. Nephrol. 13, 137–142.

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

Li, F., Aoyama, M., Shiraishi, J., Abe, H., Li, Q., Suzuki, K., Engelmann, R., Sone, S., Macmahon, H., Doi, K., 2004. Radiologists' performance for differentiating benign from malignant lung nodules on high-resolution CT using computer-estimated likelihood of malignancy. AJR Am J Roentgenol 183, 1209–1215.

Li, H., Giger, M.L., Olopade, O.I., Margolis, A., Lan, L., Chinander, M.R., 2005. Computerized texture analysis of mammographic parenchymal patterns of digitized mammograms. Acad Radiol 12, 863–873.

Li, Q., Sone, S., Doi, K., 2003. Selective enhancement filters for nodules,

vessels, and airway walls in two- and three-dimensional CT scans. Med Phys 30, 2040–2051.

Li, S.Z., 1996. Robustizing robust M-estimation using deterministic annealing. Pattern Recognition 29, 159–166.

Lieber, C.A., Mahadevan-Jansen, A., 2003. Automated method for subtraction of fluorescence from biological Raman spectra. Appl Spectrosc 57, 1363–1367.

Liney, G.P., Knowles, A.J., Manton, D.J., Turnbull, L.W., Blackband, S.J., Horsman, A., 1996a. 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 6, 603–607.

Liney, G.P., Lowry, M., Turnbull, L.W., Manton, D.J., Knowles, A.J., Blackband, S.J., Horsman, A., 1996b. Proton MR T2 maps correlate with the citrate concentration in the prostate. NMR Biomed 9, 59–64.

Liney, G.P., Turnbull, L.W., Lowry, M., Turnbull, L.S., Knowles, A.J., Horsman, A., 1997. In vivo quantification of citrate concentration and water T2 relaxation time of the pathologic prostate gland using 1H MRS and MRI. Magn Reson Imaging 15, 1177–1186.

Litjens, G., Debats, O., van de Ven, W., Karssemeijer, N., Huisman, H., 2012a. A pattern recognition approach to zonal segmentation of the prostate on MRI. Med Image Comput Comput Assist Interv 15, 413–420.

Litjens, G.J.S., Barentsz, J.O., Karssemeijer, N., Huisman, H.J., 2012b. 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, pp. 83150G–83150G–6.

Litjens, G.J.S., Vos, P.C., Barentsz, J.O., Karssemeijer, N., Huisman, H.J., 2011. Automatic computer aided detection of abnormalities in multi-parametric prostate MRI, in: Proc. SPIE 7963, Medical Imaging 2011: Computer-Aided Diagnosis, pp. 79630T–79630T–7.

Liu, P., Wang, S., Turkbey, B., Grant, K.and Pinto, P.C.P., Wood, B.J., Summers, R.M., 2013. 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, pp. 86701G–86701G–6.

Liu, W., Turkbey, B., Senegas, J., Remmele, S., Xu, S., Kruecker, J., Bernardo, M., Wood, B.J., Pinto, P.A., Choyke, P.L., 2011. Accelerated T2 mapping for characterization of prostate cancer. Magn Reson Med 65, 1400–1406.

Liu, X., Langer, D.L., Haider, M.A., Yang, Y., Wernick, M.N., Yetik, I.S., 2009. Prostate cancer segmentation with simultaneous estimation of Markov random field parameters and class. IEEE Trans Med Imaging 28, 906–915.

Lopes, R., Ayache, A., Makni, N., Puech, P., Villers, A., Mordon, S., Betrouni, N., 2011. Prostate cancer characterization on MR images using fractal features. Med Phys 38, 83–95.

Lu-Yao, G.L., Albertsen, P.C., Moore, D.F., Shih, W., Lin, Y., DiPaola, R.S., Barry, M.J., Zietman, A., O'Leary, M., Walker-Corkery, E., Yao, S.L., 2009. Outcomes of localized prostate cancer following conservative management. JAMA 302, 1202–1209.

Lv, D., Guo, X., Wang, X., Zhang, J., Fang, J., 2009. Computerized characterization of prostate cancer by fractal analysis in MR images. J Magn Reson Imaging 30, 161–168.

Ma, R.W., Chapman, K., 2009. A systematic review of the effect of diet in prostate cancer prevention and treatment. J Hum Nutr Diet 22, 187–199.

Madabhushi, A., Udupa, J., Souza, A., 2006. Generalized scale: Theory, algorithms, and application to image inhomogeneity correction . Computer Vision and Image Understanding 101, 100 – 121.

Madabhushi, A., Udupa, J.K., 2006. New methods of MR image intensity standardization via generalized scale. Med Phys 33, 3426–3434.

Maintz, J.B., Viergever, M.A., 1998. A survey of medical image registration. Med Image Anal 2, 1–36.

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

Manjon, J.V., Carbonell-Caballero, J., Lull, J.J., Garcia-Marti, G., Marti-Bonmati, L., Robles, M., 2008. MRI denoising using non-local means. Med Image Anal 12, 514–523.

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

Mazzetti, S., De Luca, M., Bracco, C., Vignati, A., Giannini, V., Stasi, M., Russo, F., Armando, E., Agliozzo, S., Regge, D., 2011. 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, pp. 79633Q–79633Q–7.

McNeal, J.E., 1981. The zonal anatomy of the prostate. Prostate 2, 35–49.

McNeal, J.E., Redwine, E.A., Freiha, F.S., Stamey, T.A., 1988. Zonal distribution of prostatic adenocarcinoma. Correlation with histologic pattern and direction of spread. Am. J. Surg. Pathol. 12, 897–906.

Middleton, D., Esposito, R., 1968. Simultaneous optimum detection and estimation of signals in noise. Information Theory, IEEE Transactions on 14, 434–444.

Mitra, J., 2012. Multimodal image registration applied to magnetic resonance and ultrasound prostatic images. Ph.D. thesis. Universitat de Girona and Université de Bourgogne.

Mitra, J., Kato, Z., Marti, R., Oliver, A., Llado, X., Sidibe, D., Ghose, S., Vilanova, J.C., Comet, J., Meriaudeau, F., 2012. A spline-based non-linear diffeomorphism for multimodal prostate registration. Med Image Anal 16, 1259–1279.

Mitra, J., Marti, R., Oliver, A., Llado, X., Vilanova, J.C., Meriaudeau, F., 2011. 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.

Mohan, J., Krishnaveni, V., Guo, Y., 2014. A survey on the magnetic resonance image denoising methods. Biomedical Signal Processing and Control 9, 56 – 69.

Moore, C.M., Ridout, A., Emberton, M., 2013. The role of MRI in active surveillance of prostate cancer. Curr Opin Urol 23, 261–267.

Morgan, R., Boxall, A., Bhatt, A., Bailey, M., Hindley, R., Langley, S., Whitaker, H.C., Neal, D.E., Ismail, M., Whitaker, H., Annels, N., Michael, A., Pandha, H., 2011. Engrailed-2 (EN2): a tumor specific urinary biomarker for the early diagnosis of prostate cancer. Clin. Cancer Res. 17, 1090–1098.

Nelder, J.A., Mead, R., 1965. A simplex method for function minimization. The Computer Journal 7, 308–313.

Niaf, E., Rouviere, O., Mege-Lechevallier, F., Bratan, F., Lartizien, C., 2012. Computer-aided diagnosis of prostate cancer in the peripheral zone using multiparametric MRI. Phys Med Biol 57, 3833–3851.

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

van Niekerk, C.G., van der Laak, J.A., Borger, M.E., Huisman, H.J., Witjes, J.A., Barentsz, J.O., Hulsbergen-van de Kaa, C.A., 2009. Computerized whole slide quantification shows increased microvascular density in pT2 prostate cancer as compared to normal prostate tissue. Prostate 69, 62–69.

van Niekerk, C.G., Witjes, J.A., Barentsz, J.O., van der Laak, J.A., Hulsbergen-van de Kaa, C.A., 2013. Microvascularity in transition zone prostate tumors resembles normal prostatic tissue. Prostate 73, 467–475.

Noguchi, M., Stamey, T.A., McNeal, J.E., Yemoto, C.M., 2001. 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. 166, 104–109.

Nowak, R., 1999. Wavelet-based Rician noise removal for magnetic resonance imaging. Image Processing, IEEE Transactions on 8, 1408–1419.

Nyul, L.G., Udupa, J.K., 1999. On standardizing the MR image intensity scale. Magn Reson Med 42, 1072–1081.

Nyul, L.G., Udupa, J.K., Zhang, X., 2000. New variants of a method of MRI scale standardization. IEEE Trans Med Imaging 19, 143–150.

Ojala, T., Pietikäinen, M., Harwood, D., 1996. A comparative study of texture measures with classification based on featured distributions. Pattern Recognition 29, 51–59.

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

Oster, G., Lamerato, L., Glass, A.G., Richert-Boe, K.E., Lopez, A., Chung, K., Richhariya, A., Dodge, T., Wolff, G.G., Balakumaran, A., Edelsberg, J., 2013. 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 21, 3279–3286.

Ozer, S., Haider, M., Langer, D.L., Van der Kwast, T.H., Evans, A., Wernick, M., Trachtenberg, J., Yetik, I., 2009. 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,

pp. 73–76.

Ozer, S., Langer, D.L., Liu, X., Haider, M.A., van der Kwast, T.H., Evans, A.J., Yang, Y., Wernick, M.N., Yetik, I.S., 2010. Supervised and unsupervised methods for prostate cancer segmentation with multispectral MRI. Med Phys 37, 1873–1883.

Padhani, A.R., 2002. Dynamic contrast-enhanced MRI in clinical oncology: current status and future directions. J Magn Reson Imaging 16, 407–422.

Padhani, A.R., 2011. Integrating multiparametric prostate MRI into clinical practice. Cancer Imaging 11 Spec No A, 27–37.

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

Parfait, S., Walker, P., Crhange, G., Tizon, X., Mitran, J., 2012. Classification of prostate magnetic resonance spectra using Support Vector Machine . Biomedical Signal Processing and Control 7, 499 – 508.

Pearson, K., 1901. On lines and planes of closest fit to systems of points in space. Philosophical Magazine 2, 559–572.

Peng, H., Long, F., Ding, C., 2005. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. Pattern Analysis and Machine Intelligence, IEEE Transactions on 27, 1226–1238.

Peng, Y., Jiang, Y., Yang, C., Brown, J., Antic, T., Sethi, I., Schmid-Tannwald, C., Giger, M., Eggener, S., Oto, A., 2013. 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 267, 787–796.

Petrick, N., Haider, M., Summers, R.M., Yeshwant, S.C., Brown, L., Iuliano, E.M., Louie, A., Choi, J.R., Pickhardt, P.J., 2008. CT colonography with computer-aided detection as a second reader: observer performance study. Radiology 246, 148–156.

Pijnappel, W., van den Boogaart, A., de Beer, R., van Ormondt, D., 1992. SVD-based quantification of magnetic resonance signals . Journal of Magnetic Resonance (1969) 97, 122 – 134.

Pizurica, A., 2002. Image denoising using wavelets and spatial context modeling. Ph.D. thesis. Universiteit Gent.

Pizurica, A., Philips, W., Lemahieu, I., Acheroy, M., 2003. A versatile wavelet domain noise filtration technique for medical imaging. IEEE Trans Med Imaging 22, 323–331.

Pluim, J., Maintz, J., Viergever, M., 2003. Mutual-information-based registration of medical images: a survey. IEEE Transcations on Medical Imaging 22, 986–1004.

Prewitt, J., 1970. Picture processing and psychohistories. Academic Press. chapter Object enhancement and extraction.

Provencher, S.W., 1993. Estimation of metabolite concentrations from localized in vivo proton NMR spectra. Magn Reson Med 30, 672–679.

Puech, P., Betrouni, N., Makni, N., Dewalle, A.S., Villers, A., Lemaitre, L., 2009. Computer-assisted diagnosis of prostate cancer using DCE-MRI data: design, implementation and preliminary results. Int J Comput Assist Radiol Surg 4, 1–10.

Quinlan, J., 1986. Induction of decision trees. Machine Learning 1, 81–106.

Quinlan, J., 1993. C4.5: Programs for machine learning. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

Quinonero-Candela, J., Girard, A., Rasmussen, C., 2002. Prediction at an Uncertain Input for Gaussian processes and relevance vector machines application to Multiple-Step ahead time-series forecasting. Technical Report. DTU Informatics.

Quint, L.E., Van Erp, J.S., Bland, P.H., Mandell, S.H., Del Buono, E.A., Grossman, H.B., Glazer, G.M., Gikas, P.W., 1991. Carcinoma of the prostate: MR images obtained with body coils do not accurately reflect tumor volume. AJR Am J Roentgenol 156, 511–516.

Rangayyan, R., Ayres, F., Desautels, J., 2007. A review of computer-aided diagnosis of breast cancer: Toward the detection of subtle signs. Journal of the Franklin Institute 344, 312 – 348.

Rasmussen, C., Williams, C., 2005. Gaussian processes for machine learning. The MIT Press.

Ratiney, H., Sdika, M., Coenradie, Y., Cavassila, S., van Ormondt, D., Graveron-Demilly, D., 2005. Time-domain semi-parametric estimation based on a metabolite basis set. NMR Biomed 18, 1–13.

Rish, I., 2001. An empirical study of the naive Bayes classifier, in: IJCAI 2001 workshop on empirical methods in artificial intelligence, pp. 41–46.

Rodriguez, C., Freedland, S.J., Deka, A., Jacobs, E.J., McCullough, M.L., Patel, A.V., Thun, M.J., Calle, E.E., 2007. Body mass index, weight change, and risk of prostate cancer in the Cancer Prevention Study II Nutrition Cohort. Cancer Epidemiol. Biomarkers Prev. 16, 63–69.

Rosenkrantz, A.B., Sabach, A., Babb, J.S., Matza, B.W., Taneja, S.S., Deng, F.M., 2013. Prostate cancer: comparison of dynamic contrast-enhanced MRI techniques for localization of peripheral zone tumor. AJR Am J Roentgenol 201, W471–478.

Roweis, S.T., Saul, L.K., 2000. Nonlinear dimensionality reduction by locally linear embedding. Science 290, 2323–2326.

Rueckert, D., Sonoda, L.I., Hayes, C., Hill, D.L., Leach, M.O., Hawkes, D.J., 1999. Nonrigid registration using free-form deformations: application to breast MR images. IEEE Trans Med Imaging 18, 712–721.

Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1988. Learning internal representations by error propagation, in: Anderson, J.A., Rosenfeld, E. (Eds.), Neurocomputing: foundations of research. MIT Press, Cambridge, MA, USA. chapter Learning Internal Representations by Error Propagation, pp. 673–695.

Saeys, Y., Inza, I., Larranaga, P., 2007. A review of feature selection techniques in bioinformatics. Bioinformatics 23, 2507–2517.

Scheidler, J., Hricak, H., Vigneron, D.B., Yu, K.K., Sokolov, D.L., Huang, L.R., Zaloudek, C.J., Nelson, S.J., Carroll, P.R., Kurhanewicz, J., 1999a. Prostate cancer: localization with three-dimensional proton MR spectroscopic imaging–clinicopathologic study. Radiology 213, 473–480.

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

Schlemmer, H.P., Merkle, J., Grobholz, R., Jaeger, T., Michel, M.S., Werner, A., Rabe, J., van Kaick, G., 2004. Can pre-operative contrast-enhanced dynamic MR imaging for prostate cancer predict microvessel density in prostatectomy specimens? Eur Radiol 14, 309–317.

Schroder, F.H., Carter, H.B., Wolters, T., van den Bergh, R.C., Gosselaar, C., Bangma, C.H., Roobol, M.J., 2008. Early detection of prostate cancer in 2007. Part 1: PSA and PSA kinetics. Eur. Urol. 53, 468–477.

Schröder, F.H., Hugosson, J., Roobol, M.J., Tammela, T.L., Ciatto, S., Nelen, V., Kwiatkowski, M., Lujan, M., Lilja, H., Zappa, M., Denis, L.J., Recker, F., Pez, A., Määttänen, L., Bangma, C.H., Aus, G., Carlsson, S., Villers, A., Rebillard, X., van der Kwast, T., Kujala, P.M., Blijenberg, B.G., Stenman, U.H., Huber, A., Taari, K., Hakama, M., Moss, S.M., de Koning, H.J., Auvinen, A., 2012. Prostate-cancer mortality at 11 years of follow-up. New England Journal of Medicine 366, 981–990.

Shapiro, L.G., Stockman, G.C., 2001. Computer vision. Prentice Hall, Upper Saddle River, NJ.

Shi, J., Malik, J., 2000. Normalized cuts and image segmentation. Pattern Analysis and Machine Intelligence, IEEE Transactions on 22, 888–905.

Shimofusa, R., Fujimoto, H., Akamata, H., Motoori, K., Yamamoto, S., Ueda, T., Ito, H., 2005. Diffusion-weighted imaging of prostate cancer. J Comput Assist Tomogr 29, 149–153.

Siegel, R., Naishadham, D., Jemal, A., 2013. Cancer statistics, 2013. CA Cancer J Clin 63, 11–30.

Sled, J.G., Zijdenbos, A.P., Evans, A.C., 1998. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Trans Med Imaging 17, 87–97.

Sobel, I., 1970. Camera models and machine perception. Technical Report. DTIC Document.

Somford, D.M., Futterer, J.J., Hambrock, T., Barentsz, J.O., 2008. Diffusion and perfusion MR imaging of the prostate. Magn Reson Imaging Clin N Am 16, 685–695.

Specht, D.F., 1988. Probabilistic neural networks for classification, mapping, or associative memory, in: Neural Networks, 1988., IEEE International Conference on, pp. 525–532 vol.1.

St Lawrence, K.S., Lee, T.Y., 1998. An adiabatic approximation to the tissue homogeneity model for water exchange in the brain: I. Theoretical derivation. J. Cereb. Blood Flow Metab. 18, 1365–1377.

Stamey, T.A., Donaldson, A.N., Yemoto, C.E., McNeal, J.E., Sozen, S., Gill, H., 1998. Histological and clinical findings in 896 consecutive prostates treated only with radical retropubic prostatectomy: epidemiologic significance of annual changes. J. Urol. 160, 2412–2417.

Staring, M., van der Heide, U.A., Klein, S., Viergever, M.A., Pluim, J.P., 2009. Registration of cervical MRI using multifeature mutual information. IEEE Trans Med Imaging 28, 1412–1421.

Steinberg, G.D., Carter, B.S., Beaty, T.H., Childs, B., Walsh, P.C., 1990. Family

42

history and the risk of prostate cancer. Prostate 17, 337–347.

Strum, S., Pogliano, D., 2005. What every doctor who treats male patients should know. PCRI Insights vol. 8, no. 2.

Styner, M., Brechbuhler, C., Szckely, G., Gerig, G., 2000. Parametric estimate of intensity inhomogeneities applied to MRI. Medical Imaging, IEEE Transactions on 19, 153–165.

Styner, M., Gerig, G., 1997. Evaluation of 2D/3D bias correction with 1+1ES-optimization. Technical Report. ETH Zürich.

Sung, Y.S., Kwon, H.J., Park, B.W., Cho, G., Lee, C.K., Cho, K.S., Kim, J.K., 2011. Prostate cancer detection on dynamic contrast-enhanced MRI: computer-aided diagnosis versus single perfusion parameter maps. AJR Am J Roentgenol 197, 1122–1129.

Suzuki, K., 2012. A review of computer-aided diagnosis in thoracic and colonic imaging. Quant Imaging Med Surg 2, 163–176.

Swanson, M.G., Vigneron, D.B., Tran, T.K., Sailasuta, N., Hurd, R.E., Kurhanewicz, J., 2001. Single-voxel oversampled J-resolved spectroscopy of in vivo human prostate tissue. Magn Reson Med 45, 973–980.

Taira, A.V., Merrick, G.S., Galbreath, R.W., Andreini, H., Taubenslag, W., Curtis, R., Butler, W.M., Adamovich, E., Wallner, K.E., 2010. Performance of transperineal template-guided mapping biopsy in detecting prostate cancer in the initial and repeat biopsy setting. Prostate Cancer Prostatic Dis. 13, 71–77.

Tipping, M., 2001. Sparse Bayesian learning and the relevance vector machine. Journal of Machine Learning Research 1, 211–244.

Tiwari, P., Kurhanewicz, J., Madabhushi, A., 2013. Multi-kernel graph embedding for detection, Gleason grading of prostate cancer via MRI/MRS. Med Image Anal 17, 219–235.

Tiwari, P., Kurhanewicz, J., Rosen, M., Madabhushi, A., 2010. Semi supervised multi kernel (SeSMiK) graph embedding: identifying aggressive prostate cancer via magnetic resonance imaging and spectroscopy. Med Image Comput Comput Assist Interv 13, 666–673.

Tiwari, P., Madabhushi, A., Rosen, M., 2007. A hierarchical unsupervised spectral clustering scheme for detection of prostate cancer from magnetic resonance spectroscopy (MRS). Med Image Comput Comput Assist Interv 10, 278–286.

Tiwari, P., Rosen, M., Madabhushi, A., 2008. Consensus-locally linear embedding (C-LLE): application to prostate cancer detection on magnetic resonance spectroscopy. Med Image Comput Comput Assist Interv 11, 330–338.

Tiwari, P., Rosen, M., Madabhushi, A., 2009a. A hierarchical spectral clustering and nonlinear dimensionality reduction scheme for detection of prostate cancer from magnetic resonance spectroscopy (MRS). Med Phys 36, 3927–3939.

Tiwari, P., Rosen, M., Reed, G., Kurhanewicz, J., Madabhushi, A., 2009b. Spectral embedding based probabilistic boosting tree (ScEPTre): classifying high dimensional heterogeneous biomedical data. Med Image Comput Comput Assist Interv 12, 844–851.

Tiwari, P., Viswanath, S., Kurhanewicz, J., Sridhar, A., Madabhushi, A., 2012. Multimodal wavelet embedding representation for data combination (MaWERiC): integrating magnetic resonance imaging and spectroscopy for prostate cancer detection. NMR Biomed 25, 607–619.

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

Tofts, P.S., 1997. Modeling tracer kinetics in dynamic Gd-DTPA MR imaging. J Magn Reson Imaging 7, 91–101.

Toth, R., Chappelow, J., Rosen, M., Pungavkar, S., Kalyanpur, A., Madabhushi, A., 2008. Multi-attribute non-initializing texture reconstruction based active shape model (MANTRA). Med Image Comput Comput Assist Interv 11, 653–661.

Toth, R., Doyle, S., Pungavkar, S., Kalyanpur, A., Madabhushi, A., 2009. A boosted ensemble scheme for accurate landmark detection for active shape models, in: SPIE Medical Imaging, Orlando, FL.

Tu, Z., 2005. Probabilistic boosting-tree: learning discriminative models for classification, recognition, and clustering, in: Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on, pp. 1589–1596 Vol. 2.

Turkbey, B., Choyke, P.L., 2012. Multiparametric MRI and prostate cancer diagnosis and risk stratification. Curr Opin Urol 22, 310–315.

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

Vapnik, V., Lerner, A., 1963. Pattern Recognition using Generalized Portrait

Method. Automation and Remote Control 24.

Verma, S., Rajesh, A., Futterer, J.J., Turkbey, B., Scheenen, T.W., Pang, Y., Choyke, P.L., Kurhanewicz, J., 2010. Prostate MRI and 3D MR spectroscopy: how we do it. AJR Am J Roentgenol 194, 1414–1426.

Verma, S., Turkbey, B., Muradyan, N., Rajesh, A., Cornud, F., Haider, M.A., Choyke, P.L., Harisinghani, M., 2012. Overview of dynamic contrast-enhanced MRI in prostate cancer diagnosis and management. AJR Am J Roentgenol 198, 1277–1288.

Vilanova, J.C., Comet, J., Barceló-Vidal, C., Barceló, J., López-Bonet, E., Maroto, A., Arzoz, M., Moreno, À., Areal, J., 2009. 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 253, 135–143.

Villers, A., Steg, A., Boccon-Gibod, L., 1991. Anatomy of the prostate: review of the different models. Eur. Urol. 20, 261–268.

Viola, P., Wells, III, W.M., 1997. Alignment by maximization of mutual information. Int. J. Comput. Vision 24, 137–154.

Viswanath, S., Bloch, B.N., Chappelow, J., Patel, P., Rofsky, N., Lenkinski, R., Genega, E., Madabhushi, A., 2011. 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.

Viswanath, S., Bloch, B.N., Genega, E., Rofsky, N., Lenkinski, R., Chappelow, J., Toth, R., Madabhushi, A., 2008a. A comprehensive segmentation, registration, and cancer detection scheme on 3 Tesla in vivo prostate DCE-MRI. Med Image Comput Comput Assist Interv 11, 662–669.

Viswanath, S., Bloch, B.N., Rosen, M., Chappelow, J., Toth, R., Rofsky, N., Lenkinski, R., Genega, E., Kalyanpur, A., Madabhushi, A., 2009. 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.

Viswanath, S., Tiwari, P., Rosen, M., Madabhushi, A., 2008b. 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, SPIE.

Viswanath, S.E., Bloch, N.B., Chappelow, J.C., Toth, R., Rofsky, N.M., Genega, E.M., Lenkinski, R.E., Madabhushi, A., 2012. 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 36, 213–224.

Vos, P.C., Barentsz, J.O., Karssemeijer, N., Huisman, H.J., 2012. Automatic computer-aided detection of prostate cancer based on multiparametric magnetic resonance image analysis. Phys Med Biol 57, 1527–1542.

Vos, P.C., Hambrock, T., Barentsz, J.O., Huisman, H.J., 2008a. Combining T2-weighted with dynamic MR images for computerized classification of prostate lesions, in: Medical Imaging 2008: Computer-Aided Diagnosis, SPIE.

Vos, P.C., Hambrock, T., Barentsz, J.O., Huisman, H.J., 2010. Computer-assisted analysis of peripheral zone prostate lesions using T2-weighted and dynamic contrast enhanced T1-weighted MRI. Phys Med Biol 55, 1719–1734.

Vos, P.C., Hambrock, T., Hulsbergen-van de Kaa, C.A., Futterer, J.J., Barentsz, J.O., Huisman, H.J., 2008b. Computerized analysis of prostate lesions in the peripheral zone using dynamic contrast enhanced MRI. Med Phys 35, 888–899.

Vovk, U., Pernus, F., Likar, B., 2007. A review of methods for correction of intensity inhomogeneity in MRI. Medical Imaging, IEEE Transactions on 26, 405–421.

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

Wang, L., Mazaheri, Y., Zhang, J., Ishill, N.M., Kuroiwa, K., Hricak, H., 2008. Assessment of biologic aggressiveness of prostate cancer: correlation of MR signal intensity with Gleason grade after radical prostatectomy. Radiology 246, 168–176.

Warfield, S.K., Zou, K.H., Wells, W.M., 2004. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. IEEE Trans Med Imaging 23, 903–921.

Wiart, M., Curiel, L., Gelet, A., Lyonnet, D., Chapelon, J.Y., Rouviere, O., 2007. Influence of perfusion on high-intensity focused ultrasound prostate ablation: a first-pass MRI study. Magn Reson Med 58, 119–127.

Ye, L., Kynaston, H.G., Jiang, W.G., 2007. Bone metastasis in prostate cancer:

molecular and cellular mechanisms (Review). Int. J. Mol. Med. 20, 103–111.

Zelhof, B., Lowry, M., Rodrigues, G., Kraus, S., Turnbull, L., 2009. Description of magnetic resonance imaging-derived enhancement variables in pathologically confirmed prostate cancer and normal peripheral zone regions. BJU Int. 104, 621–627.

Zhao, G., Ahonen, T., Matas, J., Pietikainen, M., 2012. Rotation-Invariant Image and Video Description With Local Binary Pattern Features. Image Processing, IEEE Transactions on 21, 1465–1477.

Zhu, H., Ouwerkerk, R., Barker, P.B., 2010. Dual-band water and lipid suppression for MR spectroscopic imaging at 3 Tesla. Magn Reson Med 63, 1486–1492.

Zitová, B., Flusser, J., 2003. Image registration methods: a survey. Image and Vision Computing 21, 977 – 1000.