

Computer-Aided Detection and Diagnosis for prostate cancer based on mono and multiparametric MRI/MRSI: A review

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

^aLE2I-UMR CNRS 6306, Université de Bourgogne, 12 rue de la Fonderie, 71200 Le Creusot, France

^bLE2I-UMR CNRS 6306, Université de Bourgogne, Avenue Alain Savary, 21000 Dijon, France

^cViCOROB, Universitat de Girona, Campus Montilivi, Edifici P4, 17071 Girona, Spain

^dGirona Magnetic Resonance Center, 26 Carrer Joan Maragall, 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 radiologist 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 review aimed to provide an overview of the researches carried out in this lapse of time. Computer-aided system is composed of different stages and this review aimed at reporting the different works carried out. 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,

*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)

5 synonymous with erroneous detection or interpretations of abnormalities throughout
6 clinical decisions (Giger et al. (2008), Hambrock et al. (2013)). These errors are
7 driven by two majors causes (Giger et al. (2008)). On the one hand, observer lim-
8 itations (e.g., constrained human visual perception, fatigue or distraction) are the
9 principal human issues. On the other hand, the second reason is linked to the clinical
10 cases themselves, for instance due to unbalance data (number of healthy cases more
11 abundant than malignant cases) or overlapping structures resulting from limitations
12 of imaging techniques.

13 Computer vision has given rise to many promising solutions. Instead of sup-
14 plying fully automatic computerized solutions, researchers have aimed at providing
15 computer image analysis techniques to aid radiologists in their clinical decisions
16 (Giger et al. (2008)). In fact, these investigations brought about both concepts of
17 computer-aided detection (CADe) and computer-aided diagnosis (CADx) grouped
18 under the acronym CAD. Since those first steps, evidence has shown that CAD sys-
19 tems enhance the diagnosis performance of radiologists. Chan et al. (1999) reported
20 a significant 4 % improvement in breast cancer detection, in accordance with later
21 studies (Dean and Ilvento (2006)). Similar conclusions were drawn in the case of lung
22 nodule detection (Li et al. (2004)), colon cancer (Petrick et al. (2008)) and prostate
23 cancer (CaP) as well (Hambrock et al. (2013)). Chan et al. (1999) also hypothesized
24 that CAD systems will be even more efficient assisting inexperienced radiologists to
25 senior radiologists. That hypothesis was tested by Hambrock et al. (2013) and was
26 confirmed in case of CaP detection. In this particular study, inexperienced radiol-
27 ogists obtained equivalent performance than senior radiologists, both with help of
28 CAD system whereas the accuracy of their diagnosis was significantly poorer without
29 this assistance.

30 In the late eighties, the first CAD systems were developed to detect anomalies on
31 chest radiographies and mammograms (Chan et al. (1987), Doi et al. (1990), Giger
32 et al. (1988)). In the past twenty years, extensive investigations were conducted in
33 the advancement of CAD systems, migrating from intensive time consuming algo-
34 rithms performed on reduced number of cases to “fast” processing on large medical
35 dataset. These works were focused on diverse organ cancer diagnosis making use of
36 numerous imaging modalities: micro-calcification detection in breast mammography
37 (Elter and Horsch (2009), Rangayyan et al. (2007)) and ultrasound (US) imaging
38 (Cheng et al. (2010)), lung nodules detection based on computer tomography (CT)
39 (Chan et al. (2008), Suzuki (2012)), colon tumours detection (Suzuki (2012)) and
40 melanoma detection using dermoscopy imaging (Korotkov and Garcia (2012)). Not-
41 ing the abundance of diverse CAD systems, these fields achieved a certain maturity
42 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 works.

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 progresses. 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) allowing to 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. 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.

This paper will be 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 presented as well as introduction about CAD

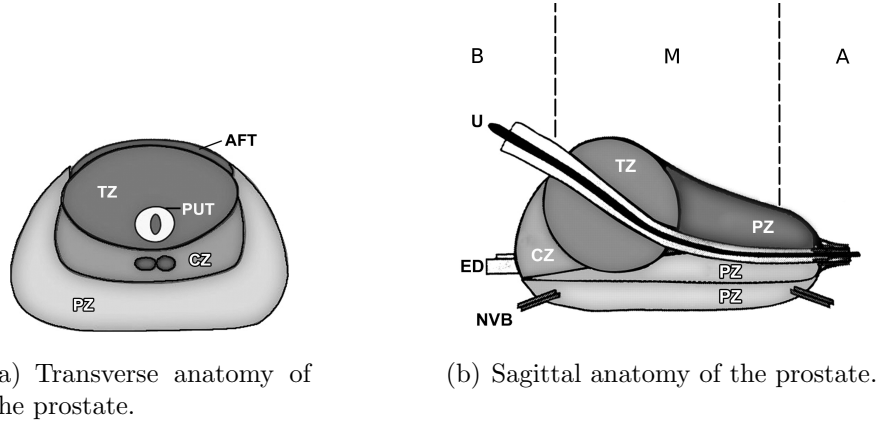


Figure 1: Prostate anatomy with division in different zones. *AFT*: anterior fibromuscular tissue, *CZ*: central zone, *ED*: ejaculatory duct, *NVB*: neurovascular bundle, *PUT*: tissue, *PZ*: peripheral zone, *U*: urethra, *TZ*: transitional zone, *B*: base, *M*: median, *A*: apex (copyright by Choi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

framework will be discussed. Sections 3 - 4 review techniques used in different steps implied in CAD work-flow which will be our main contribution. Image regularization framework including pre-processing methods (see Sect. 3.1), segmentation (see Sect. 3.2) and registration (see Sect. 3.3) will be covered as well as the image classification framework whose feature detection (see Sect.4.2), feature selection and extraction (see Sect. 4.3) and feature classification (see Sect. 4.4) are belonging to. 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 in Sect. 2.1 and 2.2. Section 2.3.1 discusses the current screening strategy for CaP and its drawbacks. MRI plays an important role in improving the current strategy and Sect 2.3.2 presents a more detailed description of MRI modalities. These different techniques are used as inputs to the CAD system which is discussed in Sect 2.3.3.

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

100 increases at two distinct stages during physical development: initially at puberty to
101 reach its normal size, then again after sixty years of age leading to benign prostatic
102 hyperplasia (BPH) (Parfait (2010)).

103 A zonal classification of the prostate, depicted in Fig. 1, was suggested by McNeal
104 (McNeal (1981)). Subsequently, this categorization was widely accepted in the lit-
105 erature (cf., Coakley and Hricak (2000), Hricak et al. (1987), Parfait (2010), Villers
106 et al. (1991)) and is used in all medical examinations (e.g., biopsy, MRI screening).
107 The classification is based on dividing the gland into three distinct regions: (i) central
108 zone (CZ) accounting for 20-25% of the whole prostate gland, (ii) transitional zone
109 (TZ) standing for 5% and (iii) peripheral zone (PZ) representing the 70%. In MRI
110 images, tissues of CZ and TZ are difficult to distinguish and are usually merged into
111 a common region, denominated central gland (CG). As part of this classification,
112 the prostate can be divided in three longitudinal portions depicted in Fig. 1(b): (i)
113 base, (ii) median gland and (iii) apex.

114 2.2. Prostate carcinoma

115 CaP was reported on a worldwide scale to be the second most frequently diag-
116 nosed cancer of men accounting for 13.6% (Ferlay et al. (2010)). Statistically, in 2008,
117 the number of new diagnosed cases was estimated to be 899,000 with no less than
118 258,100 deaths (Ferlay et al. (2010)). In United States, aside from skin cancer, CaP
119 was declared to be the most commonly diagnosed cancer among men, implying that
120 approximately one in six men will be diagnosed with CaP during their lifetime and
121 one in thirty-six will die from this disease causing CaP to be the second most com-
122 mon cause of cancer death among men (Siegel et al. (2013), American Cancer Society
123 (2013)).

124 Despite active research to determine the causes of prostate cancer, a fuzzy list of
125 risk factors has arisen (American Cancer Society (2010)). The etiology was linked to
126 the following factors (American Cancer Society (2010)): (i) family history (Giovannu-
127 cci et al. (2007), Steinberg et al. (1990)), (ii) genetic factors (Agalliu et al. (2009),
128 Amundadottir et al. (2006), Freedman et al. (2006)), (iii) race-ethnicity (Giovannucci
129 et al. (2007), Hoffman et al. (2001)), (iv) diet (Alexander et al. (2010), Giovannucci
130 et al. (2007), Ma and Chapman (2009)), (v) obesity (Giovannucci et al. (2007), Ro-
131 driguez et al. (2007)). This list of risk factors alone cannot be used to diagnose CaP
132 and in this way, screening enables early detection and treatment.

133 CaP growth is characterized by two main types of evolution (Strum and Pogliano
134 (2005)): slow-growing tumours, accounting for up to 85 % of all CaPs (Lu-Yao
135 et al. (2009)), progress slowly and usually stay confined to the prostate gland. For
136 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 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 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

174 pattern, and (ii) the grade of the second most present tumour pattern (Epstein et al.
175 (2005)). A higher Gleason score (GS) indicates a more aggressive tumour (Epstein
176 et al. (2005)). Also, it should be noted that biopsy is an invasive procedure which can
177 result in serious infection or urine retention (Chou et al. (2011), Hara et al. (2005)).

178 Although PSA screening has been shown to improve early detection of CaP (Chou
179 et al. (2011)), its lack of reliability motivates further investigations using MRI-CAD.
180 Two reliable studies, carried out in the United States (Andriole et al. (2009)) and
181 in Europe (Hugosson et al. (2010), Schröder et al. (2012)), have attempted to assess
182 the impact of early detection of CaP, with diverging outcomes (Chou et al. (2011),
183 Heidenreich et al. (2013)). The study carried out in Europe¹ concluded that PSA
184 screening reduces CaP-related mortality by 21-44% (Hugosson et al. (2010), Schröder
185 et al. (2012)), while the American² trial found no such effect (Andriole et al. (2009)).
186 However, both studies agree that PSA screening suffers from low specificity, with an
187 estimated rate of 36 % (Schroder et al. (2008)). Both studies also agree that over-
188 treatment is an issue: decision making of regarding treatment is further complicated
189 by difficulties in evaluating the aggressiveness and progression of CaP (Delpierre
190 et al. (2013)).

191 Hence, new screening methods should be developed with improved specificity of
192 detection as well as more accurate risk assessment (aggressiveness and progression).
193 Current research is focused on identifying new biological markers to replace PSA-
194 based screening (Bourdounis et al. (2010), Brenner et al. (2013), Morgan et al.
195 (2011)). Until such research comes to fruition, these needs can be met through
196 active-surveillance strategy using multi-parametric MRI techniques (Hoeks et al.
197 (2011), Moore et al. (2013)). MRI-CAD system, which is an area of active research
198 and forms the focus of this paper, can be incorporated into this screening strategy
199 allowing a more systematic and rigorous follow-up.

200 Another weakness of the current screening strategy lies in the fact that TRUS
201 biopsy does not provide trustworthy results. Due to its “blind” nature, there is a
202 chance of missing aggressive tumours or detect microfocal “cancers”, which influences
203 the aggressiveness-assessment procedure (Noguchi et al. (2001)). As a consequence,
204 over-diagnosis is estimated at up to 30 % (Haas et al. (2007)), while missed clini-
205 cally significant CaP is estimated at up to 35 % (Taira et al. (2010)). In an effort
206 to solve both issues, alternative biopsy approaches have been explored. MRI/US-

¹The 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.

²The 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.

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. Imaging techniques: MRI - MRSI

MRI provides promising imaging techniques to overcome the previous mentioned drawbacks. Unlike TRUS biopsy, MRI examination 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 later presented, 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 2.3.2 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 as shown in Fig. 3(b) (Hricak et al. (1983)). 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 as depicted in Fig. 3(c) (Akin et al. (2006), Barentsz et al. (2012)).

Table 1: Overview of the features associated with each MRI-MRSI. 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)	lower concentration	higher concentration	+
	Choline (3.21 ppm)	higher concentration	lower concentration	0
	Spermine (3.11 ppm)	lower concentration	higher concentration	+

Notes:

+ = significantly correlated.

0 = no correlation.

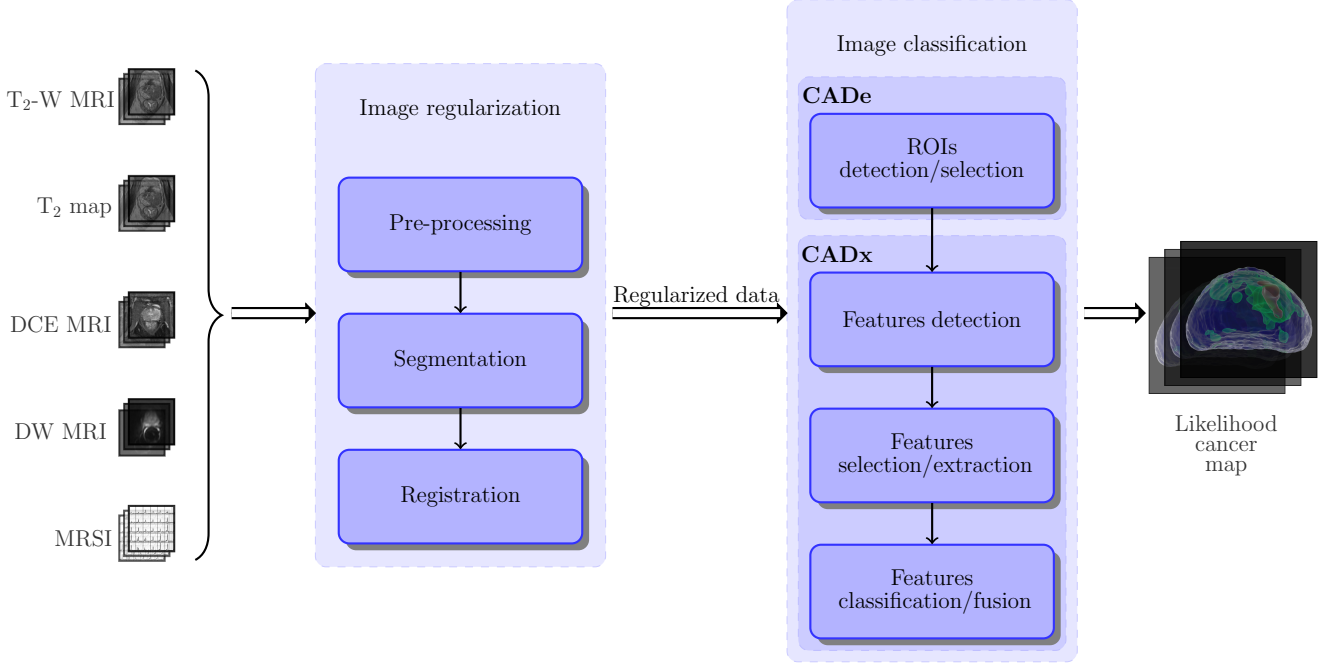
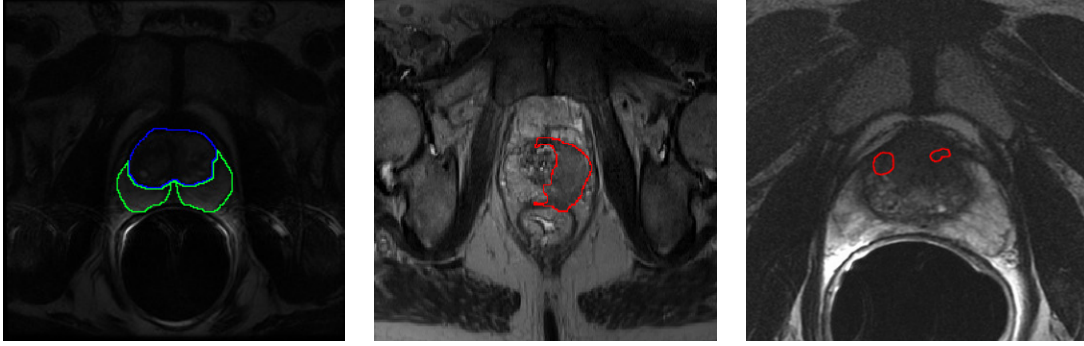


Figure 2: CAD framework using MRI images. Multiparametric MRI images are provided as inputs. These data arise from heterogeneous sources and need to be regularized. Some studies does not considered this stage as mandatory and does not implement or only partly those processes (see Tab. 2). A pre-processing 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) T₂-W-mri 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) T₂-W-mri slice of a prostate with a CaP highlighted in the PZ using a 3.0 Tesla MRI scanner.

(c) T₂-W-mri slice of a prostate with a CaP highlighted in the CG using a 3.0 Tesla MRI scanner.

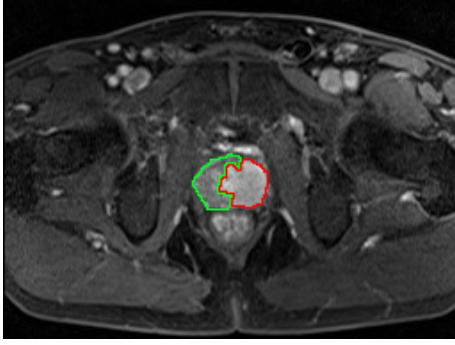
Figure 3: Rendering of T₂-W-MRI prostate image with both 1.5 and 3.0 Tesla MRI scanner.

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)).

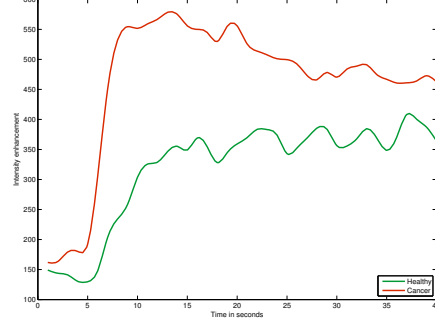
In spite of the availability of these useful and encouraging features, the T₂-W modality lacks reliability (Hoeks et al. (2011), Kirkham et al. (2006)). Sensitivity is affected by the difficulties in detecting cancers in CG (Kirkham et al. (2006)) while specificity rate is highly affected by outliers (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.

- **T₂ Map:** As previously mentioned, T₂-W MRI modality shows low sensitivity. Moreover, T₂-W MRI images are a composite of multiple effects (Hegde et al. (2013)). However, T₂ 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. (1996b, 1997)).

T₂ values are computed using the characteristics of transverse relaxation.



(a) T₁-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.

261 Transverse relaxation is formalized as shown in Eq. 1.

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

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

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

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

265 Fast Spin-Echo (FSE) sequence has been shown to be particularly well suited
266 in order to build a T_2 map and obtain accurate T_2 values (Liney et al. (1996a)).

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

269 — **DCE MRI:** DCE MRI is an imaging technique which exploits the vascular-
270 ity characteristic of tissues. Contrast media, usually gadolinium-based, is in-
271 jected intravenously into the patient. The media extravasates from vessels to
272 extravascular-extracellular space (EES) and then released back into the vascu-
273 lature before being eliminated by the kidneys (Gribbestad et al. (2005)). Further-
274 more, 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)).

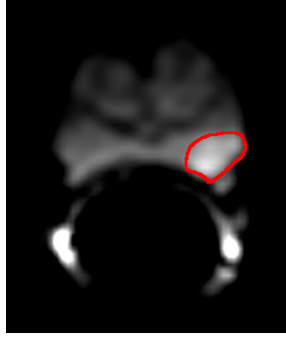
By taking the previous remarks regarding medical aspects and signal theory into account, CaPs are characterized by a signal having an earlier and faster enhancement and an earlier wash-out (cf., the rate of the contrast agent flowing out of the tissue) (see Fig. 4(b)) (Verma et al. (2012)). Three different approaches exist to analyse these signals with the aim of tagging them as corresponding to either normal or malignant tissues. Qualitative analysis is based on assessment of the signal shape (Hoeks et al. (2011)). Quantitative approaches consist in inferring pharmacokinetic 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 (see Fig. 4) (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.

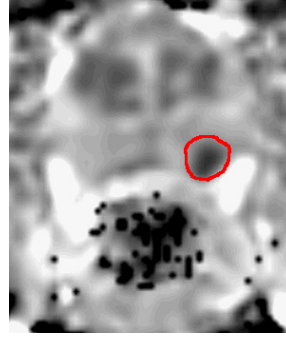
DCE MRI combined with T₂-W MRI has shown to enhance sensitivity compared to T₂-W MRI alone (Jager et al. (1997), Kim et al. (2005), Schlemmer et al. (2004), Zelhof et al. (2009)). Despite this fact, DCE MRI possesses some drawbacks. Due to its “dynamic” nature, patient motions during the image acquisition 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 illustrates the challenge of CaP detection in CG (van Niekerk et al. (2013)).

- **DW MRI:** As previously mentioned in the introduction (Sect. 1), DW MRI is the most recent MRI imaging technique developed aiming at CaP detection and diagnosis (Scheidler et al. (1999b)). This modality exploits the variations in 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



(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 images with a 1.5 Tesla MRI scanner. The cancer corresponds to the low SI region highlighted in red.

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

(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 intensity at those locations. Then, a second gradient takes place in order to cancel 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 into a more significant signal loss whereas a lower degree of random motion is synonymous with lower signal loss (Huisman (2003)). Under

368 these conditions, the MRI signal measured is formalized as in Eq. 3.

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

$$S_{\text{ADC}}(b) = \exp(-b \times \text{ADC}) . \quad (4)$$

369 where S_{ADC} refers to signal drop due to diffusion effect, ADC is the apparent
 370 diffusion coefficient and b is the attenuation coefficient depending only on gradient
 371 pulses parameters: (i) gradient intensity and (ii) gradient duration (Le Bihan et al.
 372 (1986)).

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

377 To summarize, in DW MRI images, CaPs are characterized by high-SI com-
 378 pared to normal tissues in PZ and CG as shown in Fig. 5(a) (Barentsz et al.
 379 (2012)). However, some tissues in CG can look similar to CaP with higher SI
 380 (Barentsz et al. (2012)).

381 Diagnosis using DW MRI combined with T_2 -W MRI has shown significant im-
 382 provement compared with T_2 -W MRI alone and provides highly contrasted images
 383 (Choi et al. (2007), Padhani (2011), Shimofusa et al. (2005)). As drawbacks, this
 384 modality suffers from poor spatial resolution and specificity due to false positive
 385 detection (Choi et al. (2007)).

386 With a view to eliminate these drawbacks, radiologists are extracting quanti-
 387 tative maps from DW MRI. This imaging technique is presented next.

388 – **ADC Map:** The NMR signal measured for DW MRI images is not only affected
 389 by diffusion as shown in Eq. 3. However, the signal drop (Eq. 4) is formulated
 390 such that the only variable is the acquisition parameter b (Le Bihan et al. (1986)).
 391 The ADC is considered as a “pure” diffusion coefficient and can be extracted to
 392 build a quantitative map.

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

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

396 To compute the ADC map, a minimum of two acquisitions is necessary: (i) for
 397 $b_0 = 0 \text{ s.mm}^{-2}$ where the measured signal is equal to S_0 , and (ii) $b_1 > 0 \text{ s.mm}^{-2}$

398 (typically 1000 s.mm⁻²). Then, the ADC map can be computed as:

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

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

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

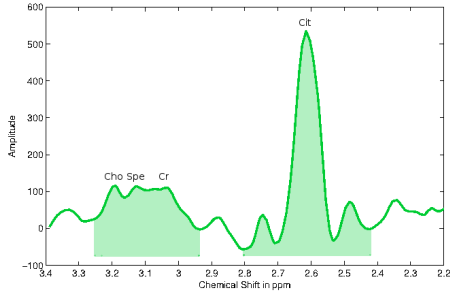
407 Similar to the gain achieved by DW MRI, diagnosis using ADC map combined
408 with T₂-W MRI significantly outperforms T₂-W MRI alone (Choi et al. (2007),
409 Doo et al. (2012)). Moreover, it has been shown that ADC is correlated with GS
410 (Hambrock et al. (2011), Itou et al. (2011), Peng et al. (2013)).

411 However, some tissues of the CG zone mimic CaP with low-SI (Kirkham et al.
412 (2006)). Image distortion can arise due to haemorrhage (Choi et al. (2007)).
413 It has also been noted that high variation of the ADC occurs between patients
414 making difficult to define a static threshold to distinguish CaP from non-malignant
415 tumours (Choi et al. (2007)).

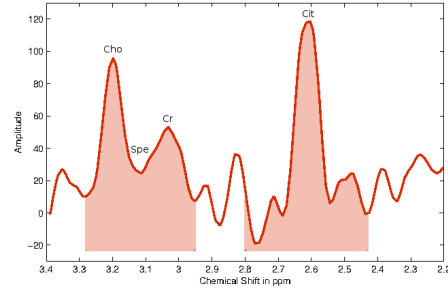
416 – **MRSI:** CaP induces metabolic changes in the prostate compared with healthy
417 tissue. Thus, CaP detection can be carried out by tracking changes of metabo-
418 lite concentration in prostate tissue. MRSI is an NMR-based technique which
419 generates spectra of relative metabolite concentration in a ROI.

420 In order to track changes of metabolite concentration, it is important to know
421 which metabolites are associated with CaP. To address this question, clinical stud-
422 ies identified three biological markers: (i) citrate, (ii) choline and (iii) polyamines
423 composed mainly of spermine, and in less abundance of spermidine and putrescine
424 (Awwad et al. (2012), Costello and Franklin (2006), Giskeodegard et al. (2013)).

425 Citrate is involved in the production and secretion of the prostatic fluid, and
426 the glandular prostate cells are associated with a high production of citrate en-
427 abled by zinc accumulation by these same cells (Costello and Franklin (2006)).
428 However, the metabolism allowing the accumulation of citrate requires a large
429 amount of energy (Costello and Franklin (2006)). In contrast, malignant cells
430 do not have high zinc levels leading to lower citrate levels due to citrate oxyda-



(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 integrated area used for quantification. Acronyms: Choline (Cho), Spermine (Spe), Creatine (Cr) and Citrate (Cit).

tion (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 phospholipid 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 Larmor 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.

Two different quantitative approaches are used to decide either or 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. 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 others 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)). Citrate and spermine concentration 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 makes this modality not normally used in clinical daily 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 addressed 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 more in details differentiate 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 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. 2.3.2.

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 that 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), 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 later processes should only investigate the prostate organ. 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 does not fully apply the methodology depicted in Fig. 2.3.2. 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, lowering the need of registering the multiparametric MRI images.

529 Once the data are regularized, it becomes possible to extract features and classify
530 the data to obtain the probabilistic maps. We referred this stage to image classification
531 where CADE and CADx are the main components.

532 In a CADE framework, possible lesions will be segmented automatically to further
533 classify them as malignant or non-malignant. We also included in CADE studies
534 which considered all voxels as ROIs but in which malignant lesions will be highlighted
535 after the classification stage. However, manual lesions segmentation are not considered
536 as part of a CADE. Therefore, ROIs manually selected or detected through a
537 CADE system are used as inputs of a CADx.

538 CADx is composed of the processes allowing to distinguish malignant from non-
539 malignant tumours. We divided CADx into three different stages. First, salient
540 features are extracted from MRI images to characterize the scene. Of course, more
541 discriminative features will be associated with robust and accurate likelihood cancer
542 map. Most frequently, the number of feature extracted can be large resulting in
543 redundant or not enough discriminative features which will negatively affect the
544 performances of the further classification. Therefore, a step consisting at selecting the
545 best features or/and reducing the number of dimensionality is commonly used. Then,
546 this modified feature vector is finally classified using different pattern recognition
547 approaches.

548 As pointed out in the introduction (see Sect. 1), performances of CaP detection
549 and diagnosis is affected by observer interpretation and limitations (Giger et al.
550 (2008), Hambrock et al. (2013)). CAD offers a possible solution in order to reduce
551 this variability. Lately, effect of CAD on the observer performance has been studied
552 (Hambrock et al. (2013)). Results shown that CADs benefit to less-experienced
553 radiologist to perform similarly as experienced radiologist in their tasks (Hambrock
554 et al. (2013)).

555 2.4. Literature classification

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		
			T ₂ -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.

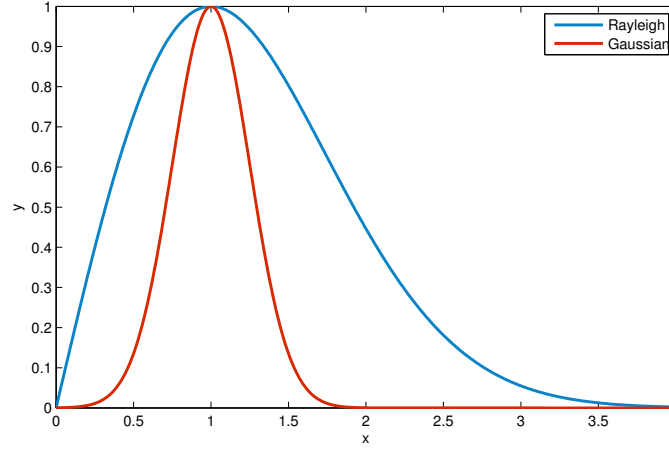


Figure 7: Illustration of a Gaussian distribution ($\mu = 1, \sigma = 0.25$) and a Rayleigh distribution ($\sigma = 2$). It can be seen that the Rayleigh distribution is suffering of a bias term when compared with the Gaussian distribution.

The review will be organized using the methodology presented in Fig. 2.3.2. Methods embedded in the image regularization framework will be presented before to focus on the image classification framework. This latter mentioned will be 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 will be discussed.

3. Image regularization framework

This section provides a review of the methods used in CADs in order to regularized 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, allowing to reduce to the prostate organ the later applied processes, 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.

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

Pre-processing operations	References
<i>MRI pre-processing:</i>	
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]
<i>MRSI pre-processing:</i>	
Phase correction	[22]
Water and lipid residuals filtering	[8]
Baseline correction	[22,31]
Frequency alignment	[31]
Normalization	[22]

573 – **Noise filtering:** The NMR signal measured and recorded in the k-space during
574 an MRI acquisition is affected by noise. This noise obeys a complex Gaussian
575 white noise mainly due to thermal noises in the patient (Nowak (1999)). Fur-
576 thermore, MRI images visualized by radiologists are in fact the magnitude images
577 resulting from the complex Fourier transform of the k-space data. The complex
578 Fourier transform, being a linear and orthogonal transform, does not affect the
579 Gaussian noise characteristics (Nowak (1999)). However, the function involved
580 in the magnitude computation is a non-linear transform (e.g., the square root
581 of the sum of squares of real and the imaginary parts), implying that the noise
582 distribution is no longer Gaussian; it indeed follows a Rician distribution mak-
583 ing the denoising task harder. Briefly, a Rician distribution can be characterized
584 as follows: in low-SI region (low SNR), it can be approximated with a Rayleigh
585 distribution while in high-SI region (high SNR), it is similar to a Gaussian distri-
586 bution (see Fig. 7) (Manjon et al. (2008)). Reviews of all denoising methods can
587 be found in the work of Buades et al. (2005) and Mohan et al. (2014).

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

596 More rational approaches were proposed to overcome this problem. A com-
597 mon method used to denoise MRI images is based on *wavelet-based filtering*. This
598 filtering exploits the sparsity property of the wavelet decomposition. The projec-
599 tion of a noisy signal from the spatial-domain to the wavelet-domain implies that
600 only few wavelet coefficients contribute to the “signal-free noise” while all wavelet
601 coefficients contribute to the noise (Donoho and Johnstone (1994)). Therefore,
602 denoising is performed by thresholding/attenuating the insignificant wavelet coef-
603 ficients to enforce the sparsity in the wavelet-domain. Investigations focus on the
604 strategies to perform the most adequate coefficient shrinkage method (e.g., using
605 thresholding, singularity property or Bayesian framework) (Pizurica (2002)).

606 Ampeliotis et al. (2007, 2008) performed wavelet shrinkage to denoise magni-
607 tude MRI images (cf., T_2 -W-MRI and DCE-MRI) using some thresholding tech-
608 niques (Mallat (2008)). However, since that the wavelet transform is an orthogonal
609 transform, the Rician distribution of the noise is preserved in the wavelet-domain.
610 Hence, for low SNR, the wavelet and scaling coefficients still suffer from a bias

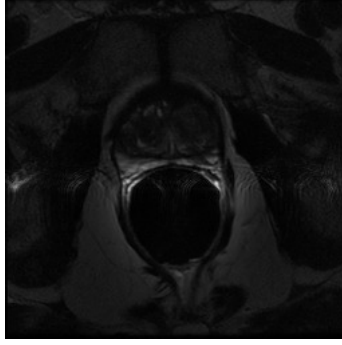


Figure 8: Example of artefacts with high SI due to perturbation from the endorectal coil which create inhomogeneity.

due to this specific noise distribution (Nowak (1999)).

Lopes et al. (2011) used a technique proposed by Pizurica et al. (2003) to denoise T₂-W-MRI. Pizurica et al. (2003) proposed a filtering technique based on the joint detection and estimation theory (Middleton and Esposito (1968)). The wavelet coefficients “free-to-noise” are estimated from the noisy wavelet coefficients using the 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 was used instead of the magnitude MRI image as proposed by Nowak (1999): it involves that the Rician distribution will change to a scaled non-central Chi-square distribution. It implies that the wavelet coefficients are unbiased estimators and the scaling coefficients are unbiased estimators but up to a constant C as defined in Eq. 7 which need to be subtracted from each scaling coefficient.

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

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

- **Bias correction:** Besides being corrupted by noise, MRI images are also affected by the inhomogeneity of the MRI field more commonly named bias field (Styner et al. (2000)). This bias field results in a smooth variation of the SI through the

image (see Fig. 8). As a consequence, the SI of identical tissues varies depending of 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 all bias correction methods can be found in Vovk et al. (2007).

The model of image formation is usually formalized such that:

$$s(\mathbf{x}) = o(\mathbf{x})b(\mathbf{x}) + \eta(\mathbf{x}) . \quad (8)$$

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

Hence, the task of bias correction results in estimated 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 T₂-W-MRI using a *parametric Legendre polynomial* model proposed by Styner et al. (2000) and available in the Insight Segmentation and Registration Toolkit (ITK) library³. Styner et al. (2000) chose to model the bias field by using a linear combination of Legendre polynomials such as:

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

where \hat{b} is the bias estimation with the images coordinates $\mathbf{x} = \{x, y\}$ and the m coefficients of the linear combination $\mathbf{p} = p_{11}, \dots, p_{ij}$; m can be defined such 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 model allows to obtain an image with smooth inhomogeneity 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:

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

³The ITK library is available at: <http://www.itk.org/>

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

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 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 T₂-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:

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

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:

$$S(s) = B(s) * O(s) . \quad (13)$$

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 result 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 latter estimate the corresponding smooth field b .

Lv et al. (2009) corrected the inhomogeneity in T₂-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

⁴The N3 algorithm implementation is available at: <http://www.bic.mni.mcgill.ca/software/N3/>

684 generalized scale (g -scale) and estimating a bias field function based on a second-
685 order polynomial model. First the background of the MRI image is eliminated by
686 thresholding. The threshold value is commonly equal to the mean SI of the con-
687 sidered image. Then, a seeded region growing algorithm is performed considering
688 every thresholded pixel as a potential seed. However, pixels already assigned to
689 a region will not be considered any more as seed. As in seeded region growing
690 algorithm (Shapiro and Stockman (2001)), two criteria are taken into account to
691 expand the region. First, the region will grow using a connected-neighbourhood,
692 initially defined by the user. Then, the homogeneity of SI is based on a fuzzy
693 membership function taking into account the absolute difference of the SIs of two
694 pixels. Depending on the membership value (cf., a threshold has to be defined),
695 the pixel considered is merged or not to the region. Once this segmentation per-
696 formed, the largest region R is used as a mask to select pixels of the original image
697 and the mean SI, μ_R , is computed. The background variation $b(\mathbf{x})$ is estimated
698 such as:

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

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

700 Finally, a second order polynomial $\hat{b}_\Theta(\mathbf{x})$ is fitted in a least-squares sense (Eq.
701 15).

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

702 Finally, the whole original MRI image is corrected by dividing it by the es-
703 timated bias field function $\hat{b}_\Theta(\mathbf{x})$. This process is repeated until the number of
704 pixels in the largest region R is not changing significantly between two iterations.

705 – ***SI normalization/standardization:***

706 As it will be mentioned in the later sections, segmentation or classification
707 tasks are usually performed by first learning from a training set of patients. Hence,
708 one can emphasized the desire to perform MRI examinations with a high repeata-
709 bility or in other words; one would ensure to obtain similar MRI images (cf.,
710 similar SIs) for patients of the same group (cf., healthy patients *vs.* patients with
711 CaP), for a similar sequence.

712 However, it is a known fact that variability between patients occurred dur-
713 ing the MRI examinations even using the same scanner, protocol or sequence
714 parameters (Nyul and Udupa (1999)). Hence, the aim of normalization or stan-
715 dardization 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 T₂-W, DCE and DW MRI images by computing the *standard score* (also called *z-score*) of the pixels of the PZ such as:

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

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, μ_{pz} is the mean-SI of the PZ and σ_{pz} is the SI standard deviation in the PZ.

This transformation enforce the image PDF to have a zero mean and a unit standard deviation.

In a similar way, Liu et al. (2013) normalized T₂-W-MRI by making use of the median and interquartile range for all the pixels.

Lv et al. (2009) scaled the SI of T₂-W-MRI images using the method proposed by Nyul et al. (2000) based on *PDFs matching*. This approach is based on the assumption that the 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 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:

$$\begin{aligned} \Phi_0 &= \{\phi_0^1, \phi_0^2, \dots, \phi_0^N\} , \\ \Phi_{25} &= \{\phi_{25}^1, \phi_{25}^2, \dots, \phi_{25}^N\} , \\ \Phi_{50} &= \{\phi_{50}^1, \phi_{50}^2, \dots, \phi_{50}^N\} , \\ \Phi_{75} &= \{\phi_{75}^1, \phi_{75}^2, \dots, \phi_{75}^N\} , \\ \Phi_{100} &= \{\phi_{100}^1, \phi_{100}^2, \dots, \phi_{100}^N\} . \end{aligned} \quad (17)$$

where $\phi_{n^{\text{th}}}^{i^{\text{th}}}$ is the n^{th} quantile of the i^{th} 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 performed, a linear transformation by parts $\mathcal{T}(\cdot)$ can be computed (Eq. 18) 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

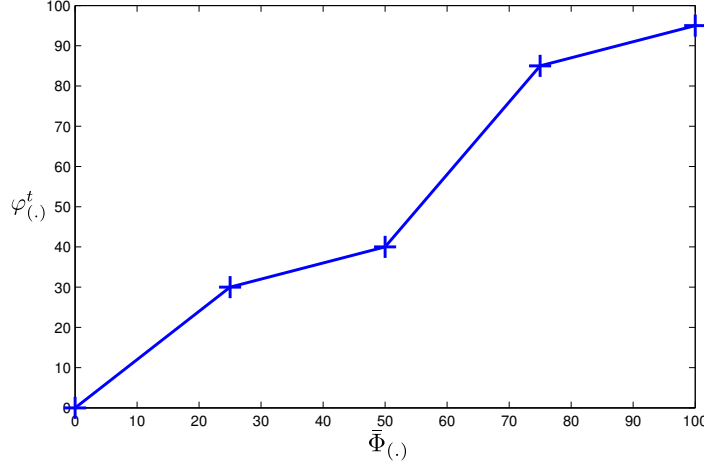


Figure 9: Example of linear mapping by parts as proposed by Nyul et al. (2000).

mapping is also depicted in Fig. 9.

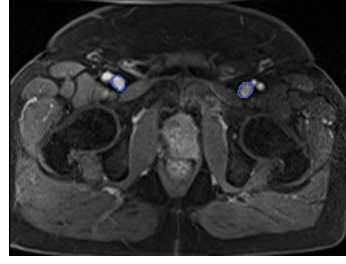
$$\mathcal{T}(s(\mathbf{x})) = \begin{cases} \lceil \bar{\Phi}_0 + (s(\mathbf{x}) - \varphi_0^t) \left(\frac{\bar{\Phi}_{25} - \bar{\Phi}_0}{\varphi_{25}^t - \varphi_0^t} \right) \rceil, & \text{if } \varphi_0^t \leq s(\mathbf{x}) < \varphi_{25}^t, \\ \lceil \bar{\Phi}_{25} + (s(\mathbf{x}) - \varphi_{25}^t) \left(\frac{\bar{\Phi}_{50} - \bar{\Phi}_{25}}{\varphi_{50}^t - \varphi_{25}^t} \right) \rceil, & \text{if } \varphi_{25}^t \leq s(\mathbf{x}) < \varphi_{50}^t, \\ \lceil \bar{\Phi}_{50} + (s(\mathbf{x}) - \varphi_{50}^t) \left(\frac{\bar{\Phi}_{75} - \bar{\Phi}_{50}}{\varphi_{75}^t - \varphi_{50}^t} \right) \rceil, & \text{if } \varphi_{50}^t \leq s(\mathbf{x}) < \varphi_{75}^t, \\ \lceil \bar{\Phi}_{75} + (s(\mathbf{x}) - \varphi_{75}^t) \left(\frac{\bar{\Phi}_{100} - \bar{\Phi}_{75}}{\varphi_{100}^t - \varphi_{75}^t} \right) \rceil, & \text{if } \varphi_{75}^t \leq s(\mathbf{x}) \leq \varphi_{100}^t. \end{cases} \quad (18)$$

Viswanath et al. (2011, 2009, 2012) are using a variant of this previous approach presented in the work of Madabhushi and Udupa (2006) aiming to standardize the T₂-W-MRI images. Instead of computing the PDF of an entire image, a pre-segmentation of the foreground is carried out via *g*-scale which was already 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. 18) takes place in order to align PDFs of the foreground of the MRI images.

The methods described above were statistical-based methods. However, this 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 T₂-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 T₁-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



(a) Illustration and location of the bladder on a T₂-W-MRI image acquire with a 3.0 Tesla MRI scanner



(b) Illustration and location of the femoral arteries on a T₁-W-MRI image acquire with a 3.0 Tesla MRI scanner

Figure 10: Illustration of the two organs used by Niaf et al. (2012, 2011) to normalize T₂-W and T₁-W MRI images.

femoral arteries (see Fig. 10(b)) as proposed by Wiart et al. (2007).

3.1.2. MRSI spectra

As presented 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 such that:

$$\begin{aligned}\Re(o(\xi)) &= \Re(s(\xi)) \cos(\Phi(\xi)) - \Im(\xi) \sin(\Phi(\xi)) , \\ \Im(o(\xi)) &= \Im(s(\xi)) \cos(\Phi(\xi)) + \Re(\xi) \sin(\Phi(\xi)) , \\ \Phi(\xi) &= \phi_0 + \phi_1 \frac{\xi}{N} .\end{aligned}\tag{19}$$

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, the simplex Nelder-Mead optimization

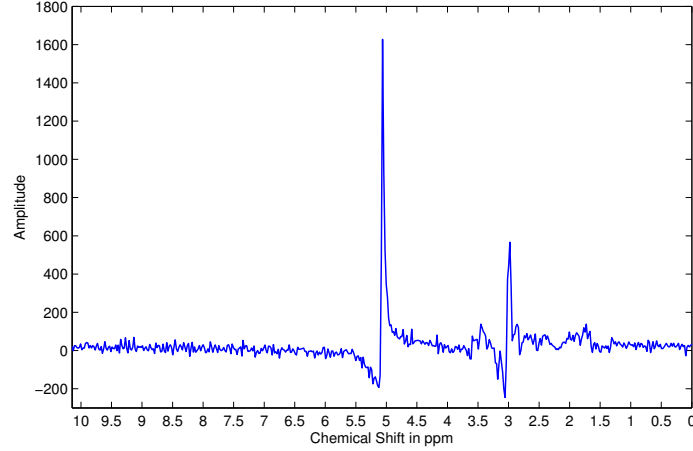


Figure 11: Illustration of phase misalignment in an MRSI spectra acquire with a 3.0 Tesla MRSI scanner.

method was used to minimize the following cost function based on the *Shannon entropy* formulation:

$$\hat{\Phi} = \arg \min_{\Phi} \left[- \sum \Re(s'(\xi)) \ln \Re(s'(\xi)) + \lambda \|\Re(s(\xi))\|_2 \right]. \quad (20)$$

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. 19.

— **Water and lipid residuals filtering:** The water and lipid metabolites occur in much higher concentrations than 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 to enhance the quality of the MRSI spectra by removing these residuals have been proposed.

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

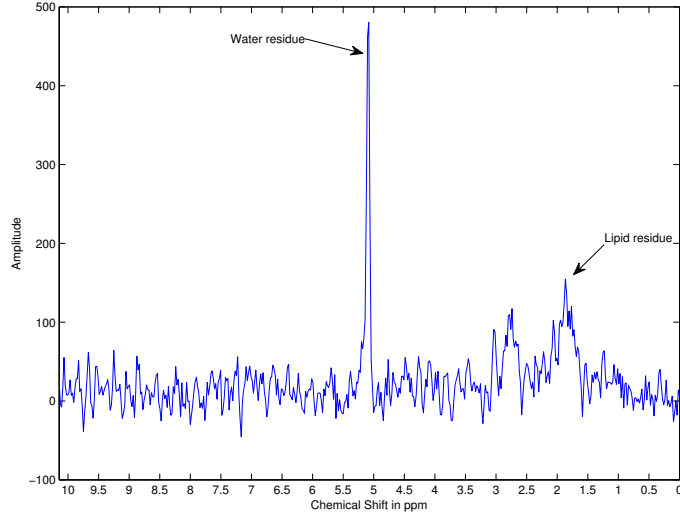


Figure 12: Illustration of the residues of water and fat even after their suppression during the acquisition protocol. The acquisition was carried out with a 3.0 Tesla MRI.

791 exponentially damped sinusoids such that:

$$s(t) = \sum_{k=1}^K a_k \exp(i\phi_k) \exp(-d_k + i2\pi f_k)t + \eta(t) . \quad (21)$$

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

795 Pijnappel et al. (1992) showed that the “signal-free noise” can be found using
 796 the singular value decomposition (SVD) decomposition. First the noisy signal
 797 is reorganized inside an Hankel matrix H . It can be shown that if the signal
 798 considered would be a “signal-free noise”, the rank of H would be equal to rank
 799 K . However, due to the presence of noise, H is in fact a full rank matrix. Thus,
 800 to recover the “signal-free noise”, the rank of H can be truncated to K using
 801 its SVD decomposition. Hence, knowing the cut off frequency of the water (cf.,
 802 4.7 ppm) and lipid (cf., 2.2 ppm) metabolites, their corresponding peaks can be
 803 reconstructed and subtracted from the original signal (Laudadio et al. (2002)).

804 – **Baseline correction:** Sometimes, the problem exposed in the above section re-
 805 garding the lipid molecules is not addressed simultaneously than the water resid-

806 uals suppression. Lipids and macromolecules are known to affect the baseline of
807 the MRSI spectra. They could cause errors during further fitting processes aiming
808 to quantify the metabolites, especially regarding the citrate metabolite.

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

$$p_f(x) = \begin{cases} p(x) , & \text{if } p(x) \leq s(x) , \\ s(x) , & \text{if } p(x) > s(x) . \end{cases} \quad (22)$$

816 Finally, this procedure of fitting and re-assignment is iteratively repeated on
817 $p_f(x)$ until a stopping criterion is reached. The final polynomial function can be
818 subtracted to the original signal $s(x)$ to correct it.

819 Parfait et al. (2012) modified this algorithm by convolving a Gaussian kernel
820 to smooth the MRSI signal instead of fitting a polynomial function, keeping the
821 rest of the algorithm identical.

822 Unlike Lieber and Mahadevan-Jansen (2003), Devos et al. (2004) proposed
823 to correct the baseline in the time domain by multiplying the MRSI signal by a
824 decreasing exponential function as:

$$c(t) = \exp(-\beta t) . \quad (23)$$

825 Having a typical value for β of 0.15.

826 However, Parfait et al. (2012) concluded that the method proposed by Lieber
827 and Mahadevan-Jansen (2003) outperforms the one of Devos et al. (2004).

828 In the contemporary work of Tiwari et al. (2012), authors detected the baseline
829 using a local non-linear fitting method avoiding regions with significant peaks
830 which were detected using a experimentally parametrised signal-to-noise ratio
831 (i.e. a value larger than 5).

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

835 Tiwari et al. (2012) correct the frequency shift by first detecting known metabo-
836 lite peaks such as choline, creatine and citrate. The frequency shift is corrected by

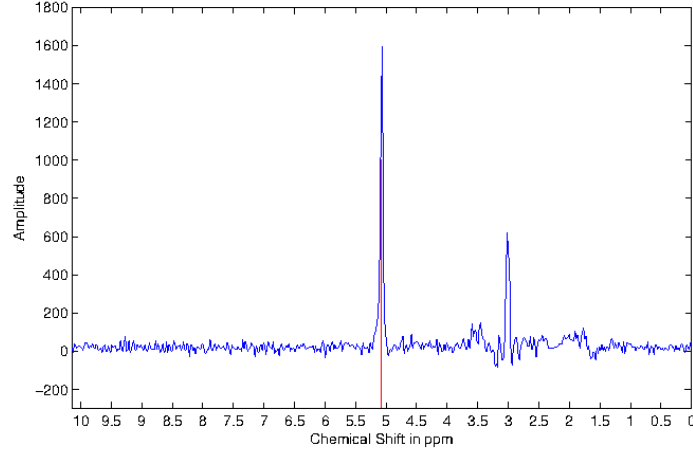


Figure 13: Illustration of frequency misalignment in an MRSI spectra acquire with a 3.0 Tesla MRSI scanner. The water peak is known to be aligned at 4.65 ppm. However, it can be seen that the peak on this spectra is aligned at around 5.1 ppm.

837 minimizing the frequency error between the experimental and theoretical values
838 of each of these peaks.

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

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

844 The first approach to obtain the normalization factor is based on an estimation
845 of the water concentration. It is required to have an additional MRSI sequence
846 where the water metabolites are unsuppressed. Using this sequence, an estimation
847 of the water concentration can be performed using the previously reported HSVD
848 algorithm. The second approach to normalization is based on using the L_2 norm
849 $\|s(\xi)\|_2$ of the MRSI spectra. It should be noted that both Parfait et al. (2012)
850 and Devos et al. (2004) concluded that the L_2 normalization was more efficient in
851 their framework.

852 3.2. Segmentation

853 The segmentation task consists in delineating the prostate boundaries in the MRI.
854 This procedure is of particular importance for focusing the posterior processing on the
855 organ of interest (Ghose et al. (2012)). In this section, only the segmentation methods

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

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

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 CHECK AGAIN, maybe forgot some CAD for this item. 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)).
- **Region-based segmentation:** Litjens et al. (2012b) used a multi-atlas-based segmentation using multi-modal images (e.g., T₂-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 in three different steps: (i) the registration between each atlas and the multi-modal images, (ii) the atlas selection and finally (iii) the classification of the prostate 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 aligned 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 T₂-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. (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 in 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 artefact and the contour between the zones were smooth using the 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 $\{I_1, I_2, \dots, I_k\}$ 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:

$$f_l = \{I_1(\mathcal{N}(l)), I_2(\mathcal{N}(l)), \dots, I_k(\mathcal{N}(l))\} . \quad (24)$$

where $I_k(\mathcal{N}(l))$ represent a feature vector obtained by sampling the k^{th} texture map using the neighbourhood $\mathcal{N}(l)$.

By generating multiple landmark in the same fashion than with the 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).

Posteriorly, 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. First, 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 shape.

3.2.2. MRSI-based segmentation

Tiwari et al. (2009a) localize 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:

$$W(\mathbf{x}, \mathbf{y}) = \begin{cases} \exp\left(-\frac{\|s(\mathbf{x}) - s(\mathbf{y})\|_2^2}{\sigma^2}\right) & \text{if } \|\mathbf{x} - \mathbf{y}\|_2 < \epsilon, \\ 0 & \text{if } \|\mathbf{x} - \mathbf{y}\|_2 > \epsilon. \end{cases} \quad (25)$$

where $s(\mathbf{x})$ and $s(\mathbf{y})$ are the MRSI spectra for the voxels \mathbf{x} and \mathbf{y} respectively, σ is

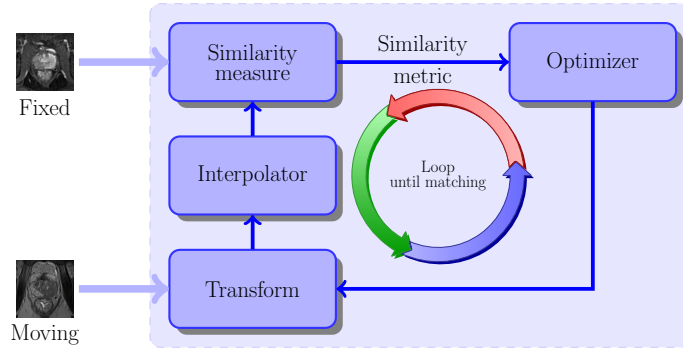


Figure 14: Typical framework involved to solve the registration problem. An iterative procedure takes place to infer the geometric transformation (parametric or non-parametric), via an optimizer, which maximizes the similarity between the two images.

the standard deviation of the Gaussian similarity measure and ϵ is the parameter to defined 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 larger cluster is assimilated to be the cluster corresponding to non-prostate voxels and is eliminated. 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 predominant in CAD systems using multi-parametric MRI images. As it will be discussed in Sect. 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-modalities MRI images ahead of performing any classification scheme.

Image registration is the procedure consisting in 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.

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.

✓!: suggested.

✓: used or implemented.

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 transformation. Regarding the transformation, from all the 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 their degree of freedom that they offer.

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

$$\begin{aligned}
 \mathcal{T}_R(\mathbf{x}) &= \begin{bmatrix} R & \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix} \mathbf{x}, \\
 &= \begin{bmatrix} \cos \theta & -\sin \theta & t_x \\ \sin \theta & \cos \theta & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}.
 \end{aligned} \tag{26}$$

984 where θ is the rotation angle and $\{t_x, t_y\}$ represents the translation along $\{x, y\}$
 985 respectively.

986 In the case of 3D registration using volume, an additional component z have to
 987 be taken into account such that $\mathbf{x} = (x, y, z)$. Thus, the rotation matrix \mathbf{R} becomes
 988 of size 3×3 whereas the translation vector \mathbf{t} consists in a vector of three elements.
 989 Hence, the geometric transformation $\mathcal{T}_R(\cdot)$ is embedded into a matrix of size 4×4 .

990 Affine transformations provide more degree of freedom managing rotations and
 991 translation as with the rigid transformations but also shearing and scaling. Hence,
 992 for a 2D space where $\mathbf{x} = (x, y) \in \mathbb{R}^2$ an affine transformation \mathcal{T}_A is formalized such
 993 as:

$$\begin{aligned} \mathcal{T}_A(\mathbf{x}) &= \begin{bmatrix} A & \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix} \mathbf{x} , \\ &= \begin{bmatrix} a_{11} & a_{12} & t_x \\ a_{21} & a_{22} & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} . \end{aligned} \quad (27)$$

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

996 Regarding volume registration, the same remark as previously mentioned can be
 997 recall in this case. Thus the geometric transformation $\mathcal{T}_A(\cdot)$ is of size 4×4 .

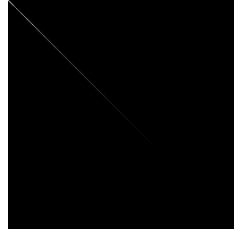
998 Finally, the last group of transformations is known as elastic transformations and
 999 offer the advantage to handle local distortions. In the reviewed CAD systems, the
 1000 radial basis functions are used to formalize the local distortions such as:

$$\mathcal{T}_E(\mathbf{x}) = \frac{a_{11}x - a_{12}y + t_x + \sum_i c_i g(\|\mathbf{x} - p_i\|)}{a_{21}x + a_{22}y + t_y + \sum_i c_i g(\|\mathbf{x} - p_i\|)} . \quad (28)$$

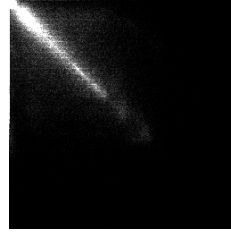
1001 where \mathbf{x} are the control points in both images and $g(\cdot)$ is the actual radial basis
 1002 function.

1003 Two radial basis functions are used: (i) the TPS and B-splines. Apart from the
 1004 formalism, these two approaches have a main difference. With B-splines, the control
 1005 points are usually uniformly and densely placed on a grid where as with TPS, the
 1006 control points correspond to detected or selected key points. By using TPS, Mitra
 1007 et al. (2011) obtained more accurate and time efficient results than with the B-splines
 1008 strategy (Mitra et al. (2012)).

1009 It is reasonable to point out that usually only rigid or affine registration are
 1010 used to register multi-parametric images from a same protocol. Elastic registration



(a) Illustration of a joint histogram between aligned images.



(b) Illustration of a joint histogram between misaligned images.

Figure 15: Difference observed in joint histogram between aligned and misaligned images. The joint measure will be more concentrated in the diagonal of the histogram in the case that the images are aligned and more randomly distributed in the case that both images are more misaligned.

method 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 MSE of the SI of MRI images. For a couple of images I and J , the MSE is formalized as:

$$\text{MSE} = \frac{1}{N} \sum_x \sum_y (I(x, y) - J(x, y))^2 . \quad (29)$$

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). 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:

$$MI(A; B) = S(A) + S(B) - S(A, B) . \quad (30)$$

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 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 common way to estimate these PDFs.

A generalized form of MI, CMI, was proposed by Chappelow et al. (2011). CMI allows to encompass interdependent information such as texture and gradient to 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:

$$CMI(\epsilon_n^A; \epsilon_m^B) = S(\epsilon_n^A) + S(\epsilon_m^B) - S(\epsilon_n^A, \epsilon_m^B) . \quad (31)$$

From Eq. 31, it can be seen that CMI is provided using high dimensional data and that histogram-based method to estimate the PDFs are not suited any more Chappelow et al. (2011). However, other approaches can be used as such 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 -nearest neighbour (k -NN) inside the high dimensional feature space lately 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

As shown in Fig. show image without interpolation, the registration procedure involves to transform an image and pixels placed on non-integer point have to 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 (Mittra (2012)).

3.3.5. Review of the methods used in CAD system

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

Ampeliotis et al. (2007, 2008) did not use the straightforward approach as presented in Fig. 14 to register 2D T₂-W and DCE images. By using image symmetries and the MSE metric, they find the parameters of an affine transformation.

Giannini et al. (2013) used also a in-house development registration method to register 2D T₂-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 their contours. Then, the images are registered via optimization approach and a deformation field is applied to the images.

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 T₂-W and DCE images using the 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 that 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 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 T₂-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, T₂-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 such problem but mentioned that Nelder-Mead simplex or quasi-Newton method could be other solutions.

4. Image classification framework

4.1. CADe: ROIs detection/selection

As discussed in the introduction and shown in Fig. 2.3.2, the image classification framework is composed of eventually a CADe and a CADx. In this section, attention will be paid on studies embedding a CADe in their framework. Let us remind the characteristics allowing to be or not tagged as a CADe system in our classification.

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]

CAD systems with manual delineations of potential lesions were excluded from the CAdE classification.

Thus, only two others approaches were considered to be CAdE which are summarized in Tab. 6. The first strategy, which concern 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 specific processing to delineate potential lesions to further obtain a diagnosis through the CADx.

Vos et al. (2012) highlighted candidate lesions by detecting blob in the ADC map follow by a suppression stage based on *a priori* criterion. As mentioned in Sect. 2.3.2 (see also Tab. 2.3.2), CaP can be interpreted as region of lower SI in ADC map. Hence, blob detector are suitable to highlight these kinds of region. Then, blobs were 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} to be part of a blob at the scale σ is given such as:

$$P(\mathbf{x}, \sigma) = \begin{cases} \frac{\|\lambda_{\sigma,3}(\mathbf{x})\|^2}{\|\lambda_{\sigma,1}(\mathbf{x})\|} , & \text{if } \lambda_{\sigma,k}(\mathbf{x}) > 0 \text{ with } k = \{1, 2, 3\} , \\ 0 , & \text{otherwise .} \end{cases} \quad (32)$$

The fusion of the different scales is computed such as:

$$L(\mathbf{x}) = \max P(\mathbf{x}, \sigma), \forall \sigma . \quad (33)$$

The resulting map $L(\mathbf{x})$ is then filtered depending on its appearance (cf. maximum of the likelihood of the region, diameter of the lesion) and their SI in ADC and T₂-W images. The detected regions were then used as inputs for the CADx.

1120 Litjens et al. (2011) used a pattern recognition approach in order to delineate
1121 the ROIs. A blobness map was calculated in the same manner as previously men-
1122 tioned using the multi-resolution Hessian blob detector on the ADC map, T₂-W and
1123 pharmacokinetic parameters maps (see Sect. 4.2 for details about those parameters).
1124 Additionally, the position of the voxel $\mathbf{x} = \{x, y, z\}$ was used as features as well as
1125 the Euclidean distance of the voxel to the prostate center. Hence, the feature vectors
1126 was composed of eight features and a support vector machine (SVM) classifier was
1127 trained using a radial basis function (RBF) kernel (see Sect. 4.4 for more details).

1128 Posteriorly, Litjens et al. (2012b) modified this approach by including only fea-
1129 tures related to the blob detection on the different maps as well as the original SIs
1130 of the parametric images. Two new maps were introduced based on texture. Instead
1131 of a SVM classifier, a k -NN classifier was performed. The candidate regions are
1132 then extracted by performing a local maxima detection follow by a post-processing
1133 region-growing and morphological operations.

1134 4.2. Feature detection

1135 Discriminative features allowing to recognize CaP from healthy tissue have to be
1136 first detected. This section will summarize the different strategies employed for this
1137 task. The feature type used is summarized in Tab. 7 while the Tab. 8 summed up
1138 which strategies were used by the different studies reviewed.

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

Feature detection methods	Indexes
MRI image:	
<i>Voxel-wise detection</i>	
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
<i>Region-wise detection</i>	
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

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

Indexes	A	B ₁	B ₂	B ₃	B ₄	C ₁	C ₂	C ₃	C ₄	D	E ₁	E ₂	F ₁	F ₂	F ₃	F ₄	G	H	I	J ₁	J ₂	J ₃	J ₄	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.

4.2.1. Image-based features

This section will focus on the 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 or (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. 2.3.2) can be discerned due to SI changes. Hence, *Intensity-based features* are one of the most common feature used to build the feature vector which has to be classified and have been widely used (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 in the SI of each voxel of the different MRI.

SI changes can be viewed as heterogeneous regions and *Edge-based features* are used in that regard. Each feature will be 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 as shown in Eq. 34-36.

$$K^P = \begin{cases} K_{xx}^P = \begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix} & K_{yy}^P = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix} \\ K_{xy}^P = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & -1 \\ 0 & -1 & -1 \end{bmatrix} & K_{yx}^P = \begin{bmatrix} -1 & -1 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \end{cases}, \quad (34)$$

$$K^S = \begin{cases} K_{xx}^S = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix} & K_{yy}^S = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} \\ K_{xy}^S = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 0 & -1 \\ 0 & -1 & -2 \end{bmatrix} & K_{yx}^S = \begin{bmatrix} -2 & -1 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & 2 \end{bmatrix} \end{cases}, \quad (35)$$

$$K^K = \begin{cases} K_{xx}^K = \begin{bmatrix} 5 & 0 & -3 \\ 5 & 0 & -3 \\ 5 & 0 & -3 \end{bmatrix} & K_{yy}^K = \begin{bmatrix} 5 & 5 & 5 \\ 0 & 0 & 0 \\ -3 & -3 & -3 \end{bmatrix} \\ K_{xy}^K = \begin{bmatrix} 5 & 5 & 0 \\ 5 & 0 & -3 \\ 0 & -3 & -3 \end{bmatrix} & K_{yx}^K = \begin{bmatrix} -3 & -3 & 0 \\ -3 & 0 & 5 \\ 0 & 5 & 5 \end{bmatrix} \end{cases}. \quad (36)$$

1158 The magnitude and gradient direction is commonly computed deriving from
1159 the edge detection such as:

$$G(\mathbf{x}) = \sqrt{(K_{xx} * I(\mathbf{x}))^2 + (K_{yy} * I(\mathbf{x}))^2} \quad (37)$$

$$\Theta(\mathbf{x}) = \arctan \left(\frac{K_{yy} * I(\mathbf{x})}{K_{xx} * I(\mathbf{x})} \right). \quad (38)$$

1160 where $K(\cdot)$ is either the Prewitt kernel or Sobel kernel and I is the input image.

1161 These features are commonly incorporated in the feature vector for further
1162 classification in the CAD systems reviewed (Niaf et al. (2012, 2011), Tiwari et al.
1163 (2013, 2010, 2009b), Viswanath et al. (2011, 2008b)).

1164 Gabor filters offer another approach to extract information related to edges
1165 and texture and were found to performed similar processing than in the human
1166 visual system (Daugman (1985), Gabor (1946)). A Gabor filter is defined by the
1167 modulation of a Gaussian function with a sinusoid which can be further rotated.

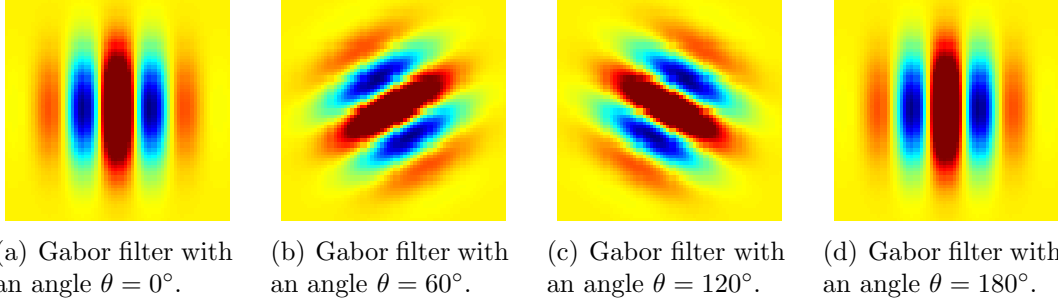


Figure 16: Illustration of four different Gabor filters varying by their orientations.

Hence, a Gabor filter g can be formalized as:

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

with

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

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

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. 1(b)) 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 system (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 voxel based on a defined distance and angle. Then, using this co-occurrence matrix, a set of features can be computed based on the statistics intending to describe the texture around each voxel. Computation of these features are reported in Tab. 9.

Fractal analysis and more precisely a local estimation of the fractal dimension (Benassi et al. (1998)) describing the texture roughness at this specific location

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 $Q(i, j) = \sum_k \frac{p(i, k)p(j, k)}{p_x(i)p_y(k)}$.

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 texture using the frequency signature via the DCT (Ahmed et al. (1974)). 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:

$$C_{k_1, k_2} = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} p_{m,n} \cos \left[\frac{\pi}{M} \left(m + \frac{1}{2} \right) k_1 \right] \cos \left[\frac{\pi}{N} \left(n + \frac{1}{2} \right) k_2 \right] . \quad (40)$$

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

In the same spirit, Viswanath et al. (2012) projected T₂-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.

Chan et al. (2003) computed these features for a neighbourhood of 7×7 pixels for each the modalities that they used.

The *position* of a voxel within the prostate was also considered as 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) embedding the same information but this time using cylindrical coordinate r , θ and z which are the radius, azimuth and altitude 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 method*. Full map corresponding to feature types presented in the previous section are computed. Then, ROIs are defined and statistics are deduced 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 n^{th} percentile will be used as feature. n can take any value between 0 and 100 but is usually selected because of its discriminative factor to differentiate CaP from healthy tissue. Identically, 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)).

Anatomical features characterizing the region defined were also used as by Litjens et al. (2012b) and Matulewicz et al. (2013). Litjens et al. (2012b) computes 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 the percentage present of PZ, CG, periurethral region or outside prostate region for the considered ROI.

Liu et al. (2013) introduced four different types of *histogram-based* features. The first type correspond to the histogram of the SI of the image. The second type is the HOG (Dalal and Triggs (2005)). HOG is descriptor describing the local shape of the object of interest by using edge directions distribution. This descriptor is extracted mainly in three steps. First the gradient image and its corresponding magnitude and direction are computed (see Eq. 37 - 38). Then, the ROI is divided into cells and an oriented-based histogram will be generated for each cell. At each pixel location, the orientation of the gradient will vote for a bin of the histogram and this vote will be weighted by the magnitude of the same gradient. Finally, The cells are grouped into block and each block will be 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 object of interest. First, a set of 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 features extracted is based on the framework of Zhao et al. (2012) which is using the Fourier transform of the histogram created via LBP (Ojala et al. (1996)). LBP is generating by comparing the value central pixel with its 8-connected neighbours. Then, in the ROI, the histogram of the LBP distribution is then computed. However, this representation is not rotational invariant. Hence, the discrete fourier transform (DFT) is taken 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 will be a statistical characteristic of the image roughness. The second fractal dimension 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

Table 10: Parameters used as features for a DCE semi-quantitative analysis in CAD systems.

Semi-quantitative features	Explanations
<i>Amplitude features:</i>	
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
<i>Time features:</i>	
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
<i>Derivatives and integral features:</i>	
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

1259 using wavelet decomposition.

1260 4.2.2. DCE-based features

1261 DCE-MRI is more commonly based on a SI analysis over time as presented in
1262 Sect. 2.3.2. In this section, a description of the features extracted for this analysis
1263 are used in the CAD systems reviewed.

- 1264 – **Whole-spectra approach:** Some studies are using the whole DCE time series
1265 as feature vector such as Ampeliotis et al. (2007, 2008), Tiwari et al. (2012) and
1266 Viswanath et al. (2008a,b). In some cases, the high-dimensional feature space will
1267 be reduced using dimension reduction methods as it will be presented in the next
1268 section (see Sect. 4.3).

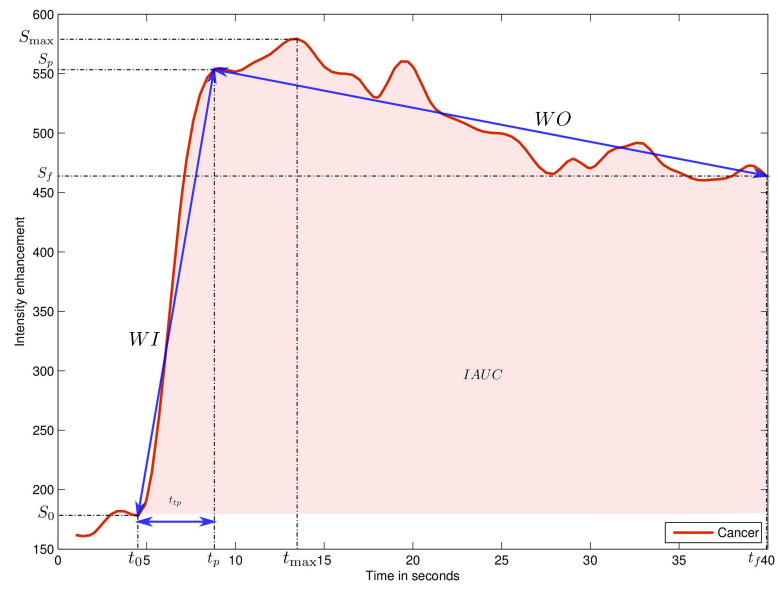


Figure 17: Graphical representation of the different semi-quantitative features used for DCE-MRI analysis.

1269 – ***Semi-quantitative approach:*** Semi-quantitative approaches are based on mod-
 1270 elling mathematically the DCE time series. The parameters modelling the signal
 1271 are commonly used mainly due to the simplicity to compute them. Parame-
 1272 ters included in semi-quantitative analysis are summarized in Tab. 10 and also
 1273 graphically depicted in Fig. 17. A set of time features corresponding to specific
 1274 amplitude level (start, maximum and end) are extracted. Then, derivative and
 1275 integral are also considered as discriminative and are commonly computed.

1276 – ***Quantitative approach:*** As presented in Sect. 2.3.2, quantitative approaches
 1277 correspond to mathematical-pharmacokinetic models based on physiological ex-
 1278 changes. Four different models have been used in CAD systems.

1279 The most common model reviewed was the *Brix model* (Artan et al. (2010,
 1280 2009), Liu et al. (2009), Ozer et al. (2009, 2010), Sung et al. (2011)). This model
 1281 is formalized such as:

$$\frac{S(t)}{S(0)} = 1 + Ak_{ep} \left(\frac{\exp(-k_{ep}t) - \exp(-k_{el}t)}{k_{el} - k_{ep}} \right) . \quad (41)$$

1282 where $S(\cdot)$ is the DCE signal, A is the parameter simulating the tissue proper-
 1283 ties, k_{el} is the parameter related to the first-order elimination from the plasma
 1284 compartment, k_{ep} is the parameter of the transvascular permeability.

1285 Thus, the parameters k_{ep} , k_{el} and A are computed and used as features.

1286 *Tofts model* (Tofts (1997)) was used by Giannini et al. (2013), Langer et al.
 1287 (2009), Mazzetti et al. (2011), Niaf et al. (2012, 2011). The DCE signal relative
 1288 to the concentration is then modelled such as:

$$C_t(t) = v_p C_p(t) + K_{trans} \int_0^t C_p(\tau) \exp(-k_{ep}(t - \tau)) d\tau . \quad (42)$$

1289 where $C_t(\cdot)$ is the concentration of the medium, $C_p(\cdot)$ is the AIF which have to be
 1290 estimated apart, K_{trans} is the parameter related to the diffuse transport of media
 1291 across the capillary endothelium, k_{ep} is the parameter related to the exchanges
 1292 back into the vascular space and v_e is the extravascular-extracellular space fraction
 1293 such as $v_e = 1 - v_p$.

1294 Thus, the parameters K_{trans} , k_{ep} and v_e are computed and used as features.

1295 Mazzetti et al. (2011) and Giannini et al. (2013) employed two others empirical
 1296 models. A *Weibull function* was used and can be formalized such as:

$$S(t) = At \exp(-t^B) . \quad (43)$$

1297 where A and B are the two parameters which have to be inferred.

The other empirical model is called phenomenological universalities model (Castorina et al. (2006)) which can be formalized:

$$S(t) = \exp \left(rt + \frac{1}{\beta} a_0 - r(\exp(\beta t) - 1) \right) . \quad (44)$$

where the parameters β , a_0 and r are inferred by using curve fitting approach.

4.2.3. MRSI-based features

– **Whole spectra approach:** As for the DCE analysis, one common approach is to incorporate the whole MRSI spectra in the feature vector to classify (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 perform to reduce the complexity during the classification as it will be presented in Sect. 4.3.

– **Quantification approach:** We cab recall 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, quantification of these metabolite concentrations are considered as feature for further classification. In order to perform this quantification, four different approaches have been used. QUEST (Ratney et al. (2005)), AMARES (Vanhamme et al. (1997)) and VARPRO (Coleman and Li (1993)) were used by Kelm et al. (2007) which 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 previous method, 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. 45 and 46) and used them as features. However, Parfait et al. (2012) used each metabolite concentrations individually.

$$R_1 = \frac{[\text{Cho}] + [\text{Cr}]}{[\text{Cit}]} . \quad (45)$$

$$R_2 = \frac{[\text{Cit}]}{[\text{Cho}] + [\text{Cr}] + [\text{Cit}]} . \quad (46)$$

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

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

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

- 1324 – **Wavelet decomposition approach:** Tiwari et al. (2012) performed a wavelet
1325 packet decomposition (Coifman and Wickerhauser (1992)) by using the Haar
1326 wavelet basis function and use the coefficients of this decomposition as features
1327 for further classification.

1328 4.3. Feature selection and feature extraction

1329 As presented in the previous section, a wide variety of features can be computed
1330 (see Tab. 7). Using multi-parametric data as well as multiple features lead to a high
1331 complexity feature space which might mislead or corrupt the classifier which have to
1332 be trained. Thus, one will be interesting to reduce the number of dimension of the
1333 feature space before to proceed to any classification task. The strategy used can be
1334 group into two groups: (i) feature selection and (ii) feature extraction. The methods
1335 used for CAD system are summarized in Tab. 11.

1336 4.3.1. Feature selection

1337 The feature selection strategy corresponds in selecting the most discriminative
1338 feature dimension of the high-dimensional space. Thus, the low-dimensional space is
1339 then composed of a subset of the original features detected. In this section, methods
1340 employed in the studies reviewed will be briefly presented. More extensive review
1341 specific to feature selection can be found in Saeys et al. (2007).

1342 Niaf et al. (2012, 2011) make use of the *p-value* by using the independent two-
1343 sample *t-test* with equal mean for each feature dimension. In this statistical test,
1344 it is assumed two classes, CaP against healthy. 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 tested is based on the fact that these both distribution means are equal. Thus, the t-statistic related to verify the null hypothesis is then formalized such that:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_{X_1 X_2} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} , \quad (47)$$

$$s_{X_1 X_2} = \sqrt{\frac{(n_1 - 1)s_{X_1}^2 + (n_2 - 1)s_{X_2}^2}{n_1 + n_2 - 2}} .$$

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

From Eq. 47, it can be seen that the more the means of the class distribution diverge, 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 to separate.

The p -value statistic can be deduced from the t -test and corresponds to the probability to obtain such an extreme test assuming that the null hypothesis is true Goodman (1999). Hence, smaller is the p -value, more likely is to reject the null hypothesis and more relevant the feature is likely to be.

Finally, the feature can be ranked and the most significant features can be selected, by defining the number of features wanted. However, this technique suffers from a main drawback since that it is assumed that each feature are independent which is unlikely to happen and introduce a high redundancy in the feature 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 variance between classes to the variance within classes such that:

$$F_r = \frac{\bar{X}_1 - \bar{X}_2}{s_{X_1}^2 + s_{X_2}^2} . \quad (48)$$

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

MI can also be used to select the subset of feature dimension. Definition of the MI was presented in Sect. 3.3.2 and formalized in 30. 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 method such as Parzen windows.

Peng et al. (2005) introduced two main criteria to select the features dimensions and then combined: (i) maximal relevance and (ii) minimum redundancy.

Maximal relevance criterion is based on the paradigm that the classes and the feature dimension which have to be selected have to share a maximal MI and can be formalized such that:

$$\begin{aligned} & \arg \max Rel(\mathbf{x}, c) , \\ Rel(\mathbf{x}, c) &= \frac{1}{|\mathbf{x}|} \sum_{x_i \in \mathbf{x}} MI(x_i, c) . \end{aligned} \quad (49)$$

where $\mathbf{x} = \{x_i, i = 1, \dots, d\}$ is a feature vector of d dimensions, c is the class considered and $MI(.)$ is the MI.

As in the previous method, using maximal relevance criterion alone will imply the independence between each feature dimension which is usually not true.

Minimal redundancy criterion will force to select a new feature dimension which shares as less as possible MI with previously selected feature dimension. It can be formalized as:

$$\begin{aligned} & \arg \min Red(\mathbf{x}) , \\ Red(\mathbf{x}) &= \frac{1}{|\mathbf{x}|^2} \sum_{x_i, x_j \in \mathbf{x}} MI(x_i, x_j) . \end{aligned} \quad (50)$$

Combination of these two previous criteria is known as maximal relevance minimal redundancy (mRMR)⁵ (Peng et al. (2005)) and can be computed such as a difference or a quotient of the Eq. 49 and 50.

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 and are not selecting discriminative features. Instead, these methods aimed at mapping

⁵mRMR implementation can be found at: <http://penglab.janelia.org/proj/mRMR/>

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 from a linear or a non-linear manner. Only methods employed in CAD system will be reviewed in this section. We guide the reader to the review of Fodor (2002) for a full review of feature extraction techniques.

Linear mapping method used to reduce the dimensionality in CAD system is usually the PCA.

PCA is a method allowing to find 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 k^{th} greatest variances will lie on the k^{th} principal components (Jolliffe (2002)).

The principal components can then be computed using the eigenvectors-eigenvalues decomposition on the covariance matrix. Let's define the \mathbf{x} being the data matrix. Then the covariance matrix is defined such as:

$$\Sigma = \mathbf{x}^T \mathbf{x} . \quad (51)$$

The eigenvectors-eigenvalues decomposition can be formalized such as:

$$\mathbf{v}^{-1} \Sigma \mathbf{v} = \Lambda . \quad (52)$$

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

It is then possible to find the new low-dimensional space by sorting the eigenvectors using the eigenvalues and finally select the largest eigenvalues. The total variation being equal to the sum of the eigenvalues of the covariance matrix (Fodor (2002)), usually the number of principal components correspond 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 and mainly based on Laplacian eigenmaps and locally linear embedding (LLE) methods.

Laplacian eigenmaps⁶ or also referred as spectral clustering in computer vision aimed 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. In like manner, two far away data points in

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

1425 the high-dimensional space should be also distant in the low-dimensional space. To
 1426 compute this projection, an adjacency matrix is defined such as:

$$W(i, j) = \exp \|x_i - x_j\|_2 . \quad (53)$$

1427 Then, the low-dimensional space will be found by solving the generalized eigenvectors-
 1428 eigenvalues problem such as:

$$(D - W)\mathbf{y} = \lambda D\mathbf{y} . \quad (54)$$

1429 where D is a diagonal matrix such that $D(i, i) = \sum_j W(j, i)$.

1430 Finally the low-dimensional space is defined by the k eigenvectors of the k smallest
 1431 eigenvalues (Belkin and Niyogi (2001)). Tiwari et al. (2007, 2009a,b), Viswanath
 1432 et al. (2008b) used this spectral clustering to project their feature vector into a low-
 1433 dimensional space. The feature space in these studies is usually composed of features
 1434 extracted from a single or multiple modalities and then concatenated before applying
 1435 the Laplacian eigenmaps dimension reduction technique.

1436 Tiwari et al. (2013, 2009a) used a slightly different approach by combining the
 1437 Laplacian eigenmaps techniques with a prior multi-kernel learning strategy. First,
 1438 multiple feature were extracted for multiple modalities. The features of a single
 1439 modalities were then mapped to an higher dimensional space via the Kernel trick
 1440 (Aizerman et al. (1964)) and more precisely using a Gaussian kernel. Then, each
 1441 kernel associated with each modality are linearly combined to obtain a combined
 1442 kernel K . Then, the computation of the adjacency matrix W takes place and the
 1443 same scheme as in Laplacian eigenmaps is performed. However, in order to used
 1444 the combined kernel, Eq. 54 becomes as shown in Eq. 55 and can be solved as a
 1445 generalized eigenvectors-eigenvalues problem as previously.

$$K(D - W)K^T\mathbf{y} = \lambda KDK^T\mathbf{y} . \quad (55)$$

1446 Viswanath et al. (2011) used Laplacian eigenmaps inside a bagging framework
 1447 in which multiple embeddings are generated by successively selecting feature dimen-
 1448 sions.

1449 LLE⁷ is another non-linear dimension technique broadly known firstly proposed
 1450 by Roweis and Saul (2000). LLE is based on the fact that a data point in the feature
 1451 space can be characterized by its neighbours. Thus, it was proposed to represent each

⁷LLE implementation is available at: <http://www.cs.nyu.edu/~roweis/lle/code.html>

1452 data point in the high-dimensional space as the linear combination of its k -nearest
 1453 neighbours. This can be expressed such as:

$$\hat{\mathbf{x}}_i = \sum_j W(i, j) \mathbf{x}_j . \quad (56)$$

1454 where \mathbf{x}_i and \mathbf{x}_j are the data point considered and its neighbours data points, re-
 1455 spectively.

1456 Hence, this problem which have to be solved at this stage is to estimate the weight
 1457 matrix W . This problem can be tackled using a least square optimization scheme by
 1458 optimizing the following objective function:

$$\begin{aligned} \hat{W} &= \arg \min_W \sum_i |\mathbf{x}_i - \sum_j W(i, j) \mathbf{x}_j|^2 , \\ &\text{subject to } \sum_j W(i, j) = 1 , \end{aligned} \quad (57)$$

1459 Then, the essence of LLE is to project the data into a low-dimensional keeping
 1460 the data organization. Thus, the projection into the low-dimensional space can be
 1461 seen as an optimization problem such that:

$$\hat{\mathbf{y}} = \arg \min_{\mathbf{y}} \sum_i |\mathbf{y}_i - \sum_j W(i, j) \mathbf{y}_j|^2 . \quad (58)$$

1462 This optimization can be performed as an eigenvectors-eigenvalues problem by
 1463 finding the k^{th} eigenvectors corresponding the k^{th} smallest eigenvalues of the sparse
 1464 matrix $(I - W)^T(I - W)$. Tiwari et al. (2008, 2009a), Viswanath et al. (2008a,b)
 1465 used LLE as dimension reduction technique to reduce the complexity of their feature
 1466 vector.

1467 Tiwari et al. (2008) used a modified version of the LLE algorithm in which they
 1468 applied LLE in a bagging approach with multiple size of neighbourhood. The em-
 1469 bedding obtained are then fusion using the maximum likelihood estimation.

1470 4.4. Classification

1471 4.4.1. Classifier

1472 Once that the feature vector was extracted and eventually the complexity re-
 1473 duced, it is possible to take decision and classify this feature vector to belong to CaP
 1474 or healthy tissue. Classification methods used in CAD system to distinguish this two
 1475 classes are summarized in Tab. 12. A full review of classification methods used in
 1476 pattern recognition can be found in Bishop (2006).

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

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

1477 – *Simple rules:*

1478 Lv et al. (2009) make use of a decision stump classifier to distinguish CaP and
1479 healthy classes.

1480 Puech et al. (2009) detect CaP by implementing a score medical decision
1481 making approach. The feature values are compared with a pre-defined threshold.
1482 Then, at each comparison, the final score is incremented or not, depending on the
1483 threshold and the final decision is taken depending of the final score.

1484 – *Clustering methods:*

1485 k -nearest neighbour (k -NN) is one of the simplest supervised machine learning
1486 classification method. In this method, a new unlabelled vector is assigned to the
1487 most represented class from its k nearest-neighbours in the feature space. The
1488 parameter k is usually an odd number in order to avoid any tie case.

1489 k -NN was one of the method used by Niaf et al. (2012, 2011) mainly to make
1490 a comparison with different machine learning techniques. Litjens et al. (2012b)
1491 used this method to roughly detect potential CaP voxel before to perform a region-
1492 based classification.

1493 k -means is an unsupervised clustering methods in which the data have to be
1494 partitioned into k clusters. The discovery of the clusters is an iterative procedure.
1495 First k random centroids are defined in the feature space and each data point is
1496 assigned to the nearest centroid. Then, the centroid position for each cluster is
1497 updated by computing the mean of all the data points belonging to this particular
1498 cluster. Both assignment and updating are repeated until the centroids are stable.
1499 The number of cluster k is usually defined as the number of classes. This algorithm
1500 can also be used incrementally. In case that new data have to be incorporated,
1501 the initial centroid positions can corresponds to the results of a previous k -means
1502 followed by the assignment-updating stage.

1503 Tiwari et al. (2007, 2009a) used k -means in an iterative procedure. Three clus-
1504 ters were defined corresponding to CaP, healthy and non-prostate, respectively.
1505 k -means was applied iteratively and the voxels corresponding to the larger cluster
1506 were excluded under the assumption that it is assigned to “non-prostate” cluster.
1507 The iteration stop until the number of voxels composing the clusters is smaller
1508 than a threshold.

1509 Tiwari et al. (2008), Viswanath et al. (2008a,b) used k -means in a repetitive
1510 manner to be less sensitive to the centroids initialisation. Thus, k clusters will be
1511 generated T times. The final assignment is performed by majority voting using a
1512 co-association matrix as proposed by Fred and Jain (2005).

1513 – *Linear model classifiers:*

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

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

where \mathbf{x} is a feature vector which the class have to be determined, 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 probabilities.

To perform the classification, a sample \mathbf{x}_i will be assigned to the class which maximize the discriminant function:

$$C(\mathbf{x}_i) = \arg \max_k \delta_k(\mathbf{x}_i) . \quad (60)$$

covariance matrix Σ_k specific at each class is computed 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:

$$p(c_1|\mathbf{x}_i) = \frac{1}{1 + \exp(-\mathbf{w}^T \mathbf{x}_i)} . \quad (61)$$

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:

$$C(\mathbf{x}_i) = \arg \max_k p(C = k|\mathbf{x}_i) . \quad (62)$$

From Eq. 61, one can see that the key of classification using logistic regression model is to infer the set of parameter \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.

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.

1541 – ***Non-linear model classifier:***

1542 Viswanath et al. (2012) used the most general quadratic discriminant analysis
 1543 (QDA) instead of LDA. Unlike in LDA in which one assumes that the class
 1544 covariance matrices Σ is identical for all the classes, in QDA, it is assumed that a
 1545 covariance matrix Σ_k specific at each class is computed. Thus, Eq. 59 becomes:

$$\delta_k(\mathbf{x}_i) = \mathbf{x}_i^T \Sigma_k^{-1} \mu_k - \frac{1}{2} \mu_k^T \Sigma_k^{-1} \mu_k + \log(\pi_k) . \quad (63)$$

1546 The classification rules in the case of the QDA is identical to Eq. 60.

1547 – ***Probabilistic classifier:***

1548 The most commonly used classifier is the naive Bayes classifier which is a
 1549 probabilistic classifier assuming the independence between each feature dimension
 1550 (Rish (2001)). This classifier is based on the Bayes' theorem:

$$p(C = k|\mathbf{x}) = \frac{p(C)p(\mathbf{x}|C)}{p(\mathbf{x})} . \quad (64)$$

1551 where $p(C = k|\mathbf{x})$ is the posterior probability, $p(C)$ is the prior probability, $p(\mathbf{x}|C)$
 1552 is the likelihood and $p(C)$ is the evidence.

1553 However, the evidence term is usually discarded since, it is not class dependent
 1554 and play the role of a normalization term. Hence, in a classification scheme, an
 1555 unlabelled observation will be classified to the class which maximizes the posterior
 1556 probability such as:

$$C(\mathbf{x}_i) = \arg \max_k p(C = k|\mathbf{x}_i) , \quad (65)$$

$$p(C = k|\mathbf{x}_i) = p(C = k) \prod_{j=1}^n p(x_{ij}, |C = k) . \quad (66)$$

1557 where n is the number of dimension of the feature vector $\mathbf{x}_i = \{x_{i1}, \dots, x_{in}\}$.

1558 Usually, a model have to be affected for the prior and likelihood probabilities
 1559 and it is common to affect an equal prior probability for each class or eventually a
 1560 value depending of the relative frequency derived from the training set. Regarding
 1561 the likelihood probability, it is common to choose a Normal distribution to
 1562 characterize each class. Thus, each class will be characterized by two parameters:
 1563 (i) the mean and (ii) the standard deviation. These parameters can be inferred
 1564 from the training set by using the maximum likelihood approach.

1565 Giannini et al. (2013), Mazzetti et al. (2011), Niaf et al. (2012, 2011) used the
 1566 naive Bayes classifier to classify their feature vector either malignant or healthy.

1567 The model used to characterized the likelihood probability using the Normal dis-
 1568 tribution.

1569 – ***Ensemble learner classifiers:***

1570 AdaBoost is an adaptive method based on ensemble learner initially proposed
 1571 by Freund and Schapire (1997). AdaBoost linearly combines several weak learners
 1572 resulting into a final strong classifier. A weak learner is defined as a classification
 1573 method performing slightly better than random classification. Popular choices
 1574 regarding the weak learner methods are: decision stump, decision tree learners
 1575 (cf., iterative dichotomiser 3 (ID3) (Quinlan (1986)), C4.5 (Quinlan (1993)), clas-
 1576 sification and regression tree (CART) (Breiman et al. (1984))).

1577 AdaBoost is considered as an adaptive method in the way that the weak learner
 1578 are selected. The selection is performed in an iterative manner. At each iteration
 1579 t , the weak learner selected h_t corresponds to the one minimizing the classification
 1580 error on a distribution of weights D_t , that is associated to the training sample. A
 1581 weight α_t is affected at each weak learner such that:

$$\alpha_t = \frac{1}{2} \ln \frac{1 - \epsilon_t}{\epsilon_t} . \quad (67)$$

1582 where ϵ_t corresponds to the classification error rate of the weak learner on the
 1583 distribution of weight D_t .

1584 Before performing a new iteration, the distribution of weight D_t is updated
 1585 such that the weight associated with the misclassified samples by h_t will increase
 1586 and the weights of well classified samples will decrease as shown in Eq. 68. This
 1587 procedure allows to select a weak learner at the next iteration $t + 1$ which will
 1588 classify in priority previous misclassified samples.

$$D_{t+1}(i) = \frac{D_t(i) \exp(-\alpha_t y_i h_t(\mathbf{x}_i))}{Z_t} . \quad (68)$$

1589 where \mathbf{x}_i is the i^{th} sample corresponding to class y_i and Z_t is a normalization
 1590 factor to impose that D_{t+1} to be a probability distribution.

1591 Thus, after T iterations, the final strong classifier corresponds to the linear
 1592 combination of the weak learners and the classification is performed such that:

$$C(\mathbf{x}_i) = \text{sign} \left(\sum_{t=1}^T \alpha_t h_t(\mathbf{x}_i) \right) . \quad (69)$$

1593 Lopes et al. (2011) make use of AdaBoost technique to perform their classifi-
 1594 cation.

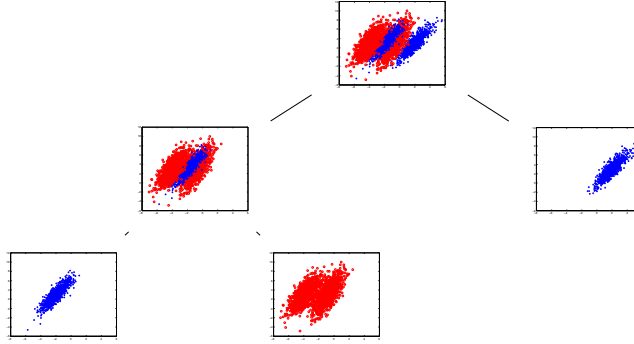


Figure 18: Representation of the capabilities of the probabilistic boosting tree algorithm to split at each node of the tree the positive and negative samples.

1595 Random forest⁸ is a classification method which is based on creating an en-
 1596 semble of decision trees and was introduced by Breiman (2001).

1597 In the learning stage, multiple decision tree learners (Breiman et al. (1984))
 1598 will be trained. However, each decision tree will be trained with a different dataset.
 1599 Each of these dataset corresponds to a bootstrap sample generated by randomly
 1600 choosing n samples with replacements from the N available in the original set
 1601 (Efron (1979)). Then, randomization is also part of the decision tree growth.
 1602 At each node of the decision tree, from the bootstrap sample of D dimensions, a
 1603 number of $d \ll D$ dimensions will be randomly selected. Finally, the d^{th} dimension
 1604 splitting the best the data will be used. This best splitting can be evaluated using
 1605 MI (see Sect. 3.3.2). Finally, each tree is grown as much as possible without using
 1606 any pruning procedure.

1607 In the prediction stage, a sample is introduced in each tree and each of the
 1608 tree will assign a class to this sample. Finally, it is common to use a majority
 1609 voting approach to choose the final class labelling.

1610 Kelm et al. (2007), Tiwari et al. (2013, 2012), Viswanath et al. (2009) make
 1611 use of the random forest classifier to classify their feature vector.

1612 Probabilistic boosting-tree are an ensemble classifier method which is sharing
 1613 principles from AdaBoost and using them inside a decision tree (Tu (2005)). In
 1614 the training stage, the probabilistic boosting-tree method is growing a decision
 1615 tree and at each node a strong classifier is learnt in an almost comparable scheme

⁸Random forest implementation can be found at: http://www.stat.berkeley.edu/~breiman/RandomForests/cc_software.htm

1616 than AdaBoost. Once the strong learner is trained, the training set will be split
 1617 into two subsets which will be used to train the next strong classifiers in the next
 1618 descending nodes. Thus, three cases are conceivable to decide which branch to
 1619 propagate each sample \mathbf{x}_i :

- 1620 • if $q(+1, \mathbf{x}_i) - \frac{1}{2} > \epsilon$ then \mathbf{x}_i is propagated to the right branch set and a weight
 1621 $w_i = 1$ is assigned.
- 1622 • if $q(-1, \mathbf{x}_i) - \frac{1}{2} > \epsilon$ then \mathbf{x}_i is propagated to the left branch set and a weight
 1623 $w_i = 1$ is assigned.
- 1624 • else \mathbf{x}_i will be propagated in both branches with $w_i = q(+1, \mathbf{x}_i)$ in the right
 1625 branch and $w_i = q(-1, \mathbf{x}_i)$ in the left branch.

1626 with $\mathbf{w} = w_i, i = \{1, \dots, N\}$ corresponding to distribution of weight, N the
 1627 number of samples as in AdaBoost and $q(\cdot)$ is defined such as:

$$q(+1, \mathbf{x}_i) = \frac{\exp(2H(\mathbf{x}_i))}{1 + \exp(2H(\mathbf{x}_i))} , \quad (70)$$

$$q(-1, \mathbf{x}_i) = \frac{\exp(-2H(\mathbf{x}_i))}{1 + \exp(-2H(\mathbf{x}_i))} . \quad (71)$$

1628 Employing such a scheme tends to divide the data in such a way that positive
 1629 and negative samples are naturally split as shown in Fig. 18.

1630 In the classification stage, the sample \mathbf{x} is propagated through the trees, where
 1631 at each node, it will be classified by the strong classifier previously learned and
 1632 where an estimation of the posterior distribution will be computed. The posterior
 1633 distribution will correspond to the sum of the posterior distribution at each node
 1634 of the tree.

1635 Tiwari et al. (2010, 2009b, 2012), Viswanath et al. (2011) make use of the
 1636 probabilistic boosting-tree classifier to determine which of the class to assign to
 1637 their feature vectors.

1638 – **Kernel method:**

1639 Gaussian process⁹ for classification is a kernel method in which it is assumed
 1640 that the data can be represented by a single sample from a multivariate Gaus-
 1641 sian distribution (Rasmussen and Williams (2005)). In the case of linear logistic
 1642 regression for classification, the posterior probability can be expressed as:

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

$$\begin{aligned} p(y_i|\mathbf{x}_i, \mathbf{w}) &= \sigma(y_i f(\mathbf{x}_i)) , \\ f(\mathbf{x}_i) &= \mathbf{x}_i^T \mathbf{w} . \end{aligned} \quad (72)$$

where $\sigma(\cdot)$ is the logistic function.

Thus, the essence of Gaussian process for the classification is based on affecting a Gaussian process prior over the function $f(\mathbf{x})$ which will be characterized by a mean \bar{f} and covariance functions K . Thus, in the training stage, the best mean and covariance function 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:

$$\begin{aligned} p(f(\mathbf{x}_*)|f(\mathbf{x})) &= \mathcal{N}(K_* K^{-1} \bar{f}, K_{**} - K_*(K')^{-1} K_*^T) , \\ K' &= K + W^{-1} , \\ W &= \nabla \nabla \log p(\mathbf{y}|f(\mathbf{x})) . \end{aligned} \quad (73)$$

where K_{**} is the covariance on testing, K_* is the covariance on training-testing, \mathbf{x} is the training data.

Then, the function $f(\mathbf{x}_*)$ is squashed using the sigmoid function and the probability of the class membership can be defined such that:

$$C(\mathbf{x}_*) = \sigma \left(\frac{\bar{f}(\mathbf{x}_*)}{\sqrt{1 + \text{var}(f(\mathbf{x}_*))}} \right) . \quad (74)$$

Kelm et al. (2007) is using Gaussian process for classification in order to distinguish CaP in MRSI data.

– *Sparse kernel methods:*

In Gaussian process, 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 few example in order to label new observations (Bishop (2006)).

Support vector machine (SVM)¹⁰ is a sparse kernel method which aimed at finding the best linear hyperplane which separates two classes such as the margin

¹⁰SVM implementation can be found at: <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>

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 no points 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 $\|\mathbf{w}\|$. Hence, This problem solved by an optimization approach and formalized such that:

$$\begin{aligned} \arg \min_{\mathbf{w}} \quad & \frac{1}{2} \|\mathbf{w}\|^2, \\ \text{subject to} \quad & y_i(\mathbf{w} \cdot \mathbf{x}_i - b) \geq 1, \quad i = 1, \dots, n. \end{aligned} \quad (75)$$

From Eq. 75, 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 problems using Lagrange multipliers α . All points which are not lying on the margin will be assigned a corresponding $\alpha_i = 0$. This can be formalized such as:

$$\arg \min_{\mathbf{w}, b} \max_{\alpha \geq 0} \left\{ \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^n \alpha_i [y_i(\mathbf{w} \cdot \mathbf{x}_i - b) - 1] \right\}. \quad (76)$$

The different parameters can be inferred using quadratic programming. This version of SVM is known as hard-margin since that no point can lie in the margin area. However, this 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 into the margin but at the cost of penalty which will be minimized in the optimization process such that:

$$\arg \min_{\mathbf{w}, \xi, b} \max_{\alpha, \beta} \left\{ \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n \xi_i - \sum_{i=1}^n \alpha_i [y_i(\mathbf{w} \cdot \mathbf{x}_i - b) - 1 + \xi_i] - \sum_{i=1}^n \beta_i \xi_i \right\}. \quad (77)$$

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

$$C(\mathbf{x}_i) = \text{sign} \left(\sum_{n=1}^N \alpha_n (\mathbf{x}_n \cdot \mathbf{x}_i) + b_0 \right). \quad (78)$$

where $\mathbf{x}_n | n = \{1, \dots, N\}$, N 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 in an higher dimensional space in which it is assumed that a linear hyperplane will split the classes.

Different kernels are popular such as RBF kernel, polynomial kernels or Gaussian kernel.

In CAD system, SVM is the most popular classification method and was used in a multitude of research: 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 following covariance function (Quinero-Candela et al. (2002)):

$$K_{RVM}(\mathbf{x}_p, \mathbf{x}_q) = \sum_{j=1}^M \frac{1}{\alpha_j} \Phi_j(\mathbf{x}_p) \Phi_j(\mathbf{x}_q) . \quad (79)$$

where $\phi(.)$ is a Gaussian basis function, $\mathbf{x}_n | n = \{1, \dots, N\}$ are the N training points and $\boldsymbol{\alpha}$ are the weights.

As mentioned in Quinero-Candela et al. (2002), the sparsity regarding the relevance vector arises if $j\alpha_j^{-1} = 0$. The set of parameters $\boldsymbol{\alpha}$ 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 network consider as well as the most successful model of this kind in pattern recognition (Bishop (2006)). The most well spread model used is based on a two layers model where a prediction of an observation is computed such as:

$$C(\mathbf{x}_n, w_{ij}^{(1)}, w_{kj}^{(2)}) = \sigma \left[\sum_{j=0}^M w_{kj}^{(2)} h \left(\sum_{i=0}^D w_{ij}^{(1)} x_{in} \right) \right] . \quad (80)$$

where $h(.)$ and $\sigma(.)$ are two activations 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 hidden unit, respectively.

A graphical representation of this network is presented in Fig. 19. Relating Fig. 19 with Eq. 80, it can be noted that this network is composed of some successive non-linear mapping of the input data. First, a linear combination of

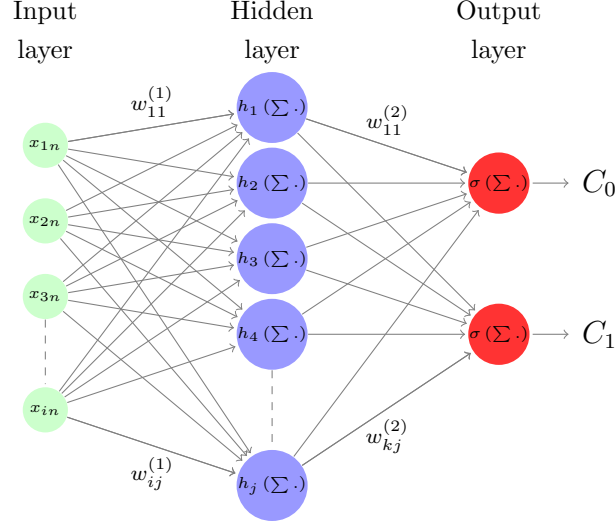


Figure 19: Representation of a neural network of the multilayer perceptron family.

the input vector \mathbf{x}_i is performed into some hidden units through a set of weight $w_{ij}^{(1)}$. This combination becomes non-linear by the use of the activation function $h(\cdot)$ which is usually chosen to be a sigmoid function. Then, a linear combination of these hidden units is performed into the output of the neural network through a set of weight $w_{kj}^{(2)}$. This combination is also mapped non-linearly using an activation function $\sigma(\cdot)$ which is usually a logistic function.

Thus, the training of such a network resides in finding the best weights $w_{ij}^{(1)}$ and $w_{kj}^{(2)}$ which will model our data the best. The error perform of this model can be computed such that:

$$E(w_{ij}^{(1)}, w_{kj}^{(2)}) = \frac{1}{2} \sum_{n=1}^N \left(C(\mathbf{x}_n, w_{ij}^{(1)}, w_{kj}^{(2)}) - y(\mathbf{x}_n) \right)^2 . \quad (81)$$

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

Thus the best set of weight can be inferred in an optimization framework such that:

$$\arg \min_{w_{ij}^{(1)}, w_{kj}^{(2)}} E(w_{ij}^{(1)}, w_{kj}^{(2)}) . \quad (82)$$

This optimization can be performed using a gradient descent method where

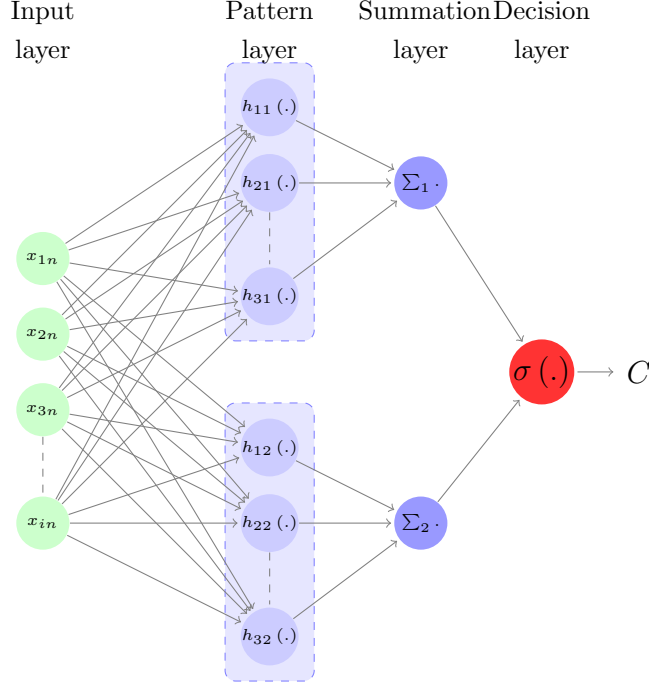


Figure 20: Representation of a neural network of the probabilistic neural network family.

the derivative of Eq. 81 can be computed using the backpropagation algorithm proposed by Rumelhart et al. (1988).

Matulewicz et al. (2013), Parfait et al. (2012) used this method to classify MRSI spectra.

Probabilistic neural network is another type of feed-forward which can be derived from the multilayer perceptron case and was proposed by Specht (1988). This classifier can be modelled by affecting the activation function $h(.)$ in Eq. 80 to an exponential function such that:

$$h(\mathbf{x}_n) = \exp \left(-\frac{(\mathbf{w}_j - \mathbf{x})^T (\mathbf{w}_j - \mathbf{x})}{2\sigma^2} \right) . \quad (83)$$

where σ is a free parameter.

The other difference of the probabilistic neural network with the multilayer perceptron resides in the architecture as shown in Fig. 20. This network is formed by two hidden layer. The first hidden layer corresponds to the pattern layer which is the mapping done using Eq. 83. This pattern layer is sub-divided into groups corresponding to the class. The second hidden layer corresponds to the

summation layer which simply sum 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 field can also be used as a lesion segmentation method to detect CaP. First, we can 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 estimate the best $\hat{\omega}$ which maximizes the posterior probability $p(\omega | \mathcal{F})$. Thus, using a Bayesian approach, this can be formulated such that:

$$p(\omega | \mathcal{F}) = \arg \max_{\omega} \prod_{s \in I} p(f_s | \omega_s) p(\omega) . \quad (84)$$

It is generally assumed that $p(f_s | \omega_s)$ follow 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 supposed to be a Markov random field thus:

$$p(\omega) = \frac{1}{Z} \exp(-U(\omega)) . \quad (85)$$

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

Thus the segmentation problem can be solved as an optimization problem where the energy function $U(.)$ has to be minimized. There are different possibilities to define the energy function $U(.)$. However, this is common to define the energy function such that it combines two type of potential function: (i) a local term relative to the pixel itself and (ii) a smoothing prior which embeds neighbourhood information which will more or less penalizes the energy function to obtain more or less homogeneous region. This optimization of such function can be performed algorithm such as iterated conditional modes (Kato and Pong (2001)).

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

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]

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

$$p(\omega|\mathcal{F}) = \frac{1}{Z} \exp \left[- \sum_{s \in I} V_{C1}(\omega_s|\mathcal{F}) - \sum_{\{s,r\} \in C} V_{C2}(\omega_s, \omega_r|\mathcal{F}) \right] . \quad (86)$$

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 validate and the other cases are used to train. This manipulation is repeated until that each patient was used as validation. This technique is popular when working with medical data due to the restricted number of patients included in datasets. Thus, it is allowing to train on a fair number of patients even with a small dataset. However, this technique suffer from high variance and can considered as a non reliable estimate (Efron (1983)).

The other very well known technique used for assessing classifier is the k 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 that it does not suffer from large variance. However, the number of patients in the dataset need to be large enough to apply such technique.

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]

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 into a class. 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 computes the ratio of true detection over 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 are giving 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. 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.

These both 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 to detect also more false negative. However, this fact is interesting in CAD since that it could be possible to obtain a high sensitivity

and to ensure that no cancers were missed even if more false alarm will have to be investigated. A statistic derive 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 comparison between models.

The previous method could have been classified on pixel-based evaluation method. However, a cancer can be also considered as a region. Free-response receiver operating characteristic (FROC) extends the ROC analysis but at region-based level. The same confusion matrix can be computed were the sample are not pixels any more but refer to a lesion. However, this 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 and the one inside the ground-truth are overlapping “enough”. “Enough” is defined by each researcher and can correspond to one pixel only!!!

Finally, Dice’s coefficient is sometimes computed which is corresponding to the similarity of a lesion between the ground-truth and the output of the classifier. This coefficient corresponds to the ratio of twice the number of pixels in common over the sum of the pixels of the lesions in the ground-truth and the output of the classifier.

5. Discussion

5.1. Brief comparison

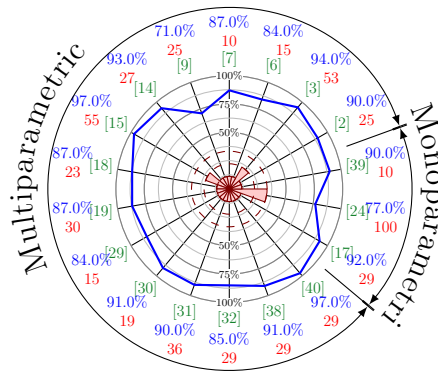
As discussed previously in Sect. 4.4.3, different metrics have been used to report results. A comparison of the different methods reviewed is given depending on the metric used in each research and also the type of MRI scanner used (cf., 1.5 *versus* 3.0 Tesla). For each research, the experiment obtaining the best result was reported into these figures.

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

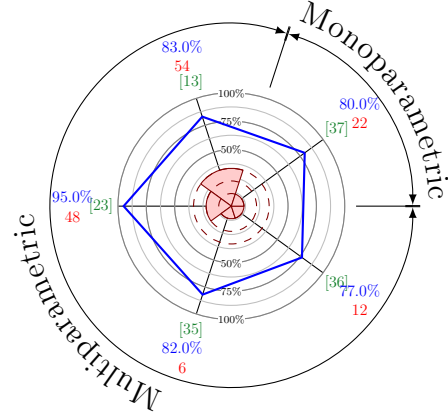
The results related with sensitivity and specificity are reported in Fig. 22. In the case that the data were collected with a 1.5 Tesla MRI scanner, the sensitivity is ranging from 74% and 100% and the specificity from 43% and 93%. For the experiments carried out with a 3.0 Tesla MRI scanner, the sensitivity is varying 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. The results are ranging 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



(a) Comparison in term of AUC-ROC of the methods using data from 1.5 Tesla MRI scanner.



(b) Comparison in term of AUC-ROC of the methods using data from 3.0 Tesla MRI scanner.

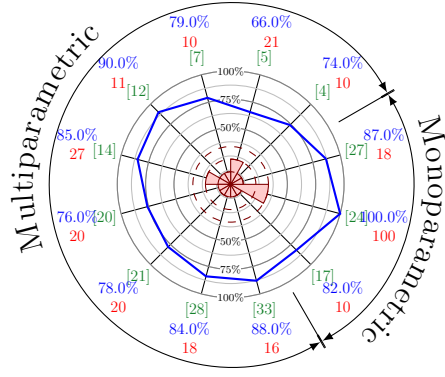
Figure 21: Comparison of the results in term of AUC for 1.5 and 3.0 Tesla MRI scanner. The blue value represent the metric and are graphically reported in the blue rose in the center of the figure. The red value correspond to the number of patients in the dataset and is also reported in the center of the figure. The numbers between brackets correspond to the reference as reported in Tab. 2.

implying different complexity involved and set of input parameter during the data acquisition. To our mind, the only to provide a real and fair comparison would be to provide a common working dataset where those algorithms could be tested.

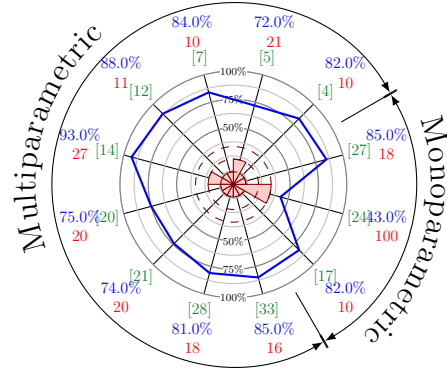
5.2. General discussion

This review leads to some general discussion 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 compare fairly the different algorithm designed over years. Also, the availability of a full multi-parametric dataset (cf., MRI and MRSI), could lead to algorithm which are using all the different modalities available currently. Recalling Tab. 2, it can be noted that none of the current work provides an algorithm using at the same time the four different modalities. We can also mentioned that all the algorithms are focused on one type of scanner only, either 1.5 Tesla and 3.0 Tesla. A dataset grouping these both types of imaging could allow to develop more generic algorithms.

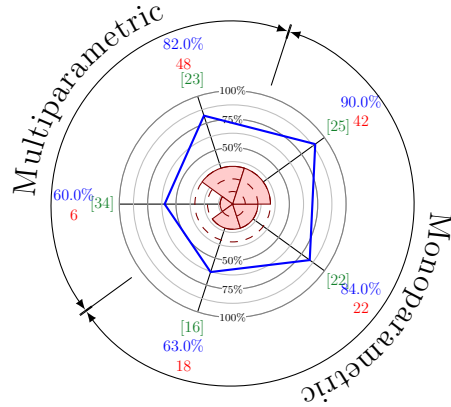
Then, by analysing the different stage of the CAD work-flow, it can be notice that the actual CAD systems do not include all the different pre-processing steps. It could be interesting to evaluate the gain of these pre-processing steps on the final results obtained after the classification in order to know if any significant improvement



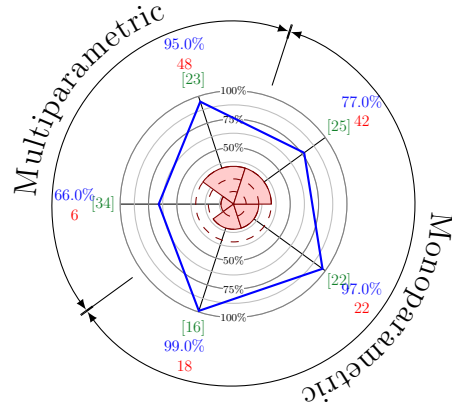
(a) Comparison in term of sensitivity of the methods using data from 1.5 Tesla MRI scanner.



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



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



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

Figure 22: Comparison of the results in term of sensitivity and specificity for 1.5 and 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.

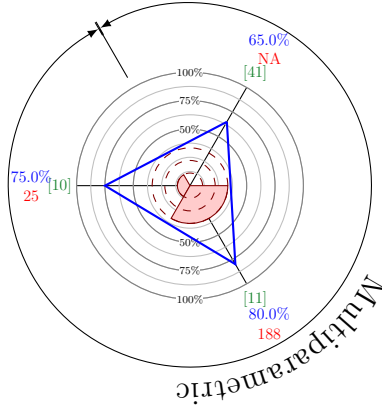


Figure 23: Comparison in term of FROC of the methods using data from 3.0 Tesla MRI scanner. The blue value represent the metric and are graphically reported in the blue curve in the center of the figure. The red value correspond to the number of patients in the dataset and is also reported in the center of the figure. The numbers between brackets correspond to the reference as reported in Tab. 2.

can be observed. Regarding segmentation and registration of the prostate, CAD systems could highly benefit from specific research in these area which could lead to a better automation of those system. Moreover, methods specific to segmentation and registration which are not actually used in CAD systems can performed better than the one used currently in CADs.

Regarding the classification framework, it seems that the current well-known pattern recognition method have been widely study. However, more investigations can 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 incorporated. Only, one study Liu et al. (2013) used some of these features.

Finally, an important point allowing a fair comparison between methods resides in the fact that no universal evaluation model and metric was defined in 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 would benefit from settling a standard metric which represents the most fairly the performance of the algorithms designed.

6. Conclusion

This review presented an overview and gave a staging of the research related to CAD development for CaP using multi-parametric MRI and MRSI data. We aimed

1899 at providing an introduction

1900 **7. Acknowledgement**

1901 We would like to acknowledge Sharad Nagappa for all the discussions and his
1902 precious advices regarding the redaction of this current review.

1903 **References**

- 1904 Agalliu, I., Gern, R., Leanza, S., Burk, R.D., 2009. Associations of high-grade
1905 prostate cancer with BRCA1 and BRCA2 founder mutations. *Clin. Cancer Res.*
1906 15, 1112–1120.
- 1907 Ahmed, N., Natarajan, T., Rao, K., 1974. Discrete cosine transform. *Computers,*
1908 *IEEE Transactions on C-23*, 90–93. doi:10.1109/T-C.1974.223784.
- 1909 Aizerman, M.A., Braverman, E.A., Rozonoer, L., 1964. Theoretical foundations of
1910 the potential function method in pattern recognition learning, in: *Automation and*
1911 *Remote Control*, pp. 821–837.
- 1912 Akin, O., Sala, E., Moskowitz, C.S., Kuroiwa, K., Ishill, N.M., Pucar, D., Scardino,
1913 P.T., Hricak, H., 2006. Transition zone prostate cancers: features, detection,
1914 localization, and staging at endorectal MR imaging. *Radiology* 239, 784–792.
- 1915 Alexander, D.D., Mink, P.J., Cushing, C.A., Scurman, B., 2010. A review and
1916 meta-analysis of prospective studies of red and processed meat intake and prostate
1917 cancer. *Nutr J* 9, 50.
- 1918 Amadasun, M., King, R., 1989. Textural features corresponding to textural prop-
1919 erties. *Systems, Man and Cybernetics, IEEE Transactions on* 19, 1264–1274.
1920 doi:10.1109/21.44046.
- 1921 American Cancer Society, A.C., 2010. Cancer Facts and Figures 2010. URL: [http:](http://www.cancer.org/research/cancerfactsfigures)
1922 [//www.cancer.org/research/cancerfactsfigures](http://www.cancer.org/research/cancerfactsfigures).
- 1923 American Cancer Society, A.C., 2013. Cancer Facts and Figures 2013. URL: [http:](http://www.cancer.org/research/cancerfactsfigures)
1924 [//www.cancer.org/research/cancerfactsfigures](http://www.cancer.org/research/cancerfactsfigures).
- 1925 Ampeliotis, D., Anonakoudi, A., Berberidis, K., Psarakis, E.Z., 2007. Computer
1926 aided detection of prostate cancer using fused information from dynamic contrast
1927 enhanced and morphological magnetic resonance images, in: *IEEE International*
1928 *Conference on Signal Processing and Communications*, pp. 888–891.

- 1929 Ampeliotis, D., Anonakoudi, A., Berberidis, K., Psarakis, E.Z., Kounoudes, A., 2008.
1930 A computer-aided system for the detection of prostate cancer based on magnetic
1931 resonance image analysis, in: International Symposium on Communications, Con-
1932 trol and Signal Processing.
- 1933 Amundadottir, L.T., Sulem, P., Gudmundsson, J., Helgason, A., Baker, A., Agnars-
1934 son, B.A., Sigurdsson, A., Benediktsdottir, K.R., Cazier, J.B., Sainz, J., Jakobs-
1935 dottir, M., Kostic, J., Magnúsdottir, D.N., Ghosh, S., Agnarsson, K., Birgisdot-
1936 tir, B., Le Roux, L., Olafsdottir, A., Blondal, T., Andresdottir, M., Gretarsdot-
1937 tir, O.S., Bergthorsson, J.T., Gudbjartsson, D., Gylfason, A., Thorleifsson, G.,
1938 Manolescu, A., Kristjansson, K., Geirsson, G., Isaksson, H., Douglas, J., Johans-
1939 son, J.E., Balter, K., Wiklund, F., Montie, J.E., Yu, X., Suarez, B.K., Ober, C.,
1940 Cooney, K.A., Gronberg, H., Catalona, W.J., Einarsson, G.V., Barkardottir, R.B.,
1941 Gulcher, J.R., Kong, A., Thorsteinsdottir, U., Stefansson, K., 2006. A common
1942 variant associated with prostate cancer in European and African populations. *Nat.*
1943 *Genet.* 38, 652–658.
- 1944 Andriole, G.L., Crawford, E.D., Grubb, R.L., Buys, S.S., Chia, D., Church, T.R.,
1945 Fouad, M.N., Gelmann, E.P., Kvale, P.A., Reding, D.J., Weissfeld, J.L., Yokochi,
1946 L.A., O'Brien, B., Clapp, J.D., Rathmell, J.M., Riley, T.L., Hayes, R.B., Kramer,
1947 B.S., Izmirlian, G., Miller, A.B., Pinsky, P.F., Prorok, P.C., Gohagan, J.K., Berg,
1948 C.D., 2009. Mortality results from a randomized Prostate-cancer screening trial.
1949 *New England Journal of Medicine* 360, 1310–1319.
- 1950 Antic, T., Peng, Y., Jiang, Y., Giger, M.L., Eggener, S., Oto, A., 2013. A study of
1951 T2-weighted MR image texture features and diffusion-weighted MR image features
1952 for computer-aided diagnosis of prostate cancer , 86701H–86701H–6.
- 1953 Artan, Y., Haider, M.A., Langer, D.L., van der Kwast, T.H., Evans, A.J., Yang, Y.,
1954 Wernick, M.N., Trachtenberg, J., Yetik, I.S., 2010. Prostate cancer localization
1955 with multispectral MRI using cost-sensitive support vector machines and condi-
1956 tional random fields. *IEEE Trans Image Process* 19, 2444–2455.
- 1957 Artan, Y., Langer, D., Haider, M., Van der Kwast, T.H., Evans, A., Wernick, M.,
1958 Yetik, I., 2009. Prostate cancer segmentation with multispectral MRI using cost-
1959 sensitive Conditional Random Fields, in: Biomedical Imaging: From Nano to
1960 Macro, 2009. ISBI '09. IEEE International Symposium on, pp. 278–281.
- 1961 Awwad, H.M., Geisel, J., Obeid, R., 2012. The role of choline in prostate cancer.
1962 *Clin. Biochem.* 45, 1548–1553.

- 1963 Barentsz, J.O., Richenberg, J., Clements, R., Choyke, P., Verma, S., Villeirs, G.,
1964 Rouviere, O., Logager, V., Futterer, J.J., 2012. ESUR prostate MR guidelines
1965 2012. *Eur Radiol* 22, 746–757.
- 1966 Belkin, M., Niyogi, P., 2001. Laplacian eigenmaps and spectral techniques for em-
1967 bedding and clustering, in: *Advances in Neural Information Processing Systems*
1968 14, MIT Press. pp. 585–591.
- 1969 Belongie, S., Malik, J., Puzicha, J., 2002. Shape matching and object recognition us-
1970 ing shape contexts. *Pattern Analysis and Machine Intelligence, IEEE Transactions*
1971 *on* 24, 509–522. doi:10.1109/34.993558.
- 1972 Benassi, A., Cohen, S., Istas, J., 1998. Identifying the multifractional function
1973 of a Gaussian process. *Statistics & Probability Letters* 39, 337 – 345. URL:
1974 <http://www.sciencedirect.com/science/article/pii/S0167715298000789>,
1975 doi:[http://dx.doi.org/10.1016/S0167-7152\(98\)00078-9](http://dx.doi.org/10.1016/S0167-7152(98)00078-9).
- 1976 Bishop, C.M., 2006. *Pattern recognition and machine learning*. Springer-Verlag New
1977 York, Inc., Secaucus, NJ, USA.
- 1978 Bookstein, F.L., 1989. Principal warps: thin-plate splines and the decomposition of
1979 deformations. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*
1980 11, 567–585. doi:10.1109/34.24792.
- 1981 Boser, B.E., Guyon, I.M., Vapnik, V.N., 1992. A training algorithm for opti-
1982 mal margin classifiers, in: *Proceedings of the Fifth Annual Workshop on Com-*
1983 *putational Learning Theory*, ACM, New York, NY, USA. pp. 144–152. URL:
1984 <http://doi.acm.org/10.1145/130385.130401>, doi:10.1145/130385.130401.
- 1985 Bourdounis, A., Papatsoris, A.G., Chrisofos, M., Efstathiou, E., Skolarikos, A.,
1986 Deliveliotis, C., 2010. The novel prostate cancer antigen 3 (PCA3) biomarker. *Int*
1987 *Braz J Urol* 36, 665–668.
- 1988 Breiman, L., 2001. Random forests. *Machine Learning* 45, 5–32.
- 1989 Breiman, L., Friedman, J., Olshen, R., Stone, C., 1984. *Classification and regression*
1990 *trees*. Wadsworth and Brooks, Monterey, CA.
- 1991 Brenner, J., Chinnaiyan, A., Tomlins, S., 2013. ETS fusion genes in prostate cancer,
1992 in: Tindall, D.J. (Ed.), *Prostate Cancer*. Springer New York. volume 16 of *Protein*
1993 *Reviews*, pp. 139–183.

- 1994 Buades, A., Coll, B., Morel, J., 2005. A review of image denoising algorithms, with
1995 a new one. *Simul* 4, 490–530.
- 1996 Buckley, D.L., Roberts, C., Parker, G.J., Logue, J.P., Hutchinson, C.E., 2004.
1997 Prostate cancer: evaluation of vascular characteristics with dynamic contrast-
1998 enhanced T1-weighted MR imaging—initial experience. *Radiology* 233, 709–715.
- 1999 Byrd, R.H., Lu, P., Nocedal, J., Zhu, C., 1995. A limited memory algorithm for
2000 bound constrained optimization. *SIAM J. Sci. Comput.* 16, 1190–1208. URL:
2001 <http://dx.doi.org/10.1137/0916069>, doi:10.1137/0916069.
- 2002 Carmeliet, P., Jain, R.K., 2000. Angiogenesis in cancer and other diseases. *Nature*
2003 407, 249–257.
- 2004 Carrol, C.L., Sommer, F.G., McNeal, J.E., Stamey, T.A., 1987. The abnormal
2005 prostate: MR imaging at 1.5 T with histopathologic correlation. *Radiology* 163,
2006 521–525.
- 2007 Castorina, P., Delsanto, P.P., Guiot, C., 2006. Classification scheme for phenomeno-
2008 logical universalities in growth problems in physics and other sciences. *Phys. Rev.*
2009 *Lett.* 96, 188701. URL: [http://link.aps.org/doi/10.1103/PhysRevLett.96.](http://link.aps.org/doi/10.1103/PhysRevLett.96.188701)
2010 188701, doi:10.1103/PhysRevLett.96.188701.
- 2011 Chan, H.P., Doi, K., Galhotra, S., Vyborny, C.J., MacMahon, H., Jokich, P.M.,
2012 1987. Image feature analysis and computer-aided diagnosis in digital radiography.
2013 I. Automated detection of microcalcifications in mammography. *Med Phys* 14,
2014 538–548.
- 2015 Chan, H.P., Hadjiiski, L., Zhou, C., Sahiner, B., 2008. Computer-aided diagnosis of
2016 lung cancer and pulmonary embolism in computed tomography—a review. *Acad*
2017 *Radiol* 15, 535–555.
- 2018 Chan, H.P., Sahiner, B., Helvie, M.A., Petrick, N., Roubidoux, M.A., Wilson, T.E.,
2019 Adler, D.D., Paramagul, C., Newman, J.S., Sanjay-Gopal, S., 1999. Improvement
2020 of radiologists’ characterization of mammographic masses by using computer-aided
2021 diagnosis: an ROC study. *Radiology* 212, 817–827.
- 2022 Chan, I., Wells, W., Mulkern, R.V., Haker, S., Zhang, J., Zou, K.H., Maier, S.E.,
2023 Tempany, C.M., 2003. Detection of prostate cancer by integration of line-scan dif-
2024 fusion, T2-mapping and T2-weighted magnetic resonance imaging; a multichannel
2025 statistical classifier. *Med Phys* 30, 2390–2398.

- 2026 Chappelow, J., Bloch, B.N., Rofsky, N., Genega, E., Lenkinski, R., DeWolf, W.,
2027 Madabhushi, A., 2011. Elastic registration of multimodal prostate MRI and his-
2028 tology via multiattribute combined mutual information. *Med Phys* 38, 2005–2018.
- 2029 Chen, L., Weng, Z., Goh, L., Garland, M., 2002. An efficient algorithm for automatic
2030 phase correction of {NMR} spectra based on entropy minimization . *Journal of*
2031 *Magnetic Resonance* 158, 164 – 168. URL: [http://www.sciencedirect.com/](http://www.sciencedirect.com/science/article/pii/S1090780702000691)
2032 [science/article/pii/S1090780702000691](http://www.sciencedirect.com/science/article/pii/S1090780702000691), doi:[http://dx.doi.org/10.1016/](http://dx.doi.org/10.1016/S1090-7807(02)00069-1)
2033 [S1090-7807\(02\)00069-1](http://dx.doi.org/10.1016/S1090-7807(02)00069-1).
- 2034 Cheng, H.D., Shan, J., Ju, W., Guo, Y., Zhang, L., 2010. Automated breast cancer
2035 detection and classification using ultrasound images: A survey. *Pattern Recogn.*
2036 43, 299–317.
- 2037 Choi, Y.J., Kim, J.K., Kim, N., Kim, K.W., Choi, E.K., Cho, K.S., 2007. Functional
2038 MR imaging of prostate cancer. *Radiographics* 27, 63–75.
- 2039 Chou, R., Croswell, J.M., Dana, T., Bougatsos, C., Blazina, I., Fu, R., Gleitsmann,
2040 K., Koenig, H.C., Lam, C., Maltz, A., Rugge, J.B., Lin, K., 2011. Screening for
2041 prostate cancer: a review of the evidence for the U.S. Preventive Services Task
2042 Force. *Ann. Intern. Med.* 155, 762–771.
- 2043 Coakley, F.V., Hricak, H., 2000. Radiologic anatomy of the prostate gland: a clinical
2044 approach. *Radiol. Clin. North Am.* 38, 15–30.
- 2045 Cohen, R.J., Shannon, B.A., Phillips, M., Moorin, R.E., Wheeler, T.M., Garrett,
2046 K.L., 2008. Central zone carcinoma of the prostate gland: a distinct tumor type
2047 with poor prognostic features. *J. Urol.* 179, 1762–1767.
- 2048 Coifman, R., Wickerhauser, M., 1992. Entropy-based algorithms for best basis selec-
2049 tion. *Information Theory, IEEE Transactions on* 38, 713–718. doi:10.1109/18.
2050 119732.
- 2051 Coleman, T., Li, Y., 1993. An interior trust region approach for nonlinear minimiza-
2052 tion subject to bounds. Technical Report. Cornell University.
- 2053 Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J., 1995. Active shape mod-
2054 els—Their training and application. *Comput. Vis. Image Underst.* 61, 38–
2055 59. URL: <http://dx.doi.org/10.1006/cviu.1995.1004>, doi:10.1006/cviu.
2056 1995.1004.

- 2057 Cortes, C., Vapnik, V., 1995. Support-Vector networks. *Machine Learning* 20, 273–
2058 297. URL: [http://dx.doi.org/10.1023/A:](http://dx.doi.org/10.1023/A%3A1022627411411)
2059 1022627411411.
- 2060 Costello, L.C., Franklin, R.B., 2006. The clinical relevance of the metabolism of
2061 prostate cancer; zinc and tumor suppression: connecting the dots. *Mol. Cancer* 5,
2062 17.
- 2063 Cruz, M., Tsuda, K., Narumi, Y., Kuroiwa, Y., Nose, T., Kojima, Y., Okuyama, A.,
2064 Takahashi, S., Aozasa, K., Barentsz, J.O., Nakamura, H., 2002. Characterization
2065 of low-intensity lesions in the peripheral zone of prostate on pre-biopsy endorectal
2066 coil MR imaging. *Eur Radiol* 12, 357–365.
- 2067 Dalal, N., Triggs, B., 2005. Histograms of oriented gradients for human detection,
2068 in: *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer*
2069 *Society Conference on*, pp. 886–893 vol. 1. doi:10.1109/CVPR.2005.177.
- 2070 Daugman, J.G., 1985. Uncertainty relation for resolution in space, spatial frequency,
2071 and orientation optimized by two-dimensional visual cortical filters. *J Opt Soc*
2072 *Am A* 2, 1160–1169.
- 2073 Dean, J.C., Ilvento, C.C., 2006. Improved cancer detection using computer-aided
2074 detection with diagnostic and screening mammography: prospective study of 104
2075 cancers. *AJR Am J Roentgenol* 187, 20–28.
- 2076 Delongchamps, N.B., Peyromaure, M., Schull, A., Beuvon, F., Bouazza, N., Flam,
2077 T., Zerbib, M., Muradyan, N., Legman, P., Cornud, F., 2013. Prebiopsy magnetic
2078 resonance imaging and prostate cancer detection: comparison of random and tar-
2079 geted biopsies. *J. Urol.* 189, 493–499.
- 2080 Delpierre, C., Lamy, S., Kelly-Irving, M., Molinie, F., Velten, M., Tretarre, B.,
2081 Woronoff, A.S., Buemi, A., Lapotre-Ledoux, B., Bara, S., Guizard, A.V., Colonna,
2082 M., Grosclaude, P., 2013. Life expectancy estimates as a key factor in over-
2083 treatment: the case of prostate cancer. *Cancer Epidemiol* 37, 462–468.
- 2084 Devos, A., Lukas, L., Suykens, J.A., Vanhamme, L., Tate, A.R., Howe, F.A., Majos,
2085 C., Moreno-Torres, A., van der Graaf, M., Arus, C., Van Huffel, S., 2004. Classi-
2086 fication of brain tumours using short echo time 1H MR spectra. *J. Magn. Reson.*
2087 170, 164–175.
- 2088 Doi, K., Chan, H.P., Giger, M., 1990. Method and system for enhancement and
2089 detection of abnormal anatomic regions in a digital image.

- 2090 Donoho, D.L., Johnstone, J.M., 1994. Ideal spatial adaptation by wavelet shrinkage.
2091 *Biometrika* 81, 425–455. doi:10.1093/biomet/81.3.425.
- 2092 Doo, K.W., Sung, D.J., Park, B.J., Kim, M.J., Cho, S.B., Oh, Y.W., Ko, Y.H., Yang,
2093 K.S., 2012. Detectability of low and intermediate or high risk prostate cancer with
2094 combined T2-weighted and diffusion-weighted MRI. *Eur Radiol* 22, 1812–1819.
- 2095 Efron, B., 1979. Bootstrap methods: Another look at the jackknife. *The Annals of*
2096 *Statistics* 7, 1–26. URL: <http://dx.doi.org/10.1214/aos/1176344552>, doi:10.
2097 1214/aos/1176344552.
- 2098 Efron, B., 1983. Estimating the error rate of a prediction rule: Improvement on
2099 cross-validation. *Journal of the American Statistical Association* 78, pp. 316–331.
2100 URL: <http://www.jstor.org/stable/2288636>.
- 2101 Elter, M., Horsch, A., 2009. CADx of mammographic masses and clustered micro-
2102 calcifications: a review. *Med Phys* 36, 2052–2068.
- 2103 Epstein, J.I., Allsbrook, W.C., Amin, M.B., Egevad, L.L., 2005. The 2005 Interna-
2104 tional Society of Urological Pathology (ISUP) Consensus Conference on Gleason
2105 Grading of Prostatic Carcinoma. *Am. J. Surg. Pathol.* 29, 1228–1242.
- 2106 Etzioni, R., Penson, D.F., Legler, J.M., di Tommaso, D., Boer, R., Gann, P.H., Feuer,
2107 E.J., 2002. Overdiagnosis due to prostate-specific antigen screening: lessons from
2108 U.S. prostate cancer incidence trends. *J. Natl. Cancer Inst.* 94, 981–990.
- 2109 Ferlay, J., Shin, H.R., Bray, F., Forman, D., Mathers, C., Parkin, D.M., 2010. Esti-
2110 mates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int. J. Cancer*
2111 127, 2893–2917.
- 2112 Fodor, I., 2002. A survey of dimension reduction techniques.
- 2113 Fred, A., Jain, A., 2005. Combining multiple clusterings using evidence accumula-
2114 tion. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 27, 835–
2115 850. doi:10.1109/TPAMI.2005.113.
- 2116 Freedman, M.L., Haiman, C.A., Patterson, N., McDonald, G.J., Tandon, A., Wal-
2117 iszewska, A., Penney, K., Steen, R.G., Ardlie, K., John, E.M., Oakley-Girvan, I.,
2118 Whittemore, A.S., Cooney, K.A., Ingles, S.A., Altshuler, D., Henderson, B.E., Re-
2119 ich, D., 2006. Admixture mapping identifies 8q24 as a prostate cancer risk locus
2120 in African-American men. *Proc. Natl. Acad. Sci. U.S.A.* 103, 14068–14073.

- 2121 Freund, Y., Schapire, R., 1997. A decision-theoretic generalization of on-line learn-
2122 ing and an application to boosting. *Journal of Computer and System Sciences*
2123 55, 119 – 139. URL: [http://www.sciencedirect.com/science/article/pii/](http://www.sciencedirect.com/science/article/pii/S002200009791504X)
2124 [S002200009791504X](http://www.sciencedirect.com/science/article/pii/S002200009791504X), doi:<http://dx.doi.org/10.1006/jcss.1997.1504>.
- 2125 Friedman, J., 1989. Regularized discriminant analysis. *Journal of the American*
2126 *Statistical Association* 84, pp. 165–175. URL: [http://www.jstor.org/stable/](http://www.jstor.org/stable/2289860)
2127 [2289860](http://www.jstor.org/stable/2289860).
- 2128 Gabor, D., 1946. Theory of communication. Part 1: The analysis of information.
2129 *Electrical Engineers - Part III: Radio and Communication Engineering*, *Journal of*
2130 *the Institution of* 93, 429–441. doi:10.1049/ji-3-2.1946.0074.
- 2131 Ghose, S., Oliver, A., Marti, R., Llado, X., Vilanova, J.C., Freixenet, J., Mitra, J.,
2132 Sidibe, D., Meriaudeau, F., 2012. A survey of prostate segmentation methodologies
2133 in ultrasound, magnetic resonance and computed tomography images. *Comput*
2134 *Methods Programs Biomed* 108, 262–287.
- 2135 Giannini, V., Vignati, A., Mazzetti, S., De Luca, M., Bracco, C., Stasi, M., Russo, F.,
2136 Armando, E., Regge, D., 2013. A prostate CAD system based on multiparametric
2137 analysis of DCE T1-w, and DW automatically registered images , 86703E–86703E–
2138 6.
- 2139 Gibbs, P., Tozer, D.J., Liney, G.P., Turnbull, L.W., 2001. Comparison of quantitative
2140 T2 mapping and diffusion-weighted imaging in the normal and pathologic prostate.
2141 *Magn Reson Med* 46, 1054–1058.
- 2142 Giger, M.L., Chan, H.P., Boone, J., 2008. Anniversary paper: History and status
2143 of CAD and quantitative image analysis: the role of Medical Physics and AAPM.
2144 *Med Phys* 35, 5799–5820.
- 2145 Giger, M.L., Doi, K., MacMahon, H., 1988. Image feature analysis and computer-
2146 aided diagnosis in digital radiography. 3. Automated detection of nodules in pe-
2147 ripheral lung fields. *Med Phys* 15, 158–166.
- 2148 Giovannucci, E., Liu, Y., Platz, E.A., Stampfer, M.J., Willett, W.C., 2007. Risk
2149 factors for prostate cancer incidence and progression in the health professionals
2150 follow-up study. *Int. J. Cancer* 121, 1571–1578.
- 2151 Giskeodegard, G.F., Bertilsson, H., Selnaes, K.M., Wright, A.J., Bathen, T.F., Viset,
2152 T., Halgunset, J., Angelsen, A., Gribbestad, I.S., Tessem, M.B., 2013. Spermine

2153 and citrate as metabolic biomarkers for assessing prostate cancer aggressiveness.
 2154 PLoS ONE 8, e62375.

2155 Gleason, D.F., 1977. Urologic pathology: The prostate. Lea and Febiger.. chapter
 2156 The Veteran's Administration Cooperative Urologic Research Group: histologic
 2157 grading and clinical staging of prostatic carcinoma. p. 171198.

2158 Goodman, S.N., 1999. Toward evidence-based medical statistics. 1: The P value
 2159 fallacy. Ann. Intern. Med. 130, 995–1004.

2160 van der Graaf, M., Schipper, R.G., Oosterhof, G.O., Schalken, J.A., Verhofstad,
 2161 A.A., Heerschap, A., 2000. Proton MR spectroscopy of prostatic tissue focused on
 2162 the detection of spermine, a possible biomarker of malignant behavior in prostate
 2163 cancer. MAGMA 10, 153–159.

2164 Gribbestad, I., Gjesdal, K., Nilsen, G., Lundgren, S., Hjelstuen, M., Jackson, A.,
 2165 2005. An introduction to dynamic contrast-enhanced MRI in oncology, in: Jack-
 2166 son, A., Buckley, D., Parker, G. (Eds.), Dynamic Contrast-Enhanced Magnetic
 2167 Resonance Imaging in Oncology. Springer Berlin Heidelberg. Medical Radiology,
 2168 pp. 1–22.

2169 Haacke, E., Brown, R., Thompson, M., Venkatesan, R., 1999. Magnetic resonance
 2170 imaging: Physical principles and sequence design. Wiley. URL: [http://books.](http://books.google.es/books?id=BnOvQgAACAAJ)
 2171 [google.es/books?id=BnOvQgAACAAJ](http://books.google.es/books?id=BnOvQgAACAAJ).

2172 Haas, G.P., Delongchamps, N.B., Jones, R.F., Chandan, V., Serio, A.M., Vickers,
 2173 A.J., Jumbelic, M., Threatte, G., Korets, R., Lilja, H., de la Roza, G., 2007.
 2174 Needle biopsies on autopsy prostates: sensitivity of cancer detection based on true
 2175 prevalence. J. Natl. Cancer Inst. 99, 1484–1489.

2176 Hambrock, T., Somford, D.M., Huisman, H.J., van Oort, I.M., Witjes, J.A.,
 2177 Hulsbergen-van de Kaa, C.A., Scheenen, T., Barentsz, J.O., 2011. Relationship
 2178 between apparent diffusion coefficients at 3.0-T MR imaging and Gleason grade
 2179 in peripheral zone prostate cancer. Radiology 259, 453–461.

2180 Hambrock, T., Vos, P.C., Hulsbergen-van de Kaa, C.A., Barentsz, J.O., Huisman,
 2181 H.J., 2013. Prostate cancer: computer-aided diagnosis with multiparametric 3-T
 2182 MR imaging—effect on observer performance. Radiology 266, 521–530.

2183 Hara, N., Okuizumi, M., Koike, H., Kawaguchi, M., Bilim, V., 2005. Dynamic
 2184 contrast-enhanced magnetic resonance imaging (DCE-MRI) is a useful modality
 2185 for the precise detection and staging of early prostate cancer. Prostate 62, 140–147.

2186 Haralick, R., Shanmugam, K., Dinstein, I., 1973. Textural features for image classi-
2187 fication. *Systems, Man and Cybernetics*, IEEE Transactions on SMC-3, 610–621.
2188 doi:10.1109/TSMC.1973.4309314.

2189 Hegde, J.V., Mulkern, R.V., Panych, L.P., Fennessy, F.M., Fedorov, A., Maier,
2190 S.E., Tempany, C.M., 2013. Multiparametric MRI of prostate cancer: an update
2191 on state-of-the-art techniques and their performance in detecting and localizing
2192 prostate cancer. *J Magn Reson Imaging* 37, 1035–1054.

2193 Heidenreich, A., Abrahamsson, P.A., Artibani, W., Catto, J., Montorsi, F., Van Pop-
2194 pel, H., Wirth, M., Mottet, N., 2013. Early Detection of Prostate Cancer: Euro-
2195 pean Association of Urology Recommendation. *Eur. Urol.* .

2196 Hero, A., Ma, B., Michel, O., Gorman, J., 2002. Applications of entropic spanning
2197 graphs. *Signal Processing Magazine*, IEEE 19, 85–95. doi:10.1109/MSP.2002.
2198 1028355.

2199 Hoeks, C.M., Barentsz, J.O., Hambrock, T., Yakar, D., Somford, D.M., Heijmink,
2200 S.W., Scheenen, T.W., Vos, P.C., Huisman, H., van Oort, I.M., Witjes, J.A.,
2201 Heerschap, A., Futterer, J.J., 2011. Prostate cancer: multiparametric MR imaging
2202 for detection, localization, and staging. *Radiology* 261, 46–66.

2203 Hoffman, R.M., Gilliland, F.D., Eley, J.W., Harlan, L.C., Stephenson, R.A., Stan-
2204 ford, J.L., Albertson, P.C., Hamilton, A.S., Hunt, W.C., Potosky, A.L., 2001.
2205 Racial and ethnic differences in advanced-stage prostate cancer: the Prostate Can-
2206 cer Outcomes Study. *J. Natl. Cancer Inst.* 93, 388–395.

2207 Hricak, H., Doms, G.C., McNeal, J.E., Mark, A.S., Marotti, M., Avallone, A.,
2208 Pelzer, M., Proctor, E.C., Tanagho, E.A., 1987. MR imaging of the prostate
2209 gland: normal anatomy. *AJR Am J Roentgenol* 148, 51–58.

2210 Hricak, H., Williams, R.D., Spring, D.B., Moon, K.L., Hedgcock, M.W., Watson,
2211 R.A., Crooks, L.E., 1983. Anatomy and pathology of the male pelvis by magnetic
2212 resonance imaging. *AJR Am J Roentgenol* 141, 1101–1110.

2213 Huch Boni, R.A., Boner, J.A., Lutolf, U.M., Trinkler, F., Pestalozzi, D.M., Krestin,
2214 G.P., 1995. Contrast-enhanced endorectal coil MRI in local staging of prostate
2215 carcinoma. *J Comput Assist Tomogr* 19, 232–237.

2216 Hugosson, J., Carlsson, S., Aus, G., Bergdahl, S., Khatami, A., Lodding, P., Pihl,
2217 C.G., Stranne, J., Holmberg, E., Lilja, H., 2010. Mortality results from the

- 2218 Göteborg randomised population-based prostate-cancer screening trial. *Lancet*
2219 *Oncol.* 11, 725–732.
- 2220 Huisman, H., Vos, P., Litjens, G., Hambrock, T., Barentsz, J., 2010. Computer aided
2221 detection of prostate cancer using T2, DWI and DCE MRI: methods and clinical
2222 applications, in: *Proceedings of the 2010 international conference on Prostate*
2223 *cancer imaging: computer-aided diagnosis, prognosis, and intervention*, Springer-
2224 Verlag, Berlin, Heidelberg. pp. 4–14.
- 2225 Huisman, T.A., 2003. Diffusion-weighted imaging: basic concepts and application in
2226 cerebral stroke and head trauma. *Eur Radiol* 13, 2283–2297.
- 2227 Itou, Y., Nakanishi, K., Narumi, Y., Nishizawa, Y., Tsukuma, H., 2011. Clinical
2228 utility of apparent diffusion coefficient (ADC) values in patients with prostate
2229 cancer: can ADC values contribute to assess the aggressiveness of prostate cancer?
2230 *J Magn Reson Imaging* 33, 167–172.
- 2231 Jager, G.J., Ruijter, E.T., van de Kaa, C.A., de la Rosette, J.J., Oosterhof, G.O.,
2232 Thornbury, J.R., Ruijs, S.H., Barentsz, J.O., 1997. Dynamic TurboFLASH sub-
2233 traction technique for contrast-enhanced MR imaging of the prostate: correlation
2234 with histopathologic results. *Radiology* 203, 645–652.
- 2235 Jolliffe, I.T., 2002. *Principal Component Analysis*. Second ed., Springer. URL:
2236 [http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20&path=](http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20&path=ASIN/0387954422)
2237 [ASIN/0387954422](http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20&path=ASIN/0387954422).
- 2238 Jungke, M., Von Seelen, W., Bielke, G., Meindl, S., Grigat, M., Pfannenstiel, P.,
2239 1987. A system for the diagnostic use of tissue characterizing parameters in NMR-
2240 tomography, in: *Proc. of Information Processing in Medical Imaging*, pp. 471–481.
- 2241 Kaji, Y., Kurhanewicz, J., Hricak, H., Sokolov, D.L., Huang, L.R., Nelson, S.J.,
2242 Vigneron, D.B., 1998. Localizing prostate cancer in the presence of postbiopsy
2243 changes on MR images: role of proton MR spectroscopic imaging. *Radiology* 206,
2244 785–790.
- 2245 Kato, Z., Pong, T., 2001. A Markov random field image segmentation model using
2246 combined color and texture features, in: Skarbek, W. (Ed.), *Computer Analy-*
2247 *sis of Images and Patterns*. Springer Berlin Heidelberg. volume 2124 of *Lecture*
2248 *Notes in Computer Science*, pp. 547–554. URL: [http://dx.doi.org/10.1007/](http://dx.doi.org/10.1007/3-540-44692-3_66)
2249 [3-540-44692-3_66](http://dx.doi.org/10.1007/3-540-44692-3_66), doi:10.1007/3-540-44692-3_66.

- 2250 Kelm, B.M., Menze, B.H., Zechmann, C.M., Baudendistel, K.T., Hamprecht, F.A.,
2251 2007. Automated estimation of tumor probability in prostate magnetic resonance
2252 spectroscopic imaging: pattern recognition vs quantification. *Magn Reson Med*
2253 57, 150–159.
- 2254 Kety, S., 1951. The theory and applications of the exchange of inert gas at the lungs
2255 and tissues. *Pharmacol. Rev.* 3, 1–41.
- 2256 Kim, J.K., Hong, S.S., Choi, Y.J., Park, S.H., Ahn, H., Kim, C.S., Cho, K.S.,
2257 2005. Wash-in rate on the basis of dynamic contrast-enhanced MRI: usefulness for
2258 prostate cancer detection and localization. *J Magn Reson Imaging* 22, 639–646.
- 2259 Kirkham, A.P., Emberton, M., Allen, C., 2006. How good is MRI at detecting and
2260 characterising cancer within the prostate? *Eur. Urol.* 50, 1163–1174.
- 2261 Kirsch, R., 1971. Computer determination of the constituent structure of biological
2262 images . *Computers and Biomedical Research* 4, 315 – 328. URL: [http://](http://www.sciencedirect.com/science/article/pii/0010480971900346)
2263 www.sciencedirect.com/science/article/pii/0010480971900346, doi:[http:](http://dx.doi.org/10.1016/0010-4809(71)90034-6)
2264 [//dx.doi.org/10.1016/0010-4809\(71\)90034-6](http://dx.doi.org/10.1016/0010-4809(71)90034-6).
- 2265 Koh, D.M., Collins, D.J., 2007. Diffusion-weighted MRI in the body: applications
2266 and challenges in oncology. *AJR Am J Roentgenol* 188, 1622–1635.
- 2267 Korotkov, K., Garcia, R., 2012. Computerized analysis of pigmented skin lesions: a
2268 review. *Artif Intell Med* 56, 69–90.
- 2269 Kurhanewicz, J., Vigneron, D.B., Hricak, H., Narayan, P., Carroll, P., Nelson, S.J.,
2270 1996. Three-dimensional H-1 MR spectroscopic imaging of the in situ human
2271 prostate with high (0.24-0.7-cm³) spatial resolution. *Radiology* 198, 795–805.
- 2272 Langer, D.L., van der Kwast, T.H., Evans, A.J., Trachtenberg, J., Wilson, B.C.,
2273 Haider, M.A., 2009. Prostate cancer detection with multi-parametric MRI: logistic
2274 regression analysis of quantitative T2, diffusion-weighted imaging, and dynamic
2275 contrast-enhanced MRI. *J Magn Reson Imaging* 30, 327–334.
- 2276 Larsson, H.B., Fritz-Hansen, T., Rostrup, E., Sondergaard, L., Ring, P., Henriksen,
2277 O., 1996. Myocardial perfusion modeling using MRI. *Magn Reson Med* 35, 716–
2278 726.
- 2279 Laudadio, T., Mastronardi, N., Vanhamme, L., Hecke, P.V., Huffel, S.V., 2002. Im-
2280 proved Lanczos algorithms for blackbox {MRS} data quantitation. *Journal of*

- 2281 Magnetic Resonance 157, 292 – 297. URL: [http://www.sciencedirect.com/](http://www.sciencedirect.com/science/article/pii/S1090780702925930)
 2282 [science/article/pii/S1090780702925930](http://www.sciencedirect.com/science/article/pii/S1090780702925930), doi:[http://dx.doi.org/10.1006/](http://dx.doi.org/10.1006/jmre.2002.2593)
 2283 [jmre.2002.2593](http://dx.doi.org/10.1006/jmre.2002.2593).
- 2284 Le Bihan, D., Breton, E., Lallemand, D., Aubin, M.L., Vignaud, J., Laval-Jeantet,
 2285 M., 1988. Separation of diffusion and perfusion in intravoxel incoherent motion
 2286 MR imaging. *Radiology* 168, 497–505.
- 2287 Le Bihan, D., Breton, E., Lallemand, D., Grenier, P., Cabanis, E., Laval-Jeantet,
 2288 M., 1986. MR imaging of intravoxel incoherent motions: application to diffusion
 2289 and perfusion in neurologic disorders. *Radiology* 161, 401–407.
- 2290 Leissner, K.H., Tisell, L.E., 1979. The weight of the human prostate. *Scand. J. Urol.*
 2291 *Nephrol.* 13, 137–142.
- 2292 Lemaître, G., 2011. Absolute quantification at 3 T. Master’s thesis. Université de
 2293 Bourgogne, Heriot-Watt University, Universitat de Girona.
- 2294 Li, F., Aoyama, M., Shiraishi, J., Abe, H., Li, Q., Suzuki, K., Engelmann, R., Sone,
 2295 S., Macmahon, H., Doi, K., 2004. Radiologists’ performance for differentiating be-
 2296 nign from malignant lung nodules on high-resolution CT using computer-estimated
 2297 likelihood of malignancy. *AJR Am J Roentgenol* 183, 1209–1215.
- 2298 Li, H., Giger, M.L., Olopade, O.I., Margolis, A., Lan, L., Chinander, M.R., 2005.
 2299 Computerized texture analysis of mammographic parenchymal patterns of digi-
 2300 tized mammograms. *Acad Radiol* 12, 863–873.
- 2301 Li, Q., Sone, S., Doi, K., 2003. Selective enhancement filters for nodules, vessels, and
 2302 airway walls in two- and three-dimensional CT scans. *Med Phys* 30, 2040–2051.
- 2303 Li, S.Z., 1996. Robustizing robust M-estimation using deterministic annealing. *Pat-*
 2304 *tern Recognition* 29, 159–166.
- 2305 Lieber, C.A., Mahadevan-Jansen, A., 2003. Automated method for subtraction of
 2306 fluorescence from biological Raman spectra. *Appl Spectrosc* 57, 1363–1367.
- 2307 Liney, G.P., Knowles, A.J., Manton, D.J., Turnbull, L.W., Blackband, S.J., Hors-
 2308 man, A., 1996a. Comparison of conventional single echo and multi-echo sequences
 2309 with a fast spin-echo sequence for quantitative T2 mapping: application to the
 2310 prostate. *J Magn Reson Imaging* 6, 603–607.

2311 Liney, G.P., Lowry, M., Turnbull, L.W., Manton, D.J., Knowles, A.J., Blackband,
2312 S.J., Horsman, A., 1996b. Proton MR T2 maps correlate with the citrate concen-
2313 tration in the prostate. *NMR Biomed* 9, 59–64.

2314 Liney, G.P., Turnbull, L.W., Lowry, M., Turnbull, L.S., Knowles, A.J., Horsman, A.,
2315 1997. In vivo quantification of citrate concentration and water T2 relaxation time
2316 of the pathologic prostate gland using ^1H MRS and MRI. *Magn Reson Imaging*
2317 15, 1177–1186.

2318 Litjens, G., Debats, O., van de Ven, W., Karssemeijer, N., Huisman, H., 2012a. A
2319 pattern recognition approach to zonal segmentation of the prostate on MRI. *Med*
2320 *Image Comput Comput Assist Interv* 15, 413–420.

2321 Litjens, G.J.S., Barentsz, J.O., Karssemeijer, N., Huisman, H.J., 2012b. Automated
2322 computer-aided detection of prostate cancer in MR images: from a whole-organ
2323 to a zone-based approach , 83150G–83150G–6.

2324 Litjens, G.J.S., Vos, P.C., Barentsz, J.O., Karssemeijer, N., Huisman, H.J., 2011.
2325 Automatic computer aided detection of abnormalities in multi-parametric prostate
2326 MRI , 79630T–79630T–7.

2327 Liu, P., Wang, S., Turkbey, B., Grant, K. and Pinto, P.C.P., Wood, B.J., Summers,
2328 R.M., 2013. A prostate cancer computer-aided diagnosis system using multimodal
2329 magnetic resonance imaging and targeted biopsy labels , 86701G–86701G–6.

2330 Liu, W., Turkbey, B., Senegas, J., Remmele, S., Xu, S., Kruecker, J., Bernardo,
2331 M., Wood, B.J., Pinto, P.A., Choyke, P.L., 2011. Accelerated T2 mapping for
2332 characterization of prostate cancer. *Magn Reson Med* 65, 1400–1406.

2333 Liu, X., Langer, D.L., Haider, M.A., Yang, Y., Wernick, M.N., Yetik, I.S., 2009.
2334 Prostate cancer segmentation with simultaneous estimation of Markov random
2335 field parameters and class. *IEEE Trans Med Imaging* 28, 906–915.

2336 Lopes, R., Ayache, A., Makni, N., Puech, P., Villers, A., Mordon, S., Betrouni, N.,
2337 2011. Prostate cancer characterization on MR images using fractal features. *Med*
2338 *Phys* 38, 83–95.

2339 Lu-Yao, G.L., Albertsen, P.C., Moore, D.F., Shih, W., Lin, Y., DiPaola, R.S., Barry,
2340 M.J., Zietman, A., O’Leary, M., Walker-Corkery, E., Yao, S.L., 2009. Outcomes of
2341 localized prostate cancer following conservative management. *JAMA* 302, 1202–
2342 1209.

- 2343 Lv, D., Guo, X., Wang, X., Zhang, J., Fang, J., 2009. Computerized characterization
2344 of prostate cancer by fractal analysis in MR images. *J Magn Reson Imaging* 30,
2345 161–168.
- 2346 Ma, R.W., Chapman, K., 2009. A systematic review of the effect of diet in prostate
2347 cancer prevention and treatment. *J Hum Nutr Diet* 22, 187–199.
- 2348 Madabhushi, A., Udupa, J., Souza, A., 2006. Generalized scale: Theory, algo-
2349 rithms, and application to image inhomogeneity correction . *Computer Vision and*
2350 *Image Understanding* 101, 100 – 121. URL: [http://www.sciencedirect.com/](http://www.sciencedirect.com/science/article/pii/S1077314205001244)
2351 [science/article/pii/S1077314205001244](http://www.sciencedirect.com/science/article/pii/S1077314205001244), doi:[http://dx.doi.org/10.1016/](http://dx.doi.org/10.1016/j.cviu.2005.07.010)
2352 [j.cviu.2005.07.010](http://dx.doi.org/10.1016/j.cviu.2005.07.010).
- 2353 Madabhushi, A., Udupa, J.K., 2006. New methods of MR image intensity standard-
2354 ization via generalized scale. *Med Phys* 33, 3426–3434.
- 2355 Maintz, J.B., Viergever, M.A., 1998. A survey of medical image registration. *Med*
2356 *Image Anal* 2, 1–36.
- 2357 Mallat, S., 2008. A wavelet tour of signal processing, Third Edition: The sparse way.
2358 3rd ed., Academic Press.
- 2359 Manjon, J.V., Carbonell-Caballero, J., Lull, J.J., Garcia-Marti, G., Marti-Bonmati,
2360 L., Robles, M., 2008. MRI denoising using non-local means. *Med Image Anal* 12,
2361 514–523.
- 2362 Matulewicz, L., Jansen, J.F., Bokacheva, L., Vargas, H.A., Akin, O., Fine, S.W.,
2363 Shukla-Dave, A., Eastham, J.A., Hricak, H., Koutcher, J.A., Zakian, K.L., 2013.
2364 Anatomic segmentation improves prostate cancer detection with artificial neu-
2365 ral networks analysis of 1H magnetic resonance spectroscopic imaging. *Journal*
2366 *of Magnetic Resonance Imaging* , n/a–n/aURL: [http://dx.doi.org/10.1002/](http://dx.doi.org/10.1002/jmri.24487)
2367 [jmri.24487](http://dx.doi.org/10.1002/jmri.24487), doi:10.1002/jmri.24487.
- 2368 Mazzetti, S., De Luca, M., Bracco, C., Vignati, A., Giannini, V., Stasi, M., Russo,
2369 F., Armando, E., Agliozzo, S., Regge, D., 2011. A CAD system based on multi-
2370 parametric analysis for cancer prostate detection on DCE-MRI , 79633Q–79633Q–
2371 7.
- 2372 McNeal, J.E., 1981. The zonal anatomy of the prostate. *Prostate* 2, 35–49.

- 2373 McNeal, J.E., Redwine, E.A., Freiha, F.S., Stamey, T.A., 1988. Zonal distribution
2374 of prostatic adenocarcinoma. Correlation with histologic pattern and direction of
2375 spread. *Am. J. Surg. Pathol.* 12, 897–906.
- 2376 Middleton, D., Esposito, R., 1968. Simultaneous optimum detection and estimation
2377 of signals in noise. *Information Theory, IEEE Transactions on* 14, 434–444. doi:10.
2378 1109/TIT.1968.1054139.
- 2379 Mitra, J., 2012. Multimodal image registration applied to magnetic resonance and
2380 ultrasound prostatic images. Ph.D. thesis. Universitat de Girona and Université
2381 de Bourgogne.
- 2382 Mitra, J., Kato, Z., Marti, R., Oliver, A., Llado, X., Sidibe, D., Ghose, S., Vilanova,
2383 J.C., Comet, J., Meriaudeau, F., 2012. A spline-based non-linear diffeomorphism
2384 for multimodal prostate registration. *Med Image Anal* 16, 1259–1279.
- 2385 Mitra, J., Marti, R., Oliver, A., Llado, X., Vilanova, J.C., Meriaudeau, F., 2011.
2386 A comparison of thin-plate splines with automatic correspondences and B-splines
2387 with uniform grids for multimodal prostate registration, in: *Society of Photo-
2388 Optical Instrumentation Engineers (SPIE) Conference Series*. doi:10.1117/12.
2389 877956.
- 2390 Mohan, J., Krishnaveni, V., Guo, Y., 2014. A survey on the magnetic res-
2391 onance image denoising methods. *Biomedical Signal Processing and Control*
2392 9, 56 – 69. URL: [http://www.sciencedirect.com/science/article/pii/
2393 S1746809413001407](http://www.sciencedirect.com/science/article/pii/S1746809413001407), doi:<http://dx.doi.org/10.1016/j.bspc.2013.10.007>.
- 2394 Moore, C.M., Ridout, A., Emberton, M., 2013. The role of MRI in active surveillance
2395 of prostate cancer. *Curr Opin Urol* 23, 261–267.
- 2396 Morgan, R., Boxall, A., Bhatt, A., Bailey, M., Hindley, R., Langley, S., Whitaker,
2397 H.C., Neal, D.E., Ismail, M., Whitaker, H., Annels, N., Michael, A., Pandha,
2398 H., 2011. Engrailed-2 (EN2): a tumor specific urinary biomarker for the early
2399 diagnosis of prostate cancer. *Clin. Cancer Res.* 17, 1090–1098.
- 2400 Nelder, J.A., Mead, R., 1965. A simplex method for func-
2401 tion minimization. *The Computer Journal* 7, 308–313.
2402 arXiv:<http://comjnl.oxfordjournals.org/content/7/4/308.full.pdf+html>.
- 2403 Niaf, E., Rouviere, O., Mege-Lechevallier, F., Bratan, F., Lartizien, C., 2012.
2404 Computer-aided diagnosis of prostate cancer in the peripheral zone using mul-
2405 tiparametric MRI. *Phys Med Biol* 57, 3833–3851.

- 2406 Niaf, E., Rouvire, O., Lartizien, C., 2011. Computer-aided diagnosis for prostate
2407 cancer detection in the peripheral zone via multisequence MRIs , 79633P–79633P–
2408 8.
- 2409 van Niekerk, C.G., van der Laak, J.A., Borger, M.E., Huisman, H.J., Witjes, J.A.,
2410 Barentsz, J.O., Hulsbergen-van de Kaa, C.A., 2009. Computerized whole slide
2411 quantification shows increased microvascular density in pT2 prostate cancer as
2412 compared to normal prostate tissue. *Prostate* 69, 62–69.
- 2413 van Niekerk, C.G., Witjes, J.A., Barentsz, J.O., van der Laak, J.A., Hulsbergen-
2414 van de Kaa, C.A., 2013. Microvasculature in transition zone prostate tumors re-
2415 sembles normal prostatic tissue. *Prostate* 73, 467–475.
- 2416 Noguchi, M., Stamey, T.A., McNeal, J.E., Yemoto, C.M., 2001. Relationship between
2417 systematic biopsies and histological features of 222 radical prostatectomy speci-
2418 mens: lack of prediction of tumor significance for men with nonpalpable prostate
2419 cancer. *J. Urol.* 166, 104–109.
- 2420 Nowak, R., 1999. Wavelet-based Rician noise removal for magnetic resonance imag-
2421 ing. *Image Processing, IEEE Transactions on* 8, 1408–1419. doi:10.1109/83.
2422 791966.
- 2423 Nyul, L.G., Udupa, J.K., 1999. On standardizing the MR image intensity scale.
2424 *Magn Reson Med* 42, 1072–1081.
- 2425 Nyul, L.G., Udupa, J.K., Zhang, X., 2000. New variants of a method of MRI scale
2426 standardization. *IEEE Trans Med Imaging* 19, 143–150.
- 2427 Ojala, T., Pietikäinen, M., Harwood, D., 1996. A comparative study of texture
2428 measures with classification based on featured distributions. *Pattern Recognition*
2429 29, 51–59. URL: [http://dx.doi.org/10.1016/0031-3203\(95\)00067-4](http://dx.doi.org/10.1016/0031-3203(95)00067-4), doi:10.
2430 1016/0031-3203(95)00067-4.
- 2431 Osorio-Garcia, M., Croitor Sava, A., Sima, D.M., Nielsen, F., Himmelreich, U.,
2432 Van Huffel, S., 2012. Magnetic Resonance Spectroscopy. InTech. chapter Quan-
2433 tification improvements of 1H MRS Signals. pp. 1–27.
- 2434 Oster, G., Lamerato, L., Glass, A.G., Richert-Boe, K.E., Lopez, A., Chung, K.,
2435 Richhariya, A., Dodge, T., Wolff, G.G., Balakumaran, A., Edelsberg, J., 2013.
2436 Natural history of skeletal-related events in patients with breast, lung, or prostate
2437 cancer and metastases to bone: a 15-year study in two large US health systems.
2438 *Support Care Cancer* .

2439 Ozer, S., Haider, M., Langer, D.L., Van der Kwast, T.H., Evans, A., Wernick, M.,
2440 Trachtenberg, J., Yetik, I., 2009. Prostate cancer localization with multispectral
2441 MRI based on Relevance Vector Machines, in: Biomedical Imaging: From Nano
2442 to Macro, 2009. ISBI '09. IEEE International Symposium on, pp. 73–76.

2443 Ozer, S., Langer, D.L., Liu, X., Haider, M.A., van der Kwast, T.H., Evans, A.J.,
2444 Yang, Y., Wernick, M.N., Yetik, I.S., 2010. Supervised and unsupervised methods
2445 for prostate cancer segmentation with multispectral MRI. *Med Phys* 37, 1873–
2446 1883.

2447 Padhani, A.R., 2002. Dynamic contrast-enhanced MRI in clinical oncology: current
2448 status and future directions. *J Magn Reson Imaging* 16, 407–422.

2449 Padhani, A.R., 2011. Integrating multiparametric prostate MRI into clinical practice.
2450 *Cancer Imaging* 11 Spec No A, 27–37.

2451 Parfait, S., 2010. Classification de spectres et recherche de biomarqueurs en spec-
2452 troscopie par résonance magnétique nucléaire du proton dans les tumeurs prosta-
2453 tiques. Ph.D. thesis. Université de Bourgogne.

2454 Parfait, S., Walker, P., Crhange, G., Tizon, X., Mitran, J., 2012. Classification of
2455 prostate magnetic resonance spectra using Support Vector Machine . *Biomedical*
2456 *Signal Processing and Control* 7, 499 – 508.

2457 Pearson, K., 1901. On lines and planes of closest fit to systems of points in space.
2458 *Philosophical Magazine* 2, 559–572.

2459 Peng, H., Long, F., Ding, C., 2005. Feature selection based on mutual information
2460 criteria of max-dependency, max-relevance, and min-redundancy. *Pattern Analysis*
2461 *and Machine Intelligence, IEEE Transactions on* 27, 1226–1238. doi:10.1109/
2462 TPAMI.2005.159.

2463 Peng, Y., Jiang, Y., Yang, C., Brown, J., Antic, T., Sethi, I., Schmid-Tannwald, C.,
2464 Giger, M., Eggener, S., Oto, A., 2013. Quantitative analysis of multiparametric
2465 prostate MR images: differentiation between prostate cancer and normal tissue
2466 and correlation with Gleason score—a computer-aided diagnosis development study.
2467 *Radiology* 267, 787–796.

2468 Petrick, N., Haider, M., Summers, R.M., Yeshwant, S.C., Brown, L., Iuliano, E.M.,
2469 Louie, A., Choi, J.R., Pickhardt, P.J., 2008. CT colonography with computer-
2470 aided detection as a second reader: observer performance study. *Radiology* 246,
2471 148–156.

- 2472 Pijnappel, W., van den Boogaart, A., de Beer, R., van Ormondt, D., 1992.
2473 SVD-based quantification of magnetic resonance signals . Journal of Mag-
2474 netic Resonance (1969) 97, 122 – 134. URL: [http://www.sciencedirect.com/](http://www.sciencedirect.com/science/article/pii/002223649290241X)
2475 [science/article/pii/002223649290241X](http://www.sciencedirect.com/science/article/pii/002223649290241X), doi:[http://dx.doi.org/10.1016/](http://dx.doi.org/10.1016/0022-2364(92)90241-X)
2476 [0022-2364\(92\)90241-X](http://dx.doi.org/10.1016/0022-2364(92)90241-X).
- 2477 Pizurica, A., 2002. Image denoising using wavelets and spatial context modeling.
2478 Ph.D. thesis. Universiteit Gent.
- 2479 Pizurica, A., Philips, W., Lemahieu, I., Acheroy, M., 2003. A versatile wavelet
2480 domain noise filtration technique for medical imaging. IEEE Trans Med Imaging
2481 22, 323–331.
- 2482 Pluim, J., Maintz, J., Viergever, M., 2003. Mutual-information-based registration of
2483 medical images: a survey. IEEE Transactions on Medical Imaging 22, 986–1004.
- 2484 Prewitt, J., 1970. Picture processing and psychohistories. Academic Press. chapter
2485 Object enhancement and extraction.
- 2486 Provencher, S.W., 1993. Estimation of metabolite concentrations from localized in
2487 vivo proton NMR spectra. Magn Reson Med 30, 672–679.
- 2488 Puech, P., Betrouni, N., Makni, N., Dewalle, A.S., Villers, A., Lemaitre, L., 2009.
2489 Computer-assisted diagnosis of prostate cancer using DCE-MRI data: design, im-
2490 plementation and preliminary results. Int J Comput Assist Radiol Surg 4, 1–10.
- 2491 Quinlan, J., 1986. Induction of decision trees. Machine Learning 1, 81–106. URL:
2492 <http://dx.doi.org/10.1007/BF00116251>, doi:10.1007/BF00116251.
- 2493 Quinlan, J., 1993. C4.5: Programs for machine learning. Morgan Kaufmann Pub-
2494 lishers Inc., San Francisco, CA, USA.
- 2495 Quinonero-Candela, J., Girard, A., Rasmussen, C., 2002. Prediction at an Uncer-
2496 tain Input for Gaussian processes and relevance vector machines application to
2497 Multiple-Step ahead time-series forecasting. Technical Report.
- 2498 Quint, L.E., Van Erp, J.S., Bland, P.H., Mandell, S.H., Del Buono, E.A., Grossman,
2499 H.B., Glazer, G.M., Gikas, P.W., 1991. Carcinoma of the prostate: MR images
2500 obtained with body coils do not accurately reflect tumor volume. AJR Am J
2501 Roentgenol 156, 511–516.

- 2502 Rangayyan, R., Ayres, F., Desautels, J., 2007. A review of computer-aided diagnosis
2503 of breast cancer: Toward the detection of subtle signs. *Journal of the Franklin*
2504 *Institute* 344, 312 – 348.
- 2505 Rasmussen, C., Williams, C., 2005. Gaussian processes for machine learning. The
2506 MIT Press.
- 2507 Ratiney, H., Sdika, M., Coenradie, Y., Cavassila, S., van Ormondt, D., Graveron-
2508 Demilly, D., 2005. Time-domain semi-parametric estimation based on a metabolite
2509 basis set. *NMR Biomed* 18, 1–13.
- 2510 Rish, I., 2001. An empirical study of the naive Bayes classifier, in: *IJCAI 2001*
2511 *workshop on empirical methods in artificial intelligence*, pp. 41–46.
- 2512 Rodriguez, C., Freedland, S.J., Deka, A., Jacobs, E.J., McCullough, M.L., Patel,
2513 A.V., Thun, M.J., Calle, E.E., 2007. Body mass index, weight change, and risk
2514 of prostate cancer in the Cancer Prevention Study II Nutrition Cohort. *Cancer*
2515 *Epidemiol. Biomarkers Prev.* 16, 63–69.
- 2516 Rosenkrantz, A.B., Sabach, A., Babb, J.S., Matza, B.W., Taneja, S.S., Deng, F.M.,
2517 2013. Prostate cancer: comparison of dynamic contrast-enhanced MRI techniques
2518 for localization of peripheral zone tumor. *AJR Am J Roentgenol* 201, W471–478.
- 2519 Roweis, S.T., Saul, L.K., 2000. Nonlinear dimensionality reduction by locally
2520 linear embedding. *Science* 290, 2323–2326. URL: <http://www.sciencemag.org/content/290/5500/2323.abstract>, doi:10.1126/science.290.5500.2323,
2521 arXiv:<http://www.sciencemag.org/content/290/5500/2323.full.pdf>.
2522
- 2523 Rueckert, D., Sonoda, L.I., Hayes, C., Hill, D.L., Leach, M.O., Hawkes, D.J., 1999.
2524 Nonrigid registration using free-form deformations: application to breast MR im-
2525 ages. *IEEE Trans Med Imaging* 18, 712–721.
- 2526 Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1988. *Neurocomputing: founda-*
2527 *tions of research*, MIT Press, Cambridge, MA, USA. chapter Learning Internal
2528 Representations by Error Propagation, pp. 673–695. URL: <http://dl.acm.org/citation.cfm?id=65669.104449>.
2529
- 2530 Saeys, Y., Inza, I., Larranaga, P., 2007. A review of fea-
2531 ture selection techniques in bioinformatics. *Bioinformatics* 23,
2532 2507–2517. URL: <http://bioinformatics.oxfordjournals.org/content/23/19/2507.abstract>,
2533 doi:10.1093/bioinformatics/btm344,
2534 arXiv:<http://bioinformatics.oxfordjournals.org/content/23/19/2507.full.pdf+html>.

- 2535 Scheidler, J., Hricak, H., Vigneron, D.B., Yu, K.K., Sokolov, D.L., Huang, L.R.,
2536 Zaloudek, C.J., Nelson, S.J., Carroll, P.R., Kurhanewicz, J., 1999a. Prostate
2537 cancer: localization with three-dimensional proton MR spectroscopic imaging–
2538 clinicopathologic study. *Radiology* 213, 473–480.
- 2539 Scheidler, J., Petsch, R., Muller-Lisse, U., Heuck, A., Reiser, M., 1999b. Echo-planar
2540 diffusion-weighted MR imaging of the prostate, in: *Proceedings of the 7th Annual*
2541 *Meeting of ISMRM Philadelphia*, p. 1103.
- 2542 Schlemmer, H.P., Merkle, J., Grobholz, R., Jaeger, T., Michel, M.S., Werner, A.,
2543 Rabe, J., van Kaick, G., 2004. Can pre-operative contrast-enhanced dynamic MR
2544 imaging for prostate cancer predict microvessel density in prostatectomy speci-
2545 mens? *Eur Radiol* 14, 309–317.
- 2546 Schroder, F.H., Carter, H.B., Wolters, T., van den Bergh, R.C., Gosselaar, C.,
2547 Bangma, C.H., Roobol, M.J., 2008. Early detection of prostate cancer in 2007.
2548 Part 1: PSA and PSA kinetics. *Eur. Urol.* 53, 468–477.
- 2549 Schröder, F.H., Hugosson, J., Roobol, M.J., Tammela, T.L., Ciatto, S., Nelen, V.,
2550 Kwiatkowski, M., Lujan, M., Lilja, H., Zappa, M., Denis, L.J., Recker, F., Pez,
2551 A., Määttänen, L., Bangma, C.H., Aus, G., Carlsson, S., Villers, A., Rebillard,
2552 X., van der Kwast, T., Kujala, P.M., Blijenberg, B.G., Stenman, U.H., Huber, A.,
2553 Taari, K., Hakama, M., Moss, S.M., de Koning, H.J., Auvinen, A., 2012. Prostate-
2554 cancer mortality at 11 years of follow-up. *New England Journal of Medicine* 366,
2555 981–990.
- 2556 Shapiro, L.G., Stockman, G.C., 2001. *Computer vision*. Prentice Hall, Upper Saddle
2557 River, NJ. URL: <http://opac.inria.fr/record=b1128947>.
- 2558 Shi, J., Malik, J., 2000. Normalized cuts and image segmentation. *Pattern Analysis*
2559 *and Machine Intelligence, IEEE Transactions on* 22, 888–905. doi:10.1109/34.
2560 868688.
- 2561 Shimofusa, R., Fujimoto, H., Akamata, H., Motoori, K., Yamamoto, S., Ueda, T.,
2562 Ito, H., 2005. Diffusion-weighted imaging of prostate cancer. *J Comput Assist*
2563 *Tomogr* 29, 149–153.
- 2564 Siegel, R., Naishadham, D., Jemal, A., 2013. Cancer statistics, 2013. *CA Cancer J*
2565 *Clin* 63, 11–30.

- 2566 Sled, J.G., Zijdenbos, A.P., Evans, A.C., 1998. A nonparametric method for auto-
2567 matic correction of intensity nonuniformity in MRI data. *IEEE Trans Med Imaging*
2568 17, 87–97.
- 2569 Sobel, I., 1970. Camera models and machine perception. Technical Report. DTIC
2570 Document.
- 2571 Somford, D.M., Futterer, J.J., Hambrock, T., Barentsz, J.O., 2008. Diffusion and
2572 perfusion MR imaging of the prostate. *Magn Reson Imaging Clin N Am* 16, 685–
2573 695.
- 2574 Specht, D.F., 1988. Probabilistic neural networks for classification, mapping, or
2575 associative memory, in: *Neural Networks, 1988., IEEE International Conference*
2576 on, pp. 525–532 vol.1. doi:10.1109/ICNN.1988.23887.
- 2577 St Lawrence, K.S., Lee, T.Y., 1998. An adiabatic approximation to the tissue homo-
2578 geneity model for water exchange in the brain: I. Theoretical derivation. *J. Cereb.*
2579 *Blood Flow Metab.* 18, 1365–1377.
- 2580 Stamey, T.A., Donaldson, A.N., Yemoto, C.E., McNeal, J.E., Sozen, S., Gill,
2581 H., 1998. Histological and clinical findings in 896 consecutive prostates treated
2582 only with radical retropubic prostatectomy: epidemiologic significance of annual
2583 changes. *J. Urol.* 160, 2412–2417.
- 2584 Staring, M., van der Heide, U.A., Klein, S., Viergever, M.A., Pluim, J.P., 2009.
2585 Registration of cervical MRI using multifeature mutual information. *IEEE Trans*
2586 *Med Imaging* 28, 1412–1421.
- 2587 Steinberg, G.D., Carter, B.S., Beaty, T.H., Childs, B., Walsh, P.C., 1990. Family
2588 history and the risk of prostate cancer. *Prostate* 17, 337–347.
- 2589 Strum, S., Pogliano, D., 2005. What every doctor who treats male patients should
2590 know. *PCRI Insights* vol. 8, no. 2.
- 2591 Styner, M., Brechbuhler, C., Szckely, G., Gerig, G., 2000. Parametric estimate of
2592 intensity inhomogeneities applied to MRI. *Medical Imaging, IEEE Transactions*
2593 on 19, 153–165. doi:10.1109/42.845174.
- 2594 Styner, M., Gerig, G., 1997. Evaluation of 2D/3D bias correction with 1+1ES-
2595 optimization. Technical Report. ETH Zürich.

2596 Sung, Y.S., Kwon, H.J., Park, B.W., Cho, G., Lee, C.K., Cho, K.S., Kim, J.K.,
2597 2011. Prostate cancer detection on dynamic contrast-enhanced MRI: computer-
2598 aided diagnosis versus single perfusion parameter maps. *AJR Am J Roentgenol*
2599 197, 1122–1129.

2600 Suzuki, K., 2012. A review of computer-aided diagnosis in thoracic and colonic
2601 imaging. *Quant Imaging Med Surg* 2, 163–176.

2602 Swanson, M.G., Vigneron, D.B., Tran, T.K., Sailasuta, N., Hurd, R.E., Kurhanewicz,
2603 J., 2001. Single-voxel oversampled J-resolved spectroscopy of in vivo human
2604 prostate tissue. *Magn Reson Med* 45, 973–980.

2605 Taira, A.V., Merrick, G.S., Galbreath, R.W., Andreini, H., Taubenslag, W., Curtis,
2606 R., Butler, W.M., Adamovich, E., Wallner, K.E., 2010. Performance of transper-
2607 ineal template-guided mapping biopsy in detecting prostate cancer in the initial
2608 and repeat biopsy setting. *Prostate Cancer Prostatic Dis.* 13, 71–77.

2609 Tipping, M., 2001. Sparse Bayesian learning and the relevance vector machine.
2610 *Journal of Machine Learning Research* 1, 211–244.

2611 Tiwari, P., Kurhanewicz, J., Madabhushi, A., 2013. Multi-kernel graph embedding
2612 for detection, Gleason grading of prostate cancer via MRI/MRS. *Med Image Anal*
2613 17, 219–235.

2614 Tiwari, P., Kurhanewicz, J., Rosen, M., Madabhushi, A., 2010. Semi supervised
2615 multi kernel (SeSMiK) graph embedding: identifying aggressive prostate cancer
2616 via magnetic resonance imaging and spectroscopy. *Med Image Comput Comput*
2617 *Assist Interv* 13, 666–673.

2618 Tiwari, P., Madabhushi, A., Rosen, M., 2007. A hierarchical unsupervised spec-
2619 tral clustering scheme for detection of prostate cancer from magnetic resonance
2620 spectroscopy (MRS). *Med Image Comput Comput Assist Interv* 10, 278–286.

2621 Tiwari, P., Rosen, M., Madabhushi, A., 2008. Consensus-locally linear embedding
2622 (C-LLE): application to prostate cancer detection on magnetic resonance spec-
2623 troscopy. *Med Image Comput Comput Assist Interv* 11, 330–338.

2624 Tiwari, P., Rosen, M., Madabhushi, A., 2009a. A hierarchical spectral clustering and
2625 nonlinear dimensionality reduction scheme for detection of prostate cancer from
2626 magnetic resonance spectroscopy (MRS). *Med Phys* 36, 3927–3939.

- 2627 Tiwari, P., Rosen, M., Reed, G., Kurhanewicz, J., Madabhushi, A., 2009b. Spectral
2628 embedding based probabilistic boosting tree (ScEPTre): classifying high dimen-
2629 sional heterogeneous biomedical data. *Med Image Comput Comput Assist Interv*
2630 12, 844–851.
- 2631 Tiwari, P., Viswanath, S., Kurhanewicz, J., Sridhar, A., Madabhushi, A., 2012.
2632 Multimodal wavelet embedding representation for data combination (MaWERiC):
2633 integrating magnetic resonance imaging and spectroscopy for prostate cancer de-
2634 tection. *NMR Biomed* 25, 607–619.
- 2635 Tofts, P., 2010. T1-weighted DCE imaging concepts: modelling, acquisition and
2636 analysis, in: *Magneton Flash*. Siemens.
- 2637 Tofts, P.S., 1997. Modeling tracer kinetics in dynamic Gd-DTPA MR imaging. *J*
2638 *Magn Reson Imaging* 7, 91–101.
- 2639 Toth, R., Chappelow, J., Rosen, M., Pungavkar, S., Kalyanpur, A., Madabhushi, A.,
2640 2008. Multi-attribute non-initializing texture reconstruction based active shape
2641 model (MANTRA). *Med Image Comput Comput Assist Interv* 11, 653–661.
- 2642 Toth, R., Doyle, S., Pungavkar, S., Kalyanpur, A., Madabhushi, A., 2009. A boosted
2643 ensemble scheme for accurate landmark detection for active shape models, in: *SPIE*
2644 *Medical Imaging*, Orlando, FL.
- 2645 Tu, Z., 2005. Probabilistic boosting-tree: learning discriminative models for classifi-
2646 cation, recognition, and clustering, in: *Computer Vision, 2005. ICCV 2005. Tenth*
2647 *IEEE International Conference on*, pp. 1589–1596 Vol. 2. doi:10.1109/ICCV.2005.
2648 194.
- 2649 Turkbey, B., Choyke, P.L., 2012. Multiparametric MRI and prostate cancer diagnosis
2650 and risk stratification. *Curr Opin Urol* 22, 310–315.
- 2651 Vanhamme, L., van den Boogaart, A., Van Huffel, S., 1997. Improved method for
2652 accurate and efficient quantification of MRS data with use of prior knowledge. *J.*
2653 *Magn. Reson.* 129, 35–45.
- 2654 Vapnik, V., Lerner, A., 1963. Pattern Recognition using Generalized Portrait
2655 Method. *Automation and Remote Control* 24.
- 2656 Verma, S., Rajesh, A., Futterer, J.J., Turkbey, B., Scheenen, T.W., Pang, Y.,
2657 Choyke, P.L., Kurhanewicz, J., 2010. Prostate MRI and 3D MR spectroscopy:
2658 how we do it. *AJR Am J Roentgenol* 194, 1414–1426.

2659 Verma, S., Turkbey, B., Muradyan, N., Rajesh, A., Cornud, F., Haider, M.A.,
2660 Choyke, P.L., Harisinghani, M., 2012. Overview of dynamic contrast-enhanced
2661 MRI in prostate cancer diagnosis and management. *AJR Am J Roentgenol* 198,
2662 1277–1288.

2663 Villers, A., Steg, A., Boccon-Gibod, L., 1991. Anatomy of the prostate: review of
2664 the different models. *Eur. Urol.* 20, 261–268.

2665 Viola, P., Wells, III, W.M., 1997. Alignment by maximization of mutual information.
2666 *Int. J. Comput. Vision* 24, 137–154.

2667 Viswanath, S., Bloch, B.N., Chappelow, J., Patel, P., Rofsky, N., Lenkinski, R.,
2668 Genega, E., Madabhushi, A., 2011. Enhanced multi-protocol analysis via intel-
2669 ligent supervised embedding (EMPrAvISE): detecting prostate cancer on multi-
2670 parametric MRI , 79630U–79630U–15URL: +[http://dx.doi.org/10.1117/12.](http://dx.doi.org/10.1117/12.878312)
2671 878312, doi:10.1117/12.878312.

2672 Viswanath, S., Bloch, B.N., Genega, E., Rofsky, N., Lenkinski, R., Chappelow, J.,
2673 Toth, R., Madabhushi, A., 2008a. A comprehensive segmentation, registration,
2674 and cancer detection scheme on 3 Tesla in vivo prostate DCE-MRI. *Med Image*
2675 *Comput Comput Assist Interv* 11, 662–669.

2676 Viswanath, S., Bloch, B.N., Rosen, M., Chappelow, J., Toth, R., Rofsky, N., Lenk-
2677 inski, R., Genega, E., Kalyanpur, A., Madabhushi, A., 2009. Integrating struc-
2678 tural and functional imaging for computer assisted detection of prostate cancer on
2679 multi-protocol in vivo 3 Tesla MRI, in: *Society of Photo-Optical Instrumentation*
2680 *Engineers (SPIE) Conference Series*.

2681 Viswanath, S., Tiwari, P., Rosen, M., Madabhushi, A., 2008b. A meta-classifier for
2682 detecting prostate cancer by quantitative integration of *In Vivo* magnetic reso-
2683 nance spectroscopy and magnetic resonance imaging, in: *Medical Imaging 2008:*
2684 *Computer-Aided Diagnosis, SPIE*.

2685 Viswanath, S.E., Bloch, N.B., Chappelow, J.C., Toth, R., Rofsky, N.M., Genega,
2686 E.M., Lenkinski, R.E., Madabhushi, A., 2012. Central gland and peripheral zone
2687 prostate tumors have significantly different quantitative imaging signatures on 3
2688 Tesla endorectal, in vivo T2-weighted MR imagery. *J Magn Reson Imaging* 36,
2689 213–224.

2690 Vos, P.C., Barentsz, J.O., Karssemeijer, N., Huisman, H.J., 2012. Automatic
2691 computer-aided detection of prostate cancer based on multiparametric magnetic
2692 resonance image analysis. *Phys Med Biol* 57, 1527–1542.

2693 Vos, P.C., Hambrock, T., Barenstz, J.O., Huisman, H.J., 2008a. Combining T2-
2694 weighted with dynamic MR images for computerized classification of prostate le-
2695 sions, in: Medical Imaging 2008: Computer-Aided Diagnosis, SPIE.

2696 Vos, P.C., Hambrock, T., Barenstz, J.O., Huisman, H.J., 2010. Computer-assisted
2697 analysis of peripheral zone prostate lesions using T2-weighted and dynamic con-
2698 trast enhanced T1-weighted MRI. *Phys Med Biol* 55, 1719–1734.

2699 Vos, P.C., Hambrock, T., Hulsbergen-van de Kaa, C.A., Futterer, J.J., Barentsz,
2700 J.O., Huisman, H.J., 2008b. Computerized analysis of prostate lesions in the
2701 peripheral zone using dynamic contrast enhanced MRI. *Med Phys* 35, 888–899.

2702 Vovk, U., Pernus, F., Likar, B., 2007. A review of methods for correction of intensity
2703 inhomogeneity in MRI. *Medical Imaging, IEEE Transactions on* 26, 405–421.
2704 doi:10.1109/TMI.2006.891486.

2705 Walker, P., Crehange, G., Parfait, S., Cochet, A., Maignon, P., Cormier, L.,
2706 Brunotte, F., 2010. Absolute quantification in 1H MRSI of the prostate at 3T, in:
2707 ISMRM Annual Meeting 2010.

2708 Wang, L., Mazaheri, Y., Zhang, J., Ishill, N.M., Kuroiwa, K., Hricak, H., 2008.
2709 Assessment of biologic aggressiveness of prostate cancer: correlation of MR signal
2710 intensity with Gleason grade after radical prostatectomy. *Radiology* 246, 168–176.

2711 Warfield, S.K., Zou, K.H., Wells, W.M., 2004. Simultaneous truth and performance
2712 level estimation (STAPLE): an algorithm for the validation of image segmentation.
2713 *IEEE Trans Med Imaging* 23, 903–921.

2714 Wiart, M., Curiel, L., Gelet, A., Lyonnet, D., Chapelon, J.Y., Rouviere, O., 2007.
2715 Influence of perfusion on high-intensity focused ultrasound prostate ablation: a
2716 first-pass MRI study. *Magn Reson Med* 58, 119–127.

2717 Ye, L., Kynaston, H.G., Jiang, W.G., 2007. Bone metastasis in prostate cancer:
2718 molecular and cellular mechanisms (Review). *Int. J. Mol. Med.* 20, 103–111.

2719 Zelhof, B., Lowry, M., Rodrigues, G., Kraus, S., Turnbull, L., 2009. Description
2720 of magnetic resonance imaging-derived enhancement variables in pathologically
2721 confirmed prostate cancer and normal peripheral zone regions. *BJU Int.* 104, 621–
2722 627.

- 2723 Zhao, G., Ahonen, T., Matas, J., Pietikainen, M., 2012. Rotation-Invariant Image
2724 and Video Description With Local Binary Pattern Features. Image Processing,
2725 IEEE Transactions on 21, 1465–1477. doi:10.1109/TIP.2011.2175739.
- 2726 Zhu, H., Ouwerkerk, R., Barker, P.B., 2010. Dual-band water and lipid suppression
2727 for MR spectroscopic imaging at 3 Tesla. Magn Reson Med 63, 1486–1492.
- 2728 Zitová, B., Flusser, J., 2003. Image registration methods: a survey. Image and
2729 Vision Computing 21, 977 – 1000. URL: [http://www.sciencedirect.com/
2730 science/article/pii/S0262885603001379](http://www.sciencedirect.com/science/article/pii/S0262885603001379), doi:[http://dx.doi.org/10.1016/
2731 S0262-8856\(03\)00137-9](http://dx.doi.org/10.1016/S0262-8856(03)00137-9).