

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

Guillaume Lemaître<sup>a,c,\*</sup>, Robert Martí<sup>c</sup>, Jordi Freixenet<sup>c</sup>, Joan C. Vilanova<sup>d</sup>,  
Paul M. Walker<sup>b</sup>, Fabrice Meriaudeau<sup>a</sup>

<sup>a</sup>LE2I-UMR CNRS 6306, Université de Bourgogne, 12 rue de la Fonderie, 71200 Le Creusot, France

<sup>b</sup>LE2I-UMR CNRS 6306, Université de Bourgogne, Avenue Alain Savary, 21000 Dijon, France

<sup>c</sup>ViCOROB, Universitat de Girona, Campus Montilivi, Edifici P4, 17071 Girona, Spain

<sup>d</sup>Girona 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 ([75, 86]). These errors are driven by two majors causes ([75]).  
7 On the one hand, observer limitations (e.g., constrained human visual perception,  
8 fatigue or distraction) are the principal human issues. On the other hand, the sec-  
9 ond reason is linked to the clinical cases themselves, for instance due to unbalance  
10 data (number of healthy cases more abundant than malignant cases) or overlapping  
11 structures resulting from limitations of imaging techniques.

12 Computer vision has given rise to many promising solutions. Instead of sup-  
13 plying fully automatic computerized solutions, researchers have aimed at providing  
14 computer image analysis techniques to aid radiologists in their clinical decisions  
15 ([75]). In fact, these investigations brought about both concepts of computer-aided  
16 detection (CADE) and computer-aided diagnosis (CADx) grouped under the acronym  
17 CAD. Since those first steps, evidence has shown that CAD systems enhance the  
18 diagnosis performance of radiologists. [36] reported a significant 4 % improvement  
19 in breast cancer detection, in accordance with later studies ([53]). Similar conclu-  
20 sions were drawn in the case of lung nodule detection ([121]), colon cancer ([176])  
21 and prostate cancer (CaP) as well ([86]). [36] also hypothesized that CAD systems  
22 will be even more efficient assisting inexperienced radiologists to senior radiologists.  
23 That hypothesis was tested by [86] and was confirmed in case of CaP detection.  
24 In this particular study, inexperienced radiologists obtained equivalent performance  
25 than senior radiologists, both with help of CAD system whereas the accuracy of their  
26 diagnosis was significantly poorer without this assistance.

27 In the late eighties, the first CAD systems were developed to detect anomalies on  
28 chest radiographies and mammograms ([34, 57, 76]). In the past twenty years, exten-  
29 sive investigations were conducted in the advancement of CAD systems, migrating  
30 from intensive time consuming algorithms performed on reduced number of cases to  
31 “fast” processing on large medical dataset. These works were focused on diverse or-  
32 gan cancer diagnosis making use of numerous imaging modalities: micro-calcification  
33 detection in breast mammography ([62, 188]) and ultrasound (US) imaging ([40]),  
34 lung nodules detection based on computer tomography (CT) ([35, 219]), colon tu-  
35 mours detection ([219]) and melanoma detection using dermoscopy imaging ([112]).  
36 Noting the abundance of diverse CAD systems, these fields achieved a certain ma-  
37 turity which can be explained by the imaging techniques employed. Indeed, x-rays,  
38 US as well as CT are medical imaging techniques developed all before the 1970s and  
39 were subject to intensive works.

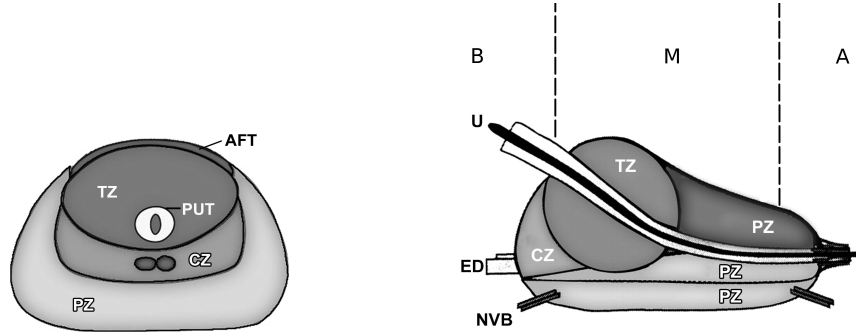
40 In contradiction with the aforementioned statement, CaP detection using CAD  
41 is a young technology due to the fact that magnetic resonance imaging (MRI) is  
42 the keystone medical imaging technique ([89]). Four distinct MRI modalities are

employed in CAD for CaP and were mainly developed after the mid-1990s: (i)  $T_2$  Weighted ( $T_2$ -W) MRI ([95]), (ii) dynamic contrast-enhanced (DCE) MRI ([96]), (iii) magnetic resonance spectroscopy imaging (MRSI) ([113]) and (iv) diffusion weighted (DW) MRI ([199]). 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 ([220]) 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 [37]. 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 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.



(a) Transverse anatomy of the prostate.

(b) Sagittal anatomy of the prostate.

Figure 1: Prostate anatomy with division in different zones. *AFT*: anterior fibromuscular tissue, *CZ*: central zone, *ED*: ejaculatory duct, *NVB*: neurovascular bundle, *PUT*: tissue, *PZ*: peripheral zone, *U*: urethra, *TZ*: transitional zone, *B*: base, *M*: median, *A*: apex (copyright by [41]). [Ask for permission.](#)

## 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 ([119]). The prostate size increases at two distinct stages during physical development: initially at puberty to reach its normal size, then again after sixty years of age leading to benign prostatic hyperplasia (BPH) ([171]).

A zonal classification of the prostate, depicted in Fig. 1, was suggested by McNeal ([146]). Subsequently, this categorization was widely accepted in the literature (cf., [43, 94, 171, 240]) and is used in all medical examinations (e.g., biopsy, MRI screening). The classification is based on dividing the gland into three distinct regions: (i) central zone (CZ) accounting for 20-25% of the whole prostate gland, (ii)

101 transitional zone (TZ) standing for 5% and (iii) peripheral zone (PZ) representing  
 102 the 70%. In MRI images, tissues of CZ and TZ are difficult to distinguish and are  
 103 usually merged into a common region, denominated central gland (CG). As part of  
 104 this classification, the prostate can be divided in three longitudinal portions depicted  
 105 in Fig. 1(b): (i) base, (ii) median gland and (iii) apex.

## 106 2.2. Prostate carcinoma

107 CaP was reported on a worldwide scale to be the second most frequently diag-  
 108 nosed cancer of men accounting for 13.6% ([65]). Statistically, in 2008, the number of  
 109 new diagnosed cases was estimated to be 899,000 with no less than 258,100 deaths  
 110 ([65]). In United States, aside from skin cancer, CaP was declared to be the most  
 111 commonly diagnosed cancer among men, implying that approximately one in six  
 112 men will be diagnosed with CaP during their lifetime and one in thirty-six will die  
 113 from this disease causing CaP to be the second most common cause of cancer death  
 114 among men ([206], [8]).

115 Despite active research to determine the causes of prostate cancer, a fuzzy list of  
 116 risk factors has arisen ([7]). The etiology was linked to the following factors ([7]):  
 117 (i) family history ([77, 214]), (ii) genetic factors ([1, 11, 68]), (iii) race-ethnicity  
 118 ([77, 93]), (iv) diet ([5, 77, 138]), (v) obesity ([77, 192]). This list of risk factors alone  
 119 cannot be used to diagnose CaP and in this way, screening enables early detection  
 120 and treatment.

121 CaP growth is characterized by two main types of evolution ([215]): slow-growing  
 122 tumours, accounting for up to 85 % of all CaPs ([136]), progress slowly and usually  
 123 stay confined to the prostate gland. For such cases, treatment can be substituted  
 124 with active surveillance. In contrast, the second variant of CaPs develops rapidly  
 125 and metastasises from prostate gland to others organs, primarily the bones ([166]).  
 126 Bone metastases, being an incurable disease, significantly affects the morbidity and  
 127 mortality rate ([256]). Hence, the results of the surveillance have to be trustworthy  
 128 in order to distinguish aggressive from slow-growing CaP.

129 CaP is more likely to come into being in specific regions of the prostate. In that  
 130 respect, around 70-80 % of CaPs originate in PZ whereas 10-20 % in TZ ([32, 147,  
 131 212]). Only about 5 % of CaPs occur in CZ ([44, 147]). However, those cancers  
 132 appear to be more aggressive and more likely to invade other organs due to their  
 133 location ([44]).

## 134 2.3. CaP screening and imaging techniques

### 135 2.3.1. Current CaP screening

136 Current CaP screening consists of three different stages. First, prostate-specific  
 137 antigen (PSA) control is performed to distinguish between low and high risk CaP.

138 Then, for confirmation, samples are taken during prostate biopsy and finally analysed  
139 to evaluate the prognosis and the stage of CaP. In this section, we present a detailed  
140 description of the current screening as well as drawbacks.

141 Since its introduction in mid-1980s, PSA is widely used for CaP screening ([64]).  
142 A higher-than-normal level of PSA can indicate an abnormality of the prostate either  
143 as a BPH or a cancer ([92]). However, other factors can lead to an increased PSA  
144 level such as prostate infections, irritations, a recent ejaculation or a recent rectal  
145 examination ([171]). PSA can be found in the bloodstream in two different forms:  
146 free PSA (about 10%), and linked to another protein (about 90%). A level of PSA  
147 higher than  $10 \text{ ng.mL}^{-1}$  is considered to be at risk ([171]). If the PSA level is between  
148  $10 \text{ ng.mL}^{-1}$  and  $4 \text{ ng.mL}^{-1}$ , the patient is considered as suspicious ([17]). In that  
149 case, the ratio of free PSA to total PSA is computed; if the ratio is higher than 15%,  
150 the case is considered as pathological ([171]).

151 A transrectal ultrasound (TRUS) biopsy is carried out for cases which are con-  
152 sidered as pathological. At least six different samples are taken randomly from the  
153 right and left parts of three different zones: apex, median and base. These  
154 samples are further evaluated using the Gleason grading system ([79]). The scoring  
155 scheme to characterize the biopsy sample is composed of five different patterns which  
156 correspond to grades ranging from 1 to 5. Higher grades are associated with poor  
157 prognosis ([63]). Then, in the Gleason system, two scores are assigned correspond-  
158 ing to (i) the grade of the most present tumour pattern, and (ii) the grade of the  
159 second most present tumour pattern ([63]). A higher Gleason score (GS) indicates  
160 a more aggressive tumour ([63]). Also, it should be noted that biopsy is an invasive  
161 procedure which can result in serious infection or urine retention ([42, 87]).

162 Although PSA screening has been shown to improve early detection of CaP ([42]),  
163 its lack of reliability motivates further investigations using MRI-CAD. Two reliable  
164 studies, carried out in the United States ([12]) and in Europe ([97, 202]), have at-  
165 tempted to assess the impact of early detection of CaP, with diverging outcomes  
166 ([42, 90]). The study carried out in Europe<sup>1</sup> concluded that PSA screening reduces  
167 CaP-related mortality by 21-44% ([97, 202]), while the American<sup>2</sup> trial found no  
168 such effect ([12]). However, both studies agree that PSA screening suffers from low  
169 specificity, with an estimated rate of 36 % ([201]). Both studies also agree that over-  
170 treatment is an issue: decision making of regarding treatment is further complicated

---

<sup>1</sup>The European Randomized Study of Screening for Prostate Cancer (ERSSPC) started in the 1990s in order to evaluate the effect of PSA screening on mortality rate.

<sup>2</sup>The Prostate Lung Colorectal and Ovarian (PLCO) cancer screening trial is carried out in the United States and intends to ascertain the effects of screening on mortality rate.

171 by difficulties in evaluating the aggressiveness and progression of CaP ([55]).

172 Hence, new screening methods should be developed with improved specificity of  
173 detection as well as more accurate risk assessment (aggressiveness and progression).  
174 Current research is focused on identifying new biological markers to replace PSA-  
175 based screening ([24, 27, 154]). Until such research comes to fruition, these needs can  
176 be met through active-surveillance strategy using multi-parametric MRI techniques  
177 ([92, 153]). MRI-CAD system, which is an area of active research and forms the  
178 focus of this paper, can be incorporated into this screening strategy allowing a more  
179 systematic and rigorous follow-up.

180 Another weakness of the current screening strategy lies in the fact that TRUS  
181 biopsy does not provide trustworthy results. Due to its “blind” nature, there is a  
182 chance of missing aggressive tumours or detect microfocal “cancers”, which influences  
183 the aggressiveness-assessment procedure ([160]). As a consequence, over-diagnosis is  
184 estimated at up to 30 % ([84]), while missed clinically significant CaP is estimated  
185 at up to 35 % ([221]). In an effort to solve both issues, alternative biopsy approaches  
186 have been explored. MRI/US-guided biopsy has been shown to outperform standard  
187 TRUS biopsy ([54]). There, multimodal MRI images are fused with US images in  
188 order to improve localization and aggressiveness assessment to carry out biopsies.  
189 Human interaction plays a major role in biopsy sampling which can lead to low  
190 repeatability; by reducing potential human errors at this stage, the CAD framework  
191 can be used to improve repeatability of examination.

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

### 195 2.3.2. *Imaging techniques: MRI - MRSI*

196 MRI provides promising imaging techniques to overcome the previous mentioned  
197 drawbacks. Unlike TRUS biopsy, MRI examination is a non-invasive protocol and has  
198 been shown to be the most acute and harmless technique available currently ([235]).  
199 In this section, we review different MRI techniques developed for CaP detection  
200 and diagnosis. Features strengthening each modality later presented, will receive  
201 particular attention together with their drawbacks. Commonly, these features form  
202 the basis for developing analytic tools and automatic algorithms. However, we refer  
203 the reader to Sect. 4.2 for more details on automatic feature detection methods since  
204 they are part and parcel of the CAD framework. Table 1 provides an overview of  
205 the following discussion.

- 206 –  **$T_2$ -W MRI:**  $T_2$ -W MRI was the first MRI-modality used to perform CaP diag-  
207 nosis using MRI ([95]). Nowadays, radiologists make use of it for CaP detection,

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 <sub>2</sub> -W MRI	SI	low-SI	intermediate to high-SI	+
T <sub>2</sub> 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.



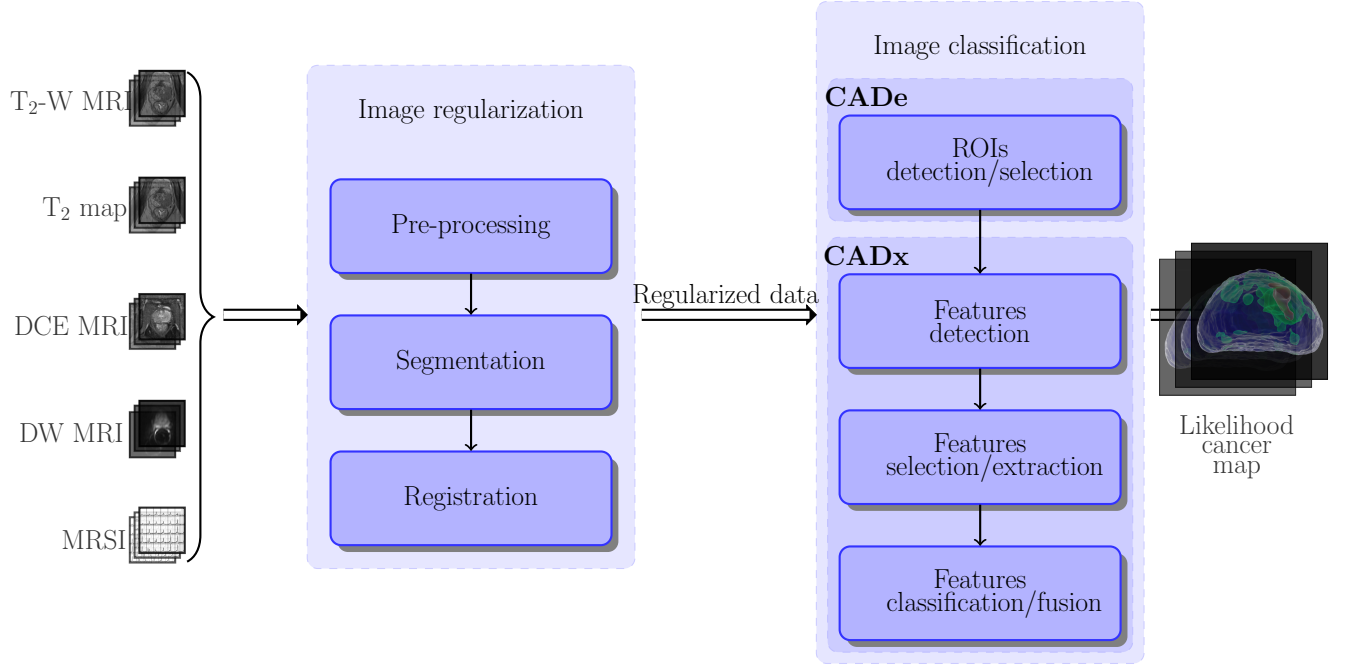
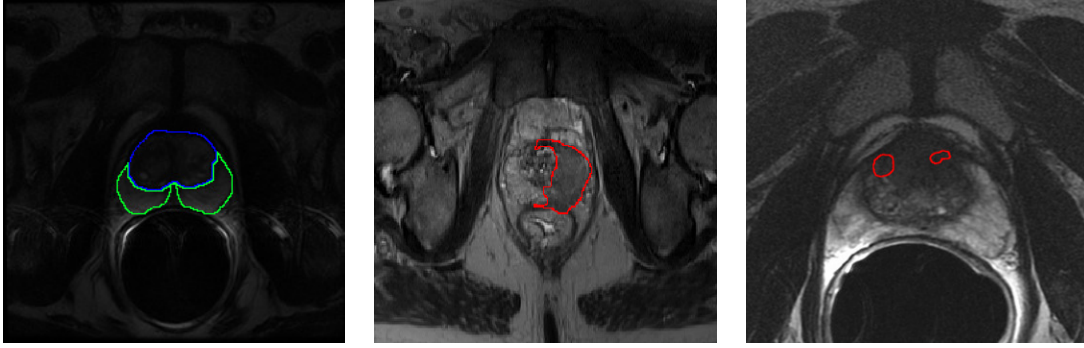


Figure 2: CAD framework using MRI images. Multiparametric MRI images are provided as inputs. These data arise from heterogeneous sources and need to be regularized. Some studies does not considered this stage as mandatory and does not implement or only partly those processes (see Tab. 2). A pre-processing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

(c) T<sub>2</sub>-W-mri slice of a prostate with a CaP highlighted in the CG using a 3.0 Tesla MRI scanner.

Figure 3: Rendering of T<sub>2</sub>-W-MRI prostate image with both 1.5 and 3.0 Tesla MRI scanner.

localization and staging purposes. This imaging technique is well suited to render zonal anatomy of the prostate ([17]).

This modality relies on a sequence based on setting a long repetition time (TR), reducing the T<sub>1</sub> effect in nuclear magnetic resonance (NMR) signal measured, and fixing the echo time (TE) to sufficiently large values in order to enhance the T<sub>2</sub> 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 ([94]). 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) ([95]). Detecting CaP in CG is more challenging. In fact both normal CG tissue and malignant tissue, have a low-SI in T<sub>2</sub>-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) ([4, 17]).

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 ([253]).

In spite of the availability of these useful and encouraging features, the T<sub>2</sub>-W modality lacks reliability ([92, 109]). Sensitivity is affected by the difficulties in detecting cancers in CG ([109]) while specificity rate is highly affected by outliers ([17]). In fact, various conditions emulate patterns of CaP such as BPH, post-

228 biopsy haemorrhage, atrophy, scars and post-treatment ([17, 50, 94, 187, 199]).  
 229 These issues can be partly addressed using more innovative and advanced modal-  
 230 ities.

231 –  **$T_2$  Map:** As previously mentioned,  $T_2$ -W MRI modality shows low sensitivity.  
 232 Moreover,  $T_2$ -W MRI images are a composite of multiple effects ([89]). However,  
 233  $T_2$  values alone have been shown to be more discriminative ([133]) and highly  
 234 correlated with citrate concentration, a biological marker in CaP ([127, 128]).

235  $T_2$  values are computed using the characteristics of transverse relaxation.  
 236 Transverse relaxation is formalized as shown in Eq. 1.

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

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

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

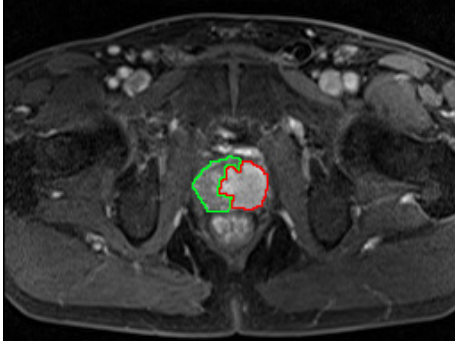
$$\ln \left[ \frac{M_{x,y}(t)}{M_{x,y}(0)} \right] = -\frac{t}{T_2} . \quad (2)$$

240 Fast Spin-Echo (FSE) sequence has been shown to be particularly well suited  
 241 in order to build a  $T_2$  map and obtain accurate  $T_2$  values ([126]).

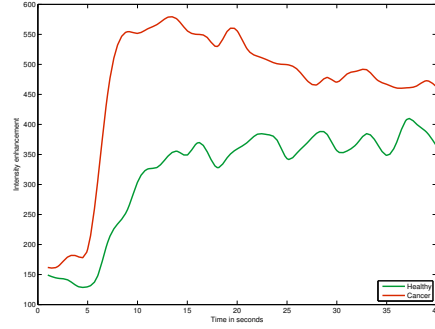
242 Such as  $T_2$ -W MRI,  $T_2$  values associated with CaP are significantly lower than  
 243 those of healthy tissues ([74, 127]).

244 – **DCE MRI:** DCE MRI is an imaging technique which exploits the vascular-  
 245 ity characteristic of tissues. Contrast media, usually gadolinium-based, is in-  
 246 jected intravenously into the patient. The media extravasates from vessels to  
 247 extravascular-extracellular space (EES) and then released back into the vascula-  
 248 ture before being eliminated by the kidneys ([82]). Furthermore, the diffusion  
 249 speed of the contrast agent may vary due to several parameters: (i) the perme-  
 250 ability of the micro-vessels, (ii) their surface area and (iii) the blood flow ([169]).

251 Healthy PZ is mainly made up of glandular tissue, around 70 % ([41]), which  
 252 implies a reduced interstitial space restricting exchanges between vessels and EES  
 253 ([29, 158]). Normal CG has a more disorganised structure, composed of mainly  
 254 fibrous tissue ([41, 92]), which facilitates the arrival of the contrast agent in EES  
 255 ([159]). To understand the difference between contrast media kinetic in malignant  
 256 tumours and the two previous behaviours mentioned, one has to focus on the pro-  
 257 cess known as angiogenesis ([31]). In order to ensure growth, malignant tumours



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



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

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

produce and release angiogenic promoter substances ([31]). These molecules stimulate the creation of new vessels towards the tumour ([31]). However, the new vessel networks in tumours differ from those present in healthy tissue ([82]). They are more porous due to the fact that their capillary walls have a large number of “openings” ([41, 82]). In contrast to healthy cases, this increased vascular permeability results in increased contrast agent exchanges between vessels and EES ([239]).

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 ([239]). 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) ([230]).

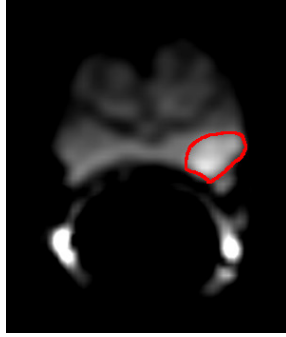
By taking the previous remarks regarding medical aspects and signal theory into account, CaPs are characterized by a signal having an earlier and faster enhancement and an earlier wash-out (cf., the rate of the contrast agent flowing out of the tissue) (see Fig. 4(b)) ([239]). Three different approaches exist to analyse these signals with the aim of tagging them as corresponding to either normal or

279 malignant tissues. Qualitative analysis is based on assessment of the signal shape  
280 ([92]). Quantitative approaches consist in inferring pharmacokinetic parameter  
281 values ([230]). Those parameters are part of mathematical-pharmacokinetic mod-  
282 els which are directly based on physiological exchanges between vessels and EES.  
283 Several pharmacokinetic models were proposed such as the Kety model ([107]),  
284 the Tofts model ([231]) and mixed models ([115, 211]). The last family of meth-  
285 ods mixed both approaches and are grouped together under the heading of semi-  
286 quantitative methods. They rely on shape characterization using mathematical  
287 modelling to extract a set of parameters such as wash-in gradient, wash-out, in-  
288 tegral under the curve, maximum signal intensity, time-to-peak enhancement and  
289 start of enhancement (see Fig. 4) ([92, 239]). It was shown that semi-quantitative  
290 and quantitative methods improve localization of CaP when compared with qual-  
291 itative methods ([193]). Section 4.2.2 provides a full description of quantitative  
292 and semi-quantitative approaches.

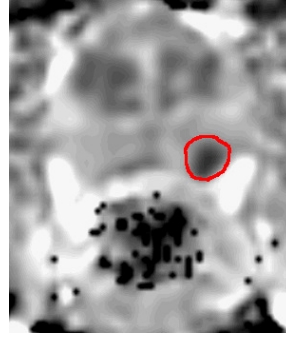
293 DCE MRI combined with T<sub>2</sub>-W MRI has shown to enhance sensitivity com-  
294 pared to T<sub>2</sub>-W MRI alone ([101, 108, 200, 257]). Despite this fact, DCE MRI  
295 possesses some drawbacks. Due to its “dynamic” nature, patient motions dur-  
296 ing the image acquisition lead to spatial misregistration of the image set ([239]).  
297 Furthermore, it has been suggested that malignant tumours are difficult to dis-  
298 tinguish from prostatitis located in PZ and BPH located in CG ([92, 239]). These  
299 pairs of tissues tend to have similar appearances. Later studies have shown that  
300 CaPs in CG do not always manifest in homogeneous fashion. Indeed, tumours in  
301 this zone can present both hypo-vascularization and hyper-vascularization which  
302 illustrates the challenge of CaP detection in CG ([159]).

303 – **DW MRI:** As previously mentioned in the introduction (Sect. 1), DW MRI  
304 is the most recent MRI imaging technique developed aiming at CaP detection  
305 and diagnosis ([199]). This modality exploits the variations in motion of water  
306 molecules in different tissues ([111, 117]).

307 From a physiological point of view, the following facts can be claimed. On  
308 the one hand, PZ, as previously mentioned, is mainly glandular and tubular in  
309 structure allowing water molecules to move freely ([41, 92]). On the other hand,  
310 CG is made up of muscular or fibrous tissue causing the motion of the water  
311 molecules to be more constrained and heterogeneous than in PZ ([92]). Then, CaP  
312 growth leads to the destruction of normal glandular structure and is associated  
313 with an increase in cellular density ([92, 111, 209]). Furthermore, these factors  
314 both have been shown to be inversely correlated with water diffusion ([111, 209]):  
315 higher cellular density implies a restricted water diffusion. Thus, 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.

in CaP will be more restricted than both healthy PZ and CG ([92, 111]).

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 ([99]). 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 ([99]). Under these conditions, the MRI signal measured is

332 formalized as in Eq. 3.

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

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

333 where  $S_{\text{ADC}}$  refers to signal drop due to diffusion effect, ADC is the apparent  
 334 diffusion coefficient and  $b$  is the attenuation coefficient depending only on gradient  
 335 pulses parameters: (i) gradient intensity and (ii) gradient duration ([118]).

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

340 To summarize, in DW MRI images, CaPs are characterized by high-SI com-  
 341 pared to normal tissues in PZ and CG ([17]). However, some tissues in CG can  
 342 look similar to CaP with higher SI ([17]). **Images and reference to theses images**  
 343 **should be included here.**

344 Diagnosis using DW MRI combined with  $T_2$ -W MRI has shown significant  
 345 improvement compared with  $T_2$ -W MRI alone and provides highly contrasted  
 346 images ([41, 170, 205]). As drawbacks, this modality suffers from poor spatial  
 347 resolution and specificity due to false positive detection ([41]).

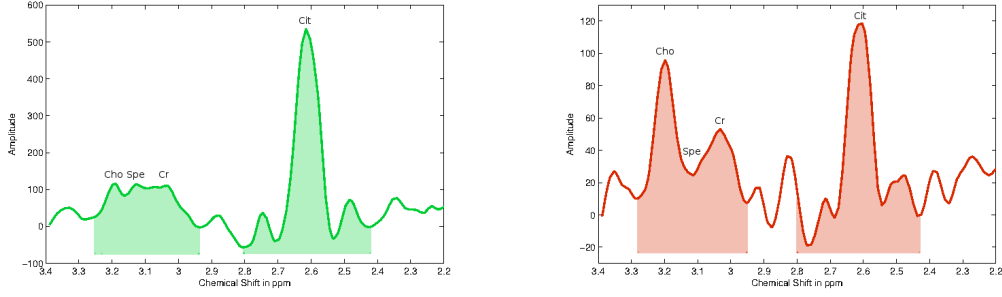
348 With a view to eliminate these drawbacks, radiologists are extracting quanti-  
 349 tative maps from DW MRI. This imaging technique is presented next.

350 – **ADC Map:** The NMR signal measured for DW MRI images is not only affected  
 351 by diffusion as shown in Eq. 3. However, the signal drop (Eq. 4) is formulated  
 352 such that the only variable is the acquisition parameter  $b$  ([118]). The ADC  
 353 is considered as a “pure” diffusion coefficient and can be extracted to build a  
 354 quantitative map.

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

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

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



(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).

360 (typically  $1000 \text{ s.mm}^{-2}$ ). Then, the ADC map can be computed as:

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

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

364 Regarding the appearance of the ADC maps, it was previously stated that by  
 365 increasing the value of  $b$ , the signal of CaP tissue increases significantly. From  
 366 Eq. 6, it can be shown that tissue appearance in the ADC map will be the reverse  
 367 of DW MRI images. Then, CaP tissue is associated with low-SI whereas healthy  
 368 tissue appears brighter ([17]). **Add the some figures and references to the figures.**

369 Similar to the gain achieved by DW MRI, diagnosis using ADC map combined  
 370 with  $T_2$ -W MRI significantly outperforms  $T_2$ -W MRI alone ([41, 59]). Moreover,  
 371 it has been shown that ADC is correlated with GS ([85, 100, 175]).

372 However, some tissues of the CG zone mimic CaP with low-SI ([109]). Image  
 373 distortion can arise due to haemorrhage ([41]). It has also been noted that high  
 374 variation of the ADC occurs between patients making difficult to define a static  
 375 threshold to distinguish CaP from non-malignant tumours ([41]).

376 – **MRSI:** CaP induces metabolic changes in the prostate compared with healthy



tissue. Thus, CaP detection can be carried out by tracking changes of metabolite concentration in prostate tissue. MRSI is an NMR-based technique which generates spectra of relative metabolite concentration in a ROI.

In order to track changes of metabolite concentration, it is important to know which metabolites are associated with CaP. To address this question, clinical studies identified three biological markers: (i) citrate, (ii) choline and (iii) polyamines composed mainly of spermine, and in less abundance of spermidine and putrescine ([16, 49, 78]).

Citrate is involved in the production and secretion of the prostatic fluid, and the glandular prostate cells are associated with a high production of citrate enabled by zinc accumulation by these same cells ([49]). However, the metabolism allowing the accumulation of citrate requires a large amount of energy ([49]). In contrast, malignant cells do not have high zinc levels leading to lower citrate levels due to citrate oxydation ([49]). Furthermore, this change results in a more energy-efficient metabolism enabling malignant cells to grow and spread ([49]).

An increased concentration of choline is related to CaP ([16]). 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 ([16]).

Spermine is also considered as a biological marker in CaP ([78, 81]). In CaP, reduction of the ductal volume due to shifts in polyamine homeostasis might lead to a reduced spermine concentration ([81]).

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 ([83]). 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 ([83]). A phenomenon known as the chemical shift effect (CSE) ([171]). 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 ([83, 171]). 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) ([171]).

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 ([238]). 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 ([238]). 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 ([120]). 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) ([81, 92]). 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 ([120]). 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 ([41]). Furthermore, it has been shown that combining MRSI with MRI improves detection and diagnosis performance ([104, 198]). Citrate and spermine concentration are inversely correlated with the GS allowing us to distinguish low from high grade CaPs ([78]). However, choline concentration does not provide the same properties ([78]).

Unfortunately, MRSI also presents several drawbacks. First, MRSI acquisition is time consuming which makes this modality not normally used in clinical daily practise ([17]). 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 ([78]). Finally, a high variability of the relative concentrations between patients was observed ([41]). The same observation was made depending on the zones studied (cf., PZ, CG, base, mid-gland, apex) ([120, 252]). 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 ([75]). 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 ([75]). 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.

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. 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.

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

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

CADx is composed of the processes allowing to distinguish malignant from non-malignant tumours. We divided CADx into three different stages. First, salient

features are extracted from MRI images to characterize the scene. Of course, more discriminative features will be associated with robust and accurate likelihood cancer map. Most frequently, the number of feature extracted can be large resulting in redundant or not enough discriminative features which will negatively affect the performances of the further classification. Therefore, a step consisting at selecting the best features or/and reducing the number of dimensionality is commonly used. Then, this modified feature vector is finally classified using different pattern recognition approaches.

As pointed out in the introduction (see Sect. 1), performances of CaP detection and diagnosis is affected by observer interpretation and limitations ([75, 86]). CAD offers a possible solution in order to reduce this variability. Lately, effect of CAD on the observer performance has been studied ([86]). Results shown that CADs benefit to less-experienced radiologist to perform similarly as experienced radiologist in their tasks ([86]).

#### 2.4. Literature classification

The review will be organized using the methodology presented in Fig. 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.

- **Noise filtering:** The NMR signal measured and recorded in the k-space during an MRI acquisition is affected by noise. This noise obeys a complex Gaussian

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 st	
			T <sub>2</sub> -W MRI	DCE MRI	DW MRI	MRSI	1.5 T	3.0 T	PZ	CG	Img. Reg.	CAD
[1]	[9]	25	✓	✓	✗	✗	✓	✗	✓	✗	✓!	✗
[2]	[10]	25	✓	✓	✗	✗	✓	✗	✓	✗	✓!	✗
[3]	[13]	53	✓	✗	✓	✗	✓	✗	✓	✓	✗	✗
[4]	[15]	10	✓	✓	✓	✗	✓	✗	✓	✗	✗	✓
[5]	[14]	21	✓	✓	✓	✗	✓	✗	✓	✗	✓!	✓
[6]	[37]	15	✓	✗	✓	✗	✓	✗	✓	✗	✗	✗
[7]	[73]	10	✓	✓	✓	✗	✓	✗	✓	✗	✓	✓
[8]	[106]	24	✗	✗	✗	✓	✓	✗	✓	✓	✓!	✓
[9]	[114]	25	✓	✓	✓	✗	✓	✗	✓	✗	✓!	✗
[10]	[131]	188	✓	✓	✓	✗	✗	✓	✓	✗	✓!	✓
[11]	[130]	288	✓	✓	✓	✗	✗	✓	✓	✓	✓!	✓
[12]	[134]	11	✓	✓	✓	✗	✓	✗	✓	✗	✓!	✓
[13]	[132]	54	✓	✓	✓	✗	✗	✓	✓	✓	✓!	✗
[14]	[135]	27	✓	✗	✗	✗	✓	✗	✓	✗	✓!	✓
[15]	[137]	55	✓	✗	✗	✗	✓	✗	✓	✗	✓!	✗
[16]	[144]	18	✗	✗	✗	✓	✗	✓	✓	✓	✗	✓
[17]	[145]	10	✗	✓	✗	✗	✓	✗	✓	✗	✓!	✓
[18]	[157]	23	✓	✓	✓	✗	✓	✗	✓	✗	✓!	✗
[19]	[156]	30	✓	✓	✓	✗	✓	✗	✓	✗	✓!	✗
[20]	[167]	20	✓	✓	✓	✗	✓	✗	✓	✗	✓!	✓
[21]	[168]	20	✓	✓	✓	✗	✓	✗	✓	✗	✓!	✓
[22]	[172]	22	✗	✗	✗	✓	✗	✓	✓	✓	✓!	✓
[23]	[175]	48	✓	✓	✓	✗	✗	✓	✓	✓	✗	✗
[24]	[183]	100	✗	✓	✗	✗	✓	✗	✓	✓	✗	✗
[25]	[218]	42	✗	✓	✗	✗	✗	✓	✓	✓	✗	✓
[26]	[225]	14	✗	✗	✗	✓	✓	✗	✓	✓	✓!	✓
[27]	[226]	18	✗	✗ <sup>21</sup>	✗	✓	✓	✗	✓	✓	✓!	✓
[28]	[227]	18	✗	✗	✗	✓	✓	✗	✓	✓	✓!	✓
[29]	[228]	15	✓	✗	✗	✓	✓	✗	✓	✓	✓!	✓
[30]	[224]	19	✓	✗	✗	✓	✓	✗	✓	✓	✓!	✓
[31]	[229]	36	✓	✗	✗	✓	✓	✗	✓	✓	✗	✓
[32]	[223]	29	✓	✗	✗	✓	✓	✗	✓	✓	✓!	✓

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]

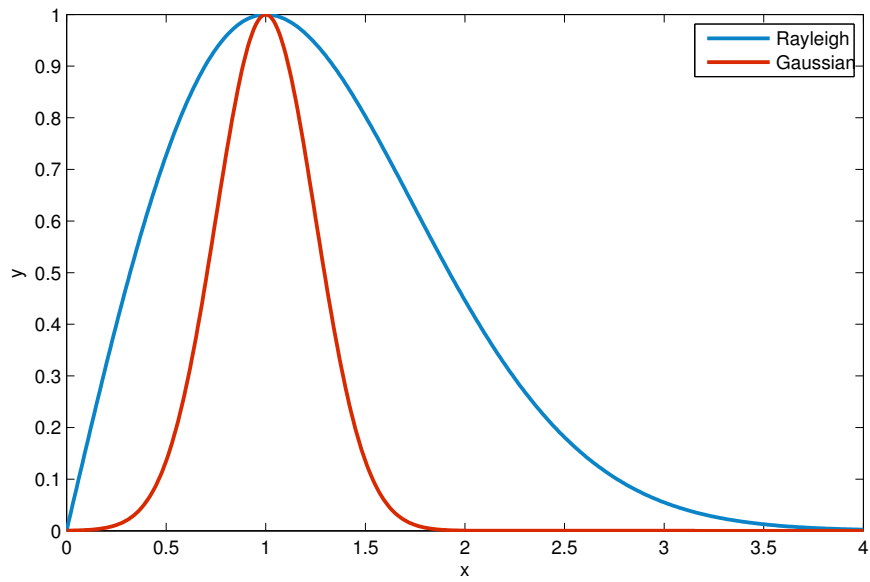


Figure 7: Illustration of a Gaussian distribution ( $\mu = 1, \sigma = 0.25$ ) and a Rayleigh distribution ( $\sigma = 2$ ). It can be seen that the Rayleigh distribution is suffering of a bias term when compared with the Gaussian distribution.

white noise mainly due to thermal noises in the patient ([161]). Furthermore, MRI images visualized by radiologists are in fact the magnitude images resulting from the complex Fourier transform of the k-space data. The complex Fourier transform, being a linear and orthogonal transform, does not affect the Gaussian noise characteristics ([161]). However, the function involved in the magnitude computation is a non-linear transform (e.g., the square root of the sum of squares of real and the imaginary parts), implying that the noise distribution is no longer Gaussian; it indeed follows a Rician distribution making the denoising task harder. Briefly, a Rician distribution can be characterized as follows: in low-SI region (low SNR), it can be approximated with a Rayleigh distribution while in high-SI region (high SNR), it is similar to a Gaussian distribution (see Fig. 7) ([143]). Reviews of all denoising methods can be found in the work of [28] and [152].

*Median filtering* is the simplest approach used to address the denoising issue in MRI images ([167, 168]). In both studies, Ozer et al. used a square kernel of size  $5 \times 5$  pixels with the image resolutions ranging from  $320 \times 256$  (cf., T<sub>2</sub>-W MRI) to  $256 \times 128$  (cf., T<sub>2</sub> map, DCE and DW MRI) and a field of view (FOV) ranging from 14 cm (cf, T<sub>2</sub>-W and DW MRI) to 20 cm (cf, T<sub>2</sub> map and DCE MRI). However, from a theoretical point of view, this simple filtering method is not well formalized to address the issues risen regarding the noise distribution in MRI images.

More rational approaches were proposed to overcome this problem. A common method used to denoise MRI images is based on *wavelet-based filtering*. This filtering exploits the sparsity property of the wavelet decomposition. The projection of a noisy signal from the spatial-domain to the wavelet-domain implies that only few wavelet coefficients contribute to the “signal-free noise” while all wavelet coefficients contribute to the noise ([58]). Therefore, denoising is performed by thresholding/attenuating the insignificant wavelet coefficients to enforce the sparsity in the wavelet-domain. Investigations focus on the strategies to perform the most adequate coefficient shrinkage method (e.g., using thresholding, singularity property or Bayesian framework) ([178]).

[9, 10] performed wavelet shrinkage to denoise magnitude MRI images (cf., T<sub>2</sub>-W-MRI and DCE-MRI) using some thresholding techniques ([142]). However, since that the wavelet transform is an orthogonal transform, the Rician distribution of the noise is preserved in the wavelet-domain. Hence, for low SNR, the wavelet and scaling coefficients still suffer from a bias due to this specific noise distribution ([161]).

[135] used a technique proposed by [179] to denoise T<sub>2</sub>-W-MRI. [179] proposed a filtering technique based on the joint detection and estimation theory ([148]).



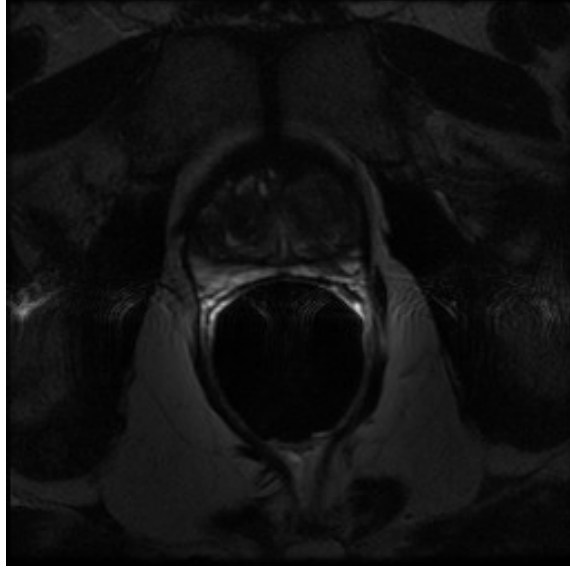


Figure 8: Example of artefacts with high SI due to perturbation from the endorectal coil which create inhomogeneity.

563 The wavelet coefficients “free-to-noise” are estimated from the noisy wavelet co-  
 564 efficients using the maximum a posteriori (MAP) estimate. Furthermore, the  
 565 estimator designed takes spatial context into account by including both local and  
 566 global information in the prior probabilities. The different probabilities needed  
 567 by the MAP are empirically estimated by using mask images representing the  
 568 locations of the significant wavelet coefficients. These mask images are computed  
 569 by thresholding the detail images obtained from the wavelet decomposition. To  
 570 remove the bias from the wavelet and scaling coefficients, the squared magnitude  
 571 MRI image was used instead of the magnitude MRI image as proposed by [161]: it  
 572 involves that the Rician distribution will change to a scaled non-central Chi-square  
 573 distribution. It implies that the wavelet coefficients are unbiased estimators and  
 574 the scaling coefficients are unbiased estimators but up to a constant  $C$  as defined  
 575 in Eq. 7 which need to be subtracted from each scaling coefficient.

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

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

578 — ***Bias correction:*** Besides being corrupted by noise, MRI images are also affected

by the inhomogeneity of the MRI field more commonly named bias field ([216]). 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 ([103, 251]). A review of all bias correction methods can be found in [251].

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”  $o(\mathbf{x})$ .

[244] performed bias correction on T<sub>2</sub>-W-MRI using a *parametric Legendre polynomial* model proposed by [216] and available in the Insight Segmentation and Registration Toolkit (ITK) library<sup>3</sup>. [216] 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  $\mu_k$  being the mean SI and a variance  $\sigma_k^2$  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  $\mu_k$ . Hence, the cost function is defined as:

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

---

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

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

606 where  $\rho_k(\cdot)$  is a M-estimator allowing estimations to be less sensitive to outliers  
 607 than usual square distance ([124]).

608 Finally, estimation of the parameters  $\mathbf{p}$  results in finding the minimum of the  
 609 cost function  $C(\mathbf{p})$ . This optimization was performed using the non-linear (1 + 1)  
 610 Evolution Strategy (ES) optimizer ([217]).

611 In a later publication, [246] make use of the well known *N3 algorithm*<sup>4</sup> to cor-  
 612 rect T<sub>2</sub>-W-MRI developed by [207]. To estimate the bias function, [207] proposed  
 613 to estimate the probability density functions (PDFs) of the signal and bias.

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

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

616 where  $\hat{o}(\mathbf{x})$  is the signal only degraded by noise. [207] shows that Eq. 12 can be  
 617 related to PDFs such that:

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

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

619 Restoring the corrupted signal  $s$  is carried out by finding the multiplicative  
 620 field  $b$  which maximizes the frequency content of the distribution  $O$ . [207] argue  
 621 that a search through all possible fields  $b$  and selection of the one which maximizes  
 622 the high frequency content of  $O$  could be carried out but result in an exhaustive  
 623 search. However, they show that the bias field distribution can be assimilated to a  
 624 near Gaussian distribution. Using this fact as *a priori*, it is then possible to infer  
 625 the distribution  $O$  using Wiener deconvolution given  $B$  and  $S$  and latter estimate  
 626 the corresponding smooth field  $b$ .

627 [137] corrected the inhomogeneity in T<sub>2</sub>-W-MRI images by using the method  
 628 proposed by [139]. In this method, the MRI images are corrected iteratively by  
 629 successively detecting the image foreground via generalized scale ( $g$ -scale) and

---

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

estimating a bias field function based on a second-order polynomial model. First the background of the MRI image is eliminated by thresholding. The threshold value is commonly equal to the mean SI of the considered image. Then, a seeded region growing algorithm is performed considering every thresholded pixel as a potential seed. However, pixels already assigned to a region will not be considered any more as seed. As in seeded region growing algorithm ([203]), two criteria are taken into account to expand the region. First, the region will grow using a connected-neighbourhood, initially defined by the user. Then, the homogeneity of SI is based on a fuzzy membership function taking into account the absolute difference of the SIs of two pixels. Depending on the membership value (cf., a threshold has to be defined), the pixel considered is merged or not to the region. Once this segmentation performed, the largest region  $R$  is used as a mask to select pixels of the original image and the mean SI,  $\mu_R$ , is computed. The background variation  $b(\mathbf{x})$  is estimated such as:

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

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

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

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

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

#### – **SI normalization/standardization:**

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

However, it is a known fact that variability between patients occurred during the MRI examinations even using the same scanner, protocol or sequence parameters ([162]). Hence, the aim of normalization or standardization of the MRI data is to remove the variability between patients and enforce the repeatability of the MRI examinations.

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

[14, 15] as well as [167, 168] standardized T<sub>2</sub>-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,  $\mu_{pz}$  is the mean-SI of the PZ and  $\sigma_{pz}$  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, [132] normalized T<sub>2</sub>-W-MRI by making use of the median and interquartile range for all the pixels.

[137] scaled the SI of T<sub>2</sub>-W-MRI images using the method proposed by [163] 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 25<sup>th</sup> quantile, the median, the 75<sup>th</sup> 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  $n^{\text{th}}$  quantile of the  $i^{\text{th}}$  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

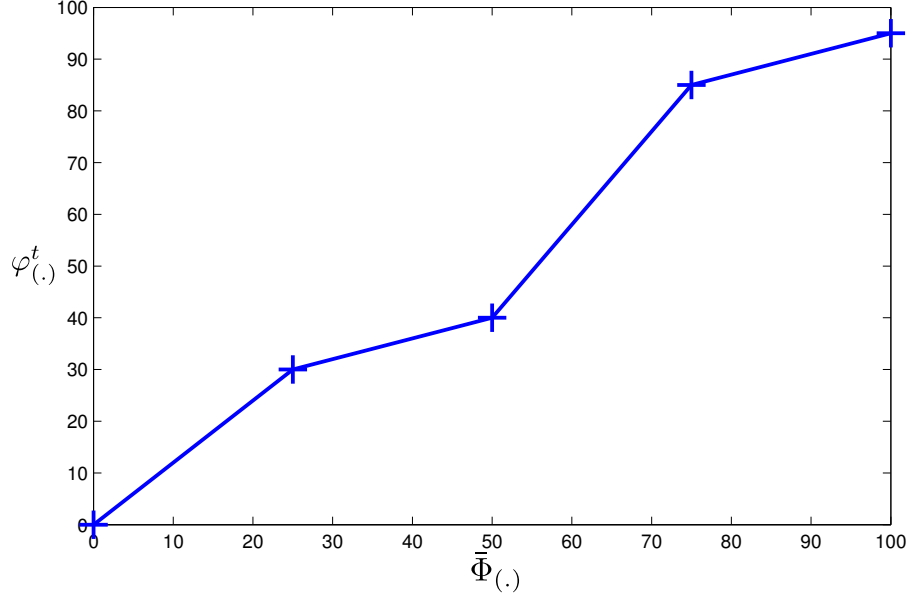


Figure 9: Example of linear mapping by parts as proposed by [163].

mapping is also depicted in Fig. 9.

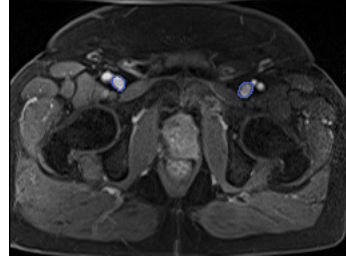
$$\mathcal{T}(s(\mathbf{x})) = \begin{cases} \lceil \bar{\Phi}_0 + (s(\mathbf{x}) - \varphi_0^t) \left( \frac{\bar{\Phi}_{25} - \bar{\Phi}_0}{\varphi_{25}^t - \varphi_0^t} \right) \rceil, & \text{if } \varphi_0^t \leq s(\mathbf{x}) < \varphi_{25}^t, \\ \lceil \bar{\Phi}_{25} + (s(\mathbf{x}) - \varphi_{25}^t)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

[242, 244, 246] are using a variant of this previous approach presented in the work of [140] aiming to standardize the T<sub>2</sub>-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. [156, 157] normalized T<sub>2</sub>-W-MRI images by dividing the original SI of the images by the mean SI of the bladder



(a) Illustration and location of the bladder on a T<sub>2</sub>-W-MRI image acquire with a 3.0 Tesla MRI scanner



(b) Illustration and location of the femoral arteries on a T<sub>1</sub>-W-MRI image acquire with a 3.0 Tesla MRI scanner

Figure 10: Illustration of the two organs used by [156, 157] to normalize T<sub>2</sub>-W and T<sub>1</sub>-W MRI images.

(see Fig. 10(a)). Likewise, [157] standardized the T<sub>1</sub>-W-MRI images using the arterial input function (AIF). They computed the AIF by taking the mean of the SI in the most enhanced part of the common femoral arteries (see Fig. 10(b)) as proposed by [255].

### 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 ([39, 165]).

[172] used a method proposed by [39] 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,  $\phi_0$  and  $\phi_1$  are the zero-order and first-order phase correction terms respectively and  $N$  is the total number of samples of the MRSI signal.

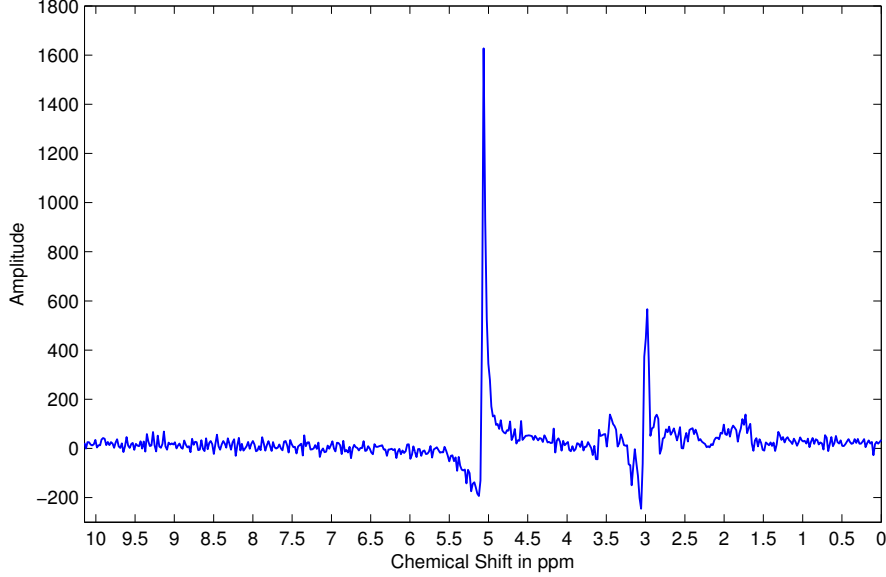


Figure 11: Illustration of phase misalignment in an MRSI spectra acquire with a 3.0 Tesla MRSI scanner.

715 [39] tackled this problem using an optimization framework where  $\phi_0$  and  $\phi_1$  had  
 716 to be inferred. Hence, the simplex Nelder-Mead optimization method was used to  
 717 minimize the following cost function based on the *Shannon entropy* formulation:

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

718 where  $s'(\xi)$  is the first derivative of the corrupted signal  $s(\xi)$  and  $\lambda$  is a regular-  
 719 ization parameter.

720 Once the best parameter  $\Phi$  is obtained, the MRSI signal is corrected using Eq.  
 721 19.

722 — **Water and lipid residuals filtering:** The water and lipid metabolites occur in  
 723 much higher concentrations than the metabolites of interests (cf., choline, creatine  
 724 and citrate) ([165, 259]). Fortunately, specific MRSI sequences were developed in  
 725 order to suppress water and lipid metabolites using pre-saturation techniques  
 726 ([259]). However, these techniques do not perfectly remove water and lipids peaks  
 727 and some residuals are still present in the MRSI spectra as shown in Fig. 12.  
 728 Therefore, different post-processing methods to enhance the quality of the MRSI



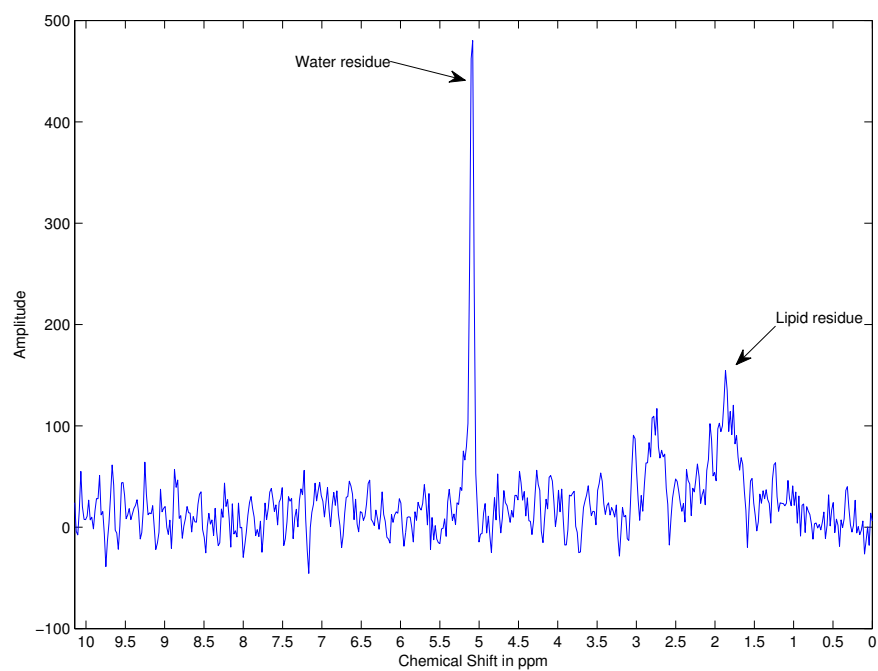


Figure 12: Illustration of the residues of water and fat even after their suppression during the acquisition protocol. The acquisition was carried out with a 3.0 Tesla MRI.

729 spectra by removing these residuals have been proposed.

730 [106] used the well known *HSVD algorithm* proposed by [177]. In the time  
731 domain, a MRSI signal  $s(t)$  is modelled by a sum of  $K$  exponentially damped  
732 sinusoids such that:

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

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

736 [177] showed that the “signal-free noise” can be found using the singular value  
737 decomposition (SVD) decomposition. First the noisy signal is reorganized inside  
738 an Hankel matrix  $H$ . It can be shown that if the signal considered would be a  
739 “signal-free noise”, the rank of  $H$  would be equal to rank  $K$ . However, due to the  
740 presence of noise,  $H$  is in fact a full rank matrix. Thus, to recover the “signal-free  
741 noise”, the rank of  $H$  can be truncated to  $K$  using its SVD decomposition. Hence,  
742 knowing the cut off frequency of the water (cf., 4.7 ppm) and lipid (cf., 2.2 ppm)  
743 metabolites, their corresponding peaks can be reconstructed and subtracted from  
744 the original signal ([116]).

745 – **Baseline correction:** Sometimes, the problem exposed in the above section re-  
746 garding the lipid molecules is not addressed simultaneously than the water resid-  
747 uals suppression. Lipids and macromolecules are known to affect the baseline of  
748 the MRSI spectra. They could cause errors during further fitting processes aiming  
749 to quantify the metabolites, especially regarding the citrate metabolite.

750 [172] made the comparison of two different methods to detect the baseline and  
751 correct the MRSI spectra which are based on the work of [125] and [56] [125]  
752 addressed the problem of baseline detection in the frequency domain by *fitting a*  
753 *polynomial of low degree*  $p(x)$  (e.g., second or third degree) to the MRSI signal  $s(x)$   
754 in a least-squares sense. Then, the values of the fitted polynomial are re-assigned  
755 such as:

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

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

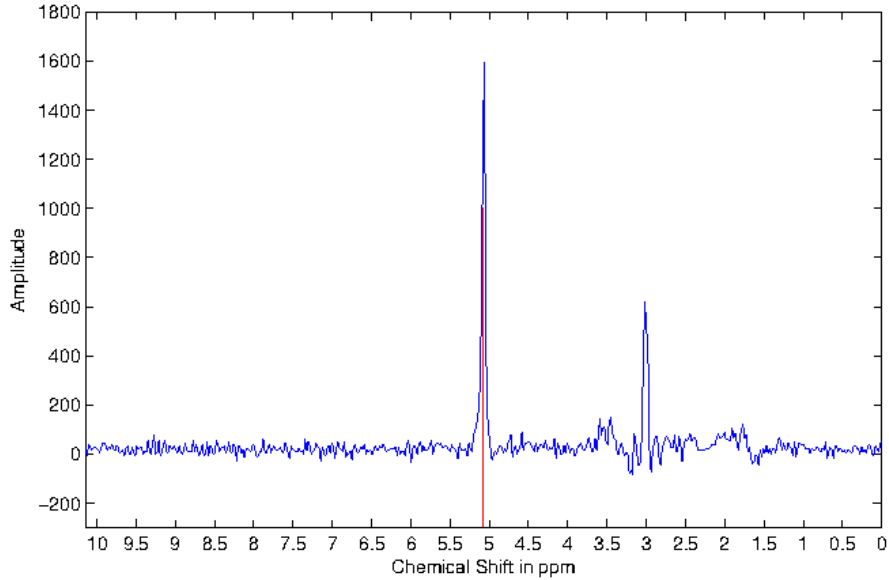


Figure 13: Illustration of frequency misalignment in an MRSI spectra acquire with a 3.0 Tesla MRSI scanner. The water peak is known to be aligned at 4.65 ppm. However, it can be seen that the peak on this spectra is aligned at around 5.1 ppm.

[172] modified this algorithm by convolving a Gaussian kernel to smooth the MRSI signal instead of fitting a polynomial function, keeping the rest of the algorithm identical.

Unlike [125], [56] proposed to correct the baseline in the time domain by multiplying the MRSI signal by a decreasing exponential function as:

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

Having a typical value for  $\beta$  of 0.15.

However, [172] concluded that the method proposed by [125] outperforms the one of [56].

In the contemporary work of [229], authors detected the baseline using a local non-linear fitting method avoiding regions with significant peaks which were detected using a experimentally parametrised signal-to-noise ratio (i.e. a value larger than 5).

- **Frequency alignment:** Due to variations of the experimental conditions, a frequency shift can be observed in the MRSI spectra as shown in Fig. 13 ([39,

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]

165]).  
[229] correct the frequency shift by first detecting known metabolite peaks such as choline, creatine and citrate. The frequency shift is corrected by minimizing the frequency error between the experimental and theoretical values of each of these peaks.

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

[172] as [56] compared two methods to normalize MRSI signal. In each method, a normalization factor is obtained and divided to the original MRSI spectra, similar to the intensity normalization described earlier.

The first approach to obtain the normalization factor is based on an estimation of the water concentration. It is required to have an additional MRSI sequence where the water metabolites are unsuppressed. Using this sequence, an estimation of the water concentration can be performed using the previously reported HSVD algorithm. The second approach to normalization is based on using the  $L_2$  norm  $\|s(\xi)\|_2$  of the MRSI spectra. It should be noted that both [172] and [56] concluded that the  $L_2$  normalization was more efficient in their framework.

### 3.2. Segmentation

The segmentation task consists in delineating the prostate boundaries in the MRI. This procedure is of particular importance for focusing the posterior processing on the organ of interest ([72]). In this section, only the segmentation methods used in CAD systems are presented and summarized in Tab. 4. An exhaustive review of prostate segmentation methods in MRI can be found in [72].

### 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 ([14, 15, 144, 156, 157, 167, 168, 183, 248, 250]).

— **Region-based segmentation:** [130] used a multi-atlas-based segmentation using multi-modal images (e.g., T<sub>2</sub>-W-MRI and ADC map) to segment the prostate with an additional pattern recognition method to differentiate CG and PZ as proposed in [129]. 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 T<sub>2</sub>-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 ([254]).

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<sub>2</sub> map. The texture features were composed of five different features: homogeneity, correlation ([6]), entropy, texture strength ([122]) and local binary pattern (LBP) ([164]). Finally, morphological operations were applied to remove artefact and the contour between the zones were smooth using the thin plate spline (TPS) ([22]).

— **Model-based segmentation:** [243, 244] used the multi-attribute non-initializing texture reconstruction based active shape model (MANTRA) method as proposed by [232]. MANTRA is closely related to the active shape model (ASM) from [47]. 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 ([47]). 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  $k^{\text{th}}$  texture map using the neighbourhood  $\mathcal{N}(l)$ .

By generating multiple landmark in the same fashion than with the ASM, principal components analysis (PCA) ([173]) 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 [47].

Posteriorly, the same authors ([246]) used the weighted ensemble of regional image textures for active shape model segmentation (WERITAS) method also proposed by [233]. 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.

[131] and [247] used an approach proposed by [98] 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. [131] used only ADC

map to encode the appearance whereas [247] 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

[227] 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 [204]. To proceed, a similarity matrix  $W$  is computed using a Gaussian similarity measure from the Euclidean distance ([18]) 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,  $\sigma$  is the standard deviation of the Gaussian similarity measure and  $\epsilon$  is the parameter to defined an  $\epsilon$ -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.

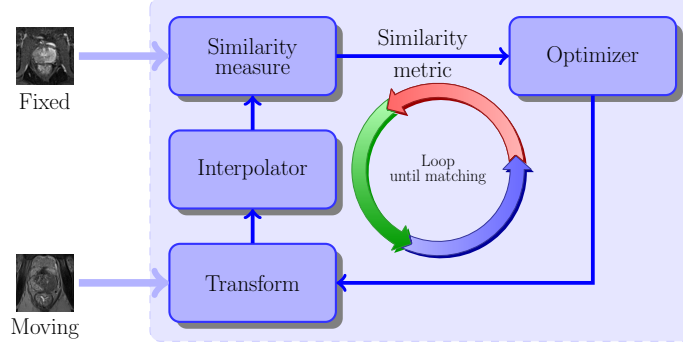


Figure 14: Typical framework involved to solve the registration problem. An iterative procedure takes place to infer the geometric transformation (parametric or non-parametric), via an optimizer, which maximizes the similarity between the two images.

From Sect. 3.3.1 to 3.3.4, we individually review the different components of a typical registration framework (Fig 14). Section 3.3.5 will summarize the combinations of these components especially for the frameworks used in CAD systems. Exhaustive reviews covering all registration methods in computer science and medical fields can be found in [141] and [260].

### 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:

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

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



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			Optimization	
			Rigid	Affine	Elastic	MSE	MI	CMI	GD	L-BFGS-
[1-2]	T <sub>2</sub> -W - DCE	2D	—	✓	—	✓	—	—	—	—
[7]	T <sub>2</sub> -W - DW	2D	—	✓	✓	—	—	—	—	—
[7]	T <sub>2</sub> -W - DCE	2D	—	✓	✓	—	✓	—	✓	—
[34-35]	T <sub>2</sub> -W - DCE	2D	—	✓	—	—	✓	—	—	—
[36]	T <sub>2</sub> -W - DCE - DW	3D	—	✓	—	—	—	✓	✓	✓!
[38]	T <sub>2</sub> -W - DCE	3D	—	✓	—	—	✓	—	—	—
[40]	T <sub>2</sub> -W - DCE	3D	—	✓	✓	—	✓	—	—	✓

Notes:

—: not used or not mentioned.

✓!: suggested.

✓: used or implemented.

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

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

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

925 Hence the four parameters  $\{a_{11}, a_{12}, a_{21}, a_{22}\}$  of the affine matrix and  $\{t_x, t_y\}$  of

926 the translation transform encode an affine transformation.

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

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

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

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

934 Two radial basis functions are used: (i) the TPS and B-splines. Apart from the  
 935 formalism, these two approaches have a main difference. With B-splines, the control  
 936 points are usually uniformly and densely placed on a grid where as with TPS, the  
 937 control points correspond to detected or selected key points. By using TPS, [151]  
 938 obtained more accurate and time efficient results than with the B-splines strategy  
 939 ([150]).

940 It is reasonable to point out that usually only rigid or affine registration are  
 941 used to register multi-parametric images from a same protocol. Elastic registration  
 942 method are more commonly used to register multi-protocol images (e.g., histopathol-  
 943 ogy with MRI images) ([232, 233]).

### 944 3.3.2. Similarity measure

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

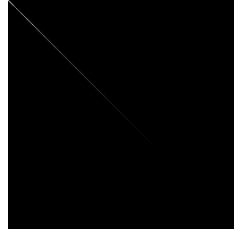
949 The most naive similarity measure is the MSE of the SI of MRI images. For a  
 950 couple of images  $I$  and  $J$ , the MSE is formalized as:

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

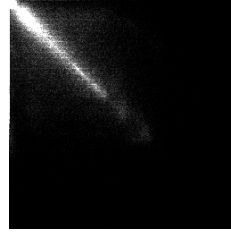
951 where  $N$  is the total number of pixels.

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

954 In that regard, MI was introduced as a registration measure in the late 1990's by  
 955 [180]. MI measure finds its foundation in the assumption that a homogeneous region  
 956 in the first modality image should also appear as a homogeneous region in the second



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



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

Figure 15: Difference observed in joint histogram between aligned and misaligned images. The joint measure will be more concentrated 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.

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 [38]. 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  $\epsilon_n^A$  and  $\epsilon_m^B$  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 [38]. However, other approaches can be used as such the one employed by [213] to

976 compute the  $\alpha$ -MI ([91]) which is based on the construction of entropic graphs using  
 977  $k$ -nearest neighbour ( $k$ -NN) inside the high dimensional feature space lately used  
 978 to estimate the MI.

### 979 3.3.3. Optimization methods

980 Registration is usually regarded as an optimization problem where the parame-  
 981 ters of the geometric transformation model have to be inferred by minimizing the  
 982 similarity measure. Iterative estimation methods are commonly used: Nelder-Mead  
 983 simplex method ([155]), L-BFGS-B quasi-Newton method ([30]) and gradient de-  
 984 scent ([241]). During our review, we noticed that authors do not usually linger over  
 985 optimizer choice.

### 986 3.3.4. Interpolation

987 As shown in Fig. show image without interpolation, the registration procedure  
 988 involves to transform an image and pixels placed on non-integer point have to be  
 989 approximated using interpolation methods. As for the optimization methods, we no-  
 990 tice that little attention has been paid on the choice of those interpolations methods.  
 991 However, commonly used methods are bilinear, nearest-neighbour, bi-cubic, spline  
 992 and inverse-distance weighting method ([149]).

### 993 3.3.5. Review of the methods used in CAD system

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

996 [9, 10] did not use the straightforward approach as presented in Fig. 14 to register  
 997 2D T<sub>2</sub>-W and DCE images. By using image symmetries and the MSE metric, they  
 998 find the parameters of an affine transformation.

999 [73] used also a in-house development registration method to register 2D T<sub>2</sub>-W  
 1000 and DW images. To register both images using an affine transformation model,  
 1001 the bladder is first segmented in both modalities in order to obtain their contours.  
 1002 Then, the images are registered via optimization approach and a deformation field  
 1003 is applied to the images.

1004 [73] and also [249] used the same framework ([195]) which is based on finding an  
 1005 affine transformation to register the T<sub>2</sub>-W and DCE images using the MI. Then, an  
 1006 elastic registration using B-spline takes place using the affine parameters to initialize  
 1007 the geometric model with the same similarity measure. However, the approaches  
 1008 differ regarding the choice of the optimizer since that [73] used a gradient descent and  
 1009 [249] tackle the optimization via a quasi-Newton method. Moreover, [73] performed  
 1010 a 2D registration whereas [249] registered volumes.

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]

[243, 244] as well as [248] performed an affine registration using the MI as similarity measure to correct the misalignment between T<sub>2</sub>-W and DCE images. The choice of the optimizer was not specified. [243, 244] focused on 2D registration while [248] performed 3D registration.

Finally, [242] performed a 3D registration with the three modalities, T<sub>2</sub>-W and DCE and DW MRI, by using an affine transformation model combined with the CMI similarity measure as presented in [38]. Moreover, [38] 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, 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. 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.

[247] 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

1038 Tab. 1), CaP can be interpreted as region of lower SI in ADC map. Hence, blob  
 1039 detector are suitable to highlight these kinds of region. Then, blobs were detected in  
 1040 a multi-resolution scheme by computing the three main eigenvalues  $\{\lambda_{\sigma,1}, \lambda_{\sigma,2}, \lambda_{\sigma,3}\}$   
 1041 of the Hessian matrix for each voxel location of the ADC map at a specific scale  $\sigma$   
 1042 as proposed by [123]. The probability  $p$  of a voxel  $\mathbf{x}$  to be part of a blob at the scale  
 1043  $\sigma$  is given such as:

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

1044 The fusion of the different scales is computed such as:

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

1045 The resulting map  $L(\mathbf{x})$  is then filtered depending on its appearance (cf. maxi-  
 1046 mum of the likelihood of the region, diameter of the lesion) and their SI in ADC and  
 1047 T<sub>2</sub>-W images. The detected regions were then used as inputs for the CADx.

1048 [131] used a pattern recognition approach in order to delineate the ROIs. A  
 1049 blobness map was calculated in the same manner as previously mentioned using the  
 1050 multi-resolution Hessian blob detector on the ADC map, T<sub>2</sub>-W and pharmacokinetic  
 1051 parameters maps (see Sect. 4.2 for details about those parameters). Additionally,  
 1052 the position of the voxel  $\mathbf{x} = \{x, y, z\}$  was used as features as well as the Euclidean  
 1053 distance of the voxel to the prostate center. Hence, the feature vectors was composed  
 1054 of eight features and a support vector machine (SVM) classifier was trained using a  
 1055 radial basis function (RBF) kernel (see Sect. 4.4 for more details).

1056 Posteriorly, [130] modified this approach by including only features related to the  
 1057 blob detection on the different maps as well as the original SIs of the parametric  
 1058 images. Two new maps were introduced based on texture. Instead of a SVM classi-  
 1059 fier, a  $k$ -NN classifier was performed. The candidate regions are then extracted by  
 1060 performing a local maxima detection follow by a post-processing region-growing and  
 1061 morphological operations.

#### 1062 4.2. Feature detection

1063 Discriminative features allowing to recognize CaP from healthy tissue have to be  
 1064 first detected. This section will summarize the different strategies employed for this  
 1065 task. The feature type used is summarized in Tab. 7 while the Tab. 8 summed up  
 1066 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
<b>MRI image:</b>	
<i>Voxel-wise detection</i>	
Intensity-based	A
Edge-based	
Prewitt operator	B <sub>1</sub>
Sobel operator	B <sub>2</sub>
Kirsch operator	B <sub>3</sub>
Gabor filtering	B <sub>4</sub>
Texture-based	
Haralick features	C <sub>1</sub>
Fractal analysis	C <sub>2</sub>
Discrete cosine transform (DCT)	C <sub>3</sub>
Wavelet-based features	C <sub>4</sub>
Position-based	D
<i>Region-wise detection</i>	
Statistical-based	
Percentiles	E <sub>1</sub>
Statistical-moments	E <sub>2</sub>
Histogram-based	
PDF	F <sub>1</sub>
Histogram of oriented gradient (HOG)	F <sub>2</sub>
Shape context	F <sub>3</sub>
LBP	F <sub>4</sub>
Anatomical-based	G
<b>DCE signal:</b>	
Whole spectra approach	H
Semi-quantitative approach	I
Quantitative approach	
Toft model	J <sub>1</sub>
Brix model	J <sub>2</sub>
Weibull function	J <sub>3</sub>
Phenomenological universalities model	J <sub>4</sub>
<b>MRSI signal:</b>	
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 <sub>1</sub>	B <sub>2</sub>	B <sub>3</sub>	B <sub>4</sub>	C <sub>1</sub>	C <sub>2</sub>	C <sub>3</sub>	C <sub>4</sub>	D	E <sub>1</sub>	E <sub>2</sub>	F <sub>1</sub>	F <sub>2</sub>	F <sub>3</sub>	F <sub>4</sub>
[1]	$\triangle--$											$\triangle--$				
[2]	$\triangle--$											$\triangle--$				
[3]						$\triangle--$					$--\diamond$	$--\diamond$				
[4]	$\triangle-\diamond$															
[5]	$\triangle-\diamond$															
[6]	$\triangle-\diamond$							$\triangle-\diamond$		✓						
[7]	$--\diamond$															
[8]																
[9]	$\triangle-\diamond$															
[10]	$\triangle-\diamond$									✓	$\triangle\bullet\Diamond$	$\triangle\bullet\Diamond$				
[11]	$\triangle-\diamond$					$\triangle\bullet\Diamond$				✓	$\triangle\bullet\Diamond$	$\triangle\bullet\Diamond$				
[12]	$\triangle-\diamond$															
[13]													$\triangle\bullet\Diamond$	$\triangle\bullet\Diamond$	$\triangle\bullet\Diamond$	$\triangle\bullet\Diamond$
[14]							$\triangle--$									
[15]							$\triangle--$									
[16]																
[17]																
[18]	$\triangle\bullet\Diamond$		$\triangle\bullet\Diamond$	$\triangle\bullet\Diamond$		$\triangle\bullet\Diamond$					$\triangle\bullet\Diamond$	$\triangle\bullet\Diamond$				
[19]	$\triangle\bullet\Diamond$		$\triangle\bullet\Diamond$	$\triangle\bullet\Diamond$		$\triangle\bullet\Diamond$					$\triangle\bullet\Diamond$	$\triangle\bullet\Diamond$				
[20]	$\triangle-\diamond$															
[21]	$\triangle-\diamond$															
[22]																
[23]											$--\diamond$	$\triangle-\diamond$				
[24]																
[25]																
[26]																
[27]																
[28]																
[29]		$\triangle--$	$\triangle--$	$\triangle--$		$\triangle--$						$\triangle--$				
[30]		$\triangle--$	$\triangle--$	$\triangle--$		$\triangle--$						$\triangle--$				
[31]					$\triangle--$											
[32]		$\triangle--$	$\triangle--$	$\triangle--$		$\triangle--$						$\triangle--$				
[33]		$\triangle--$	$\triangle--$	$\triangle--$	$\triangle--$	$\triangle--$						$\triangle--$				
[34]																
[35]			$\triangle--$	$\triangle--$		$\triangle--$						$\triangle--$				



#### 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. 1) 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 ([9, 10, 14, 15, 37, 114, 134, 156, 157, 242, 243]). 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 ([181]), (ii) Sobel operator ([208]) and (iii) Kirsch operator ([110]). Results obtained with these operators vary, due to their different kernels as shown in Eq. 34-36.

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

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

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

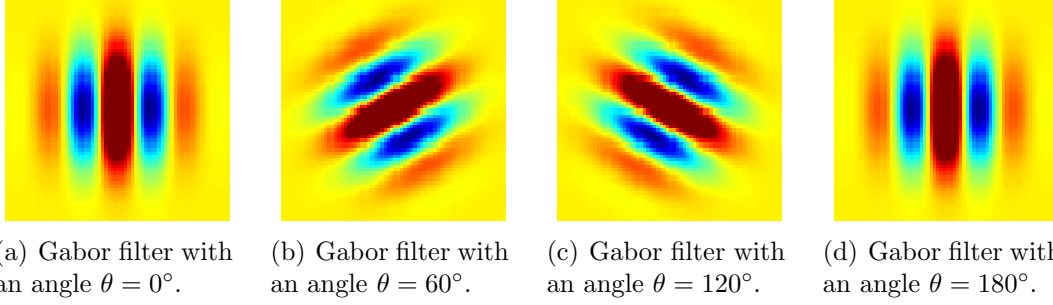


Figure 16: Illustration of four different Gabor filters varying by their orientations.

1084 The magnitude and gradient direction is commonly computed deriving from  
 1085 the edge detection such as:

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

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

1086 where  $K(\cdot)$  is either the Prewitt kernel or Sobel kernel and  $I$  is the input image.

1087 These features are commonly incorporated in the feature vector for further  
 1088 classification in the CAD systems reviewed ([156, 157, 223, 224, 228, 242, 245]).

1089 Gabor filters offer another approach to extract information related to edges  
 1090 and texture and were found to performed similar processing than in the human  
 1091 visual system ([52, 71]). A Gabor filter is defined by the modulation of a Gaussian  
 1092 function with a sinusoid which can be further rotated. Hence, a Gabor filter  $g$   
 1093 can be formalized as:

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

1094 with

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

where  $\lambda$  is the wavelength of the sinusoidal factor,  $\theta$  represents the orientation of the Gabor filter,  $\psi$  is the phase offset,  $\sigma$  is the standard deviation of the Gaussian envelope,  $\gamma$  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. [243, 246] and [229] 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 [88] and is commonly used in CAD system ([13, 156, 157, 223, 224, 228, 242, 243, 245, 246]). 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 ([20]) describing the texture roughness at this specific location was used in [135]. 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.

[37] aimed to describe texture using the frequency signature via the DCT ([2]). 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, [246] projected T<sub>2</sub>-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.

[37] computed these features for a neighbourhood of  $7 \times 7$  pixels for each the modalities that they used.

The *position* of a voxel within the prostate was also considered as possible feature. [131] computed the Euclidean distance from each voxel to the prostate center as well as the individual distance in the three directions  $x$ ,  $y$  and  $z$ . [37]

Table 9: The fourteen statistical features for texture analysis commonly computed from the gray level co-occurrence matrix  $p$  as presented by [88].

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

embedding the same information but this time using cylindrical coordinate  $r$ ,  $\theta$  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 ([13, 130, 131, 175, 223, 224, 228, 242, 243, 245–250]). In fact, once that a ROI is defined, the features corresponding the  $n^{\text{th}}$  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 ([9, 10, 13, 156, 157, 175]).

*Anatomical features* characterizing the region defined were also used as by [130] and [144]. [130] computes the volume, compactness and sphericity related to the region to integrate it in their feature vector to later classify. [144] introduced four features corresponding the percentage present of PZ, CG, periurethral region or outside prostate region for the considered ROI.

[132] 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 ([51]). 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 [132] was shape context ([19]). 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 [258] which is using the Fourier transform of the histogram created via LBP ([164]). 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. [137] 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. [135] proposed a 3D version to estimate the fractal dimension of a volume using wavelet decomposition.

#### 4.2.2. DCE-based features

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

- **Whole-spectra approach:** Some studies are using the whole DCE time series as feature vector such as [9, 10], [229] and [243, 245]. In some cases, the high-dimensional feature space will be reduced using dimension reduction methods as it will be presented in the next section (see Sect. 4.3).
- **Semi-quantitative approach:** Semi-quantitative approaches are based on modelling mathematically the DCE time series. The parameters modelling the signal are commonly used mainly due to the simplicity to compute them. Parameters included in semi-quantitative analysis are summarized in Tab. 10 and also graphically depicted in Fig. 17. A set of time features corresponding to specific amplitude level (start, maximum and end) are extracted. Then, derivative and integral are also considered as discriminative and are commonly computed.
- **Quantitative approach:** As presented in Sect. 2.3.2, quantitative approaches correspond to mathematical-pharmacokinetic models based on physiological exchanges. Four different models have been used in CAD systems.

The most common model reviewed was the *Brix model* ([14, 15, 134, 167, 168, 218]). This model is formalized such as:

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

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 $t_f$

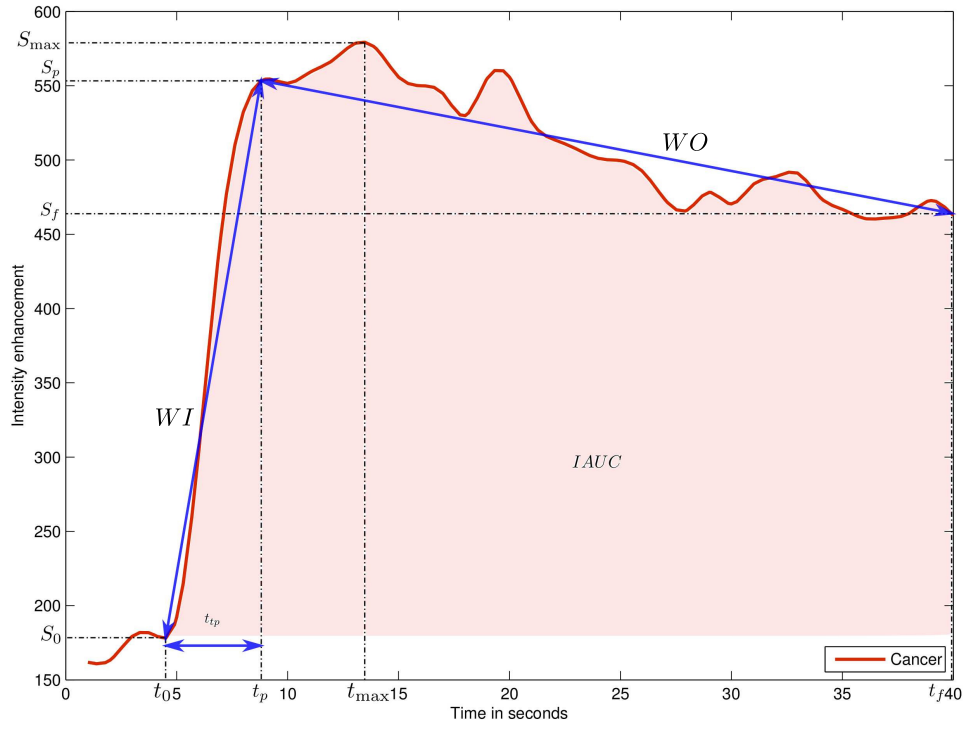


Figure 17: Graphical representation of the different semi-quantitative features used for DCE-MRI analysis.



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

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

*Tofts model* ([231]) was used by [73, 114, 145, 156, 157]. The DCE signal relative to the concentration is then modelled such as:

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

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

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

[145] and [73] employed two others empirical models. A *Weibull function* was used and can be formalized such as:

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

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

The other empirical model is called phenomenological universalities model ([33]) 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  $\beta$ ,  $a_0$  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 ([106, 144, 172, 223–225, 227, 228, 243]) 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 ([190]), AMARES ([236]) and VARPRO ([46]) were used by [106] 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, [172] used the LcModel approach ([182]) which solves the optimization problem in the frequency domain.

Once the different concentrations are computed, [106] calculated the relative concentrations (Eq. 45 and 46) and used them as features. However, [172] 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.

– **Wavelet decomposition approach:** [229] performed a wavelet packet decomposition ([45]) by using the Haar wavelet basis function and use the coefficients of this decomposition as features for further classification.

#### 4.3. Feature selection and feature extraction

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

##### 4.3.1. Feature selection

The feature selection strategy corresponds in selecting the most discriminative feature dimension of the high-dimensional space. Thus, the low-dimensional space is then composed of a subset of the original features detected. In this section, methods employed in the studies reviewed will be briefly presented. More extensive review specific to feature selection can be found in [197].

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]

[156, 157] make use of the  $p$ -value by using the independent two-sample  $t$ -test with equal mean for each feature dimension. In this statistical test, it is 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 [80]. Hence, smaller is the  $p$ -value, more likely is to reject the null hypothesis and more relevant the feature is likely to be.

1269 Finally, the feature can be ranked and the most significant features can be se-  
 1270 lected, by defining the number of features wanted. However, this technique suffers  
 1271 from a main drawback since that it is assumed that each feature are independent  
 1272 which is unlikely to happen and introduce a high redundancy in the feature selected.  
 1273 [247] employed a similar feature ranking approach but make use of the Fisher  
 1274 discriminant ratio to compute the relevance of each feature dimension. Taking the  
 1275 aforementioned formulation, the Fisher discriminant ratio is formalized as the ratio  
 1276 of the variance between classes to the variance within classes such that:

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

1277 Hence, a feature dimension can be seen as more relevant when the variance be-  
 1278 tween classes is maximum and the variance within classes in minimum. Once the  
 1279 features ordered, [247] select the feature dimensions with the larger Fisher discrimi-  
 1280 nant ratio.

1281 MI can also be used to select the subset of feature dimension. Definition of  
 1282 the MI was presented in Sect. 3.3.2 and formalized in 30. The computation of the  
 1283 entropies involves the estimation of some PDFs and the data being usually continuous  
 1284 variables, it is then necessary to estimate the PDFs using method such as Parzen  
 1285 windows.

1286 [174] introduced two main criteria to select the features dimensions and then  
 1287 combined: (i) maximal relevance and (ii) minimum redundancy.

1288 Maximal relevance criterion is based on the paradigm that the classes and the  
 1289 feature dimension which have to be selected have to share a maximal MI and can be  
 1290 formalized such that:

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

1291 where  $\mathbf{x} = \{x_i, i = 1, \dots, d\}$  is a feature vector of  $d$  dimensions,  $c$  is the class  
 1292 considered and  $MI(.)$  is the MI.

1293 As in the previous method, using maximal relevance criterion alone will imply  
 1294 the independence between each feature dimension which is usually not true.

1295 Minimal redundancy criterion will force to select a new feature dimension which  
 1296 shares as less as possible MI with previously selected feature dimension. It can be  
 1297 formalized as:

$$\arg \min Red(\mathbf{x}) ,$$

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

1298 Combination of these two previous criteria is known as maximal relevance min-  
 1299 imal redundancy (mRMR)<sup>5</sup> ([174]) and can be computed such as a difference or a  
 1300 quotient of the Eq. 49 and 50.

1301 [156, 157] make use of maximal relevance criterion alone and also of both mRMR  
 1302 difference and quotient criterion. [246] also reduced their feature vector via mRMR  
 1303 difference and quotient.

#### 1304 4.3.2. Feature extraction

1305 The feature extraction strategy is related to dimension reduction methods and  
 1306 are not selecting discriminative features. Instead, these methods aimed at mapping  
 1307 the data from the high-dimensional space into a low-dimensional space created to  
 1308 maximize the separability between the classes. The mapping can be performed from  
 1309 a linear or a non-linear manner. Only methods employed in CAD system will be  
 1310 reviewed in this section. We guide the reader to the review of [66] for a full review  
 1311 of feature extraction techniques.

1312 Linear mapping method used to reduce the dimensionality in CAD system is  
 1313 usually the PCA.

1314 PCA is a method allowing to find the orthogonal linear transform mapping the  
 1315 original data into a low-dimensional space. The space is defined such that the linear  
 1316 combinations of the original data with the  $k^{th}$  greatest variances will lie on the  $k^{th}$   
 1317 principal components ([102]).

1318 The principal components can then by computing using the eigenvectors-eigenvalues  
 1319 decomposition on the covariance matrix. Let's define the  $\mathbf{x}$  being the data matrix.  
 1320 Then the covariance matrix is defined such as:

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

1321 The eigenvectors-eigenvalues decomposition can be formalized such as:

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

---

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

1322 where  $\mathbf{v}$  is the eigenvectors matrix and  $\Lambda$  is a diagonal matrix containing the eigen-  
1323 values.

1324 It is then possible to find the new low-dimensional space by sorting the eigen-  
1325 vectors using the eigenvalues and finally select the largest eigenvalues. The total  
1326 variation being equal to the sum of the eigenvalues of the covariance matrix ([66]),  
1327 usually the number of principal components correspond to the 95% to 98% of the  
1328 cumulative sum of the eigenvalues. [226, 227, 229] used PCA in order to reduce the  
1329 dimensionality of their feature vector.

1330 Non-linear mapping was also used for dimension reduction and mainly based on  
1331 Laplacian eigenmaps and locally linear embedding (LLE) methods.

1332 Laplacian eigenmaps<sup>6</sup> or also referred as spectral clustering in computer vision  
1333 aimed to find a low-dimensional space in which the proximity of the data should  
1334 be preserved from the high-dimensional space ([18, 204]). Thus, two adjacent data  
1335 points in the high-dimensional space should also be close in the low-dimensional  
1336 space. In like manner, two far away data points in the high-dimensional space  
1337 should be also distant in the low-dimensional space. To compute this projection,  
1338 an adjacency matrix is defined such as:

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

1339 Then, the low-dimensional space will be found by solving the generalized eigenvectors-  
1340 eigenvalues problem such as:

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

1341 where  $D$  is a diagonal matrix such that  $D(i, i) = \sum_j W(j, i)$ .

1342 Finally the low-dimensional space is defined by the  $k$  eigenvectors of the  $k$  smallest  
1343 eigenvalues ([18]). [225, 227, 228, 245] used this spectral clustering to project their  
1344 feature vector into a low-dimensional space. The feature space in these studies is  
1345 usually composed of features extracted from a single or multiple modalities and  
1346 then concatenated before applying the Laplacian eigenmaps dimension reduction  
1347 technique.

1348 [223, 227] used a slightly different approach by combining the Laplacian eigen-  
1349 maps techniques with a prior multi-kernel learning strategy. First, multiple feature  
1350 were extracted for multiple modalities. The features of a single modalities were then  
1351 mapped to an higher dimensional space via the Kernel trick ([3]) and more precisely

---

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

1352 using a Gaussian kernel. Then, each kernel associated with each modality are linearly  
1353 combined to obtain a combined kernel  $K$ . Then, the computation of the adjacency  
1354 matrix  $W$  takes place and the same scheme as in Laplacian eigenmaps is performed.  
1355 However, in order to used the combined kernel, Eq. 54 becomes as shown in Eq. 55  
1356 and can be solved as a generalized eigenvectors-eigenvalues problem as previously.

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

1357 [242] used Laplacian eigenmaps inside a bagging framework in which multiple  
1358 embeddings are generated by successively selecting feature dimensions.

1359 LLE<sup>7</sup> is another non-linear dimension technique broadly known firstly proposed  
1360 by [194]. LLE is based on the fact that a data point in the feature space can be  
1361 characterized by its neighbours. Thus, it was proposed to represent each data point  
1362 in the high-dimensional space as the linear combination of its  $k$ -nearest neighbours.  
1363 This can be expressed such as:

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

1364 where  $\mathbf{x}_i$  and  $\mathbf{x}_j$  are the data point considered and its neighbours data points, re-  
1365 spectively.

1366 Hence, this problem which have to be solved at this stage is to estimate the weight  
1367 matrix  $W$ . This problem can be tackled using a least square optimization scheme by  
1368 optimizing the following objective function:

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

1369 Then, the essence of LLE is to project the data into a low-dimensional keeping  
1370 the data organization. Thus, the projection into the low-dimensional space can be  
1371 seen as an optimization problem such that:

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

---

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

This optimization can be performed as an eigenvectors-eigenvalues problem by finding the  $k^{\text{th}}$  eigenvectors corresponding the  $k^{\text{th}}$  smallest eigenvalues of the sparse matrix  $(I - W)^T(I - W)$ . [226, 227, 243, 245] used LLE as dimension reduction technique to reduce the complexity of their feature vector.

[226] used a modified version of the LLE algorithm in which they applied LLE in a bagging approach with multiple size of neighbourhood. The embedding obtained are then fusion using the maximum likelihood estimation.

#### 4.4. Classification

##### 4.4.1. Classifier

Once that the feature vector was extracted and eventually the complexity reduced, it is possible to take decision and classify this feature vector to belong to CaP or healthy tissue. Classification methods used in CAD system to distinguish this two classes are summarized in Tab. 12. A full review of classification methods used in pattern recognition can be found in [21].

##### — *Simple rules:*

[137] make use of a decision stump classifier to distinguish CaP and healthy classes.

[183] detect CaP by implementing a score medical decision making approach. The feature values are compared with a pre-defined threshold. Then, at each comparison, the final score is incremented or not, depending on the threshold and the final decision is taken depending of the final score.

##### — *Clustering methods:*

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

$k$ -NN was one of the method used by [156, 157] mainly to make a comparison with different machine learning techniques. [130] used this method to roughly detect potential CaP voxel before to perform a region-based classification.

$k$ -means is an unsupervised clustering methods in which the data have to be partitioned into  $k$  clusters. The discovery of the clusters is an iterative procedure. First  $k$  random centroids are defined in the feature space and each data point is assigned to the nearest centroid. Then, the centroid position for each cluster is updated by computing the mean of all the data points belonging to this particular cluster. Both assignment and updating are repeated until the centroids are stable. The number of cluster  $k$  is usually defined as the number of classes. This algorithm



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]
QDA	[37]
Logistic regression	[8-9]
<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]

can also be used incrementally. In case that new data have to be incorporated, the initial centroid positions can corresponds to the results of a previous  $k$ -means follow by the assignment-updating stage.

[225, 227] used  $k$ -means in an iterative procedure. Three clusters were defined corresponding to CaP, healthy and non-prostate, respectively.  $k$ -means was applied iteratively and the larger cluster excluded under the assumption that it is assigned to “non-prostate” cluster. The iteration stop until the number of cluster is smaller than a threshold.

[226, 243, 245] used  $k$ -means in a repetitive manner to be less sensitive to the centroids initialisation. Thus,  $k$  clusters will be generated  $T$  times. The final assignment is performed by majority voting using a co-association matrix as proposed by [67].

#### – **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 ([70]). 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,  $\Sigma$  is the covariance matrix of the training data,  $\mu_k$  is the mean vector of the class  $k$  and  $\pi_k$  is the prior probabilities.

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

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

[13, 37, 156, 157, 247] used LDA to classify their feature vectors defining two classes CaP *versus* healthy.

[246] used the most general quadratic discriminant analysis (QDA) instead of LDA. Unlike in LDA in which one assumes that the class covariance matrices  $\Sigma$  is identical for all the classes, in QDA, it is assumes that a covariance matrix  $\Sigma_k$  specific at each class is computed. Thus, Eq. 59 becomes:

$$\delta_k(\mathbf{x}_i) = \mathbf{x}_i^T \Sigma_k^{-1} \mu_k - \frac{1}{2} \mu_k^T \Sigma_k^{-1} \mu_k + \log(\pi_k) . \quad (61)$$

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

1437 Logistic regression can be used to perform binary classification and can provide  
 1438 the probability of an observation to belong to a class. The posterior probability  
 1439 of one of the class  $c_1$  can be written as:

$$p(c_1|\mathbf{x}_i) = \frac{1}{1 + \exp(-\mathbf{w}^T \mathbf{x}_i)} . \quad (62)$$

1440 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  
 1441 allowing to obtain a linear combination of the input feature vector  $\mathbf{x}_i$ .

1442 Thus, an unlabelled observation  $\mathbf{x}_i$  will be assigned to the class which maxi-  
 1443 mizes the posterior probability:

$$C(\mathbf{x}_i) = \arg \max_k p(C = k|\mathbf{x}_i) . \quad (63)$$

1444 From Eq. 62, one can see that the key of classification using logistic regression  
 1445 model is to infer the set of parameter  $\mathbf{w}$  through a learning stage in the training set.  
 1446 This vector of parameters  $\mathbf{w}$  can be inferred by finding the maximum likelihood  
 1447 estimates. This step can be performed through an optimization scheme, using a  
 1448 quasi-Newton method ([30]), which iteratively seeks for the local minimum in the  
 1449 derivative.

1450 [106, 183] used a logistic regression to create a linear probabilistic model in  
 1451 order to classify their feature vectors.

#### 1452 – **Probabilistic classifier:**

1453 The most commonly used classifier is the naive Bayes classifier which is a  
 1454 probabilistic classifier assuming the independence between each feature dimension  
 1455 ([191]). This classifier is based on the Bayes' theorem:

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

1456 where  $p(C = k, \mathbf{x})$  is the posterior probability,  $p(k)$  is the prior probability,  $p(\mathbf{x}, k)$   
 1457 is the likelihood and  $p(k)$  is the evidence.

1458 However, the evidence term is usually discarded since, it is not class dependent  
 1459 and play the role of a normalization term. Hence, in a classification scheme, an  
 1460 unlabelled observation will be classified to the class which maximized the posterior  
 1461 probability such as:

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

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

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

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

[73, 145, 156, 157] used the naive Bayes classifier to classify their feature vector either malignant or healthy. The model used to characterized the likelihood probability using the Normal distribution.

#### – *Ensemble learner classifiers:*

AdaBoost is an adaptive method based on ensemble learner initially proposed by [69]. AdaBoost linearly combines several weak learners resulting into a final strong classifier. A weak learner is defined as a classification method performing slightly better than random classification. Popular choices regarding the weak learner methods are: decision stump, decision tree learners (cf., iterative dichotomiser 3 (ID3) ([184]), C4.5 ([185]), classification and regression tree (CART) ([26])).

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

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

where  $\epsilon_t$  corresponds to the classification error rate of the weak learner on the distribution of weight  $D_t$ .

Before to perform a new iteration, the distribution of weight  $D_t$  is updated such that the weight associated with the misclassified samples by  $h_t$  will increased

1490 and the weights of well classified samples will decrease as shown in Eq. 68. This  
 1491 procedure allows to select a weak learner at the next iteration  $t + 1$  which will  
 1492 classified in priority previous misclassified samples.

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

1493 where  $\mathbf{x}_i$  is the  $i^{\text{th}}$  sample corresponding to class  $y_i$  and  $Z_t$  is a normalization  
 1494 factor to impose that  $D_{t+1}$  to be a probability distribution.

1495 Thus, after  $T$  iterations, the final strong classifier corresponds to the linear  
 1496 combination of the weak learners and the classification is performed such that:

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

1497 [135] make use of AdaBoost technique to perform their classification.

1498 Random forest<sup>8</sup> is a classification method which is based creating an ensemble  
 1499 of decision tree and was introduced by [25].

1500 In the learning stage, multiple decision tree learners ([26]) will be trained.  
 1501 However, each decision tree will be trained with a different dataset. Each of these  
 1502 dataset corresponds to a bootstrap sample generated by randomly choosing  $n$   
 1503 samples with replacements from the  $N$  available in the original set ([60]). Then,  
 1504 randomization is also part of the decision tree growth. At each node of the  
 1505 decision tree, from the bootstrap sample of  $D$  dimensions, a number of  $d \ll D$   
 1506 dimensions will be randomly selected. Finally, the  $d^{\text{th}}$  dimension splitting the best  
 1507 the data will be used. This best splitting can be evaluated using MI (see Sect.  
 1508 3.3.2). Finally, each tree is grown as much as possible without using any pruning  
 1509 procedure.

1510 In the prediction stage, a sample is introduced in each tree and each of the  
 1511 tree will assign a class to this sample. Finally, it is common to use a a majority  
 1512 voting approach to choose the final class labelling.

1513 [106, 223, 229, 244] make use of the random forest classifier to classify their  
 1514 feature vector.

1515 Probabilistic boosting-tree are an ensemble classifier method which is sharing  
 1516 principles from AdaBoost and using them inside a decision tree ([234]). In the  
 1517 training stage, the probabilistic boosting-tree method is growing a decision tree

---

<sup>8</sup>Random forest implementation can be found at: <http://www.stat.berkeley.edu/~breiman/RandomForests/cc-software.htm>

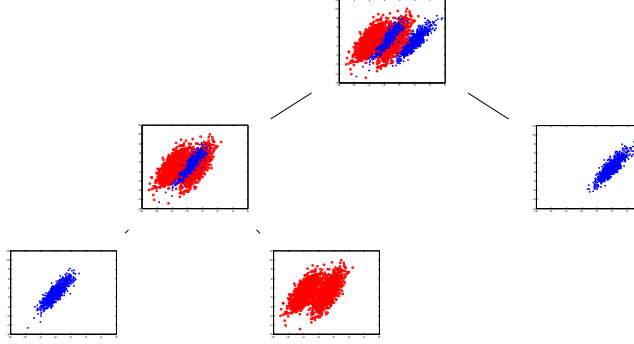


Figure 18: Representation of the capabilities of the probabilistic boosting tree algorithm to split at each node of the tree the positive and negative samples.

1518 and at each node a strong classifier is learnt in an almost comparable scheme  
 1519 than AdaBoost. Once the strong learner is trained, the training set will be split  
 1520 into two subsets which will be used to train the next strong classifiers in the next  
 1521 descending nodes. Thus, three cases are conceivable to decide which branch to  
 1522 propagate each sample  $\mathbf{x}_i$ :

- 1523 • if  $q(+1, \mathbf{x}_i) - \frac{1}{2} > \epsilon$  then  $\mathbf{x}_i$  is propagated to the right branch set and a weight  
 1524  $w_i = 1$  is assigned.
- 1525 • if  $q(-1, \mathbf{x}_i) - \frac{1}{2} > \epsilon$  then  $\mathbf{x}_i$  is propagated to the left branch set and a weight  
 1526  $w_i = 1$  is assigned.
- 1527 • else  $\mathbf{x}_i$  will be propagated in both branches with  $w_i = q(+1, \mathbf{x}_i)$  in the right  
 1528 branch and  $w_i = q(-1, \mathbf{x}_i)$  in the left branch.

1529 with  $\mathbf{w} = w_i, i = \{1, \dots, N\}$  corresponding to distribution of weight,  $N$  the  
 1530 number of samples as in AdaBoost and  $q(\cdot)$  is defined such as:

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

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

1531 Employing such a scheme tend to divide the data in such a way that positive  
 1532 and negative samples are naturally split as shown in Fig. 18.

1533 In the classification stage, the sample  $\mathbf{x}$  is propagated through the trees, where  
 1534 at each node, it will be classified by the strong classifier previously learned and  
 1535 where an estimation of the posterior distribution will be computed. The posterior

1536 distribution will correspond to the sum of the posterior distribution at each node  
 1537 of the tree.

1538 [224, 228, 229, 242] make use of the probabilistic boosting-tree classifier to  
 1539 determine which of the class to assign to their feature vectors.

1540 – **Kernel method:**

1541 Gaussian process<sup>9</sup> for classification is a kernel method in which it is assumed  
 1542 that the data can be represented as a single sample from a multivariate Gaussian  
 1543 distribution ([189]). In the case of linear logistic regression for classification, the  
 1544 posterior probability can be expressed as:

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

1545 where  $\sigma(\cdot)$  is the logistic function.

1546 Thus, the essence of Gaussian process for the classification is based on affecting  
 1547 a Gaussian process prior over the function  $f(\mathbf{x})$  which will be characterized by a  
 1548 mean  $\bar{f}$  and covariance functions  $K$ . Thus, in the training stage, the best mean  
 1549 and covariance function have to be inferred in regard to our training data using  
 1550 a Newton optimization and a Laplacian approximation.

1551 The prediction stage can be performed in two stages. First, for a new ob-  
 1552 servation  $\mathbf{x}_*$ , the corresponding probability  $p(f(\mathbf{x}_*)|f(\mathbf{x}))$  can be computed such  
 1553 that:

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

1554 where  $K_{**}$  is the covariance on testing,  $K_*$  is the covariance on training-testing,  
 1555  $\mathbf{x}$  is the training data.

1556 Then, the function  $f(\mathbf{x}_*)$  is squashed using the sigmoid function and the prob-  
 1557 ability of the class membership can be defined such that:

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

---

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

1558 [106] is using Gaussian process for classification in order to distinguish CaP in  
1559 MRSI data.

1560 – ***Sparse kernel methods:***

1561 In Gaussian process, when a prediction has to be performed, the whole train-  
1562 ing data will be used to assign a label to the new observations. That is why,  
1563 this method is also called kernel method. Sparse kernel category is composed of  
1564 methods which rely only on few example in order to label new observations ([21]).

1565 Support vector machine (SVM)<sup>10</sup> is a sparse kernel method which aimed at  
1566 finding the best linear hyperplane which separates two classes such as the margin  
1567 between the two classes is maximized ([237]). The margin is in fact the region  
1568 defined by two hyperplanes splitting the two classes, such that no points in be-  
1569 tween. The distance between these two hyperplanes is equal to  $\frac{2}{\|\mathbf{w}\|}$  where  $\mathbf{w}$  is  
1570 the normal vector of the hyperplane splitting the classes. Thus, maximizing the  
1571 margin is equivalent to minimized  $\|\mathbf{w}\|$ . This problem solved by an optimization  
1572 approach and formalized such that:

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

1573 From Eq. 75, it is important to notice that only few points from the set of  
1574  $n$  points have to be selected which will later define the hyperplane. This can be  
1575 introduce in the optimization problems using Lagrange multipliers  $\alpha$ . All points  
1576 which are not lying on the margin will be assigned a corresponding  $\alpha_i = 0$ . This  
1577 can be formalized such as:

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

1578 The different parameters can be inferred using quadratic programming. This  
1579 version of SVM is known as hard-margin since that no point can lie in the margin  
1580 area. However, this is highly probable to not find any hyperplane splitting the  
1581 classes such as specified previously. Thus, a soft-margin optimization approach  
1582 was proposed ([48]), where points can lie into the margin but at the cost of penalty

---

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



1583

which will be minimized in the optimization process such that:

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

1584

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

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

1585

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

1586

SVM can also be used as a non-linear classifier by performing a kernel trick [23]. 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.

1588

1589

1590

In CAD system, SVM is the most popular classification method and was used in a multitude of research: [14, 15, 37, 106, 130–132, 135, 156, 157, 167, 168, 172, 175, 218, 229, 247–250].

1592

1593

Relevant vector machine (RVM) is a sparse version of Gaussian process previously presented and was proposed by [222]. RVM is identical to a Gaussian process with following covariance function ([186]):

1594

1595

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

1596

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

1597

1598

As mentioned in [186], the sparsity regarding the relevance vector arises if  $j\alpha_j^{-1} = 0$ . The set of parameters  $\alpha$  is inferred using the expectation maximization algorithm.

1599

1600

1601

[167, 168] make use of RVM and make a comparison with SVM for the task of CaP detection.

1602

1603

#### – **Neural network:**

1604

Multilayer perceptron is a feed-forward neural network consider as well as the most successful model of this kind in pattern recognition ([21]). The most

1605

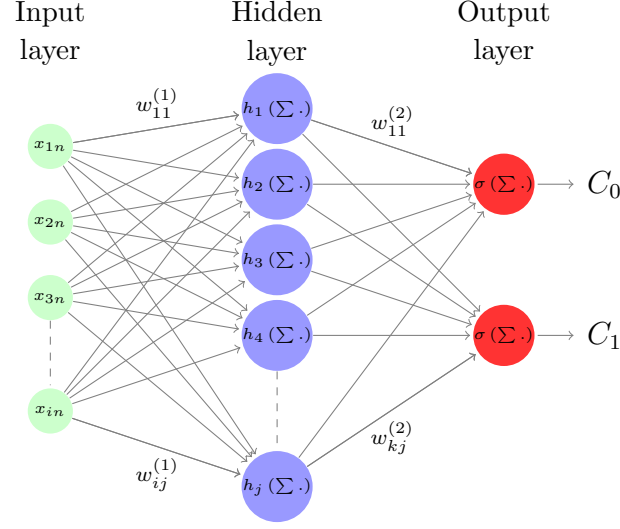


Figure 19: Representation of a neural network of the multilayer perceptron family.

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  $h(\cdot)$  and  $\sigma(\cdot)$  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 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

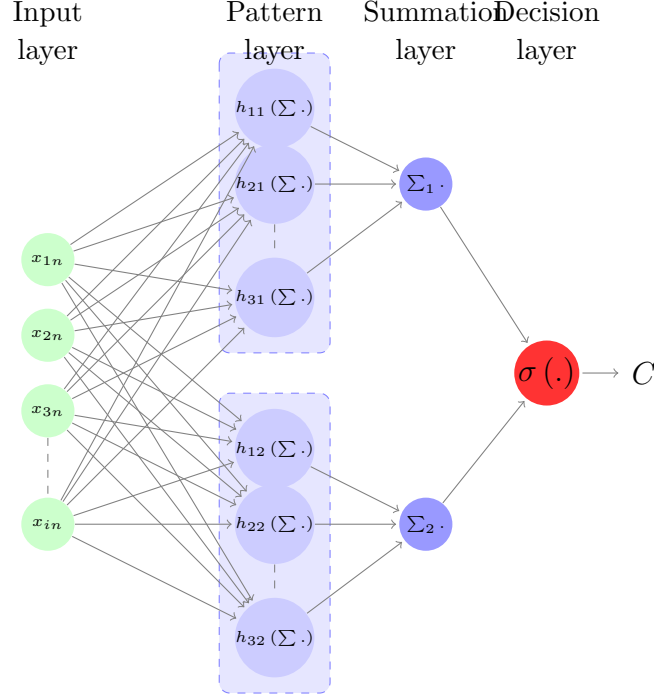


Figure 20: Representation of a neural network of the probabilistic neural network family.

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 the derivative of Eq. 81 can be computed using the backpropagation algorithm proposed by [196].

[144, 172] 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 [210]. This classifier can be modelled by affecting the activation function  $h(\cdot)$  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  $\sigma$  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 [9, 10, 242] 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  $\omega_s$ . 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  $\omega$  ([105]). Hence, lesion segmentation is equivalent to estimate the best  $\hat{\omega}$  which maximize 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  $\mu_\lambda$  and standard deviation  $\sigma_\lambda$ . Then,  $\omega$  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,  $U(\cdot)$  is the energy function.

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]

Thus the segmentation problem can be solved as an optimization problem where the energy function  $U(\cdot)$  has to be minimized. There is different possibilities to define the energy function  $U(\cdot)$ . 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 embed neighbourhood information which will more or less penalize the energy function to obtain more or less homogeneous region. Th optimization of such function can be performed algorithm such as iterated conditional modes ([105]).

[134, 168] used Markov random field as an unsupervised method to segment lesions in multi-parametric MRI images.

[14, 15] 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 case 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 ([61]).

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]

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.

#### 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 compute 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

1709 samples over the true positive added with the false negative samples. Specificity  
1710 is also named the true negative rate and is equal to the ratio of the true negative  
1711 samples over the true negative added with the false positive samples.

1712 These both statistics gave rise to the receiver operating characteristic (ROC)  
1713 analysis. This analysis represents graphically the sensitivity as a function of (1 -  
1714 specificity), which is in fact the false positive rate, by varying the discriminative  
1715 threshold of the classifier. By varying this threshold, more true negative samples  
1716 will be found but at the cost to detect also more false negative. However, this fact  
1717 is interesting in CAD since that it could be possible to obtain a high sensitivity  
1718 and to ensure that no cancers were missed even if more false alarm will have to  
1719 be investigated. A statistic derive from ROC analysis is the area under the curve  
1720 (AUC) which correspond to the area under the ROC and is a measure used to make  
1721 comparison between model.

1722 The previous method could have been classified on pixel-based evaluation method.  
1723 However, a cancer can be also considered as a region. free-response receiver operating  
1724 characteristic (FROC) extend the ROC analysis but a region-based level. The same  
1725 confusion matrix can be computed were the sample are not a pixel any more but a  
1726 lesion. However, this is important to define what is a true positive sample in that  
1727 case. Usually, a lesion is considered as a true positive sample if the region detected  
1728 by the classifier and the one inside the ground are overlapping “enough”. “Enough”  
1729 is defined by each researcher and can correspond to one pixel only.

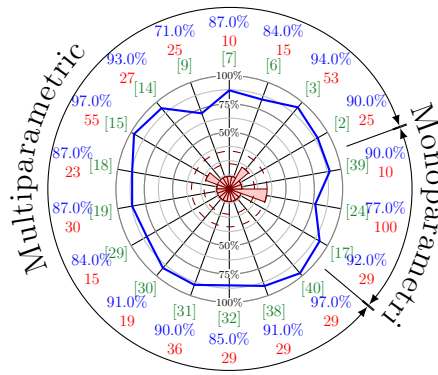
1730 Finally, Dice’s coefficient is sometimes computed which is corresponding to the  
1731 similarity of a lesion between the ground-truth and the output of the classifier. This  
1732 coefficient correspond to the ratio of the twice the number of pixels in common  
1733 over the sum of the pixels of the lesions in the ground-truth and the output of the  
1734 classifier.

## 1735 5. Discussion

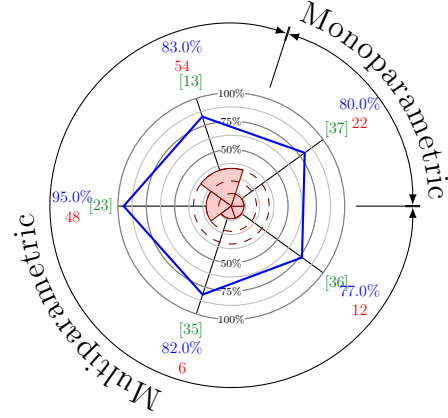
### 1736 5.1. Brief comparison

1737 As discussed previously in Sect. 4.4.3, different metrics have been used to report  
1738 results. A comparison of the different methods reviewed is given depending on the  
1739 metric used in each research and also the type of MRI scanner used (cf., 1.5 *versus*  
1740 3.0 Tesla). For each research, the experiment obtaining the best result was reported  
1741 into these figures.

1742 The results given in term of AUC-ROC are depicted in Fig. 21. The results vary  
1743 between 77% and 95% for some experiment with a 1.5 MRI Tesla scanner and 77%  
1744 and 95% with a 3.0 Tesla MRI scanner.



(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.

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%.

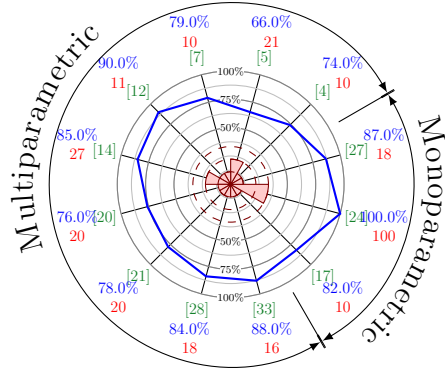
Three studies also use FROC analysis to report their results and are shown in Fig. 23. The results are ranging from 65% to 80%

However, we would like to emphasize the fact that the results obtained from these different experiments cannot be fairly compared. Different dataset were used 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.

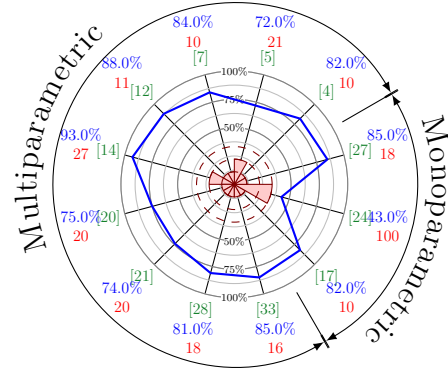
## 5.2. General discussion

This review lead 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 lead to an impossibility to compare fairly the different algorithm designed over year. Also, the availability of a full multi-parametric dataset (cf., MRI and MRSI), could lead to algorithm which are using all the different modalities

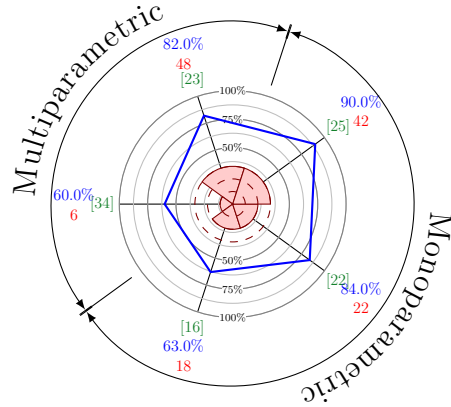




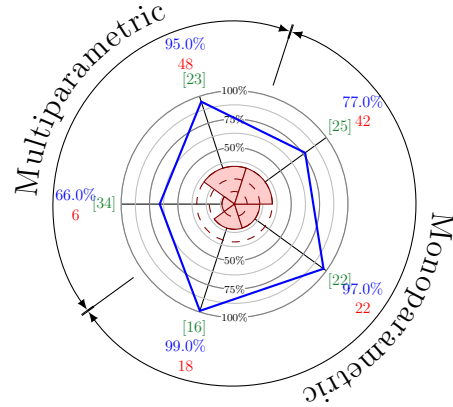
(a) Comparison in term of sensitivity of the methods using data from 1.5 Tesla MRI scanner.



(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 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.

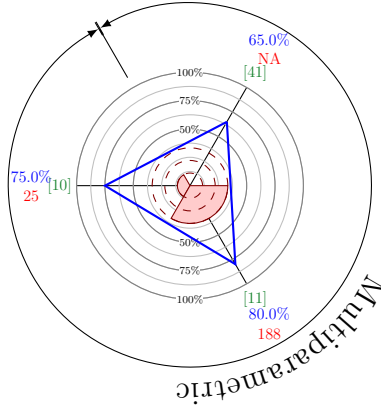


Figure 23: Comparison in term of FROC of the methods using data from 3.0 Tesla MRI scanner. The blue value represent the metric and are graphically reported in the blue rose in the center of the figure. The red value correspond to the number of patients in the dataset and is also reported in the center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

available currently. Recalling Tab. 2, it can be noted that none of the current work provide a algorithm using at the same time the four different modalities. We can also mentioned that all the algorithm 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 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 [132] used some of these features.

Finally, an important point allowing a fair comparison between methods reside in the fact that no universal evaluation model and metric was defined in the re-

1784 search community allowing such comparison. Usually, it is quite common to choose  
1785 an evaluation model which fit the dataset limitations, usually the size. Regarding  
1786 the evaluation, the community would benefit from settling a standard metric which  
1787 represents the most fairly the performance of the algorithms designed.

## 1788 6. Conclusion

1789 This review presented an overview and gave a staging of the research related to  
1790 CAD development for CaP using multi-parametric MRI and MRSI data. We aimed  
1791 at providing an introduction

## 1792 7. Acknowledgement

1793 We would like to acknowledge Sharad Nagappa for all the discussions and his  
1794 precious advices regarding the redaction of this current review.

## 1795 References

- 1796 [1] Agalliu, I., Gern, R., Leanza, S., Burk, R.D., 2009. Associations of high-grade  
1797 prostate cancer with BRCA1 and BRCA2 founder mutations. Clin. Cancer  
1798 Res. 15, 1112–1120.
- 1799 [2] Ahmed, N., Natarajan, T., Rao, K., 1974. Discrete cosine transform. Comput-  
1800 ers, IEEE Transactions on C-23, 90–93. doi:10.1109/T-C.1974.223784.
- 1801 [3] Aizerman, M.A., Braverman, E.A., Rozonoer, L., 1964. Theoretical founda-  
1802 tions of the potential function method in pattern recognition learning, in: Au-  
1803 tomation and Remote Control,, pp. 821–837.
- 1804 [4] Akin, O., Sala, E., Moskowitz, C.S., Kuroiwa, K., Ishill, N.M., Pucar, D.,  
1805 Scardino, P.T., Hricak, H., 2006. Transition zone prostate cancers: features,  
1806 detection, localization, and staging at endorectal MR imaging. Radiology 239,  
1807 784–792.
- 1808 [5] Alexander, D.D., Mink, P.J., Cushing, C.A., Scurman, B., 2010. A review  
1809 and meta-analysis of prospective studies of red and processed meat intake and  
1810 prostate cancer. Nutr J 9, 50.
- 1811 [6] Amadasun, M., King, R., 1989. Textural features corresponding to textural  
1812 properties. Systems, Man and Cybernetics, IEEE Transactions on 19, 1264–  
1813 1274. doi:10.1109/21.44046.

- 1814 [7] American Cancer Society, A.C., 2010. Cancer Facts and Figures 2010. URL:  
1815 <http://www.cancer.org/research/cancerfactsfigures>.
- 1816 [8] American Cancer Society, A.C., 2013. Cancer Facts and Figures 2013. URL:  
1817 <http://www.cancer.org/research/cancerfactsfigures>.
- 1818 [9] Ampeliotis, D., Anonakoudi, A., Berberidis, K., Psarakis, E.Z., 2007. Com-  
1819 puter aided detection of prostate cancer using fused information from dynamic  
1820 contrast enhanced and morphological magnetic resonance images, in: IEEE  
1821 International Conference on Signal Processing and Communications, pp. 888–  
1822 891.
- 1823 [10] Ampeliotis, D., Anonakoudi, A., Berberidis, K., Psarakis, E.Z., Kounoudes,  
1824 A., 2008. A computer-aided system for the detection of prostate cancer based  
1825 on magnetic resonance image analysis, in: International Symposium on Com-  
1826 munications, Control and Signal Processing.
- 1827 [11] Amundadottir, L.T., Sulem, P., Gudmundsson, J., Helgason, A., Baker, A.,  
1828 Agnarsson, B.A., Sigurdsson, A., Benediktsdottir, K.R., Cazier, J.B., Sainz,  
1829 J., Jakobsdottir, M., Kostic, J., Magnusdottir, D.N., Ghosh, S., Agnarsson,  
1830 K., Birgisdottir, B., Le Roux, L., Olafsdottir, A., Blondal, T., Andresdottir,  
1831 M., Gretarsdottir, O.S., Bergthorsson, J.T., Gudbjartsson, D., Gylfason, A.,  
1832 Thorleifsson, G., Manolescu, A., Kristjansson, K., Geirsson, G., Isaksson, H.,  
1833 Douglas, J., Johansson, J.E., Balter, K., Wiklund, F., Montie, J.E., Yu, X.,  
1834 Suarez, B.K., Ober, C., Cooney, K.A., Gronberg, H., Catalona, W.J., Einars-  
1835 son, G.V., Barkardottir, R.B., Gulcher, J.R., Kong, A., Thorsteinsdottir, U.,  
1836 Stefansson, K., 2006. A common variant associated with prostate cancer in  
1837 European and African populations. *Nat. Genet.* 38, 652–658.
- 1838 [12] Andriole, G.L., Crawford, E.D., Grubb, R.L., Buys, S.S., Chia, D., Church,  
1839 T.R., Fouad, M.N., Gelmann, E.P., Kvale, P.A., Reding, D.J., Weissfeld, J.L.,  
1840 Yokochi, L.A., O’Brien, B., Clapp, J.D., Rathmell, J.M., Riley, T.L., Hayes,  
1841 R.B., Kramer, B.S., Izmirlian, G., Miller, A.B., Pinsky, P.F., Prorok, P.C., Go-  
1842 hagan, J.K., Berg, C.D., 2009. Mortality results from a randomized Prostate-  
1843 cancer screening trial. *New England Journal of Medicine* 360, 1310–1319.
- 1844 [13] Antic, T., Peng, Y., Jiang, Y., Giger, M.L., Eggener, S., Oto, A., 2013. A study  
1845 of T2-weighted MR image texture features and diffusion-weighted MR image  
1846 features for computer-aided diagnosis of prostate cancer , 86701H–86701H–6.

- [14] Artan, Y., Haider, M.A., Langer, D.L., van der Kwast, T.H., Evans, A.J., Yang, Y., Wernick, M.N., Trachtenberg, J., Yetik, I.S., 2010. Prostate cancer localization with multispectral MRI using cost-sensitive support vector machines and conditional random fields. *IEEE Trans Image Process* 19, 2444–2455.
- [15] Artan, Y., Langer, D., Haider, M., Van der Kwast, T.H., Evans, A., Wernick, M., Yetik, I., 2009. Prostate cancer segmentation with multispectral MRI using cost-sensitive Conditional Random Fields, in: *Biomedical Imaging: From Nano to Macro, 2009. ISBI '09. IEEE International Symposium on*, pp. 278–281.
- [16] Awwad, H.M., Geisel, J., Obeid, R., 2012. The role of choline in prostate cancer. *Clin. Biochem.* 45, 1548–1553.
- [17] Barentsz, J.O., Richenberg, J., Clements, R., Choyke, P., Verma, S., Villeirs, G., Rouviere, O., Logager, V., Futterer, J.J., 2012. ESUR prostate MR guidelines 2012. *Eur Radiol* 22, 746–757.
- [18] Belkin, M., Niyogi, P., 2001. Laplacian eigenmaps and spectral techniques for embedding and clustering, in: *Advances in Neural Information Processing Systems 14*, MIT Press. pp. 585–591.
- [19] Belongie, S., Malik, J., Puzicha, J., 2002. Shape matching and object recognition using shape contexts. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 24, 509–522. doi:10.1109/34.993558.
- [20] Benassi, A., Cohen, S., Istas, J., 1998. Identifying the multifractional function of a Gaussian process. *Statistics & Probability Letters* 39, 337 – 345. URL: <http://www.sciencedirect.com/science/article/pii/S0167715298000789>, doi:[http://dx.doi.org/10.1016/S0167-7152\(98\)00078-9](http://dx.doi.org/10.1016/S0167-7152(98)00078-9).
- [21] Bishop, C.M., 2006. *Pattern recognition and machine learning*. Springer-Verlag New York, Inc., Secaucus, NJ, USA.
- [22] Bookstein, F.L., 1989. Principal warps: thin-plate splines and the decomposition of deformations. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 11, 567–585. doi:10.1109/34.24792.
- [23] Boser, B.E., Guyon, I.M., Vapnik, V.N., 1992. A training algorithm for optimal margin classifiers, in: *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, ACM, New York, NY, USA. pp. 144–152. URL: <http://doi.acm.org/10.1145/130385.130401>, doi:10.1145/130385.130401.

- 1880 [24] Bourdounis, A., Papatsoris, A.G., Chrisofos, M., Efstathiou, E., Skolarikos,  
1881 A., Deliveliotis, C., 2010. The novel prostate cancer antigen 3 (PCA3)  
1882 biomarker. *Int Braz J Urol* 36, 665–668.
- 1883 [25] Breiman, L., 2001. Random forests. *Machine Learning* 45, 5–32.
- 1884 [26] Breiman, L., Friedman, J., Olshen, R., Stone, C., 1984. Classification and  
1885 regression trees. Wadsworth and Brooks, Monterey, CA.
- 1886 [27] Brenner, J., Chinnaiyan, A., Tomlins, S., 2013. ETS fusion genes in prostate  
1887 cancer, in: Tindall, D.J. (Ed.), *Prostate Cancer*. Springer New York. volume 16  
1888 of *Protein Reviews*, pp. 139–183.
- 1889 [28] Buades, A., Coll, B., Morel, J., 2005. A review of image denoising algorithms,  
1890 with a new one. *Simul* 4, 490–530.
- 1891 [29] Buckley, D.L., Roberts, C., Parker, G.J., Logue, J.P., Hutchinson, C.E., 2004.  
1892 Prostate cancer: evaluation of vascular characteristics with dynamic contrast-  
1893 enhanced T1-weighted MR imaging—initial experience. *Radiology* 233, 709–715.
- 1894 [30] Byrd, R.H., Lu, P., Nocedal, J., Zhu, C., 1995. A limited memory algorithm for  
1895 bound constrained optimization. *SIAM J. Sci. Comput.* 16, 1190–1208. URL:  
1896 <http://dx.doi.org/10.1137/0916069>, doi:10.1137/0916069.
- 1897 [31] Carmeliet, P., Jain, R.K., 2000. Angiogenesis in cancer and other diseases.  
1898 *Nature* 407, 249–257.
- 1899 [32] Carrol, C.L., Sommer, F.G., McNeal, J.E., Stamey, T.A., 1987. The abnormal  
1900 prostate: MR imaging at 1.5 T with histopathologic correlation. *Radiology*  
1901 163, 521–525.
- 1902 [33] Castorina, P., Delsanto, P.P., Guiot, C., 2006. Classification scheme for phe-  
1903 nomenological universalities in growth problems in physics and other sciences.  
1904 *Phys. Rev. Lett.* 96, 188701. URL: [http://link.aps.org/doi/10.1103/](http://link.aps.org/doi/10.1103/PhysRevLett.96.188701)  
1905 [PhysRevLett.96.188701](http://link.aps.org/doi/10.1103/PhysRevLett.96.188701), doi:10.1103/PhysRevLett.96.188701.
- 1906 [34] Chan, H.P., Doi, K., Galhotra, S., Vyborny, C.J., MacMahon, H., Jokich,  
1907 P.M., 1987. Image feature analysis and computer-aided diagnosis in digital  
1908 radiography. I. Automated detection of microcalcifications in mammography.  
1909 *Med Phys* 14, 538–548.

- 1910 [35] Chan, H.P., Hadjiiski, L., Zhou, C., Sahiner, B., 2008. Computer-aided di-  
1911 agnosis of lung cancer and pulmonary embolism in computed tomography-a  
1912 review. *Acad Radiol* 15, 535–555.
- 1913 [36] Chan, H.P., Sahiner, B., Helvie, M.A., Petrick, N., Roubidoux, M.A., Wilson,  
1914 T.E., Adler, D.D., Paramagul, C., Newman, J.S., Sanjay-Gopal, S., 1999. Im-  
1915 provement of radiologists’ characterization of mammographic masses by using  
1916 computer-aided diagnosis: an ROC study. *Radiology* 212, 817–827.
- 1917 [37] Chan, I., Wells, W., Mulkern, R.V., Haker, S., Zhang, J., Zou, K.H., Maier,  
1918 S.E., Tempany, C.M., 2003. Detection of prostate cancer by integration of  
1919 line-scan diffusion, T2-mapping and T2-weighted magnetic resonance imaging;  
1920 a multichannel statistical classifier. *Med Phys* 30, 2390–2398.
- 1921 [38] Chappelow, J., Bloch, B.N., Rofsky, N., Genega, E., Lenkinski, R., DeWolf,  
1922 W., Madabhushi, A., 2011. Elastic registration of multimodal prostate MRI  
1923 and histology via multiattribute combined mutual information. *Med Phys* 38,  
1924 2005–2018.
- 1925 [39] Chen, L., Weng, Z., Goh, L., Garland, M., 2002. An efficient algorithm for  
1926 automatic phase correction of {NMR} spectra based on entropy minimiza-  
1927 tion . *Journal of Magnetic Resonance* 158, 164 – 168. URL: <http://www.sciencedirect.com/science/article/pii/S1090780702000691>, doi:[http://dx.doi.org/10.1016/S1090-7807\(02\)00069-1](http://dx.doi.org/10.1016/S1090-7807(02)00069-1).  
1928  
1929
- 1930 [40] Cheng, H.D., Shan, J., Ju, W., Guo, Y., Zhang, L., 2010. Automated breast  
1931 cancer detection and classification using ultrasound images: A survey. *Pattern*  
1932 *Recogn.* 43, 299–317.
- 1933 [41] Choi, Y.J., Kim, J.K., Kim, N., Kim, K.W., Choi, E.K., Cho, K.S., 2007.  
1934 Functional MR imaging of prostate cancer. *Radiographics* 27, 63–75.
- 1935 [42] Chou, R., Croswell, J.M., Dana, T., Bougatsos, C., Blazina, I., Fu, R., Gleits-  
1936 mann, K., Koenig, H.C., Lam, C., Maltz, A., Rugge, J.B., Lin, K., 2011.  
1937 Screening for prostate cancer: a review of the evidence for the U.S. Preventive  
1938 Services Task Force. *Ann. Intern. Med.* 155, 762–771.
- 1939 [43] Coakley, F.V., Hricak, H., 2000. Radiologic anatomy of the prostate gland: a  
1940 clinical approach. *Radiol. Clin. North Am.* 38, 15–30.

- [44] Cohen, R.J., Shannon, B.A., Phillips, M., Moorin, R.E., Wheeler, T.M., Garrett, K.L., 2008. Central zone carcinoma of the prostate gland: a distinct tumor type with poor prognostic features. *J. Urol.* 179, 1762–1767.
- [45] Coifman, R., Wickerhauser, M., 1992. Entropy-based algorithms for best basis selection. *Information Theory, IEEE Transactions on* 38, 713–718. doi:10.1109/18.119732.
- [46] Coleman, T., Li, Y., 1993. An interior trust region approach for nonlinear minimization subject to bounds. Technical Report. Cornell University.
- [47] Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J., 1995. Active shape models—Their training and application. *Comput. Vis. Image Underst.* 61, 38–59. URL: <http://dx.doi.org/10.1006/cviu.1995.1004>, doi:10.1006/cviu.1995.1004.
- [48] Cortes, C., Vapnik, V., 1995. Support-Vector networks. *Machine Learning* 20, 273–297. URL: <http://dx.doi.org/10.1023/A:1022627411411>, doi:10.1023/A:1022627411411.
- [49] Costello, L.C., Franklin, R.B., 2006. The clinical relevance of the metabolism of prostate cancer; zinc and tumor suppression: connecting the dots. *Mol. Cancer* 5, 17.
- [50] Cruz, M., Tsuda, K., Narumi, Y., Kuroiwa, Y., Nose, T., Kojima, Y., Okuyama, A., Takahashi, S., Aozasa, K., Barentsz, J.O., Nakamura, H., 2002. Characterization of low-intensity lesions in the peripheral zone of prostate on pre-biopsy endorectal coil MR imaging. *Eur Radiol* 12, 357–365.
- [51] Dalal, N., Triggs, B., 2005. Histograms of oriented gradients for human detection, in: *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, pp. 886–893 vol. 1. doi:10.1109/CVPR.2005.177.
- [52] Daugman, J.G., 1985. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *J Opt Soc Am A* 2, 1160–1169.
- [53] Dean, J.C., Ilvento, C.C., 2006. Improved cancer detection using computer-aided detection with diagnostic and screening mammography: prospective study of 104 cancers. *AJR Am J Roentgenol* 187, 20–28.



- 1973 [54] Delongchamps, N.B., Peyromaure, M., Schull, A., Beuvon, F., Bouazza, N.,  
1974 Flam, T., Zerbib, M., Muradyan, N., Legman, P., Cornud, F., 2013. Pre-  
1975 biopsy magnetic resonance imaging and prostate cancer detection: comparison  
1976 of random and targeted biopsies. *J. Urol.* 189, 493–499.
- 1977 [55] Delpierre, C., Lamy, S., Kelly-Irving, M., Molinie, F., Velten, M., Tretarre,  
1978 B., Woronoff, A.S., Buemi, A., Lapotre-Ledoux, B., Bara, S., Guizard, A.V.,  
1979 Colonna, M., Grosclaude, P., 2013. Life expectancy estimates as a key factor  
1980 in over-treatment: the case of prostate cancer. *Cancer Epidemiol* 37, 462–468.
- 1981 [56] Devos, A., Lukas, L., Suykens, J.A., Vanhamme, L., Tate, A.R., Howe, F.A.,  
1982 Majos, C., Moreno-Torres, A., van der Graaf, M., Arus, C., Van Huffel, S.,  
1983 2004. Classification of brain tumours using short echo time 1H MR spectra. *J.*  
1984 *Magn. Reson.* 170, 164–175.
- 1985 [57] Doi, K., Chan, H.P., Giger, M., 1990. Method and system for enhancement  
1986 and detection of abnormal anatomic regions in a digital image.
- 1987 [58] Donoho, D.L., Johnstone, J.M., 1994. Ideal spatial adaptation by wavelet  
1988 shrinkage. *Biometrika* 81, 425–455. doi:10.1093/biomet/81.3.425.
- 1989 [59] Doo, K.W., Sung, D.J., Park, B.J., Kim, M.J., Cho, S.B., Oh, Y.W., Ko, Y.H.,  
1990 Yang, K.S., 2012. Detectability of low and intermediate or high risk prostate  
1991 cancer with combined T2-weighted and diffusion-weighted MRI. *Eur Radiol*  
1992 22, 1812–1819.
- 1993 [60] Efron, B., 1979. Bootstrap methods: Another look at the jackknife. *The Annals*  
1994 *of Statistics* 7, 1–26. URL: <http://dx.doi.org/10.1214/aos/1176344552>,  
1995 doi:10.1214/aos/1176344552.
- 1996 [61] Efron, B., 1983. Estimating the error rate of a prediction rule: Improvement  
1997 on cross-validation. *Journal of the American Statistical Association* 78, pp.  
1998 316–331. URL: <http://www.jstor.org/stable/2288636>.
- 1999 [62] Elter, M., Horsch, A., 2009. CADx of mammographic masses and clustered  
2000 microcalcifications: a review. *Med Phys* 36, 2052–2068.
- 2001 [63] Epstein, J.I., Allsbrook, W.C., Amin, M.B., Egevad, L.L., 2005. The 2005  
2002 International Society of Urological Pathology (ISUP) Consensus Conference  
2003 on Gleason Grading of Prostatic Carcinoma. *Am. J. Surg. Pathol.* 29, 1228–  
2004 1242.

- 2005 [64] Etzioni, R., Penson, D.F., Legler, J.M., di Tommaso, D., Boer, R., Gann, P.H.,  
2006 Feuer, E.J., 2002. Overdiagnosis due to prostate-specific antigen screening:  
2007 lessons from U.S. prostate cancer incidence trends. *J. Natl. Cancer Inst.* 94,  
2008 981–990.
- 2009 [65] Ferlay, J., Shin, H.R., Bray, F., Forman, D., Mathers, C., Parkin, D.M., 2010.  
2010 Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int. J.*  
2011 *Cancer* 127, 2893–2917.
- 2012 [66] Fodor, I., 2002. A survey of dimension reduction techniques.
- 2013 [67] Fred, A., Jain, A., 2005. Combining multiple clusterings using evidence accu-  
2014 mulation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*  
2015 27, 835–850. doi:10.1109/TPAMI.2005.113.
- 2016 [68] Freedman, M.L., Haiman, C.A., Patterson, N., McDonald, G.J., Tandon, A.,  
2017 Waliszewska, A., Penney, K., Steen, R.G., Ardlie, K., John, E.M., Oakley-  
2018 Girvan, I., Whittemore, A.S., Cooney, K.A., Ingles, S.A., Altshuler, D., Hen-  
2019 derson, B.E., Reich, D., 2006. Admixture mapping identifies 8q24 as a prostate  
2020 cancer risk locus in African-American men. *Proc. Natl. Acad. Sci. U.S.A.* 103,  
2021 14068–14073.
- 2022 [69] Freund, Y., Schapire, R., 1997. A decision-theoretic generalization of on-line  
2023 learning and an application to boosting. *Journal of Computer and System*  
2024 *Sciences* 55, 119 – 139. URL: [http://www.sciencedirect.com/science/](http://www.sciencedirect.com/science/article/pii/S002200009791504X)  
2025 [article/pii/S002200009791504X](http://www.sciencedirect.com/science/article/pii/S002200009791504X), doi:[http://dx.doi.org/10.1006/jcss.](http://dx.doi.org/10.1006/jcss.1997.1504)  
2026 1997.1504.
- 2027 [70] Friedman, J., 1989. Regularized discriminant analysis. *Journal of the Amer-*  
2028 *ican Statistical Association* 84, pp. 165–175. URL: [http://www.jstor.org/](http://www.jstor.org/stable/2289860)  
2029 [stable/2289860](http://www.jstor.org/stable/2289860).
- 2030 [71] Gabor, D., 1946. Theory of communication. Part 1: The analysis of informa-  
2031 tion. *Electrical Engineers - Part III: Radio and Communication Engineering,*  
2032 *Journal of the Institution of* 93, 429–441. doi:10.1049/ji-3-2.1946.0074.
- 2033 [72] Ghose, S., Oliver, A., Marti, R., Llado, X., Vilanova, J.C., Freixenet, J., Mi-  
2034 tra, J., Sidibe, D., Meriaudeau, F., 2012. A survey of prostate segmentation  
2035 methodologies in ultrasound, magnetic resonance and computed tomography  
2036 images. *Comput Methods Programs Biomed* 108, 262–287.

2037 [73] Giannini, V., Vignati, A., Mazzetti, S., De Luca, M., Bracco, C., Stasi, M.,  
2038 Russo, F., Armando, E., Regge, D., 2013. A prostate CAD system based  
2039 on multiparametric analysis of DCE T1-w, and DW automatically registered  
2040 images , 86703E–86703E–6.

2041 [74] Gibbs, P., Tozer, D.J., Liney, G.P., Turnbull, L.W., 2001. Comparison of  
2042 quantitative T2 mapping and diffusion-weighted imaging in the normal and  
2043 pathologic prostate. *Magn Reson Med* 46, 1054–1058.

2044 [75] Giger, M.L., Chan, H.P., Boone, J., 2008. Anniversary paper: History and  
2045 status of CAD and quantitative image analysis: the role of Medical Physics  
2046 and AAPM. *Med Phys* 35, 5799–5820.

2047 [76] Giger, M.L., Doi, K., MacMahon, H., 1988. Image feature analysis and  
2048 computer-aided diagnosis in digital radiography. 3. Automated detection of  
2049 nodules in peripheral lung fields. *Med Phys* 15, 158–166.

2050 [77] Giovannucci, E., Liu, Y., Platz, E.A., Stampfer, M.J., Willett, W.C., 2007.  
2051 Risk factors for prostate cancer incidence and progression in the health profes-  
2052 sionals follow-up study. *Int. J. Cancer* 121, 1571–1578.

2053 [78] Giskeodegard, G.F., Bertilsson, H., Selnaes, K.M., Wright, A.J., Bathen, T.F.,  
2054 Viset, T., Halgunset, J., Angelsen, A., Gribbestad, I.S., Tessem, M.B., 2013.  
2055 Spermine and citrate as metabolic biomarkers for assessing prostate cancer  
2056 aggressiveness. *PLoS ONE* 8, e62375.

2057 [79] Gleason, D.F., 1977. Urologic pathology: The prostate. Lea and Febiger..  
2058 chapter The Veteran’s Administration Cooperative Urologic Research Group:  
2059 histologic grading and clinical staging of prostatic carcinoma. p. 171198.

2060 [80] Goodman, S.N., 1999. Toward evidence-based medical statistics. 1: The P  
2061 value fallacy. *Ann. Intern. Med.* 130, 995–1004.

2062 [81] van der Graaf, M., Schipper, R.G., Oosterhof, G.O., Schalken, J.A., Verhofs-  
2063 tad, A.A., Heerschap, A., 2000. Proton MR spectroscopy of prostatic tissue  
2064 focused on the detection of spermine, a possible biomarker of malignant be-  
2065 havior in prostate cancer. *MAGMA* 10, 153–159.

2066 [82] Gribbestad, I., Gjesdal, K., Nilsen, G., Lundgren, S., Hjelstuen, M., Jackson,  
2067 A., 2005. An introduction to dynamic contrast-enhanced MRI in oncology,  
2068 in: Jackson, A., Buckley, D., Parker, G. (Eds.), *Dynamic Contrast-Enhanced*

- 2069 Magnetic Resonance Imaging in Oncology. Springer Berlin Heidelberg. Medical  
2070 Radiology, pp. 1–22.
- 2071 [83] Haacke, E., Brown, R., Thompson, M., Venkatesan, R., 1999. Magnetic  
2072 resonance imaging: Physical principles and sequence design. Wiley. URL:  
2073 <http://books.google.es/books?id=BnOvQgAACAAJ>.
- 2074 [84] Haas, G.P., Delongchamps, N.B., Jones, R.F., Chandan, V., Serio, A.M., Vick-  
2075 ers, A.J., Jumbelic, M., Threatte, G., Korets, R., Lilja, H., de la Roza, G., 2007.  
2076 Needle biopsies on autopsy prostates: sensitivity of cancer detection based on  
2077 true prevalence. *J. Natl. Cancer Inst.* 99, 1484–1489.
- 2078 [85] Hambrock, T., Somford, D.M., Huisman, H.J., van Oort, I.M., Witjes, J.A.,  
2079 Hulsbergen-van de Kaa, C.A., Scheenen, T., Barentsz, J.O., 2011. Relationship  
2080 between apparent diffusion coefficients at 3.0-T MR imaging and Gleason grade  
2081 in peripheral zone prostate cancer. *Radiology* 259, 453–461.
- 2082 [86] Hambrock, T., Vos, P.C., Hulsbergen-van de Kaa, C.A., Barentsz, J.O., Huis-  
2083 man, H.J., 2013. Prostate cancer: computer-aided diagnosis with multiparam-  
2084 etric 3-T MR imaging—effect on observer performance. *Radiology* 266, 521–  
2085 530.
- 2086 [87] Hara, N., Okuizumi, M., Koike, H., Kawaguchi, M., Bilim, V., 2005. Dynamic  
2087 contrast-enhanced magnetic resonance imaging (DCE-MRI) is a useful modal-  
2088 ity for the precise detection and staging of early prostate cancer. *Prostate* 62,  
2089 140–147.
- 2090 [88] Haralick, R., Shanmugam, K., Dinstein, I., 1973. Textural features for image  
2091 classification. *Systems, Man and Cybernetics, IEEE Transactions on SMC-3*,  
2092 610–621. doi:10.1109/TSMC.1973.4309314.
- 2093 [89] Hegde, J.V., Mulkern, R.V., Panych, L.P., Fennessy, F.M., Fedorov, A., Maier,  
2094 S.E., Tempany, C.M., 2013. Multiparametric MRI of prostate cancer: an up-  
2095 date on state-of-the-art techniques and their performance in detecting and  
2096 localizing prostate cancer. *J Magn Reson Imaging* 37, 1035–1054.
- 2097 [90] Heidenreich, A., Abrahamsson, P.A., Artibani, W., Catto, J., Montorsi, F.,  
2098 Van Poppel, H., Wirth, M., Mottet, N., 2013. Early Detection of Prostate  
2099 Cancer: European Association of Urology Recommendation. *Eur. Urol.* .

- 2100 [91] Hero, A., Ma, B., Michel, O., Gorman, J., 2002. Applications of entropic  
2101 spanning graphs. *Signal Processing Magazine, IEEE* 19, 85–95. doi:10.1109/  
2102 MSP.2002.1028355.
- 2103 [92] Hoeks, C.M., Barentsz, J.O., Hambrock, T., Yakar, D., Somford, D.M., Hei-  
2104 jmink, S.W., Scheenen, T.W., Vos, P.C., Huisman, H., van Oort, I.M., Witjes,  
2105 J.A., Heerschap, A., Futterer, J.J., 2011. Prostate cancer: multiparametric  
2106 MR imaging for detection, localization, and staging. *Radiology* 261, 46–66.
- 2107 [93] Hoffman, R.M., Gilliland, F.D., Eley, J.W., Harlan, L.C., Stephenson, R.A.,  
2108 Stanford, J.L., Albertson, P.C., Hamilton, A.S., Hunt, W.C., Potosky, A.L.,  
2109 2001. Racial and ethnic differences in advanced-stage prostate cancer: the  
2110 Prostate Cancer Outcomes Study. *J. Natl. Cancer Inst.* 93, 388–395.
- 2111 [94] Hricak, H., Doms, G.C., McNeal, J.E., Mark, A.S., Marotti, M., Avallone, A.,  
2112 Pelzer, M., Proctor, E.C., Tanagho, E.A., 1987. MR imaging of the prostate  
2113 gland: normal anatomy. *AJR Am J Roentgenol* 148, 51–58.
- 2114 [95] Hricak, H., Williams, R.D., Spring, D.B., Moon, K.L., Hedgcock, M.W., Wat-  
2115 son, R.A., Crooks, L.E., 1983. Anatomy and pathology of the male pelvis by  
2116 magnetic resonance imaging. *AJR Am J Roentgenol* 141, 1101–1110.
- 2117 [96] Huch Boni, R.A., Boner, J.A., Lutolf, U.M., Trinkler, F., Pestalozzi, D.M.,  
2118 Krestin, G.P., 1995. Contrast-enhanced endorectal coil MRI in local staging of  
2119 prostate carcinoma. *J Comput Assist Tomogr* 19, 232–237.
- 2120 [97] Hugosson, J., Carlsson, S., Aus, G., Bergdahl, S., Khatami, A., Lodding, P.,  
2121 Pihl, C.G., Stranne, J., Holmberg, E., Lilja, H., 2010. Mortality results from the  
2122 Göteborg randomised population-based prostate-cancer screening trial. *Lancet*  
2123 *Oncol.* 11, 725–732.
- 2124 [98] Huisman, H., Vos, P., Litjens, G., Hambrock, T., Barentsz, J., 2010. Computer  
2125 aided detection of prostate cancer using T2, DWI and DCE MRI: methods and  
2126 clinical applications, in: *Proceedings of the 2010 international conference on*  
2127 *Prostate cancer imaging: computer-aided diagnosis, prognosis, and interven-*  
2128 *tion*, Springer-Verlag, Berlin, Heidelberg. pp. 4–14.
- 2129 [99] Huisman, T.A., 2003. Diffusion-weighted imaging: basic concepts and appli-  
2130 cation in cerebral stroke and head trauma. *Eur Radiol* 13, 2283–2297.

- 2131 [100] Itou, Y., Nakanishi, K., Narumi, Y., Nishizawa, Y., Tsukuma, H., 2011. Clin-  
2132 ical utility of apparent diffusion coefficient (ADC) values in patients with  
2133 prostate cancer: can ADC values contribute to assess the aggressiveness of  
2134 prostate cancer? *J Magn Reson Imaging* 33, 167–172.
- 2135 [101] Jager, G.J., Ruijter, E.T., van de Kaa, C.A., de la Rosette, J.J., Oosterhof,  
2136 G.O., Thornbury, J.R., Ruijs, S.H., Barentsz, J.O., 1997. Dynamic Tur-  
2137 boFLASH subtraction technique for contrast-enhanced MR imaging of the  
2138 prostate: correlation with histopathologic results. *Radiology* 203, 645–652.
- 2139 [102] Jolliffe, I.T., 2002. Principal Component Analysis. Second ed.,  
2140 Springer. URL: [http://www.amazon.com/exec/obidos/redirect?tag=](http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20&path=ASIN/0387954422)  
2141 [citeulike07-20&path=ASIN/0387954422](http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20&path=ASIN/0387954422).
- 2142 [103] Jungke, M., Von Seelen, W., Bielke, G., Meindl, S., Grigat, M., Pfannenstiel,  
2143 P., 1987. A system for the diagnostic use of tissue characterizing parameters  
2144 in NMR-tomography, in: *Proc. of Information Processing in Medical Imaging*,  
2145 pp. 471–481.
- 2146 [104] Kaji, Y., Kurhanewicz, J., Hricak, H., Sokolov, D.L., Huang, L.R., Nelson, S.J.,  
2147 Vigneron, D.B., 1998. Localizing prostate cancer in the presence of postbiopsy  
2148 changes on MR images: role of proton MR spectroscopic imaging. *Radiology*  
2149 206, 785–790.
- 2150 [105] Kato, Z., Pong, T., 2001. A Markov random field image segmentation model  
2151 using combined color and texture features, in: Skarbek, W. (Ed.), *Computer*  
2152 *Analysis of Images and Patterns*. Springer Berlin Heidelberg. volume 2124 of  
2153 *Lecture Notes in Computer Science*, pp. 547–554. URL: [http://dx.doi.org/](http://dx.doi.org/10.1007/3-540-44692-3_66)  
2154 [10.1007/3-540-44692-3\\_66](http://dx.doi.org/10.1007/3-540-44692-3_66), doi:10.1007/3-540-44692-3\_66.
- 2155 [106] Kelm, B.M., Menze, B.H., Zechmann, C.M., Baudendistel, K.T., Hamprecht,  
2156 F.A., 2007. Automated estimation of tumor probability in prostate magnetic  
2157 resonance spectroscopic imaging: pattern recognition vs quantification. *Magn*  
2158 *Reson Med* 57, 150–159.
- 2159 [107] Kety, S., 1951. The theory and applications of the exchange of inert gas at the  
2160 lungs and tissues. *Pharmacol. Rev.* 3, 1–41.
- 2161 [108] Kim, J.K., Hong, S.S., Choi, Y.J., Park, S.H., Ahn, H., Kim, C.S., Cho, K.S.,  
2162 2005. Wash-in rate on the basis of dynamic contrast-enhanced MRI: usefulness

- 2163 for prostate cancer detection and localization. *J Magn Reson Imaging* 22, 639–  
2164 646.
- 2165 [109] Kirkham, A.P., Emberton, M., Allen, C., 2006. How good is MRI at detecting  
2166 and characterising cancer within the prostate? *Eur. Urol.* 50, 1163–1174.
- 2167 [110] Kirsch, R., 1971. Computer determination of the constituent struc-  
2168 ture of biological images . *Computers and Biomedical Research* 4,  
2169 315 – 328. URL: [http://www.sciencedirect.com/science/article/  
2170 pii/0010480971900346](http://www.sciencedirect.com/science/article/pii/0010480971900346), doi:[http://dx.doi.org/10.1016/0010-4809\(71\)  
2171 90034-6](http://dx.doi.org/10.1016/0010-4809(71)90034-6).
- 2172 [111] Koh, D.M., Collins, D.J., 2007. Diffusion-weighted MRI in the body: applica-  
2173 tions and challenges in oncology. *AJR Am J Roentgenol* 188, 1622–1635.
- 2174 [112] Korotkov, K., Garcia, R., 2012. Computerized analysis of pigmented skin  
2175 lesions: a review. *Artif Intell Med* 56, 69–90.
- 2176 [113] Kurhanewicz, J., Vigneron, D.B., Hricak, H., Narayan, P., Carroll, P., Nelson,  
2177 S.J., 1996. Three-dimensional H-1 MR spectroscopic imaging of the in situ  
2178 human prostate with high (0.24-0.7-cm<sup>3</sup>) spatial resolution. *Radiology* 198,  
2179 795–805.
- 2180 [114] Langer, D.L., van der Kwast, T.H., Evans, A.J., Trachtenberg, J., Wilson,  
2181 B.C., Haider, M.A., 2009. Prostate cancer detection with multi-parametric  
2182 MRI: logistic regression analysis of quantitative T2, diffusion-weighted imaging,  
2183 and dynamic contrast-enhanced MRI. *J Magn Reson Imaging* 30, 327–334.
- 2184 [115] Larsson, H.B., Fritz-Hansen, T., Rostrup, E., Sondergaard, L., Ring, P., Hen-  
2185 riksen, O., 1996. Myocardial perfusion modeling using MRI. *Magn Reson Med*  
2186 35, 716–726.
- 2187 [116] Laudadio, T., Mastronardi, N., Vanhamme, L., Hecke, P.V., Huffel, S.V.,  
2188 2002. Improved Lanczos algorithms for blackbox {MRS} data quantita-  
2189 tion. *Journal of Magnetic Resonance* 157, 292 – 297. URL: [http://www.  
2190 sciencedirect.com/science/article/pii/S1090780702925930](http://www.sciencedirect.com/science/article/pii/S1090780702925930), doi:[http:  
2191 //dx.doi.org/10.1006/jmre.2002.2593](http://dx.doi.org/10.1006/jmre.2002.2593).
- 2192 [117] Le Bihan, D., Breton, E., Lallemand, D., Aubin, M.L., Vignaud, J., Laval-  
2193 Jeantet, M., 1988. Separation of diffusion and perfusion in intravoxel incoherent  
2194 motion MR imaging. *Radiology* 168, 497–505.

- 2195 [118] Le Bihan, D., Breton, E., Lallemand, D., Grenier, P., Cabanis, E., Laval-  
2196 Jeantet, M., 1986. MR imaging of intravoxel incoherent motions: application  
2197 to diffusion and perfusion in neurologic disorders. *Radiology* 161, 401–407.
- 2198 [119] Leissner, K.H., Tisell, L.E., 1979. The weight of the human prostate. *Scand.*  
2199 *J. Urol. Nephrol.* 13, 137–142.
- 2200 [120] Lemaître, G., 2011. Absolute quantification at 3 T. Master’s thesis. Université  
2201 de Bourgogne, Heriot-Watt University, Universitat de Girona.
- 2202 [121] Li, F., Aoyama, M., Shiraishi, J., Abe, H., Li, Q., Suzuki, K., Engelmann,  
2203 R., Sone, S., Macmahon, H., Doi, K., 2004. Radiologists’ performance for  
2204 differentiating benign from malignant lung nodules on high-resolution CT using  
2205 computer-estimated likelihood of malignancy. *AJR Am J Roentgenol* 183,  
2206 1209–1215.
- 2207 [122] Li, H., Giger, M.L., Olopade, O.I., Margolis, A., Lan, L., Chinander, M.R.,  
2208 2005. Computerized texture analysis of mammographic parenchymal patterns  
2209 of digitized mammograms. *Acad Radiol* 12, 863–873.
- 2210 [123] Li, Q., Sone, S., Doi, K., 2003. Selective enhancement filters for nodules,  
2211 vessels, and airway walls in two- and three-dimensional CT scans. *Med Phys*  
2212 30, 2040–2051.
- 2213 [124] Li, S.Z., 1996. Robustizing robust M-estimation using deterministic annealing.  
2214 *Pattern Recognition* 29, 159–166.
- 2215 [125] Lieber, C.A., Mahadevan-Jansen, A., 2003. Automated method for subtraction  
2216 of fluorescence from biological Raman spectra. *Appl Spectrosc* 57, 1363–1367.
- 2217 [126] Liney, G.P., Knowles, A.J., Manton, D.J., Turnbull, L.W., Blackband, S.J.,  
2218 Horsman, A., 1996a. Comparison of conventional single echo and multi-echo  
2219 sequences with a fast spin-echo sequence for quantitative T2 mapping: appli-  
2220 cation to the prostate. *J Magn Reson Imaging* 6, 603–607.
- 2221 [127] Liney, G.P., Lowry, M., Turnbull, L.W., Manton, D.J., Knowles, A.J., Black-  
2222 band, S.J., Horsman, A., 1996b. Proton MR T2 maps correlate with the citrate  
2223 concentration in the prostate. *NMR Biomed* 9, 59–64.
- 2224 [128] Liney, G.P., Turnbull, L.W., Lowry, M., Turnbull, L.S., Knowles, A.J., Hors-  
2225 man, A., 1997. In vivo quantification of citrate concentration and water T2



2226 relaxation time of the pathologic prostate gland using  $^1\text{H}$  MRS and MRI. *Magn*  
2227 *Reson Imaging* 15, 1177–1186.

2228 [129] Litjens, G., Debats, O., van de Ven, W., Karssemeijer, N., Huisman, H., 2012a.  
2229 A pattern recognition approach to zonal segmentation of the prostate on MRI.  
2230 *Med Image Comput Comput Assist Interv* 15, 413–420.

2231 [130] Litjens, G.J.S., Barentsz, J.O., Karssemeijer, N., Huisman, H.J., 2012b. Au-  
2232 tomated computer-aided detection of prostate cancer in MR images: from a  
2233 whole-organ to a zone-based approach , 83150G–83150G–6.

2234 [131] Litjens, G.J.S., Vos, P.C., Barentsz, J.O., Karssemeijer, N., Huisman,  
2235 H.J., 2011. Automatic computer aided detection of abnormalities in multi-  
2236 parametric prostate MRI , 79630T–79630T–7.

2237 [132] Liu, P., Wang, S., Turkbey, B., Grant, K. and Pinto, P.C.P., Wood, B.J., Sum-  
2238 mers, R.M., 2013. A prostate cancer computer-aided diagnosis system using  
2239 multimodal magnetic resonance imaging and targeted biopsy labels , 86701G–  
2240 86701G–6.

2241 [133] Liu, W., Turkbey, B., Senegas, J., Remmele, S., Xu, S., Kruecker, J., Bernardo,  
2242 M., Wood, B.J., Pinto, P.A., Choyke, P.L., 2011. Accelerated T2 mapping for  
2243 characterization of prostate cancer. *Magn Reson Med* 65, 1400–1406.

2244 [134] Liu, X., Langer, D.L., Haider, M.A., Yang, Y., Wernick, M.N., Yetik, I.S.,  
2245 2009. Prostate cancer segmentation with simultaneous estimation of Markov  
2246 random field parameters and class. *IEEE Trans Med Imaging* 28, 906–915.

2247 [135] Lopes, R., Ayache, A., Makni, N., Puech, P., Villers, A., Mordon, S., Betrouni,  
2248 N., 2011. Prostate cancer characterization on MR images using fractal features.  
2249 *Med Phys* 38, 83–95.

2250 [136] Lu-Yao, G.L., Albertsen, P.C., Moore, D.F., Shih, W., Lin, Y., DiPaola,  
2251 R.S., Barry, M.J., Zietman, A., O’Leary, M., Walker-Corkery, E., Yao, S.L.,  
2252 2009. Outcomes of localized prostate cancer following conservative manage-  
2253 ment. *JAMA* 302, 1202–1209.

2254 [137] Lv, D., Guo, X., Wang, X., Zhang, J., Fang, J., 2009. Computerized charac-  
2255 terization of prostate cancer by fractal analysis in MR images. *J Magn Reson*  
2256 *Imaging* 30, 161–168.

- 2257 [138] Ma, R.W., Chapman, K., 2009. A systematic review of the effect of diet in  
2258 prostate cancer prevention and treatment. *J Hum Nutr Diet* 22, 187–199.
- 2259 [139] Madabhushi, A., Udupa, J., Souza, A., 2006. Generalized scale: The-  
2260 ory, algorithms, and application to image inhomogeneity correction . *Com-  
2261 puter Vision and Image Understanding* 101, 100 – 121. URL: [http://www.  
2262 sciencedirect.com/science/article/pii/S1077314205001244](http://www.sciencedirect.com/science/article/pii/S1077314205001244), doi:[http:  
2263 //dx.doi.org/10.1016/j.cviu.2005.07.010](http://dx.doi.org/10.1016/j.cviu.2005.07.010).
- 2264 [140] Madabhushi, A., Udupa, J.K., 2006. New methods of MR image intensity  
2265 standardization via generalized scale. *Med Phys* 33, 3426–3434.
- 2266 [141] Maintz, J.B., Viergever, M.A., 1998. A survey of medical image registration.  
2267 *Med Image Anal* 2, 1–36.
- 2268 [142] Mallat, S., 2008. A wavelet tour of signal processing, Third Edition: The sparse  
2269 way. 3rd ed., Academic Press.
- 2270 [143] Manjon, J.V., Carbonell-Caballero, J., Lull, J.J., Garcia-Marti, G., Marti-  
2271 Bonmati, L., Robles, M., 2008. MRI denoising using non-local means. *Med  
2272 Image Anal* 12, 514–523.
- 2273 [144] Matulewicz, L., Jansen, J.F., Bokacheva, L., Vargas, H.A., Akin, O., Fine,  
2274 S.W., Shukla-Dave, A., Eastham, J.A., Hricak, H., Koutcher, J.A., Za-  
2275 kian, K.L., 2013. Anatomic segmentation improves prostate cancer detec-  
2276 tion with artificial neural networks analysis of 1H magnetic resonance spec-  
2277 troscopic imaging. *Journal of Magnetic Resonance Imaging* , n/a–n/aURL:  
2278 <http://dx.doi.org/10.1002/jmri.24487>, doi:10.1002/jmri.24487.
- 2279 [145] Mazzetti, S., De Luca, M., Bracco, C., Vignati, A., Giannini, V., Stasi, M.,  
2280 Russo, F., Armando, E., Agliozzo, S., Regge, D., 2011. A CAD system based on  
2281 multi-parametric analysis for cancer prostate detection on DCE-MRI , 79633Q–  
2282 79633Q–7.
- 2283 [146] McNeal, J.E., 1981. The zonal anatomy of the prostate. *Prostate* 2, 35–49.
- 2284 [147] McNeal, J.E., Redwine, E.A., Freiha, F.S., Stamey, T.A., 1988. Zonal distri-  
2285 bution of prostatic adenocarcinoma. Correlation with histologic pattern and  
2286 direction of spread. *Am. J. Surg. Pathol.* 12, 897–906.

- 2287 [148] Middleton, D., Esposito, R., 1968. Simultaneous optimum detection and es-  
2288 timation of signals in noise. *Information Theory, IEEE Transactions on* 14,  
2289 434–444. doi:10.1109/TIT.1968.1054139.
- 2290 [149] Mitra, J., 2012. Multimodal image registration applied to magnetic resonance  
2291 and ultrasound prostatic images. Ph.D. thesis. Universitat de Girona and  
2292 Université de Bourgogne.
- 2293 [150] Mitra, J., Kato, Z., Marti, R., Oliver, A., Llado, X., Sidibe, D., Ghose, S.,  
2294 Vilanova, J.C., Comet, J., Meriaudeau, F., 2012. A spline-based non-linear  
2295 diffeomorphism for multimodal prostate registration. *Med Image Anal* 16,  
2296 1259–1279.
- 2297 [151] Mitra, J., Marti, R., Oliver, A., Llado, X., Vilanova, J.C., Meriaudeau, F.,  
2298 2011. A comparison of thin-plate splines with automatic correspondences and  
2299 B-splines with uniform grids for multimodal prostate registration, in: *Society*  
2300 *of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*. doi:10.  
2301 1117/12.877956.
- 2302 [152] Mohan, J., Krishnaveni, V., Guo, Y., 2014. A survey on the magnetic res-  
2303 onance image denoising methods. *Biomedical Signal Processing and Con-*  
2304 *trol* 9, 56 – 69. URL: [http://www.sciencedirect.com/science/article/](http://www.sciencedirect.com/science/article/pii/S1746809413001407)  
2305 [pii/S1746809413001407](http://www.sciencedirect.com/science/article/pii/S1746809413001407), doi:[http://dx.doi.org/10.1016/j.bspc.2013.](http://dx.doi.org/10.1016/j.bspc.2013.10.007)  
2306 10.007.
- 2307 [153] Moore, C.M., Ridout, A., Emberton, M., 2013. The role of MRI in active  
2308 surveillance of prostate cancer. *Curr Opin Urol* 23, 261–267.
- 2309 [154] Morgan, R., Boxall, A., Bhatt, A., Bailey, M., Hindley, R., Langley, S.,  
2310 Whitaker, H.C., Neal, D.E., Ismail, M., Whitaker, H., Annels, N., Michael,  
2311 A., Pandha, H., 2011. Engrailed-2 (EN2): a tumor specific urinary biomarker  
2312 for the early diagnosis of prostate cancer. *Clin. Cancer Res.* 17, 1090–1098.
- 2313 [155] Nelder, J.A., Mead, R., 1965. A simplex method for func-  
2314 tion minimization. *The Computer Journal* 7, 308–313.  
2315 arXiv:<http://comjnl.oxfordjournals.org/content/7/4/308.full.pdf+html>.
- 2316 [156] Niaf, E., Rouviere, O., Mege-Lechevallier, F., Bratan, F., Lartizien, C., 2012.  
2317 Computer-aided diagnosis of prostate cancer in the peripheral zone using mul-  
2318 tiparametric MRI. *Phys Med Biol* 57, 3833–3851.

- 2319 [157] Niaf, E., Rouvire, O., Lartizien, C., 2011. Computer-aided diagnosis for  
2320 prostate cancer detection in the peripheral zone via multisequence MRIs ,  
2321 79633P–79633P–8.
- 2322 [158] van Niekerk, C.G., van der Laak, J.A., Borger, M.E., Huisman, H.J., Witjes,  
2323 J.A., Barentsz, J.O., Hulsbergen-van de Kaa, C.A., 2009. Computerized whole  
2324 slide quantification shows increased microvascular density in pT2 prostate can-  
2325 cer as compared to normal prostate tissue. *Prostate* 69, 62–69.
- 2326 [159] van Niekerk, C.G., Witjes, J.A., Barentsz, J.O., van der Laak, J.A.,  
2327 Hulsbergen-van de Kaa, C.A., 2013. Microvascularity in transition zone  
2328 prostate tumors resembles normal prostatic tissue. *Prostate* 73, 467–475.
- 2329 [160] Noguchi, M., Stamey, T.A., McNeal, J.E., Yemoto, C.M., 2001. Relationship  
2330 between systematic biopsies and histological features of 222 radical prostatec-  
2331 tomy specimens: lack of prediction of tumor significance for men with nonpal-  
2332 pable prostate cancer. *J. Urol.* 166, 104–109.
- 2333 [161] Nowak, R., 1999. Wavelet-based Rician noise removal for magnetic resonance  
2334 imaging. *Image Processing, IEEE Transactions on* 8, 1408–1419. doi:10.1109/  
2335 83.791966.
- 2336 [162] Nyul, L.G., Udupa, J.K., 1999. On standardizing the MR image intensity scale.  
2337 *Magn Reson Med* 42, 1072–1081.
- 2338 [163] Nyul, L.G., Udupa, J.K., Zhang, X., 2000. New variants of a method of MRI  
2339 scale standardization. *IEEE Trans Med Imaging* 19, 143–150.
- 2340 [164] Ojala, T., Pietikäinen, M., Harwood, D., 1996. A comparative study of texture  
2341 measures with classification based on featured distributions. *Pattern Recognition*  
2342 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),  
2343 doi:10.1016/0031-3203(95)00067-4.
- 2344 [165] Osorio-Garcia, M., Croitor Sava, A., Sima, D.M., Nielsen, F., Himmelreich,  
2345 U., Van Huffel, S., 2012. Magnetic Resonance Spectroscopy. InTech. chapter  
2346 Quantification improvements of 1H MRS Signals. pp. 1–27.
- 2347 [166] Oster, G., Lamerato, L., Glass, A.G., Richert-Boe, K.E., Lopez, A., Chung,  
2348 K., Richhariya, A., Dodge, T., Wolff, G.G., Balakumaran, A., Edelsberg, J.,  
2349 2013. Natural history of skeletal-related events in patients with breast, lung,  
2350 or prostate cancer and metastases to bone: a 15-year study in two large US  
2351 health systems. *Support Care Cancer* .

- 2352 [167] Ozer, S., Haider, M., Langer, D.L., Van der Kwast, T.H., Evans, A., Wernick,  
2353 M., Trachtenberg, J., Yetik, I., 2009. Prostate cancer localization with mul-  
2354 tispectral MRI based on Relevance Vector Machines, in: Biomedical Imaging:  
2355 From Nano to Macro, 2009. ISBI '09. IEEE International Symposium on, pp.  
2356 73–76.
- 2357 [168] Ozer, S., Langer, D.L., Liu, X., Haider, M.A., van der Kwast, T.H., Evans,  
2358 A.J., Yang, Y., Wernick, M.N., Yetik, I.S., 2010. Supervised and unsupervised  
2359 methods for prostate cancer segmentation with multispectral MRI. *Med Phys*  
2360 37, 1873–1883.
- 2361 [169] Padhani, A.R., 2002. Dynamic contrast-enhanced MRI in clinical oncology:  
2362 current status and future directions. *J Magn Reson Imaging* 16, 407–422.
- 2363 [170] Padhani, A.R., 2011. Integrating multiparametric prostate MRI into clinical  
2364 practice. *Cancer Imaging* 11 Spec No A, 27–37.
- 2365 [171] Parfait, S., 2010. Classification de spectres et recherche de biomarqueurs en  
2366 spectroscopie par résonance magnétique nucléaire du proton dans les tumeurs  
2367 prostatiques. Ph.D. thesis. Université de Bourgogne.
- 2368 [172] Parfait, S., Walker, P., Crhange, G., Tizon, X., Mitran, J., 2012. Classifica-  
2369 tion of prostate magnetic resonance spectra using Support Vector Machine .  
2370 *Biomedical Signal Processing and Control* 7, 499 – 508.
- 2371 [173] Pearson, K., 1901. On lines and planes of closest fit to systems of points in  
2372 space. *Philosophical Magazine* 2, 559–572.
- 2373 [174] Peng, H., Long, F., Ding, C., 2005. Feature selection based on mutual infor-  
2374 mation criteria of max-dependency, max-relevance, and min-redundancy. *Pat-  
2375 tern Analysis and Machine Intelligence, IEEE Transactions on* 27, 1226–1238.  
2376 doi:10.1109/TPAMI.2005.159.
- 2377 [175] Peng, Y., Jiang, Y., Yang, C., Brown, J., Antic, T., Sethi, I., Schmid-Tannwald,  
2378 C., Giger, M., Eggener, S., Oto, A., 2013. Quantitative analysis of multi-  
2379 parametric prostate MR images: differentiation between prostate cancer and  
2380 normal tissue and correlation with Gleason score—a computer-aided diagnosis  
2381 development study. *Radiology* 267, 787–796.
- 2382 [176] Petrick, N., Haider, M., Summers, R.M., Yeshwant, S.C., Brown, L., Iuliano,  
2383 E.M., Louie, A., Choi, J.R., Pickhardt, P.J., 2008. CT colonography with

2384 computer-aided detection as a second reader: observer performance study.  
 2385 Radiology 246, 148–156.

2386 [177] Pijnappel, W., van den Boogaart, A., de Beer, R., van Ormondt, D.,  
 2387 1992. SVD-based quantification of magnetic resonance signals . Jour-  
 2388 nal of Magnetic Resonance (1969) 97, 122 – 134. URL: <http://www.sciencedirect.com/science/article/pii/002223649290241X>, doi:[http://dx.doi.org/10.1016/0022-2364\(92\)90241-X](http://dx.doi.org/10.1016/0022-2364(92)90241-X).  
 2390

2391 [178] Pizurica, A., 2002. Image denoising using wavelets and spatial context model-  
 2392 ing. Ph.D. thesis. Universiteit Gent.

2393 [179] Pizurica, A., Philips, W., Lemahieu, I., Acheroy, M., 2003. A versatile wavelet  
 2394 domain noise filtration technique for medical imaging. IEEE Trans Med Imag-  
 2395 ing 22, 323–331.

2396 [180] Pluim, J., Maintz, J., Viergever, M., 2003. Mutual-information-based registra-  
 2397 tion of medical images: a survey. IEEE Transactions on Medical Imaging 22,  
 2398 986–1004.

2399 [181] Prewitt, J., 1970. Picture processing and psychohistories. Academic Press.  
 2400 chapter Object enhancement and extraction.

2401 [182] Provencher, S.W., 1993. Estimation of metabolite concentrations from localized  
 2402 in vivo proton NMR spectra. Magn Reson Med 30, 672–679.

2403 [183] Puech, P., Betrouni, N., Makni, N., Dewalle, A.S., Villers, A., Lemaitre, L.,  
 2404 2009. Computer-assisted diagnosis of prostate cancer using DCE-MRI data:  
 2405 design, implementation and preliminary results. Int J Comput Assist Radiol  
 2406 Surg 4, 1–10.

2407 [184] Quinlan, J., 1986. Induction of decision trees. Machine Learning 1, 81–106.  
 2408 URL: <http://dx.doi.org/10.1007/BF00116251>, doi:10.1007/BF00116251.

2409 [185] Quinlan, J., 1993. C4.5: Programs for machine learning. Morgan Kaufmann  
 2410 Publishers Inc., San Francisco, CA, USA.

2411 [186] Quinonero-Candela, J., Girard, A., Rasmussen, C., 2002. Prediction at an Un-  
 2412 certain Input for Gaussian processes and relevance vector machines application  
 2413 to Multiple-Step ahead time-series forecasting. Technical Report.

- [187] Quint, L.E., Van Erp, J.S., Bland, P.H., Mandell, S.H., Del Buono, E.A., Grossman, H.B., Glazer, G.M., Gikas, P.W., 1991. Carcinoma of the prostate: MR images obtained with body coils do not accurately reflect tumor volume. *AJR Am J Roentgenol* 156, 511–516.
- [188] Rangayyan, R., Ayres, F., Desautels, J., 2007. A review of computer-aided diagnosis of breast cancer: Toward the detection of subtle signs. *Journal of the Franklin Institute* 344, 312 – 348.
- [189] Rasmussen, C., Williams, C., 2005. Gaussian processes for machine learning. The MIT Press.
- [190] Ratiney, H., Sdika, M., Coenradie, Y., Cavassila, S., van Ormondt, D., Graveron-Demilly, D., 2005. Time-domain semi-parametric estimation based on a metabolite basis set. *NMR Biomed* 18, 1–13.
- [191] Rish, I., 2001. An empirical study of the naive Bayes classifier, in: *IJCAI 2001 workshop on empirical methods in artificial intelligence*, pp. 41–46.
- [192] Rodriguez, C., Freedland, S.J., Deka, A., Jacobs, E.J., McCullough, M.L., Patel, A.V., Thun, M.J., Calle, E.E., 2007. Body mass index, weight change, and risk of prostate cancer in the Cancer Prevention Study II Nutrition Cohort. *Cancer Epidemiol. Biomarkers Prev.* 16, 63–69.
- [193] Rosenkrantz, A.B., Sabach, A., Babb, J.S., Matza, B.W., Taneja, S.S., Deng, F.M., 2013. Prostate cancer: comparison of dynamic contrast-enhanced MRI techniques for localization of peripheral zone tumor. *AJR Am J Roentgenol* 201, W471–478.
- [194] Roweis, S.T., Saul, L.K., 2000. Nonlinear dimensionality reduction by locally linear embedding. *Science* 290, 2323–2326. URL: <http://www.sciencemag.org/content/290/5500/2323.abstract>, doi:10.1126/science.290.5500.2323, arXiv:<http://www.sciencemag.org/content/290/5500/2323.full.pdf>.
- [195] Rueckert, D., Sonoda, L.I., Hayes, C., Hill, D.L., Leach, M.O., Hawkes, D.J., 1999. Nonrigid registration using free-form deformations: application to breast MR images. *IEEE Trans Med Imaging* 18, 712–721.
- [196] Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1988. *Neurocomputing: foundations of research*, MIT Press, Cambridge, MA, USA. chapter Learning

- 2446 Internal Representations by Error Propagation, pp. 673–695. URL: <http://dl.acm.org/citation.cfm?id=65669.104449>.  
2447
- 2448 [197] Saeys, Y., Inza, I., Larranaga, P., 2007. A review of fea-  
2449 ture selection techniques in bioinformatics. *Bioinformatics* 23,  
2450 2507–2517. URL: <http://bioinformatics.oxfordjournals.org/content/23/19/2507.abstract>,  
2451 doi:10.1093/bioinformatics/btm344,  
2452 arXiv:<http://bioinformatics.oxfordjournals.org/content/23/19/2507.full.pdf+html>.
- 2453 [198] Scheidler, J., Hricak, H., Vigneron, D.B., Yu, K.K., Sokolov, D.L., Huang, L.R.,  
2454 Zaloudek, C.J., Nelson, S.J., Carroll, P.R., Kurhanewicz, J., 1999a. Prostate  
2455 cancer: localization with three-dimensional proton MR spectroscopic imaging–  
2456 clinicopathologic study. *Radiology* 213, 473–480.
- 2457 [199] Scheidler, J., Petsch, R., Muller-Lisse, U., Heuck, A., Reiser, M., 1999b. Echo-  
2458 planar diffusion-weighted MR imaging of the prostate, in: *Proceedings of the*  
2459 *7th Annual Meeting of ISMRM Philadelphia*, p. 1103.
- 2460 [200] Schlemmer, H.P., Merkle, J., Grobholz, R., Jaeger, T., Michel, M.S., Werner,  
2461 A., Rabe, J., van Kaick, G., 2004. Can pre-operative contrast-enhanced dy-  
2462 namic MR imaging for prostate cancer predict microvessel density in prostate-  
2463 ctomy specimens? *Eur Radiol* 14, 309–317.
- 2464 [201] Schroder, F.H., Carter, H.B., Wolters, T., van den Bergh, R.C., Gosselaar, C.,  
2465 Bangma, C.H., Roobol, M.J., 2008. Early detection of prostate cancer in 2007.  
2466 Part 1: PSA and PSA kinetics. *Eur. Urol.* 53, 468–477.
- 2467 [202] Schröder, F.H., Hugosson, J., Roobol, M.J., Tammela, T.L., Ciatto, S., Nelen,  
2468 V., Kwiatkowski, M., Lujan, M., Lilja, H., Zappa, M., Denis, L.J., Recker, F.,  
2469 Pez, A., Määttänen, L., Bangma, C.H., Aus, G., Carlsson, S., Villers, A., Re-  
2470 billard, X., van der Kwast, T., Kujala, P.M., Blijenberg, B.G., Stenman, U.H.,  
2471 Huber, A., Taari, K., Hakama, M., Moss, S.M., de Koning, H.J., Auvinen, A.,  
2472 2012. Prostate-cancer mortality at 11 years of follow-up. *New England Journal*  
2473 *of Medicine* 366, 981–990.
- 2474 [203] Shapiro, L.G., Stockman, G.C., 2001. *Computer vision*. Prentice Hall, Upper  
2475 Saddle River, NJ. URL: <http://opac.inria.fr/record=b1128947>.
- 2476 [204] Shi, J., Malik, J., 2000. Normalized cuts and image segmentation. *Pattern*  
2477 *Analysis and Machine Intelligence, IEEE Transactions on* 22, 888–905. doi:10.  
2478 1109/34.868688.



- 2479 [205] Shimofusa, R., Fujimoto, H., Akamata, H., Motoori, K., Yamamoto, S., Ueda,  
2480 T., Ito, H., 2005. Diffusion-weighted imaging of prostate cancer. *J Comput*  
2481 *Assist Tomogr* 29, 149–153.
- 2482 [206] Siegel, R., Naishadham, D., Jemal, A., 2013. Cancer statistics, 2013. *CA*  
2483 *Cancer J Clin* 63, 11–30.
- 2484 [207] Sled, J.G., Zijdenbos, A.P., Evans, A.C., 1998. A nonparametric method for  
2485 automatic correction of intensity nonuniformity in MRI data. *IEEE Trans Med*  
2486 *Imaging* 17, 87–97.
- 2487 [208] Sobel, I., 1970. Camera models and machine perception. Technical Report.  
2488 DTIC Document.
- 2489 [209] Somford, D.M., Futterer, J.J., Hambrock, T., Barentsz, J.O., 2008. Diffusion  
2490 and perfusion MR imaging of the prostate. *Magn Reson Imaging Clin N Am*  
2491 16, 685–695.
- 2492 [210] Specht, D.F., 1988. Probabilistic neural networks for classification, mapping,  
2493 or associative memory, in: *Neural Networks, 1988.*, IEEE International Con-  
2494 ference on, pp. 525–532 vol.1. doi:10.1109/ICNN.1988.23887.
- 2495 [211] St Lawrence, K.S., Lee, T.Y., 1998. An adiabatic approximation to the tissue  
2496 homogeneity model for water exchange in the brain: I. Theoretical derivation.  
2497 *J. Cereb. Blood Flow Metab.* 18, 1365–1377.
- 2498 [212] Stamey, T.A., Donaldson, A.N., Yemoto, C.E., McNeal, J.E., Sozen, S., Gill,  
2499 H., 1998. Histological and clinical findings in 896 consecutive prostates treated  
2500 only with radical retropubic prostatectomy: epidemiologic significance of an-  
2501 nual changes. *J. Urol.* 160, 2412–2417.
- 2502 [213] Staring, M., van der Heide, U.A., Klein, S., Viergever, M.A., Pluim, J.P.,  
2503 2009. Registration of cervical MRI using multifeature mutual information.  
2504 *IEEE Trans Med Imaging* 28, 1412–1421.
- 2505 [214] Steinberg, G.D., Carter, B.S., Beaty, T.H., Childs, B., Walsh, P.C., 1990.  
2506 Family history and the risk of prostate cancer. *Prostate* 17, 337–347.
- 2507 [215] Strum, S., Pogliano, D., 2005. What every doctor who treats male patients  
2508 should know. *PCRI Insights* vol. 8, no. 2.

- 2509 [216] Styner, M., Brechbuhler, C., Szckely, G., Gerig, G., 2000. Parametric esti-  
 2510 mate of intensity inhomogeneities applied to MRI. *Medical Imaging, IEEE*  
 2511 *Transactions on* 19, 153–165. doi:10.1109/42.845174.
- 2512 [217] Styner, M., Gerig, G., 1997. Evaluation of 2D/3D bias correction with 1+1ES-  
 2513 optimization. Technical Report. ETH Zürich.
- 2514 [218] Sung, Y.S., Kwon, H.J., Park, B.W., Cho, G., Lee, C.K., Cho, K.S., Kim,  
 2515 J.K., 2011. Prostate cancer detection on dynamic contrast-enhanced MRI:  
 2516 computer-aided diagnosis versus single perfusion parameter maps. *AJR Am J*  
 2517 *Roentgenol* 197, 1122–1129.
- 2518 [219] Suzuki, K., 2012. A review of computer-aided diagnosis in thoracic and colonic  
 2519 imaging. *Quant Imaging Med Surg* 2, 163–176.
- 2520 [220] Swanson, M.G., Vigneron, D.B., Tran, T.K., Sailasuta, N., Hurd, R.E.,  
 2521 Kurhanewicz, J., 2001. Single-voxel oversampled J-resolved spectroscopy of  
 2522 in vivo human prostate tissue. *Magn Reson Med* 45, 973–980.
- 2523 [221] Taira, A.V., Merrick, G.S., Galbreath, R.W., Andreini, H., Taubenslag, W.,  
 2524 Curtis, R., Butler, W.M., Adamovich, E., Wallner, K.E., 2010. Performance of  
 2525 transperineal template-guided mapping biopsy in detecting prostate cancer in  
 2526 the initial and repeat biopsy setting. *Prostate Cancer Prostatic Dis.* 13, 71–77.
- 2527 [222] Tipping, M., 2001. Sparse Bayesian learning and the relevance vector machine.  
 2528 *Journal of Machine Learning Research* 1, 211–244.
- 2529 [223] Tiwari, P., Kurhanewicz, J., Madabhushi, A., 2013. Multi-kernel graph em-  
 2530 bedding for detection, Gleason grading of prostate cancer via MRI/MRS. *Med*  
 2531 *Image Anal* 17, 219–235.
- 2532 [224] Tiwari, P., Kurhanewicz, J., Rosen, M., Madabhushi, A., 2010. Semi super-  
 2533 vised multi kernel (SeSMiK) graph embedding: identifying aggressive prostate  
 2534 cancer via magnetic resonance imaging and spectroscopy. *Med Image Comput*  
 2535 *Comput Assist Interv* 13, 666–673.
- 2536 [225] Tiwari, P., Madabhushi, A., Rosen, M., 2007. A hierarchical unsupervised  
 2537 spectral clustering scheme for detection of prostate cancer from magnetic re-  
 2538 sonance spectroscopy (MRS). *Med Image Comput Comput Assist Interv* 10,  
 2539 278–286.

- 2540 [226] Tiwari, P., Rosen, M., Madabhushi, A., 2008. Consensus-locally linear embed-  
 2541 ding (C-LLE): application to prostate cancer detection on magnetic resonance  
 2542 spectroscopy. *Med Image Comput Comput Assist Interv* 11, 330–338.
- 2543 [227] Tiwari, P., Rosen, M., Madabhushi, A., 2009a. A hierarchical spectral clus-  
 2544 tering and nonlinear dimensionality reduction scheme for detection of prostate  
 2545 cancer from magnetic resonance spectroscopy (MRS). *Med Phys* 36, 3927–3939.
- 2546 [228] Tiwari, P., Rosen, M., Reed, G., Kurhanewicz, J., Madabhushi, A., 2009b.  
 2547 Spectral embedding based probabilistic boosting tree (ScEPTre): classifying  
 2548 high dimensional heterogeneous biomedical data. *Med Image Comput Comput*  
 2549 *Assist Interv* 12, 844–851.
- 2550 [229] Tiwari, P., Viswanath, S., Kurhanewicz, J., Sridhar, A., Madabhushi, A., 2012.  
 2551 Multimodal wavelet embedding representation for data combination (MaW-  
 2552 ERiC): integrating magnetic resonance imaging and spectroscopy for prostate  
 2553 cancer detection. *NMR Biomed* 25, 607–619.
- 2554 [230] Tofts, P., 2010. T1-weighted DCE imaging concepts: modelling, acquisition  
 2555 and analysis, in: *Magneton Flash*. Siemens.
- 2556 [231] Tofts, P.S., 1997. Modeling tracer kinetics in dynamic Gd-DTPA MR imaging.  
 2557 *J Magn Reson Imaging* 7, 91–101.
- 2558 [232] Toth, R., Chappelow, J., Rosen, M., Pungavkar, S., Kalyanpur, A., Madab-  
 2559 hushi, A., 2008. Multi-attribute non-initializing texture reconstruction based  
 2560 active shape model (MANTRA). *Med Image Comput Comput Assist Interv*  
 2561 11, 653–661.
- 2562 [233] Toth, R., Doyle, S., Pungavkar, S., Kalyanpur, A., Madabhushi, A., 2009.  
 2563 A boosted ensemble scheme for accurate landmark detection for active shape  
 2564 models, in: *SPIE Medical Imaging*, Orlando, FL.
- 2565 [234] Tu, Z., 2005. Probabilistic boosting-tree: learning discriminative models for  
 2566 classification, recognition, and clustering, in: *Computer Vision, 2005. ICCV*  
 2567 *2005. Tenth IEEE International Conference on*, pp. 1589–1596 Vol. 2. doi:10.  
 2568 1109/ICCV.2005.194.
- 2569 [235] Turkbey, B., Choyke, P.L., 2012. Multiparametric MRI and prostate cancer  
 2570 diagnosis and risk stratification. *Curr Opin Urol* 22, 310–315.

- 2571 [236] Vanhamme, L., van den Boogaart, A., Van Huffel, S., 1997. Improved method  
2572 for accurate and efficient quantification of MRS data with use of prior knowl-  
2573 edge. *J. Magn. Reson.* 129, 35–45.
- 2574 [237] Vapnik, V., Lerner, A., 1963. Pattern Recognition using Generalized Portrait  
2575 Method. *Automation and Remote Control* 24.
- 2576 [238] Verma, S., Rajesh, A., Futterer, J.J., Turkbey, B., Scheenen, T.W., Pang, Y.,  
2577 Choyke, P.L., Kurhanewicz, J., 2010. Prostate MRI and 3D MR spectroscopy:  
2578 how we do it. *AJR Am J Roentgenol* 194, 1414–1426.
- 2579 [239] Verma, S., Turkbey, B., Muradyan, N., Rajesh, A., Cornud, F., Haider, M.A.,  
2580 Choyke, P.L., Harisinghani, M., 2012. Overview of dynamic contrast-enhanced  
2581 MRI in prostate cancer diagnosis and management. *AJR Am J Roentgenol*  
2582 198, 1277–1288.
- 2583 [240] Villers, A., Steg, A., Boccon-Gibod, L., 1991. Anatomy of the prostate: review  
2584 of the different models. *Eur. Urol.* 20, 261–268.
- 2585 [241] Viola, P., Wells, III, W.M., 1997. Alignment by maximization of mutual infor-  
2586 mation. *Int. J. Comput. Vision* 24, 137–154.
- 2587 [242] Viswanath, S., Bloch, B.N., Chappelow, J., Patel, P., Rofsky, N., Lenkinski,  
2588 R., Genega, E., Madabhushi, A., 2011. Enhanced multi-protocol analysis via  
2589 intelligent supervised embedding (EMPrAvISE): detecting prostate cancer on  
2590 multi-parametric MRI , 79630U–79630U–15URL: [+http://dx.doi.org/10.](http://dx.doi.org/10.1117/12.878312)  
2591 [1117/12.878312](http://dx.doi.org/10.1117/12.878312), doi:10.1117/12.878312.
- 2592 [243] Viswanath, S., Bloch, B.N., Genega, E., Rofsky, N., Lenkinski, R., Chappelow,  
2593 J., Toth, R., Madabhushi, A., 2008a. A comprehensive segmentation, registra-  
2594 tion, and cancer detection scheme on 3 Tesla in vivo prostate DCE-MRI. *Med*  
2595 *Image Comput Comput Assist Interv* 11, 662–669.
- 2596 [244] Viswanath, S., Bloch, B.N., Rosen, M., Chappelow, J., Toth, R., Rofsky, N.,  
2597 Lenkinski, R., Genega, E., Kalyanpur, A., Madabhushi, A., 2009. Integrating  
2598 structural and functional imaging for computer assisted detection of prostate  
2599 cancer on multi-protocol in vivo 3 Tesla MRI, in: *Society of Photo-Optical*  
2600 *Instrumentation Engineers (SPIE) Conference Series*.
- 2601 [245] Viswanath, S., Tiwari, P., Rosen, M., Madabhushi, A., 2008b. A meta-classifier  
2602 for detecting prostate cancer by quantitative integration of *In Vivo* magnetic

2603 resonance spectroscopy and magnetic resonance imaging, in: Medical Imaging  
2604 2008: Computer-Aided Diagnosis, SPIE.

2605 [246] Viswanath, S.E., Bloch, N.B., Chappelow, J.C., Toth, R., Rofsky, N.M.,  
2606 Genega, E.M., Lenkinski, R.E., Madabhushi, A., 2012. Central gland and  
2607 peripheral zone prostate tumors have significantly different quantitative imag-  
2608 ing signatures on 3 Tesla endorectal, in vivo T2-weighted MR imagery. *J Magn*  
2609 *Reson Imaging* 36, 213–224.

2610 [247] Vos, P.C., Barentsz, J.O., Karssemeijer, N., Huisman, H.J., 2012. Automatic  
2611 computer-aided detection of prostate cancer based on multiparametric mag-  
2612 netic resonance image analysis. *Phys Med Biol* 57, 1527–1542.

2613 [248] Vos, P.C., Hambrock, T., Barentsz, J.O., Huisman, H.J., 2008a. Combining T2-  
2614 weighted with dynamic MR images for computerized classification of prostate  
2615 lesions, in: Medical Imaging 2008: Computer-Aided Diagnosis, SPIE.

2616 [249] Vos, P.C., Hambrock, T., Barentsz, J.O., Huisman, H.J., 2010. Computer-  
2617 assisted analysis of peripheral zone prostate lesions using T2-weighted and  
2618 dynamic contrast enhanced T1-weighted MRI. *Phys Med Biol* 55, 1719–1734.

2619 [250] Vos, P.C., Hambrock, T., Hulsbergen-van de Kaa, C.A., Futterer, J.J., Bar-  
2620 entsz, J.O., Huisman, H.J., 2008b. Computerized analysis of prostate lesions  
2621 in the peripheral zone using dynamic contrast enhanced MRI. *Med Phys* 35,  
2622 888–899.

2623 [251] Vovk, U., Pernus, F., Likar, B., 2007. A review of methods for correction of  
2624 intensity inhomogeneity in MRI. *Medical Imaging, IEEE Transactions on* 26,  
2625 405–421. doi:10.1109/TMI.2006.891486.

2626 [252] Walker, P., Crehange, G., Parfait, S., Cochet, A., Maignon, P., Cormier, L.,  
2627 Brunotte, F., 2010. Absolute quantification in 1H MRSI of the prostate at 3T,  
2628 in: ISMRM Annual Meeting 2010.

2629 [253] Wang, L., Mazaheri, Y., Zhang, J., Ishill, N.M., Kuroiwa, K., Hricak, H.,  
2630 2008. Assessment of biologic aggressiveness of prostate cancer: correlation of  
2631 MR signal intensity with Gleason grade after radical prostatectomy. *Radiology*  
2632 246, 168–176.

2633 [254] Warfield, S.K., Zou, K.H., Wells, W.M., 2004. Simultaneous truth and per-  
2634 formance level estimation (STAPLE): an algorithm for the validation of image  
2635 segmentation. *IEEE Trans Med Imaging* 23, 903–921.

- 2636 [255] Wiart, M., Curiel, L., Gelet, A., Lyonnet, D., Chapelon, J.Y., Rouviere, O.,  
2637 2007. Influence of perfusion on high-intensity focused ultrasound prostate ab-  
2638 lation: a first-pass MRI study. *Magn Reson Med* 58, 119–127.
- 2639 [256] Ye, L., Kynaston, H.G., Jiang, W.G., 2007. Bone metastasis in prostate cancer:  
2640 molecular and cellular mechanisms (Review). *Int. J. Mol. Med.* 20, 103–111.
- 2641 [257] Zelhof, B., Lowry, M., Rodrigues, G., Kraus, S., Turnbull, L., 2009. Description  
2642 of magnetic resonance imaging-derived enhancement variables in pathologically  
2643 confirmed prostate cancer and normal peripheral zone regions. *BJU Int.* 104,  
2644 621–627.
- 2645 [258] Zhao, G., Ahonen, T., Matas, J., Pietikainen, M., 2012. Rotation-Invariant Im-  
2646 age and Video Description With Local Binary Pattern Features. *Image Process-  
2647 ing, IEEE Transactions on* 21, 1465–1477. doi:10.1109/TIP.2011.2175739.
- 2648 [259] Zhu, H., Ouwerkerk, R., Barker, P.B., 2010. Dual-band water and lipid suppres-  
2649 sion for MR spectroscopic imaging at 3 Tesla. *Magn Reson Med* 63, 1486–1492.
- 2650 [260] Zitová, B., Flusser, J., 2003. Image registration methods: a sur-  
2651 vey. *Image and Vision Computing* 21, 977 – 1000. URL: [http://www.  
2652 sciencedirect.com/science/article/pii/S0262885603001379](http://www.sciencedirect.com/science/article/pii/S0262885603001379), doi:[http:  
2653 //dx.doi.org/10.1016/S0262-8856\(03\)00137-9](http://dx.doi.org/10.1016/S0262-8856(03)00137-9).