



Available online at [www.sciencedirect.com](http://www.sciencedirect.com)

ScienceDirect

journal homepage: [www.elsevier.com/locate/bbe](http://www.elsevier.com/locate/bbe)



Original Research Article

# Multi-sequence texture analysis in classification of *in vivo* MR images of the prostate<sup>☆</sup>

Dorota Duda<sup>a,\*</sup>, Marek Kretowski<sup>a</sup>, Romain Mathieu<sup>b</sup>,  
Renaud de Crevoisier<sup>c</sup>, Johanne Bezy-Wendling<sup>c</sup>

<sup>a</sup> Faculty of Computer Science, Bialystok University of Technology, Bialystok, Poland

<sup>b</sup> Department of Urology, Pontchaillou University Hospital, Rennes, France

<sup>c</sup> LTSI, INSERM U1099, University of Rennes 1, France

ARTICLE INFO

Article history:

Received 27 January 2016

Received in revised form

15 May 2016

Accepted 22 May 2016

Available online 31 May 2016

Keywords:

Computer-aided diagnosis

Tissue characterization

Multi-image texture analysis

Feature extraction

Feature selection

Classification

ABSTRACT

The aim of the study is to investigate the potential of multi-sequence texture analysis in the characterization of prostatic tissues from *in vivo* Magnetic Resonance Images (MRI). The approach consists in simultaneous analysis of several images, each acquired under different conditions, but representing the same part of the organ. First, the texture of each image is characterized independently of the others. Then the feature values corresponding to different acquisition conditions are combined in one vector, characterizing a combination of textures derived from several sequences. Three MRI sequences are considered: T1-weighted, T2-weighted, and diffusion-weighted. Their textures are characterized using six methods (statistical and model-based). In total, 30 tissue descriptors are calculated for each sequence. The feature space is reduced using a modified Monte Carlo feature selection, combined with wrapper methods, and Principal Components Analysis.

Six classifiers were used in the work. Multi-sequence texture analysis led to better classification results than single-sequence analysis. The subsets of features selected with the Monte Carlo method guaranteed the highest classification accuracies.

© 2016 Nałęcz Institute of Biocybernetics and Biomedical Engineering of the Polish Academy of Sciences. Published by Elsevier Sp. z o.o. All rights reserved.

## 1. Introduction

According to the latest “Global Cancer Statistics” [2], prostate cancer is the second most frequently diagnosed cancer in men and the fifth leading cause of cancer-related death worldwide. In more developed countries it is even the second most frequent cause of male cancer death (behind only lung

cancer). More than 1.1 million cases of prostate cancer were recorded around the world in 2012 [3]. The American Cancer Society estimated about 220,800 new cases of prostate cancer in the United States alone in 2015 [4]. In this context, the search for methods enabling the earliest possible detection of prostate pathology and determination of its type (benign or malignant) is crucial for reducing prostate cancer mortality rates.

<sup>☆</sup> This is an extended and revised version of a paper presented at the ITiB'2014 Conference [1].

\* Corresponding author at: Faculty of Computer Science, Bialystok University of Technology, Wiejska 45A, 15-351 Bialystok, Poland.  
E-mail addresses: [d.duda@pb.edu.pl](mailto:d.duda@pb.edu.pl) (D. Duda), [m.kretowski@pb.edu.pl](mailto:m.kretowski@pb.edu.pl) (M. Kretowski).

<http://dx.doi.org/10.1016/j.bbe.2016.05.002>

0208-5216/© 2016 Nałęcz Institute of Biocybernetics and Biomedical Engineering of the Polish Academy of Sciences. Published by Elsevier Sp. z o.o. All rights reserved.

Diagnostic tools for prostate cancers [5] currently include prostate-specific antigen (PSA) serum screening, trans-rectal ultrasound imaging (TRUS), needle biopsies, or an ensemble of Magnetic Resonance Imaging (MRI) techniques enabling visualization of different prostatic tissue properties. However, the first three of these tools have many deficiencies and their use remains under discussion. For example, a 10-year experiment conducted by Andriole et al. on 76,693 men [6] revealed no significant benefits of screening for prostate cancer with PSA serum testing. According to another report, prostate cancer screening could in some cases lead to over-treatment [7]. TRUS techniques do not always differentiate between normal and cancerous tissues [8]. Furthermore, the use of needle biopsy, which is the current standard when high PSA values have been detected, carries a risk of serious complications, including prolonged bleeding from the biopsy site, difficulty urinating, or severe infections requiring hospitalization. A study on 51,321 men, conducted by Lundstrom et al. [9], showed that hospitalizations due to biopsy-related complications were on the rise. Also, a biopsy may miss a tumor if the needle is not placed in the right location.

In view of the above, great hope may be placed in correct interpretation of MRI prostate images, especially as their acquisition is not overly invasive or harmful to health. However, correct recognition of image content may be beyond the capacity of a non-equipped physician. It is therefore essential to develop appropriate tools for computer-aided diagnosis (CAD) [10].

The aim of the present study is to validate methods for texture-based analysis of MRI prostate images and to assess their usefulness in prostatic tissue classification. Particular attention is devoted here to a multi-image approach to texture characterization. It consists in simultaneous analysis of textures from several images, each acquired under different conditions and thus enhancing a different tissue property. All the textures considered together refer to the same prostate slice. In the case where each of the textures considered simultaneously corresponds to a different image sequence, we refer to this as a multi-sequence texture analysis. To the best of our knowledge, there are very few works on prostate tumor recognition based on multi-image and/or multi-sequence texture analysis of MR images. However, a few studies employing such a technique have been conducted with regard to other organs and other imaging modalities. These have mainly concerned hepatic tissue recognition from dynamic contrast-enhanced (DCE) Computed Tomography (CT) or breast tumor characterization based on a combination of DCE and T2W MRI (for details, please refer to [11,12]). Such works have already shown that a multi-image and multi-sequence approach may outperform the single-sequence and single-image techniques. The latter imply consideration of only one type of sequence and consist in finding tissue descriptors based on a texture from only one image at a time. They are still frequently applied in image-based tissue differentiation.

In the work, we explore three MRI sequences: fat-suppressed contrast-enhanced T1-weighted, T2-weighted, and diffusion-weighted (abbreviated T1W, T2W, and DW, respectively). Each pair of sequences (“T1W & T2W”, “T1W & DW”, and “T2W & DW”) is considered in a two-sequence

texture analysis. Analysis of triples of textures, composed of textures from the T1W, T2W and DW sequences, is performed as well. The classification results obtained with the multi-sequence texture analysis are compared to those obtained by the classic single-sequence and single-image approaches.

Dealing with many image sequences at a time substantially increases the number of texture characteristics that may be considered. Using too many features could make the classification algorithms less effective, the computation time too long, and the memory requirements too high. An important part of our work is therefore to test different approaches to reducing the feature space, to assess the relative importance of each feature, and to determine which subsets of features perform the best. Here, we use a slightly modified Monte Carlo (MC) feature selection, initially proposed by Draminski et al. [13], combined with wrapper methods [14] for feature subset evaluations. A second approach involves classical Principal Component Analysis (PCA), and is tested as a baseline.

The remaining sections are organized as follows. Section 2 gives an overview of prostatic tissue recognition from *in vivo* MR images. The next section is devoted to presentation of the methods used in our approach: the proposed multi-sequence texture characterization and the reduction of the feature space. Section 4 details the setup for experimental validation. This is followed by presentation of the results and discussion (Section 5). Finally, in Section 6, conclusions are drawn and possible future work is outlined.

---

## 2. Related work

Systems adapted for (semi-)automatic recognition of prostate disorders from MR images are still not widely used. Moreover, many of them do not exploit information about image texture. They are often based mainly on pharmacokinetic models, employing signal-to-time curves, to find perfusion parameters (see, for example, [15]). Such models provide some information about the propagation of the contrast product, extracted from T1-weighted DCE MRI sequences. It should be noted, however, that many works presented in the last decade have shown that detection of prostate tumors could be improved if several types of MR image sequences were analyzed simultaneously, rather than considered one sequence at a time. The remainder of this section will summarize previous studies devoted to (semi-) automatic analysis of MR prostate images.

### 2.1. Tissue recognition not based on texture

Several works comparing the potential of automated multi-sequence analysis to single-sequence analysis can be cited. The earliest of them do not consider textures. For example, Langer et al. [16] identified cancers in the peripheral zone (PZ) of the prostate using T2W imaging, DW imaging, T2-mapping, and DCE MRI. In their work, four tissue descriptors were evaluated (ADC, T2, volume transfer constant –  $K^{trans}$ , and extravascular extracellular volume fraction –  $v_e$ ), with step-wise Logistic Regression (LR) used as a classifier. The optimal multi-sequence model significantly outperformed the most effective single-sequence model, based on the ADC.

Further, Vos et al. [17] combined information from only two MRI modalities (T2W and DCE T1W) to characterize PZ prostate lesions. Their feature vectors, which contained T2 estimators and pharmacokinetic parameters, were fed into a Support Vector Machines (SVM) classifier [18]. The combination of T2W and DCE parameters significantly improved the discriminating performance of the system in comparison to that achieved with T2 values alone.

Another system (Artan et al. [19]) proposed for automated localization of prostate cancers used a segmentation method based on Conditional Random Fields and cost-sensitive SVM. Three types of images were finally considered simultaneously in further analyses: T2 maps, ADC maps, and  $k_{ep}$  (derived from DCE MRI). The study showed again that using several image types at a time significantly improves tumor localization in comparison with single-image analysis.

## 2.2. Texture analysis on T2W images only

A different approach to prostatic tissue characterization was presented in [20,21]. These studies considered only textural properties of T2W image regions (single-image and single-sequence analysis), and were shown to ensure satisfactory classification accuracies. For example, Lopes et al. [20] investigated the potential of fractal (and multifractal) textural features. The fractal dimension was calculated using the variance method, while the multifractal spectrum was estimated with a modified Brownian motion model. Finally, two classification algorithms were used to differentiate between tumorous and healthy tissue: SVM and AdaBoost (an adaptive boosting voting scheme [22]). The fractal-based features performed better than classical textural features based on co-occurrence matrices (COM), wavelets, or Gabor filters (all described in the work).

Recently, Ginsburg et al. [21] attempted to describe prostate cancer morphology using several groups of T2W MRI-based textural features. The aim of their work was to predict the probability of biochemical recurrence within 10 years of radiation therapy in sixteen patients. The features evaluated were based on first-order statistics, co-occurrence matrices, gradient operators, and Gabor wavelet decomposition [23]. Three of the features obtained with the last of these methods proved to be the best predictors of biochemical recurrence risk when LR was used as a classifier.

Some other studies have combined textural features with functional, kinetic, and/or anatomic tumor descriptors derived from multi-sequence (multispectral) MRI. However, in all of these studies texture analysis was still carried out only for T2W MR images. Other image sequences (such as DCE-MRI or DW) were used to extract non-texture-related parameters. In the work of Viswanath et al. [24], aiming at the detection of prostate cancers, functional and structural data derived from MR images were integrated. Their system performed prostate segmentation, feature extraction, and classification. DCE MRI series served here to calculate functional tissue attributes (the wash-in and wash-out of the contrast agent). The set of structural attributes (textural features) comprised standard deviation (a first-order statistical feature), edge detectors, intensity average, and second-order textural features. Random Forest [25] (RF) was used as a classifier. Experiments on a small

data set showed that integration of T2W structural textural information and functional data significantly improves prostate cancer detection.

Litjens et al. [26] developed a fully automated CAD system which was able to differentiate between patients with and without prostate cancer. Their study considered T2W, DW, DCE, and proton density-weighted (PDW) images. Thus several types of features were used in the study: intensity, pharmacokinetic, anatomical, blobness, and texture features. Texture features were derived from Gaussian textural models. The system used three classifiers: Linear Discriminant Analysis (LDA), Gentle Boost with regression stumps as weak learners, and RF. The study did not compare the results obtained with each separate group of features. However, textural features were included in the final set of selected features, ensuring the highest system performance (comparable to that achieved by radiologists).

Recently, Molina et al. [27] proposed a system that also integrated different features (anatomic, textural, and functional) in order to recognize three classes of prostatic tissue from MR images (cancerous, unhealthy but non-cancerous, and healthy). Three types of MR images were considered in the work: T2W, Dynamic-Contrast Enhanced Plasma Flow (DCE-PF) and DCE Mean Transit Time (DCE-MTT). Texture information was still extracted only from the T2W images. Experiments on a small dataset showed that the use of texture descriptors provided more relevant discriminatory information than functional parameters alone.

## 2.3. Multi-sequence texture analysis

Very few studies have used texture analysis for more than one type of MR prostate images. The earliest of these was presented by Chan et al. [28]. In their automated system for detection of PZ prostate cancers, they combined information from three different MR image types: T2W, T2-mapping, and Line Scan Diffusion Imaging (LSDI). These three image sequences were used to calculate four different sets of image intensities: T2W, ADC from LSDI, PD, and T2 Map from T2-mapping imaging. Five different feature sets were evaluated: the first included only the four signal intensities (T2W, ADC, PD and T2 Map), the second comprised four signal intensities and three anatomical features, and the next three were additionally extended with co-occurrence matrix-based textural features, textural features based on Discrete Cosine Transform (DCT), and both these types of textural features. A maximum likelihood classifier was used for single-sequence image classification, and SVM and LDA for the combination of three sequences. The classification performance was evaluated by Receiver Operating Characteristic (ROC) analysis [29]. The highest area under the ROC curve for the LDA classifier corresponded to application of the largest considered set of multi-sequence features (including COM and DCT textural features). For the SVM classifier, convergence for the three largest sets of features was not achieved, but the combination of signal intensities and anatomical features gave the best results. In turn, the best single-sequence result was obtained with the T2W sequence.

Finally, Niaf et al. [30] simultaneously analyzed three different MRI sequences (DCE-MRI, T2W, and DW images)

to discriminate between (i) malignant and benign prostatic tissues, and (ii) malignant and nonmalignant but suspicious ones. The CAD system proposed in the study combined functional parameters extracted from DCE images, together with textural features derived from the three considered sequences. In total, 140 tissue descriptors were evaluated. Four filter methods for feature selection were tested and four classifiers were applied: SVM, LDA,  $k$ -Nearest Neighbors ( $k$ -NN), and naïve Bayes. The study was conducted on a dataset of images derived from 49 patients. The system's best diagnostic performance, assessed by the area under the ROC curve, was 0.89 and 0.82, for the first (i) and second (ii) discrimination problem, respectively. No single-sequence texture analysis methods were tested as a baseline.

To the best of our knowledge, there are no other CAD systems combining texture characteristics corresponding to different MRI sequences in order to characterize prostatic tissue in the classification process. The cited work concerning this problem has shown that texture analysis can be a good tool for prostatic tissue characterization, even if it is applied to only one type of MRI sequences. Furthermore, differentiation of prostatic tissues proved to be better when features (not necessarily textural) corresponding to different image modalities were considered simultaneously. The aim of our study, therefore, is to assess the utility of a multi-sequence texture analysis in characterization of prostatic tissue from MRI, and to compare its potential to the potential of single-sequence analysis. Three image sequences (T1W, T2W, and DW) are considered in our work. Moreover, as using multiple sequences results in a large number of possible tissue descriptors, feature selection is performed in order to assess which textural features and which MRI sequences contribute the most to prostatic tissue recognition. In contrast with Niaf's approach [30], we will focus on wrapper-based (supervised) methods for evaluation of each tested subset of candidate features. With this type of method, the usefulness of a given subset of features is assessed by the quality of the classification which the subset can guarantee.

### 3. Methods

#### 3.1. Tissue classification based on texture analysis

Two stages of work of a typical, image-based CAD system can be distinguished [10]. The first, called *training* (or *learning*), consists in preparation of the system for the recognition of several predefined tissue classes. In practice, this means constructing classifiers from a database of images visualizing the organ under consideration. All the images used at this stage should represent only verified cases, that is, those for which the most accurate possible diagnosis has been made. A definitive confirmation of a diagnosis often involves the use of invasive methods. The second stage is application of the system in order to aid less invasive (or even non-invasive) diagnosis.

Our system also follows this two-stage design. What distinguishes our system from many others is that  $n$  images representing the same tissue slice can be combined in  $n$ -tuples and analyzed simultaneously. Each of the images in an  $n$ -tuple

is usually acquired under different conditions (e.g. different scanner settings or different concentrations of the contrast medium), so it is derived from a different image sequence.

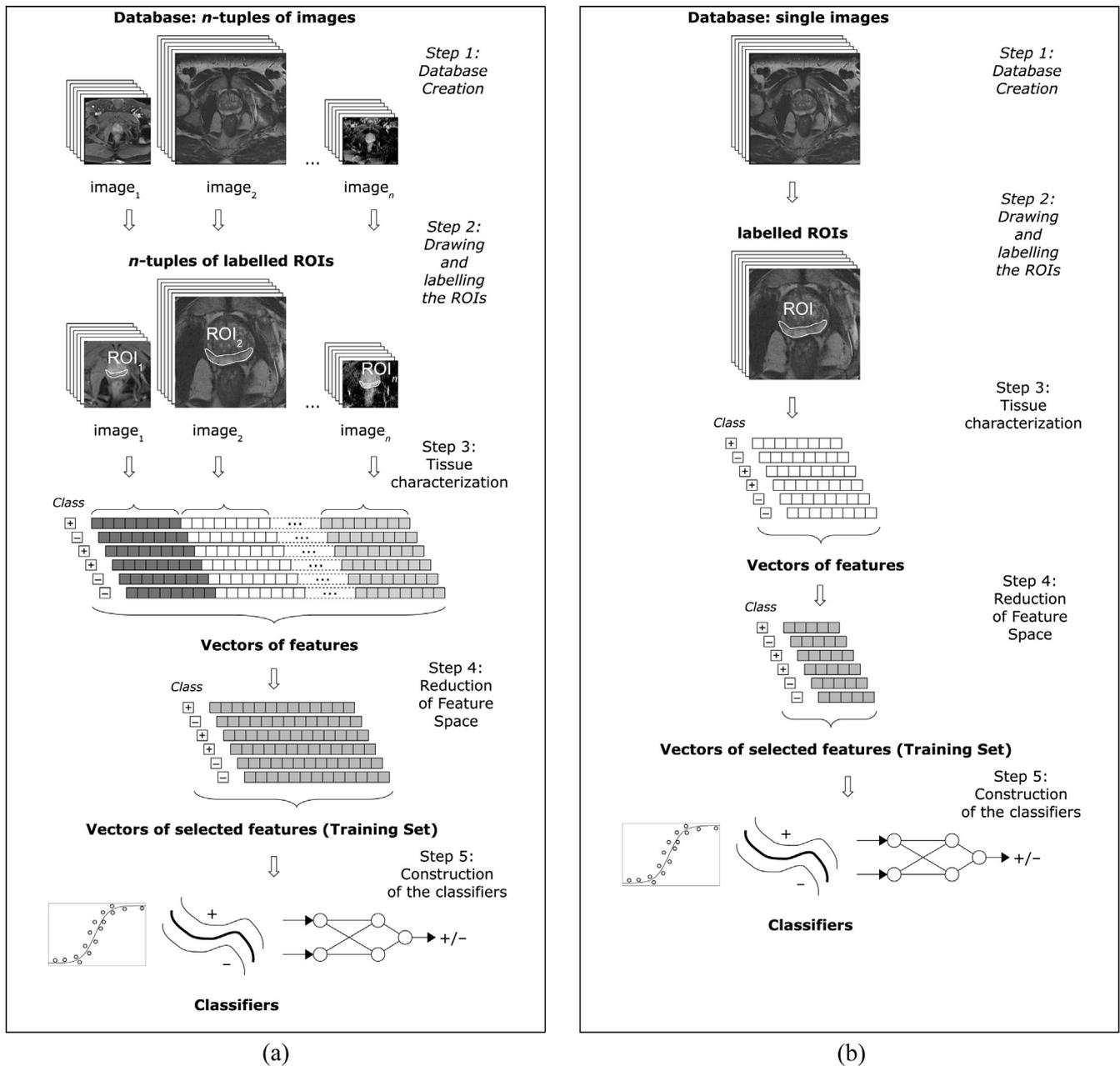
The first stage of work of the system based on multi-image texture analysis is presented in Fig. 1(a). For comparison, part (b) of the figure shows a typical training scheme based on single-image texture analysis, used in a broad range of CAD systems. In our system, after a database has been created,  $n$ -tuples of images are formed. Depending on the number of image sequences considered, the  $n$ -tuple can comprise two or more images. The order of images in each  $n$ -tuple is fixed. For example, a triple of MRI prostate images might contain a T1W, T2W, and DW image in the first, second, and third position, respectively. These initial steps are often followed by pre-processing of the images, used to improve contrast, eliminate noise or artifacts, or equalize ranges of pixel values corresponding to different studies (as in our database).

The next step is to outline image regions for analysis, called *Regions of Interest* (ROIs). On each of the images forming the  $n$ -tuple an ROI covering the same part of the organ is outlined. This can be accomplished using (semi-)automated methods incorporated in the system. ROIs can also be delineated manually. The  $n$ -tuple of corresponding ROIs is analyzed simultaneously in order to characterize the same part of the tissue. Tissue characterization consists in calculating different numerical parameters providing information about the analyzed ROIs. Very often such parameters are obtained based on texture analysis and hence are called *texture features*. They may provide information about texture coarseness, contrast, or complexity. They also allow conclusions to be drawn about the presence, frequency, and size of different texture elements, such as beams, strips, waves, edges, or spots. These texture properties may be associated with pathological processes developing in the tissue.

Characterization of texture from multi-sequence images runs in two phases. First, the same set of textural features is calculated for each ROI in the  $n$ -tuple. Next, features corresponding to different images in the  $n$ -tuple are combined in one vector, characterizing textures from several sequences. In the simplest case, such a vector is formed by concatenation of  $n$  sets of features, where each set corresponds to different sequence. Multi-sequence vectors can also be composed of parameters which are functions of several feature values obtained with the same method and corresponding to different sequences.

The definition of ROIs and characterization of their textures often requires cooperation with a physician, who specifies the tissue class (label) for each  $n$ -tuple of ROIs. The tissue class is then attributed (separately) to each vector of multi-sequence features. The label reflects the pathology affecting the organ under consideration and is determined on the basis of a verified diagnosis. The "ground truth" can be provided here, e.g. by histological analysis. Labeled feature vectors form what is known as the *training* (or *learning*) set. On the basis of this training set one or more classifiers can be constructed.

Many different texture analysis methods have been proposed [31,32]. Not all of the applicable features are equally useful in the texture-based classification process. Some can prove to be redundant, irrelevant, or noisy. The use of too many features entails the need for large computer resources,

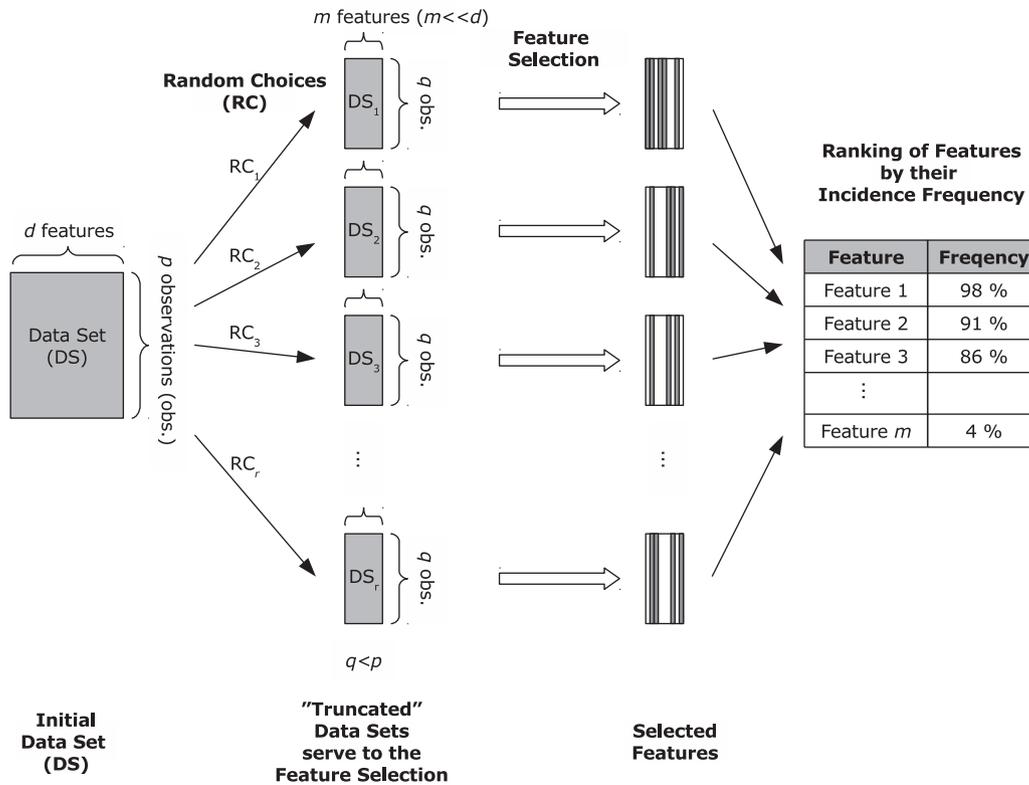


**Fig. 1 – A system for tissue classification based on (a) multi-image texture analysis or (b) single-image texture analysis. The first stage of work: the construction of classifiers from (a) a database of image  $n$ -tuples and (b) a database of single images.**

can significantly lengthen the time of calculation, and reduces the efficiency of the CAD system. An important task is therefore feature selection. This usually takes place before the construction of classifiers, or it can be embedded into the process of classifier training. It makes it possible to find the most relevant features and to reject redundant or inefficient ones, as well as to reduce the memory and computation time required for subsequent processing steps.

Once the classifiers are constructed, the system can be applied to identify new, as yet undiagnosed cases (the second stage of system work). At this stage, an  $n$ -tuple of images

representing the same part of the organ is necessary. The order of the sequences from which subsequent images are derived is the same as in the first stage of the system work. The image pre-processing and texture feature extraction techniques also remain the same. Outlining of the ROI on each of the images composing the  $n$ -tuple is followed by extraction of textural features for each ROI. Next, a multi-sequence vector characterizing the  $n$ -tuple of textures is created. If feature selection was applied in the first stage, only the features selected are used here. Finally, the classifiers available in the system are applied and the most probable tissue class is indicated.



**Fig. 2 – A modified Monte Carlo feature selection method, used to assess the relative importance of each feature in the tissue recognition process.**

### 3.2. Reduction of feature space

The number of potential texture descriptors to be used in our experiments is rather large in comparison with the number of available ROIs. This could be particularly problematic for a multi-image texture analysis, employing feature sets two or three times larger than in the case of a single-image analysis. To avoid the risk of overfitting [33], it is therefore necessary to reduce the feature space. In our work we use two approaches to reduction of the feature space.

In the first approach, we use a slightly modified Monte Carlo feature selection method, initially proposed by Draminski et al. [13]. A schematic representation of our modified version of this method is shown in Fig. 2. Here, the initial number of observations (objects) available in the data set and the initial number of features characterizing each observation are denoted by  $p$  and  $d$ , respectively. In our experiments, first a small number ( $m$ ) of all the considered  $d$  features ( $m \ll d$ ) and a certain number ( $q$ ) of observations are chosen. Then, these truncated data serve as input for feature selection. The feature selection procedure is repeated  $r$  times, each time for different, randomly chosen data, where  $r$  has a relatively large value, e.g. hundreds of thousands. Finally, based on multiple repetitions of the selection experiment, the feature incidence frequency rate (the percentage of cases in which the feature was selected with respect to the number of its occurrences in the input data sets subjected to selection) is counted and features are ranked according to their frequency rates. The most frequently

selected features are top-ranked, and these are considered the most useful in the further tissue identification process.

By taking into account a portion of the whole set of observations and repeating the feature selection procedure many times, each time with a different subset of randomly chosen observations, we obtain more general results. Reduction of the number of features before making a selection is determined by two factors. Firstly, the procedure could be excessively time-consuming when dealing with the entire set of features in each of  $r$  feature selection experiments. Secondly, we are interested in which features contribute the most to correct classification results and in the relative importance of each feature. This cannot be estimated by analyzing the entire set of features, because some features may not prove informative separately, but only when combined with other features [34]. In our work, different combinations of features are already considered before the actual selection procedures begin.

In the second approach, classical Principal Components Analysis is used to reduce the feature space. Here, the new features (principal components) are different linear combinations of features constituting the input vectors. Thus, each new feature contains a portion of information about the initial textural characteristics. Since we only need a certain number of first principal components to describe the variability of objects, PCA results in reduction of the feature space. Finally, we analyze how the classification quality changes when fewer and fewer important principal components are added to a new feature set.

**Table 1 – Key details about the images and the ROIs used in our experiments.**

	Base 1 T1W images	Base 1 T2W images	Base 1 DW images	Base 2 T2W images
Number of considered images:	84	78	54	42
Image size in pixels:	192 × 192	320 × 320 (16 studies) or 448 × 448 (3 studies)	160 × 136	384 × 384
Slice thickness:	3 mm	3 mm	6 mm	3 mm
Number of delineated ROIs:	107	100	68	80
– for tumorous tissue:	47	45	31	80
– for healthy tissue:	60	55	37	0
Average ROI size in pixels:	88	456	85	425

#### 4. Experimental setup

The aim of the experiments was to determine the usefulness of multi-image texture analysis in the recognition of prostatic tissues from MR images. It was assessed on the basis of the classification accuracy obtained when different sets of features were utilized as texture descriptors. Two tissue classes were differentiated: cancerous and healthy. As the number of potential textural features was quite large, particular emphasis was placed here on reduction of the feature space. Two approaches were used to find the smallest possible set of features providing the best possible classification results: a modified MC feature selection method and PCA.

Two groups of experiments were performed. During the first, only single-image and single-sequence texture analysis was carried out. In this case each sequence of images (T1W, T2W, and DW) was considered separately, independently of the others. The second group of experiments concerned multi-image and multi-sequence texture characterization. Here, pairs or triples of images representing the same prostatic slice, but derived from different image sequences, were analyzed simultaneously. The best results obtained with the multi-sequence texture analysis were compared to the corresponding best results achieved with the single-sequence analysis.

##### 4.1. Database description

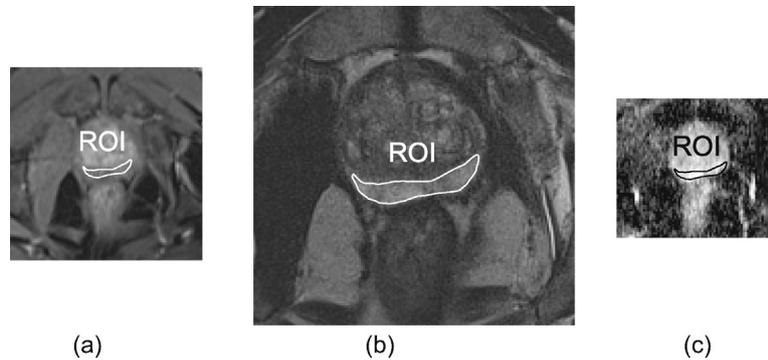
Two databases of images were available for the experimental validation. The first (“Base 1”) contained an image series acquired at Pontchaillou University Hospital in Rennes, France, between August 2009 and April 2010. Here, studies from 19 patients were considered. The second database (“Base 2”) included images gathered between January and July 2013, at Pontchaillou University Hospital and at the Eugene Marquis Anticancer Center, also in Rennes, France. In the latter case only 10 studies were available. All examinations were carried out with the approval of the local human ethics committee, as part of a routine clinical diagnostic procedure.

In the case of both databases, the images were acquired with a 3T Siemens Verio MR scanner and were initially recorded in DICOM format. Each acquisition contained at least 3 image sequences: fat-suppressed T1W, T2W, and DW. The T1W image series were contrast-enhanced with the gadolinium-based contrast agent Dotarem®. Dotarem® was intravenously injected at a standard rate of 2 ml/s. Its amount varied from 10 ml to 20 ml, depending on the patient (it was of 0.2 ml/kg of body weight). Images were captured in 30 (or sometimes

even 40) moments of contrast product propagation. For the majority of the T1W series, the ROIs corresponded to only one tissue class. Therefore it was impossible to determine separately for each patient at which moment of contrast propagation the texture characteristics for healthy and tumorous tissue would have differed most significantly. For this reason we decided to always analyze the image corresponding to the “middle” contrast-agent propagation moment (the 15th).

The key information about the images and the defined ROIs from two databases is summarized in Table 1. As regards the first database, the three aforementioned image sequences were considered and the ROIs were outlined for two tissue types, tumorous and healthy. The ROI delineation and labeling was carried out by Romain Mathieu, hospital physician (urology doctor). In most cases it was only possible to consider one tissue class (either tumorous or healthy) per study. Due to differences in slice thicknesses between DW and the other types of images (T1W and T2W), it was not possible to find a DW image with a slice location matching each of the T1W and/or T2W images. In total, we managed to create 48 image triples, composed of images acquired at exactly the same slice position. Then, 60 triples of ROIs (in total) were defined on them. The ROI tracing procedure was as follows. First, a ROI was manually traced by the physician on the DW image. Then the traced ROI was copied onto the two remaining images in the triple. Because the resolutions of the three corresponding images were different, the sizes of the copied ROIs were modified so that they covered the same part of the prostate. Some “single” ROIs were also defined for T1W or/and T2W images with no counterparts in the DW image. These were considered (additionally) in the experiments involving single-image texture characterization. For the second database, only the ROIs in the T2W images were provided to us, and these corresponded only to tumorous tissue. Here, it was Renaud de Crevoisier (professor, cancer physician and radiotherapist) that delineated the ROIs. All the ROIs from both databases concerned only PZ prostate. As for the cancerous cases – they were of Gleason score of 6 or 7. For each case, the ground truth was confirmed by histopathological analysis.

In the first group of experiments, involving single-image texture analysis, we utilized all the available ROIs derived from both databases. Multi-sequence texture analysis was only possible for the first database, as only in this base three required series of images (T1W, T2W, and DW) were available. In Fig. 3 an exemplary triple of images with marked ROIs is shown. To increase visibility, only the central parts (quarters) of the whole images are presented.



**Fig. 3 – Triple of corresponding prostate MR images derived from different MRI sequences: (a) T1W, (b) T2W, and (c) DW. The images were acquired at the same slice position. Three corresponding Regions of Interest (ROIs) cover the same prostate area.**

#### 4.2. Equalization of pixel value ranges

A certain inconvenience of our database was that the full range of pixel values (gray levels) that might occur in the images (image gray scale) could not be determined from the DICOM headers. Such range has to be known and equalized (if necessary) for all the images within a series, before performing any texture-based tissue characterization. In order to estimate the image gray scale, as well as the minimum interval of pixel values large enough to describe all the pixels belonging to the prostate ROIs, we constructed pixel value histograms. These were based on prostate ROIs and calculated separately for each patient examination and for each type of image sequence (T1W, T2W, and DW). In addition, histograms based on the values of pixels belonging to the hips were constructed as well. The hip regions were treated here as reference objects, as they were much larger than the prostate and not affected by any tumor. Hip regions comprised on average 547 pixels for T1W, 2560 pixels for T2W, and 412 pixels for DW images. In total, 1776 ROIs were outlined for the hips. Among them 678, 697, and 401 ROIs were derived from T1W, T2W, and DW sequences, respectively. The average number of ROIs per study was therefore 36, 23, and 21, respectively.

An example of pixel value histograms based on entire images, on hip ROIs and on prostate ROIs, is given in Fig. 4. Each histogram was created for T2W sequence of images. Two columns in the example (named “Study A” and “Study B”) correspond to two different examinations, from two different patients. The example demonstrates that image gray scales are obviously different for each examination. Such a problem was observed for the T1W and T2W sequences in the whole database. The widest range of pixel values was more than nine times wider than the narrowest. The range centers obtained for different studies and the same series were located in different places. However, we noticed that for each study and for each image sequence, the ratio between the range of pixel values from the whole images and the range of pixel values from the hip regions was almost the same. This gave us the idea that linear scaling of the pixel values could be an acceptable solution to the problem.

Greater differences were noticed for the ratios between the ranges for the entire images and the ranges obtained for the

available prostate ROIs. In fact, some of the studies involved only cancerous prostatic tissue, and others only healthy tissue. At that time we did not yet know whether and in what way the presence of cancer modified the texture corresponding to affected prostate regions. Therefore, assessment of image gray scales based only on prostate ROIs did not seem to be a reliable solution.

In order to equalize the ranges of pixel values corresponding to T1W and T2W sequences, the images were converted. The conversion was a linear transformation of image pixel values and was performed separately for each examination. Its objective was to obtain the same range of pixel values (the smallest possible) corresponding to ROIs covering the hips. In the assessment of the range, the 5% of pixels with marginal pixel values (the brightest and the darkest ones) were omitted. After the conversion, the entire range of gray levels sufficient to characterize all the pixels belonging to the prostate ROIs did not exceed 256 for each of the considered sequences. This allowed the images to be processed as if they were in an 8-bit BMP format.

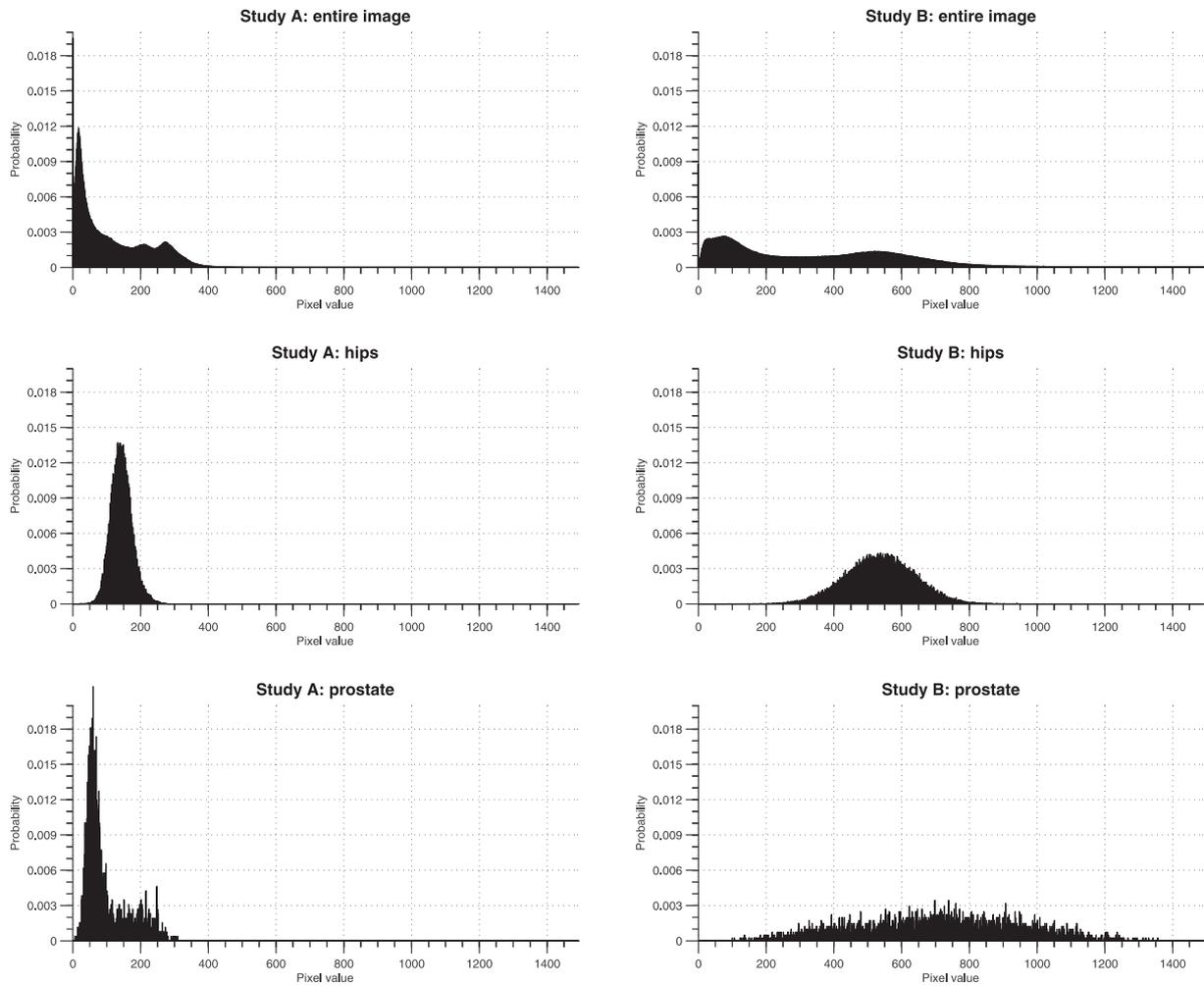
#### 4.3. Feature extraction

Features were extracted with the home-made application *Medical Image Processing*. Six methods for characterization of single-image textures were considered. They were based on the following:

- autocorrelation (AC)
- the gray-level histogram, giving first-order statistics (FOS)
- gradient matrices (GM)
- the fractional Brownian motion model (FB)
- co-occurrence matrices (COM)
- run-length matrices (RLM).

A detailed description of all of the above methods, as well as the parameters derived from them, can be found in [35]. In total, 30 features were calculated separately for each single image. Their abbreviations and full names are given in Table 2.

Since most ROIs were relatively small and narrow, only the smallest possible distances between pixel pairs (1 and 2) could be taken into account when applying the AC, FB, and



**Fig. 4 – Pixel value histograms obtained for whole sequences of T2W images, derived from two different examinations, from two different patients (named “Study A” and “Study B”). The histograms were made on the entire images, ROIs covering the hips, and available ROIs delineated for the prostate.**

COM methods. Moreover, for the COM and RLM methods, the number of gray levels was reduced from 256 to 64 and 32, respectively. This final choice of reduced numbers of gray levels was made after some preliminary classification experiments testing all the possibilities, including 256 (initially used), 128, 64, 32, and 16 gray levels. In the case of relatively small ROIs, the use of the initial number of 256 gray levels could have resulted in many zero entries in the co-occurrence and run-length matrices. Reducing this number reduced memory requirements for matrix storage and the time needed for calculation of textural features, without affecting the final classification accuracies.

The co-occurrence matrices were constructed separately for 4 standard directions (0°, 45°, 90°, and 135°) and for 2 considered distances between pixel pairs. Also, the run-length matrices corresponded to 4 aforementioned directions of pixel runs. Features obtained for different pixel distances and/or for different directions were averaged.

Normalized autocorrelation coefficients (the AC method) were also calculated separately for 4 standard directions and

for 2 pixel distances. In this case, only features corresponding to different directions of pixel arrangement were averaged.

Experiments were performed separately for each set of features listed in Table 2. In addition, two sets containing combinations of features derived from several texture analysis methods were considered:  $All_{23}$ , containing features obtained by the FOS, COM, and RLM methods, and  $All_{30}$ , containing all of the available 30 features. Moreover, features from the last set, corresponding to the three considered sequences, T1W, T2W, and DW ( $3 \times 30 = 90$  feature values in total), were subjected to selection. Different sets of selected features were evaluated.

#### 4.4. Reduction of feature space

Both feature selection and Principal Component Analysis were performed with Weka software [36]. For the feature selection, a wrapper method for evaluation of candidate subsets of features was used (the method called *WrapperSubsetEval* in Weka). This is a supervised approach in which the relevance of each candidate subset of features is evaluated by the classification accuracy that a given subset of features can

**Table 2 – Textural features considered in our experiments. The name of a feature set is created by adding the number of features (as a subscript index) to the name of the extraction method.**

Set	Feature abbreviation	Feature name
AC <sub>2</sub>	Autocorr <sub>1</sub>	Autocorrelation for a pixel distance $d = 1$
	Autocorr <sub>2</sub>	Autocorrelation for a pixel distance $d = 2$
FOS <sub>4</sub> :	Avg	Average
	Var	Variance
	Skew	Skewness
	Kurt	Kurtosis
GB <sub>4</sub> :	GradAvg	Average
	GradVar	Variance
	GradSkew	Skewness
	GradKurt	Kurtosis
FB <sub>1</sub> :	FractalDim	Fractal dimension
COM <sub>11</sub> :	AngSecMom	Angular second moment
	InvDiffMom	Inverse difference moment
	Entr	Entropy
	Corr	Correlation
	SumAvg	Sum average
	DiffAvg	Difference average
	SumVar	Sum variance
	DiffVar	Difference variance
	SumEntr	Sum entropy
	DiffEntr	Difference entropy
	Contrast	Contrast
RLM <sub>8</sub> :	ShortEmp	Short run emphasis
	LongEmp	Long run emphasis
	GLNonUni	Gray level non-uniformity
	RLNonUni	Run length non-uniformity
	Fraction	Fraction of image in runs
	LowGLREmp	Low gray level runs emphasis
	HighGLREmp	High gray level runs emphasis
	REnter	Run length entropy

ensure. We decided to use a C4.5 Decision Tree [37] (J48 in Weka) as a classifier, due to its simplicity and the very short time required for its induction. The latter property is particularly important when we consider that such trees are to be constructed hundreds of thousands of times. The classification accuracies were assessed with a 10-fold cross-validation. The space of subsets of features was searched using the *BestFirst* strategy, which performed a greedy search with backtracking. Due to time constraints, only the *Forward* searching direction was tested. The selection procedure was repeated one hundred thousand times ( $r = 100,000$ ). Each time, the input (truncated) dataset fed into the feature selection procedure was composed of two-thirds of the available observations ( $q = 2/3p$ ), which were characterized either by 10% ( $m = 0.1d$ ) or by 20% ( $m = 0.2d$ ) of all the features initially considered (two cases were tested).

#### 4.5. Classification

Classification was also performed using Weka software. Six classifiers were used to assess the potential of different sets of textural features:

- DT: Decision Tree – C4.5 (J48)
- AB: Ensemble of Classifiers with an Adaptive Boosting voting scheme (*AdaBoostM1* algorithm in Weka), using the C4.5 classifier as the underlying algorithm
- RF: Random Forest (*RandomForest* algorithm)
- LR: Logistic Regression (*Logistic* algorithm)
- NN: Neural Network [38] (*MultilayerPerceptron*) with a back-propagation and sigmoidal activating function, having one hidden layer wherein the number of neurons was equal to the average value of the number of features and the number of classes
- SVM: Support Vector Machines with the Sequential Minimal Optimization (SMO) algorithm [39]. The classifier used either a Gaussian kernel (*RBFKernel*) or a polynomial kernel (*PolyKernel*).

In the case of DT, LR, and NN classifiers, default classifier settings (proposed by Weka software) were used. For the three remaining classifiers, some classifier variations were tested. Different numbers of iterations were considered for training the AB classifier (parameter *numIterations* equal to 10, 100, 200, 500, and 1000). The RF classifier was trained with different numbers of trees to be generated (*numTrees*): 10, 50, 100, 200, and 500. For the SVM, different bandwidths (parameter *gamma*) of the Gaussian kernel were tested ( $10^{-4}$ ,  $10^{-3}$ ,  $10^{-2}$ , 0.1, 0.2, 0.3, ..., 1) and different polynomial degrees were considered (*exponent* values from 1 to 5).

#### 4.6. Statistical analysis

In all the classification experiments, classification accuracies were estimated by 10-fold cross-validation, repeated 10 times. 100 partial results obtained for each experiment were then averaged. Tests of statistical significance were performed to determine whether the classification results obtained for different sets of features were statistically significant at a significance level of 0.05. In particular, these tests were used to assess the improvement in classification accuracy ensured by multi-sequence texture analysis in relation to the best corresponding single-sequence results. First, the Shapiro–Wilk test was used to determine whether the samples of 100 partial results followed the normal distribution. Then, another test (the *T*-test for normal distribution cases and the Mann–Whitney *U*-test for abnormal ones) was applied to assess whether the means of two samples were statistically different from each other. All the statistical tests were performed with STATISTICA 10 software (StatSoft, Inc. 2010).

## 5. Results and discussion

The initial experiment concerned only the single-image texture analysis and was performed as a baseline. We limit ourselves to present only the classification results obtained separately for each series of images. They allowed us to assess what series of images (among the three considered) would provide the most useful information, in terms of prostatic tissue differentiation.

The course of experiments in the multi-sequence texture analysis was as follows. First, two methods for reduction of the

feature space were tested: modified MC feature selection and PCA. The experiment involving modified MC feature selection resulted in the creation of rankings of features according to their incidence of selection. Then classification experiments were conducted, during which more and more top features from the ranking were used to describe observations. This enabled us to assess how many features would be sufficient to yield the best recognition between healthy and cancerous prostatic tissues. The results presented and discussed in the paper correspond to the case in which observations in the input datasets, fed into the feature selection procedure, were characterized by 20% of initially calculated textural features ( $m = 0.2d$ ). For the case with 10% of initially calculated features, further classification experiments resulted in slightly worse or comparable results, so the results for this case are not commented on in the paper. In the PCA as well, more and more principal components were added to the set of components describing the observations used in the classification experiment.

5.1. Single-image texture analysis. Classification results

Table 3 presents the most interesting results obtained with different feature sets in a single-image texture analysis. As shown above, the numbers of observations recognized were 107, 180, and 68, for the T1W, T2W, and DW sequences, respectively. Experiments were performed with all the six classifiers, although we limit ourselves to show only the results obtained by the three of them: RF, NN, and SVM. In fact, the conclusions drawn from the presented results coincide closely with those from all the results obtained. The latter, however, are also commented.

The results obtained by the C4.5 Decision Tree were slightly worse than those obtained by Adaptive Boosting in the case of the T1W and T2W sequences (the differences in classification results ranged from 0.28% to 6.66%). In contrast, for the DW sequence the C4.5 Decision Tree generally provided slightly better results, but the improvements obtained were not statistically significant at a significance level of 0.05. Among the results provided by the Adaptive Boosting classifier, only those corresponding to the use of 200 iterations for classifier

training were chosen for further analyses. Indeed, increasing the number of iterations to over 200 no longer resulted in significantly better classification accuracies. In the case of the Random Forest classifier, setting the number of composing trees to 100 resulted in the best classification. As for the Support Vector Machines, it was difficult to specify only one set of parameters (one kernel function and one parameter of it –  $\gamma$  for the RBF kernel or  $\text{exponent}$  for the polynomial kernel) providing the best tissue differentiation irrespective of the set of textural features used. Smaller differences in results were observed for the polynomial kernel with different polynomial degrees than for the RBF kernel with different bandwidths. So, in this part of the work we limited ourselves to presenting the results corresponding to the application of the second-degree polynomial kernel.

We can conclude that the most useful information for the prostatic tissue classification process was extracted from the T2W and DW images. For some classification algorithms (AB, RF, and LR) the best tissue recognition was achieved with the T2W sequences irrespective of the set of features used. In such cases the results obtained for the DW images were a bit inferior (up to about 4%). For the remaining three classifiers (NN, DT, and SVM), the DW sequences sometimes proved to be slightly better. However, the advantage was not statistically significant at a significance level of 0.05 (the difference in the classification results was less than 1%). Finally, it should be noted that the results obtained may also have been affected by disproportion of the observations available for the two image sequences (T2W and DW). The ROIs corresponding to the T2W sequence were almost three times more numerous than the ROIs from the DW sequence. In fact, the more numerous the number of samples, the easier it is to avoid the risk of overfitting, and therefore it is generally easier to find a more reliable classifier [33].

The advantage of the T2W images was certainly that they were the largest (ranging in size from  $320 \times 320$  to  $448 \times 448$  pixels). Their drawback was the necessity of pre-conversion, as they initially showed the greatest differences in ranges of pixel values. Finally, taking into account the results obtained with T2W images, we may conclude that the applied image conversion probably did not affect the classification results (too) negatively. Therefore, pre-conversion could be an acceptable solution when no information on the full range of image pixel values is available in DICOM headers. The inferior results obtained for the T1W sequence may indicate the need to develop methods for choosing the most appropriate moment (in terms of tissue characterization) of contrast agent propagation.

To sum up, the best classification accuracies were: 81.00%, 97.39%, and 96.88%, for the T1W, T2W, and DW sequences, respectively. The two first results were obtained by the SVM classifier and the  $All_{23}$  feature set. The last one was achieved by the NN classifier and the  $COM_{11}$  set.

5.2. Multi-sequence texture analysis

5.2.1. Reduction of feature space

Table 4 presents the feature incidence frequency ranking obtained for three-image texture analysis (considering simultaneously triples of images derived from different sequences).

**Table 3 – Classification accuracy [%] (and standard deviation) achieved after application of single-image and single-sequence texture analysis. Three image sequences were considered: T1W, T2W, and DW. The results were obtained using three classifiers: RF, NN, and SVM with a second degree polynomial kernel.**

Classifier	Feature set	T1W	T2W	DW
RF	All <sub>30</sub>	87.86 (4.97)	96.56 (1.80)	95.29 (3.61)
	COM <sub>11</sub>	79.00 (6.05)	94.94 (2.34)	93.86 (3.78)
	RLM <sub>8</sub>	80.25 (5.95)	96.50 (2.19)	94.36 (4.21)
NN	All <sub>30</sub>	79.04 (5.81)	96.39 (2.07)	95.17 (3.47)
	COM <sub>11</sub>	75.48 (5.95)	96.00 (2.09)	96.88 (3.22)
	RLM <sub>8</sub>	79.02 (6.09)	96.72 (2.01)	92.45 (4.86)
SVM	All <sub>30</sub>	80.77 (5.53)	96.56 (2.08)	96.74 (3.67)
	COM <sub>11</sub>	77.76 (5.31)	96.50 (1.88)	93.24 (4.14)
	RLM <sub>8</sub>	75.55 (5.60)	96.78 (1.90)	92.98 (4.21)

**Table 4 – Multi-image (three-image) texture analysis: Ranking of features according to their incidence of selection obtained by the modified Monte Carlo method. Only the first 18 features are considered. The “Frequency” is the percentage of cases in which the feature was selected with respect to the number of its occurrences in the input data sets (subjected to selection). Superscript indices indicate the corresponding sequence of images: T1W, T2W, or DW.**

Rank	Feature	Frequency [%]
1	HighGLREmp <sup>DW</sup>	73.15
2	SumAvg <sup>DW</sup>	72.40
3	Avg <sup>DW</sup>	71.20
4	Avg <sup>T2W</sup>	42.35
5	SumAvg <sup>T2W</sup>	40.35
6	LowGLREmp <sup>DW</sup>	38.85
7	HighGLREmp <sup>T2W</sup>	37.25
8	Autocorr <sub>1</sub> <sup>DW</sup>	28.90
9	Autocorr <sub>2</sub> <sup>DW</sup>	26.35
10	RLNonUni <sup>T1W</sup>	21.50
11	LongEmp <sup>DW</sup>	18.65
12	GLNonUni <sup>DW</sup>	14.60
13	SumEntr <sup>T2W</sup>	14.15
14	Fraction <sup>DW</sup>	13.20
15	LowGLREmp <sup>T2W</sup>	12.40
16	Autocorr <sub>1</sub> <sup>T1W</sup>	11.05
17	Autocorr <sub>1</sub> <sup>T2W</sup>	10.70
18	Autocorr <sub>2</sub> <sup>T2W</sup>	10.60

Features are sorted in descending order of their incidence frequency in the entire MC feature selection experiment. We limit the results to the top-ranked 18 features from the ranking (of the 90 available). Each feature is followed by the percentage of cases in which it was selected. In addition, for each feature the corresponding sequence is indicated as a superscript index.

The MC feature selection experiment also suggests that the T2W and DW sequences are better than the T1W sequence in terms of prostatic tissue recognition. The features corresponding to the T2W and DW sequences can be usually found at the beginning of the ranking. This time, features derived from DW sequences seem to be generally better than their equivalents derived from T2W sequences. The three most ranked features are derived from the DW series. These are: HighGLREmp (RLM), SumAvg (COM), and Avg (FOS). For each of them, the percentage of selections exceeds 70%. In turn, the features obtained from the T1W sequence are at the end of the ranking (not presented in Table 4). We may presume that they are less relevant than those corresponding to the T2W and DW sequences. Exceptions are observed for the RLNonUni feature (RLM method) and the Autocorr<sub>1</sub> feature, for which their T1W versions were ranked as the 10th and the 16th, respectively (according to their frequency of selections).

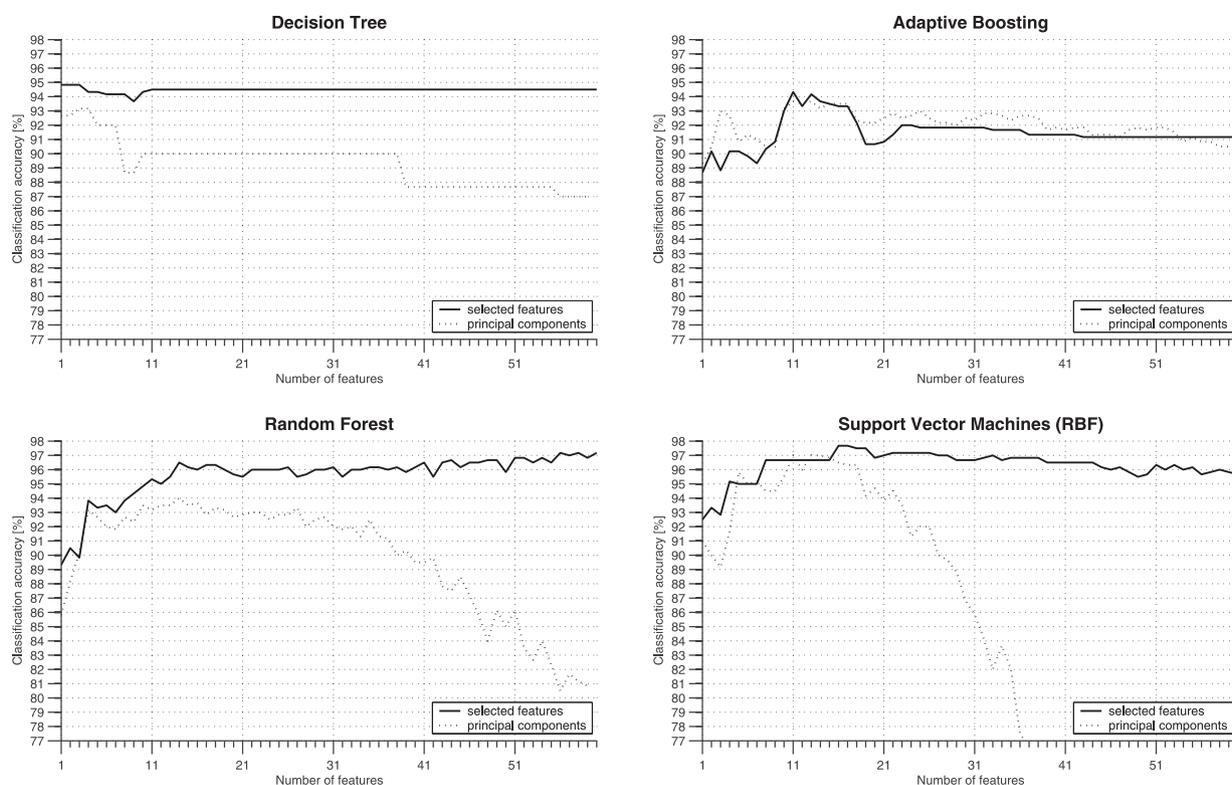
The plots in Fig. 5 show the quality of classification obtained with different numbers of the most frequently selected features (selection by the modified MC method), and with different numbers of first principal components. The plots were made for four selected classifiers: DT, AB, RF, and SVM with the RBF kernel and the  $\gamma$  parameter set to 0.5. In the case of the remaining classifiers (LR, NN, and SVM with a second-degree polynomial kernel), the plots obtained are very similar to the one obtained with the SVM-RBF classifier.

It can be observed that it is possible to achieve highly satisfactory classification results with a small subset of the whole set of features. In practice, increasing the number of selected features to over 15 does not lead to significant improvement in classification quality, nor does it substantially degrade the classification results. A different trend can be observed for the principal components. Sometimes they provide slightly better results than the selected features. However, the advantage is not statistically significant (at a significance level of 0.05). Moreover, above a certain threshold, the classification quality obtained with the principal components decreases dramatically. This threshold is different for each classifier and ranges from 3 features (DT) to 35 (NN). For this reason, one should be particularly careful when using principal components.

### 5.2.2. Comparison of classification results

Table 5 presents the most interesting classification results corresponding to the application of multi-image texture analysis, when two or three image sequences were considered simultaneously. Each line of the table refers to the same classifier and the same feature extraction method. The first column of results is given for reference and contains, for each case, the best classification result of the three obtained separately for single-image cases (for the T1W, T2W, or DW sequence). Such a result corresponds to the use of only 60 observations, the same ones that were considered for the multi-sequence texture analysis. The second column shows the best result among those obtained for two-image cases (combinations “T1W & T2W”, “T1W & DW”, and “T2W & DW”). Finally, the third column presents the result obtained for the three-image texture analysis (the combination “T1W, T2W & DW”). The table does not consider DT, AB, and LR classifiers. The results corresponding to the application of the AB classifier were comparable (in this case slightly inferior) to those obtained with the DT classifier. For those both classifiers, the multi-image texture analysis resulted in maximum classification accuracy of about 95% and little improvement in classification results (1% or even less) in comparison with the best single-image cases. Similarly, the results corresponding to the SVM-RBF classifier are not presented because they were comparable to those obtained with the SVM-polynomial classifier. For the latter,  $\gamma$  in the kernel function was set to 2. The best results obtained by the RF classifier correspond to the application of 100 composing trees.

The comments included in the present subsection are based on all the results we received, not just those shown in Table 5. Comparing the results obtained with single- and multi-image texture analysis we noticed that the best result, among those achieved with the same classifier and the same feature extraction method, always corresponded to the multi-image analysis. Moreover there were always at least one combination of two sequences that led to better tissue recognition than the best possible recognition achieved in the case of single-sequence analysis. The improvement in the quality of classification ranged from 0.33% (not statistically significant at a significance level of 0.05; obtained with a DT classifier and different sets of features: All<sub>23</sub>, All<sub>30</sub>, and RLM<sub>8</sub>) to 7.83% (statistically significant at the 0.05 level; obtained with an SVM-RBF classifier and the RLM<sub>8</sub> feature set). Exceptions to



**Fig. 5 – Multi-sequence texture analysis: classification accuracy [%] achieved with different numbers of the most frequently selected features and with different numbers of first principal components. The results were obtained by four classifiers: Decision Tree, Adaptive Boosting, Random Forest, and Support Vector Machines with the RBF kernel ( $\gamma = 0.5$ ).**

**Table 5 – Classification accuracy [%] (and standard deviation) achieved after application of single-image/single-sequence and multi-image/multi-sequence texture analysis (TA). The first column of results contains the best result among the three obtained separately for single-image cases (T1W, T2W, or DW). The name of the best sequence is given next to the result. The next column contains the best result obtained for two-image TA, followed by the name of the best combination (“T1W & T2W”, “T1W & DW”, or “T2W & DW”). The last column contains the result for three-image texture analysis (“T1W, T2W, & DW”). The results were obtained using three classifiers: RF, NN, and SVM with a second degree polynomial kernel. The number given in square brackets for the “MC selected” and “PCs” sets of features is the dimension of reduced feature space in which the classification accuracy was the highest.**

Classifier	Set of features	Single-image TA	Two-image TA	Three-image TA
RF	All <sub>30</sub>	94.50 (4.60) DW	96.33 (3.86) “T2W &DW”	97.50 (2.99)
	COM <sub>11</sub>	93.50 (4.87) DW	94.83 (4.22) “T2W &DW”	94.50 (4.60)
	RLM <sub>8</sub>	94.33 (4.47) DW	95.33 (4.29) “T2W &DW”	96.67 (3.56)
	MC selected	–	97.00 (3.63) “T2W &DW” [56]	97.67 (3.14) [76]
	PCs	–	93.17 (4.75) “T2W &DW” [5]	94.00 (4.52) [14]
NN	All <sub>30</sub>	94.00 (4.02) DW	96.33 (3.67) “T1W &DW”	96.00 (4.29)
	COM <sub>11</sub>	95.67 (4.21) DW	96.83 (3.29) “T1W &T2W”	96.50 (3.98)
	RLM <sub>8</sub>	91.50 (5.74) T2W	97.83 (2.82) “T1W &T2W”	97.33 (3.50)
	MC selected	–	98.33 (2.52) “T1W &DW” [27]	96.33 (4.20) [35]
	PCs	–	98.17 (2.62) “T1W &DW” [21]	99.00 (1.99) [14]
SVM	All <sub>30</sub>	95.67 (4.53) DW	95.83 (4.00) “T1W &DW”	98.17 (2.62)
	COM <sub>11</sub>	93.50 (5.15) DW	96.67 (3.75) “T1W &DW”	98.00 (2.97)
	RLM <sub>8</sub>	93.17 (4.75) T2W	96.67 (3.56) “T1W &T2W”	97.83 (3.06)
	MC selected	–	98.67 (2.27) “T1W &DW” [27]	98.17 (2.62) [76]
	PCs	–	96.67 (3.35) “T2W &DW” [19]	99.17 (1.83) [18]

this rule were sometimes observed when first-order (FOS) features were used (DT, AB, and LR classifiers). Generally, such features could not be very good texture descriptors because they contain no information on the relationship between neighboring pixels, the possible direction of the texture, its structure, and other properties resulting from these relationships.

Considering the three sequences at once did not always yield the best result. However, it should be taken into account that in our data set the number of features ( $3 \times 30 = 90$ ) exceeds the number of observations (60), and certain classifiers do not generalize sufficiently in such cases. Here, a good solution is the use of SVM, known for their generalization ability. On the other hand, a drawback of SVM is that the kernel function must be determined experimentally (sometimes with many trials), because the best kernel could be different for each task. The problem of proper matching of classifier parameters also arises when NN algorithms are used. But, like SVM, they can ensure highly satisfactory tissue recognition (what was observed in our experiments).

The best classification results for the multi-image texture analysis were about 99%, obtained when three considered image sequences were treated simultaneously: 99.00% for the NN classifier (the PCs), 99.33% and 99.17% for the SVM-polynomial classifier, with the All<sub>23</sub> and the PCA feature set, respectively. For comparison, the best results for the single-image texture analysis were less than 97%, and corresponded to the application of DW images. They were obtained either with the set of selected features (96.17% of correctly recognized cases with the NN and with the SVM-polynomial classifier), or with the set of first PCs (96.83% with the NN classifier).

Finally, let us comment the results obtained with different sets of features, but with the same classifier and for the same sequence of images or for the same combination of sequences. It was seen that application of the set of features selected by the modified MC method almost always led to the best result. Otherwise (in a small number of cases), the best result was obtained with the set containing some first principal components. Very often the reduced feature space had only a few dimensions. When the selected features performed worse than the PCs, the difference between the classification accuracy obtained for these two sets were not statistically insignificant (here, only one exception was noted, when three-image textures were recognized by the NN classifier). If we consider that it is much easier to interpret the meaning of each selected feature than the real meaning of a principal component, we can conclude that application of only selected features could generally be a better solution to our problem.

## 6. Conclusions and future work

Two main objectives were pursued in this work. The first was to exploit the potential of texture analysis in the recognition of prostatic tissues based on MR images. Two tissue classes were differentiated, cancerous and healthy. The study placed particular emphasis on the multi-image texture characterization. This method consists in simultaneous analysis of several images representing the same organ slice but corresponding to different acquisition conditions (for example, in the case of MR

images, yielding T1W, T2W, and DW sequences). To the best of our knowledge, there exists hardly any works using multi-image texture analysis in order to characterize prostatic tissue in the classification process. Moreover, this is the first time the potential of multi-image texture analysis in the recognition of prostatic tissues from MR images has been compared to the potential of single-image analysis.

The experiments showed that simultaneous analysis of two or three images can improve recognition of prostatic tissues in comparison with single-image analysis. In the vast majority of cases, the best results obtained for multi-image (two- or three-image) cases were better than the best corresponding results achieved for single-image cases. The best improvement of classification quality reached 7.83% and was statistically significant at a level of 0.05. The analysis of three-image textures was found to ensure the best overall classification result.

The second objective was to identify which textural features contribute the most to tissue differentiation. Here, 90 candidate features (30 for each sequence type: T1W, T2W, and DW) were analyzed by the modified Monte Carlo feature selection method. Based on this method, feature incidence frequency rankings were made. The features *HighGLREmp* (calculated by the RLM method), *SumAvg* (COM method), and *Avg* (FOS), all derived from the DW image sequence, were the most frequently selected. The combination of several best-ranked features ensured highly satisfactory results – about 99% of textures correctly recognized (in the case of multi-image analysis). Another approach for dimensionality reduction was Principal Component Analysis, also performed on sets of observations described by  $3 \times 30$  features. Its application did not significantly improve results in comparison to those obtained after Monte Carlo feature selection (the significance level was always set at 0.05).

We acknowledge that the experimental results, although promising, could also be subject to bias. This could have been avoided if a key piece of information had been available in the database, namely the full ranges of image pixel values, evidently different for different studies. As such ranges were not given, image conversion based on analysis of intervals of pixel values corresponding to another part of the body (in our case the hips) seemed to be the only solution. However, the hip regions may also be altered by the presence of various pathological processes, different for each patient. To avoid this problem in the future, either image acquisition protocols should be standardized or images should contain information on the full ranges of pixel values. Furthermore, when acquiring images, a good idea would be to place a “reference object” in view. The texture analysis of such an object could be crucial for the purposes of image conversion aimed at equalization of pixel value ranges corresponding to different studies.

Finally, it would be useful to have two types of ROIs (corresponding to cancerous and healthy tissue) delineated for each study. Such information would enable analysis of changes in texture characteristics under contrast product propagation (in T1W sequences) corresponding to both types of tissue. Based on such an analysis we could determine which moment of contrast product propagation is associated with the most significant differences in texture characteristics obtained for cancerous and healthy tissue.

In the future it would be worthwhile to repeat the experiments with more tissue classes taken into account. Other MRI sequences, such as FLAIR (fluid-attenuated inversion recovery) or proton density-weighted, could also be considered for multi-image texture analysis. Other methods for extraction of texture features (e.g. wavelet- or Gabor-based) could also be tested. It would also be interesting to find a method for characterizing texture evolution under contrast product propagation based on simultaneous analysis of many contrast-enhanced T1W images associated with different concentrations of the contrast product in prostatic vessels.

### Ethical approval

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

### Informed consent

Informed consent was obtained from all individual participants included in the study.

### Funding

This work was partially supported by grant S/WI/2/2013 from the Bialystok University of Technology, Bialystok, Poland, founded by the Polish Ministry of Science and Higher Education. The funding source had no involvement in study design; in the collection, analysis and interpretation of data; in the writing of the report; and in the decision to submit the article for publication.

### Acknowledgements

The authors thank Dr. M. Topczewska for her valuable assistance in performing the statistical analysis of the results.

### REFERENCES

- [1] Duda D, Kretowski M, Mathieu R, de Crevoisier R, Bezy-Wendling J. Multi-image texture analysis in classification of prostatic tissues from MRI. Preliminary results. *Adv Intell Syst Comput* 2014;283:139–50.
- [2] Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics. *CA Cancer J Clin* 2015;65(2):87–108.
- [3] Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser S, Mathers C, et al. GLOBOCAN 2012 v1.1, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11. Lyon, France: International Agency for Research on Cancer; 2016. <http://globocan.iarc.fr> [accessed 25.01.16].
- [4] American Cancer Society. Cancer facts & figures 2015. Atlanta: American Cancer Society; 2015.
- [5] Wolf AMD, Wender RC, Etzioni RB, Thompson IM, D'Amico AV, Volk RJ, et al. American Cancer Society Guideline for the early detection of prostate cancer: update 2010. *CA Cancer J Clin* 2010;60(2):70–98.
- [6] Andriole GL, Crawford ED, Grubb III RL, Buys SS, Chia D, Church TR, et al. Mortality results from a randomized prostate-cancer screening trial. *N Engl J Med* 2009;360:1310–9.
- [7] Greene KL, Albertsen PC, Babaian RJ, Carter HB, Gann PH, Han M, et al. Prostate specific antigen best practice statement: 2009 update. *J Urol* 2009;182(5):2232–41.
- [8] Harvey CJ, Pilcher J, Richenberg J, Patel U, Frauscher F. Applications of transrectal ultrasound in prostate cancer. *Br J Radiol* 2012;85(1). S3-17 [special issue].
- [9] Lundstrom KJ, Drevin L, Carlsson S, Garmo H, Loeb S, Stattin P, et al. Nationwide population based study of infections after transrectal ultrasound guided prostate biopsy. *J Urol* 2014;192(4):1116–22.
- [10] Petrick N, Sahiner B, Armato 3rd SG, Bert A, Correale L, Delsanto S, et al. Evaluation of computer-aided detection and diagnosis systems. *Med Phys* 2013;40(8). 087001-1-17.
- [11] Duda D. Texture analysis as a tool for medical decision support. Part 1: Recent applications for cancer early detection. *Adv Comput Sci Res* 2014;11:61–84.
- [12] Duda D. Texture analysis as a tool for medical decision support. Part 2: Classification of liver disorders based on computed tomography images. *Adv Comput Sci Res* 2014;11:85–108.
- [13] Draminski M, Rada-Iglesias A, Enroth S, Wadelius C, Koronacki J, Komorowski J. Monte Carlo feature selection for supervised classification. *Bioinformatics* 2008;24(1):110–7.
- [14] Kohavi R, John GH. Wrappers for feature subset selection. *Artif Intell* 1997;97(1/2):273–324.
- [15] Sung YS, Kwon HJ, Park BW, Cho G, Lee CK, Cho KS, et al. Prostate cancer detection on dynamic contrast-enhanced MRI: computer-aided diagnosis versus single perfusion parameter maps. *AJR Am J Roentgenol* 2011;197(5):1122–9.
- [16] Langer DL, van der Kwast TH, Evans AJ, Trachtenberg J, Wilson BC, Haider MA. Prostate cancer detection with multi-parametric MRI: logistic regression analysis of quantitative T2, diffusion-weighted imaging, and dynamic contrast-enhanced MRI. *J Magn Reson Imaging* 2009;30(2):327–34.
- [17] Vos PC, Hambroek T, Barenstz JO, Huisman HJ. Computer-assisted analysis of peripheral zone prostate lesions using T2-weighted and dynamic contrast enhanced T1-weighted MRI. *Phys Med Biol* 2010;55(6):1719–34.
- [18] Vapnik VN, editor. The nature of statistical learning theory. 2nd ed. New York: Springer; 2000.
- [19] Artan Y, Haider MA, Langer DL, van der Kwast TH, Evans AJ, Yang Y, et al. Prostate cancer localization with multispectral MRI using cost-sensitive support vector machines and conditional random fields. *IEEE Trans Image Process* 2010;19(9):2444–55.
- [20] Lopes R, Ayache A, Makni N, Puech P, Villers A, Mordon S, et al. Prostate cancer characterization on MR images using fractal features. *Med Phys* 2011;38(1):83–95.
- [21] Ginsburg SB, Rusu M, Kurhanewicz J, Madabhushi A. Computer extracted texture features on T2W MRI to predict biochemical recurrence following radiation therapy for prostate cancer. *Proc SPIE* 2014. Paper 903509.
- [22] Freund Y, Shapire R. A decision-theoretic generalization of online learning and an application to boosting. *J Comput Syst Sci* 1997;55(1):119–39.

- [23] Turner MR. Texture discrimination by Gabor functions. *Biol Cybern* 1986;55(2/3):71–82.
- [24] Viswanath S, Bloch BN, Rosen M, Chappelow J, Toth R, Rofsky N, et al. Integrating structural and functional imaging for computer assisted detection of prostate cancer on multi-protocol in vivo 3 tesla MRI. *Proc SPIE* 2009. Paper 72603I.
- [25] Breiman L. Random forests. *Mach Learn* 2001;45(1):5–32.
- [26] Litjens G, Debats O, Barentsz J, Karssemeijer N, Huisman H. Computer-aided detection of prostate cancer in MRI. *IEEE Trans Med Imaging* 2014;33(5):1083–92.
- [27] Molina JFG, Zheng L, Sertdemir M, Dinter DJ, Schonberg S, Radle M. Incremental learning with SVM for multimodal classification of prostatic adenocarcinoma. *PLoS ONE* 2014;9(4). Paper e93600.
- [28] Chan I, Wells 3rd W, Mulker RV, Haker S, Zhang J, Zou KH, et al. Detection of prostate cancer by integration of line-scan diffusion, T2-mapping and T2-weighted magnetic resonance imaging; a multichannel statistical classifier. *Med Phys* 2003;30(9):2390–8.
- [29] Metz CE. ROC methodology in radiologic imaging. *Invest Radiol* 1986;21(9):720–33.
- [30] Niaf E, Rouviere O, Mege-Lechevallier F, Bratan F, Lartizien C. Computer-aided diagnosis of prostate cancer in the peripheral zone using multiparametric MRI. *Phys Med Biol* 2012;57(12):3833–51.
- [31] Castellano G, Bonilha L, Li LM, Cendes F. Texture analysis of medical images. *Clin Radiol* 2004;59(12):1061–9.
- [32] Nailon WH. Texture analysis methods for medical image characterisation. In: Mao Y, editor. *Biomedical imaging*. InTech Open; 2010. p. 75–100.
- [33] Hastie T, Tibshirani R, Friedman J. *The elements of statistical learning: data mining, inference, and prediction*. 2nd ed. New York: Springer; 2009.
- [34] Guyon I, Elisseeff A. An introduction to variable and feature selection. *J Mach Learn Res* 2003;3:1157–82.
- [35] Lerski RA, de Certaines JD, Duda D, Klonowski W, Yang G, Coatrieux JL, et al. Application of texture analysis to muscle MRI: 2 – technical recommendations. *EPJ Nonlinear Biomed Phys* 2015;3(2):1–20.
- [36] Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. *The WEKA data mining software: an update*. *SIGKDD Explor* 2009;11(1):10–8.
- [37] Quinlan J. *C4.5: programs for machine learning*. San Francisco: Morgan Kaufmann; 1993.
- [38] Bishop CM. *Neural networks for pattern recognition*. New York: Oxford University Press; 1995.
- [39] Platt JC. Fast training of support vector machines using sequential minimal optimization. In: Scholkopf B, Burges CJC, Smola AJ, editors. *Advances in kernel methods – support vector learning*. Cambridge: MIT Press; 1998. p. 185–208.