International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

COPY RIGHT

ELSEVIER SSRN

IJIEMR Transactions, online available on 13th Dec 2019. Link :http://www.ijiemr.org/downloads.php?vol=Volume-08&issue=ISSUE-12

Title **AUTOMATED PROSTATE CANCER DETECTION VIA COMPREHENSIVE MULTI-PARAMETRIC MAGNETIC RESONANCE IMAGING TEXTURE**

Volume 08, Issue 12, Pages: 83-92.

Paper Authors

M.MAHESH, M.AMARANATH REDDY

SIR C.V. RAMAN Institute of Technology & Science, AP, India

USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper As Per UGC Guidelines We Are Providing A Electronic Bar Code

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

AUTOMATED PROSTATE CANCER DETECTION VIA COMPREHENSIVE MULTI-PARAMETRIC MAGNETIC RESONANCE IMAGING TEXTURE

M.MAHESH[1], M.AMARANATH REDDY[2]
1PG Scholar, Dept of ECE, SIR C.V. RAMAN Institute of Technology & Science, AP, India
2 Assistant Professor, Dept of ECE, SIR C.V. RAMAN Institute of Technology & Science, AP, India

ABSTRACT: Prostate cancer (CaP) is the second most diagnosed cancer in men all over the world. In the last decades, new imaging techniques based on magnetic resonance imaging (MRI) have been developed improving diagnosis. In practice, diagnosis is affected by multiple factors such as observer variability and visibility and complexity of the lesions. In this regard, computer-aided detection and diagnosis (CAD) systems are being designed to help radiologists in their clinical practice. We propose a CAD system taking advantage of all MRI modalities (i.e., T2-W-MRI, DCE-MRI, diffusion weighted (DW)-MRI, MRSI). The aim of this CAD system was to provide a probabilistic map of cancer location in the prostate. We extensively tested our proposed CAD using different fusion approaches to combine the features provided by each modality. The source code and the dataset have been released**.**

1.INTRODUCTION

Cancer has become a severe threat to human lives due to its prevalence. According to the American Cancer Society [1], the estimated number of new cases, for all types of cancer in the U.S. for the year 2013 is 1,660,290, and the estimated number of deaths from cancer is 580,350. In men, prostate cancer is the most prevalent cancer type (other prevalent cancers are lung, colorectum, and bronchus). The estimated new cases of prostate cancer is 238,590 Current prostate cancer (CaP) screening consists of 3 different stages. First, prostate-specific antigen (PSA) control is performed to distinguish between low- and high-risk CaP. To assert such diagnosis, samples are taken during prostate biopsy and analyzed to make an accurate prognosis of the CaP. Although PSA screening has been shown to improve early detection of CaP [1], its lack of reliability motivates further investigations using magnetic resonance imaging (MRI)-based computer-aided detection and diagnosis (CAD). Consequently, current research is focused on identifying new biological markers to replace PSA-based screening [2]. Until such research comes to fruition, these needs can be met through active-surveillance strategy using multi parametric MRI (mp-MRI) techniques [3].

Lemaitre et al. recently reviewed more than 50 research works that focused on CAD system for CaP [4]. These studies are based on CAD systems that consists of the following steps: (i) pre-processing, (ii) segmentation, (iii) registration, (iv) feature

detection, (v) feature selection-extraction, and (vi) finally classification.The reviewed mp-MRI-based CAD used 2 to 3 MRI modalities among T2 Weighted (T2-W)-MRI, dynamic contrast-enhanced (DCE)-MRI, and diffusion weighted (DW)-MRI, discarding the potential discriminative power of magnetic resonance spectroscopy imaging (MRSI). Furthermore, only half of these studies tackled the challenging detection of CaP in the central gland (CG). Additionally, none of the works investigated the issue related to feature balancing when developing their CAD systems. Finally, none of the datasets nor source codes used have been released, making impossible the possibilities to compare the methods.

In this work, we propose a CAD system to detect CaP in peripheral zone (PZ) and CG, using the 4 aforementioned MRI modalities. The ultimate goal of the current CAD system is to provide a probabilistic map of the cancer within the prostate. Therefore, each voxel in the prostate will be classified as healthy or cancerous. The dataset used and the source code developed are released for future comparisons and reproducibility.

The distribution of prostate cancer incidences around the world is illustrated in Figure 1.1. According to this map, we can see that prostate cancer is concentrated in North Amer-ica, Australia, and Northern Europe. Detailed information about the prostate and prostate cancer can be found. In this chapter, we summarize information about prostate cancer extracted from these studies.Prostate is a gland in the male body. As can be seen in Figure 1.2, the prostate is located next to the bladder. The function of the prostate is to create the uid to protect the sperm cells. A commonly occurring problem in the prostate is \Benign Prostatic Hyperplasia", in which the size of the prostate increases as men get older. The enlarged prostate obstructs the urethra, which makes it di cult for passing urine. Despite having similar symptoms with prostate cancer, \Benign Prostatic Hyperplasia" is not considered a cancer Figure 1.1: World map of prostate cancer incidence for the year 2008 [5]. For interpretation of the references to color in this and all other gures, the reader is referred to the electronic version of this dissertation.

2.RELATED WORK

The literature on prostate histopathological grading can be divided into two main categories, namely tissue image classification and cancer region detection. However, most of the studies address the tissue image classification problem. In this problem, the aim is to classify a tissue image into one of the two classes: normal and cancer, or into one of the multiple classes: normal, grade 3, 4 and 5 [40{44]. Only a few author have addressed the cancer detection problem, i.e., given a tissue image which includes both cancer regions and normal regions, automatically detect the cancer regions (regions of interest) in the image [3, 45, 46]. A major challenge in the cancer detection problem is that the ground truth data, namely the annotated cancer regions marked by pathologists, is not su cient to accurately evaluate an automated cancer detection method (this was previously mentioned in [45]). The ground truth, created as a hand drawn contour by pathologists not only contains cancer glands but also contains stroma (which is really the

background area). Furthermore, there are no standardized rules for the contour drawing, e.g., it may contain multiple glands or just a few glands, which is a subjective choice of the pathologist.To address the two problems mentioned above, the published studies have either used a segmentation-based approach (gland segmentation followed by gland feature extraction) or used a segmentation-free approach (directly extracting texture features from the image

Contributions

This thesis makes the following contributions to CAD systems for histopathological grading of prostate cancer (Figure 1.10):

1. A lumen-based gland segmentation method, termed nuclei-lumen-association (NLA), that considers lumen as the central component of the gland, and nd nuclei associated with the lumen. The NLA method leads to better gland classi cation results than the level set method, and comparable gland classi cation results to the manual gland segmentation method.

2. The gland features that capture rich information about the gland, including structural and contextual information. We show that these proposed gland features are (i) better than gland features used in the published studies in solving the gland classi cation problem and (ii) better than texture features used in the published studies in solving the tissue image classi cation problem.

3. A nuclei-based gland segmentation method that models the relationship between nuclei and lumina in the image by a nuclei-lumina-graph. The edges in the graph correspond to the links between nuclei and nuclei and between nuclei and lumina. The recursive normalized cut method is applied to partition the graph into di erent components, each of which corresponds to a gland. The nuclei-based method overcomes the major limitation of the conventional lumen-based methods: it is able to detect glands without lumen and glands with multiple lumina.

   A method to compute the gland-score of a segmented region that measures how similar the segmented region is to a gland unit. The gland-scores are combined with the structural-contextual features to improve the grade 3 vs grade 4 tissue image classi - cation result.

5. Both the lumen-based and nuclei-based methods are evaluated on both low magni- cation (5) and high magni cation images (20) to evaluate the e ect of image magni cation on the performance of these methods.

6. A gland-based method to compute the similarity between two tissue regions. This method can be used to search for regions similar to a region of interest (ROI) in the annotated tissue slides. The retrieved regions can serve as the references that a medical student or a technician can rely upon when grading the ROI.

7. A method to extract an important cytological feature in the prostate tissue images, i.e. the presence of nuclei with prominent nucleoli. To our knowledge, this cytological feature has not yet been exploited in any CAD system. We also demonstrate the usefulness of this cytological feature by applying it in a prostate cancer detection framework.

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

3.PROPOSED METHOD

The mp-MRI data are acquired from a cohort of patients with higher-than-normal level of PSA. Acquisition is achieved with a 3 T whole body MRI scanner (Siemens Magnetom Trio TIM, Erlangen, Germany) using sequences to obtain T2-W-MRI, DCE-MRI, DW-MRI, and MRSI. In addition of the MRI examination, these patients also have undergone a transrectal ultrasound (TRUS) guided-biopsy. The dataset is composed of 17 of which have biopsies that were positive for CaP. In all 12 patients have a CaP in the PZ, 3 patients have CaP in the CG, 2 patients have invasive CaP in both the PZ and the CG. An experienced radiologist segmented the prostate organ — on T2-W-MRI, DCE-MRI, and apparent diffusion coefficient (ADC) — as well as the prostate zones — i.e., PZ and CG —, and CaP on the T2- W-MRI. The full description and the data set are available at I2Cvb

CAD pipeline for CaP

Our mp-MRI CAD system consists of 7 different steps: pre-processing, segmentation, registration, feature detection, feature balancing, feature selection/extraction, and finally classification. The different source codes are publicly available2.

Pre-processing: Normalization is, a crucial step to reduce the inter-patient variations which allows to improve the learning during the classification stage. However, the MRI modalities provide specific type of data — static vs. dynamic information, images vs. signals — that required a dedicated pre-processing. Therefore, we pre-process differently the data: T2-W-MRI is normalized using a Rician apriori that has been shown to be better than the traditional z-score [6]. In contrast to T2-W-MRI, in ADC map the probability density function (PDF) within the prostate does not follow a known distribution and thus one cannot use a parametric model to normalize these images and a nonparametric piecewise-linear normalization [7] is the best option for this case. DCE-MRI is a dynamic sequence and the data are normalized based on a mean kinetic expression registration as proposed in [5]. Finally, the MRSI modality has been pre-processed to correct the phase, suppress the baseline, and align the frequencies

Segmentation and registration: For this work, our radiologist has manually segmented the prostate organs on the different modalities. However, the segmented prostate needs to be registered before to extract features. Therefore, the patients motion during the DCE-MRI is corrected using a rigid registration with an mean squared error (MSE) similarity metric and a gradient descent optimizer. Subsequently, the T2-W-MRI and DCE-MRI are co-registered using a rigid transformation and the delineation of the prostate gland, using the same metric and optimizer previously mentioned. ADC maps and T2-W-MRI are also co-registered with the same strategy. Additionally, volumes from all modalities have been interpolated to the resolution of T2-W-MRI. used gland features to classify the tissue image rather than explicitly classifying the individual glands.only used gland size to classify the tissue images into normal and cancer images

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

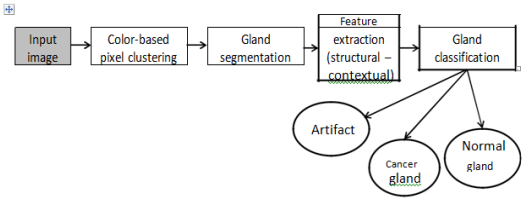


Figure .Proposed method for gland segmentation and classification

Gleason grade 3 or Gleason grade 4. By denoting the area, perimeter and boundary contour of the lumen as A; P , and B, respectively, they computed the following shape features for the lumen:

Gland Structure

A gland consists of epithelial nuclei, epithelial cytoplasm and lumen (Figure 2.2, [2], [56]). In between the glands is the stroma which can be considered as the background. Artifacts ap-pear as non-lumen white regions in the tissue. We perform gland segmentation by employing the following two steps.

Tissue Component Identification:

To facilitate segmentation, we rst identify the basic components in the tissue by using a color-based clustering procedure. By utilizing the di erences in color of the four tissue components in an H&E image [(stroma (S) color is mostly pink, lumen (Lu) is mostly white, nucleus (N) is mostly blue and cytoplasm (C) is mostly purple), we perform clustering in the Lab color space by using the k-means algorithm (k = 4). For computational e ciency, in each image, we run the k-means algorithm on 10,000 randomly selected pixels to nd 4 clusters. By using the L and a values of the cluster centers, denoted by L^0 and a^0, respectively, we are able to match the four clusters with the four tissue components (nuclei, stroma, cytoplasm, and lumen) by the two following rules (see Figure 2.8b for an illustration):

1. $L^0_N < L^0_S$; $L^0_C < L^0_{Lu}$: Nuclei cluster has the lowest L value (since nuclei have the dark-est color), while lumen cluster has the highest L value (since lumen has the brightest color).
2. $a^0_S > a^0_C$: Stroma cluster has a higher a value than cytoplasm cluster since the color of stroma is more red than cytoplasm.

Next, each pixel in the image is assigned to the tissue component corresponding to the nearest cluster center (Figure 2.8c). We apply a connected component algorithm [58] on nuclei pixels and lumen pixels to generate nuclei objects and lumen objects, respectively, which are later used for segmentation. Although the list of lumen objects also contains artifacts, we cannot discard them easily by a simple size thresholding operation because these artifacts can be as large as the lumen.

Nuclei-Lumen Association

The gland segmentation algorithm starts with lumen as the rst component of a gland and then searches for the nuclei for that gland. The proposed algorithm (Figure 2.7), which is referred to as nuclei-lumen association (NLA) algorithm, associates appropriate nuclei with each lumen to create a gland segment. Nuclei are searched along a direction normal to the lumen boundary contour. The algorithm consists of three steps:

Given n points on the lumen boundary, we sample n=3 points uniformly at equal interval, called lumen points. The number of points is selected by considering the trade-o

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

between a sparse (for computational e ciency) and dense set (for adequate search coverage).

2. A search region ($_S$) of a conical shape, centered at each lumen point, is expanded to nd nuclei on the gland boundary. When nuclei pixels are found within $_S$, we use a circular mask to select the nuclei region to join the gland boundary. This step is repeated for all lumen points to nd the complete gland boundary.
3. A pruning procedure, based on the median absolute deviation (MAD), is applied to remove outlier nuclei and generate a smoother segmentation boundary[2].

Besides gland segments, the proposed algorithm also produces a set of points located at the detected nuclei, referred to as the nuclei point set. The nuclei point set and lumen point set are used for gland feature extraction. A detailed description of the algorithm is given in Algorithm 2.1.

Since artifacts are included in the list of lumen objects, some non-gland segments created by artifacts are also present in the segmentation result (Figure 2.9a). However, instead of detecting them at this step, we identify them in the classi cation procedure, which is more reliable since it uses a larger number of features.

SVM

In machine learning, **support vector machines** are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis In addition to performing linear classification, SVMs can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces.

When data are not labeled, supervised learning is not possible, and an unsupervised learning approach is required, which attempts to find natural clustering of the data to groups, and then map new data to these formed groups. The clustering algorithm which provides an improvement to the support vector machines is called **support vector clustering** and is often used in industrial applications either when data is not labeled or when only some data is labeled as a preprocessing for a classification pass More formally, a support vector machine constructs a hyper plane or set of hyper planes in a high- or infinite-dimensional space, which can be used for classification, regression, or other tasks. Intuitively, a good separation is achieved by the hyperplane that has the largest distance to the nearest training-data point of any class (so-called functional margin), since in general the larger the margin the lower the generalization of the classifier.

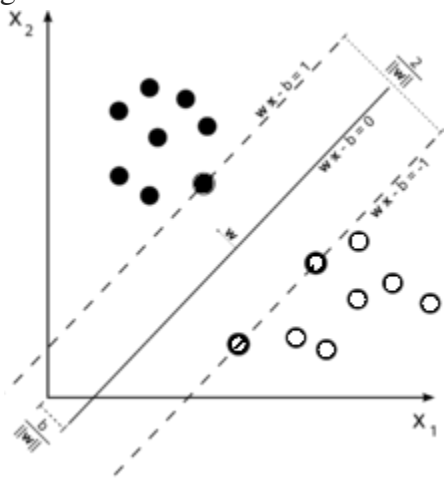


Figure 4.2 Illustrting Linear SVM along with graph

Maximum-margin hyperplane and margins for an SVM trained with samples

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

from two classes. Samples on the margin are called the support vectors where is the (notnecessarily normalized) normal vector to the hyperplane. The parameter determines the offset of the hyperplane from the origin along the normal vector

$$\vec{w} \cdot \vec{x} + b = 1$$

and

$$\vec{w} \cdot \vec{x} + b = -1.$$

Similarly to the pre-processing, specific features are extracted depending of the specificity of each MRI modality. T2-W-MRI and ADC map features Additionally to the normalized intensity, edge- and texture-based features are commonly extracted from T2-W-MRI and ADC map. The following set of filters characterizing edges have been used: (i) Kirsch, (ii) Laplacian, (iii) Prewitt, (iv) Scharr, (v) Sobel, and (vi) Gabor. Except for the Kirsch filter, the other filters are applied in 3D, taking advantage of the volume information instead of slice information, as it is usually done. Additionally, features based on phase congruency are computed [9]. To characterize the local texture, both second-order graylevel co-occurence matrix (GLCM)-based features [10] and rotation invariant and uniform local binary pattern (LBP) [11] are extracted. To encode 3D information, the 13 first Haralick features are computed for the 13 possible directions. For the same reason, the LBP codes are computed for the three-orthogonal-planes of each MRI volume. All these features are extracted at each voxel of the volume.

DCE-MRI features In brief, the entire enhanced signal, semi-quantitative [12], and quantitative-based models [13], [14], [15], [16] are computed.

MRSI features Three different techniques are used to extract discriminative features: (i) relative quantification based on metabolite quantification, (ii) relative quantification based on bounds integration, and (iii) spectra extraction from 2 ppm to 4 ppm [5].

Anatomical features Four different metrics are computed based on the relative distance to the prostate boundary as well as the prostate center, and the relative position in the Euclidean and cylindrical coordinate systems [17],[18].

Feature balancing:

Imbalanced dataset is a common problem in medical imaging. The number of cancerous voxels is much lower than the number of "healthy" voxels for a patient. This problem compromises the learning process. Solving the problem of imbalanced is equivalent to under- or over-sampling part of the dataset to obtain equal number of samples in both classes. In this regard, the imbalanced dataset was under-sampled using the different variant of near miss (NM) [19] and the instance hardness threshold (IHT) algorithm. In addition, the dataset was also balanced using over-sampling methods, namely different variant of synthetic minority over-sampling techniques (SMOTE) [21], [22]. Those algorithms were developed and made publicly available in the scikitlearn-contrib imbalanced-learn3 python package [23].

Feature selection and extraction:

Feature selection and extraction are used in our experiment. MRSI and DCEMRI are decomposed using three

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

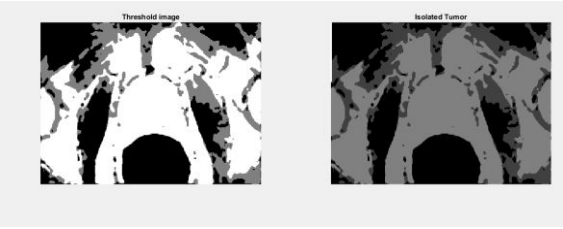
feature extraction methods: principal components analysis (PCA), sparse-PCA, and independent components analysis (ICA) are used to decompose signal-based data. Additionally to feature extraction, two methods of feature selection are used: (i) the one-way analysis of variance (ANOVA) and (ii) the Gini importance obtained while learning the random forest (RF) classifiers.The scikit-learn4 python package provides all those methods and was use in our experiment [24].

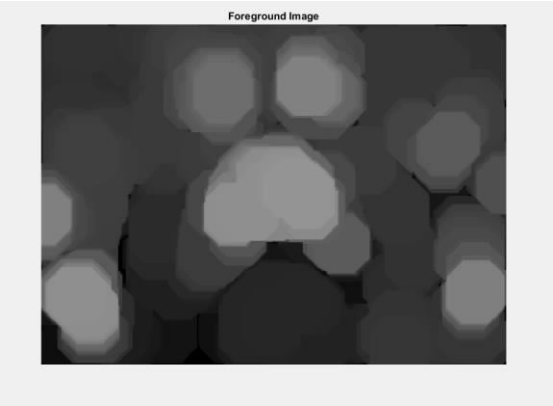
6) Classification: RF has been chosen as our base classifier to perform classification of individual modality as well as the combination of modalities. RF and more generally decision trees do not required to scale features and provide a feature selection add-on by analyzing the feature importance derived from the impurity improvement successive splits. Additionally, we use stacking to create ensemble of base learners using a meta-classifier [25], namely AdaBoost (AdB) and Gradient Boosting (GB).

Performance Evaluation

the database used in this experiment includes 48 tissue image regions at 5 magni cation (average size is 900 1,500 pixels). Given the pathologist's annotation for each tissue image region, we manually labeled 525 artifacts, 931 normal glands and 1,375 cancer glands to form the labeled (ground truth) gland database. We also implemented the state-of-the-art methods in [2] and [3] to compare their performance with the proposed segmentation (NLA) and classi cation methods. The segmentation process for all the three methods starts with the lumen objects identi ed in each image by the k-means clustering procedure discussed in section 2.3. Since the same set of lumen objects and the same ground truth information (which is not a ected by lumen objects) are used for all the three methods, the comparison among the three methods is fair.

4.SIMULATION RESULTS





5.CONCLUSION

In this paper, we presented one of the the first CAD system using all the mp-MRI modalities for prostate cancer detection. Indeed, MRSI has nearly never been used together with the other modalities. With an extensive validation approach to select the

best features, the best balancing strategy as well as the best classifier, we obtained results on a rather complicated dataset of 17 patients with an average AUC of 0:836 _ 0:083 which put our system in the state-of-the-art, even so different CADs were tested on different datasets. As avenues for future research, one could switch from voxel-based classification to super-voxel classification such that spatial structure are classified instead of voxel. In addition, the registration relies on the segmentation of the prostate gland which was provided by our doctors. To be used in a clinical environment, this step need to be fully automatized.

REFERENCES

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

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

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

[4] G. Lemaitre, R. Marti, J. Freixenet, J. C. Vilanova, P. M. Walker, and
F. Meriaudeau, "Computer-aided detection and diagnosis for prostate
cancer based on mono and multi-parametric mri: A review," Computers
in Biology and Medicine, vol. 60, pp. 8–31, 2015.

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

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

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

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

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

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

[11] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," IEEE Transactions on pattern analysis and machine intelligence, vol. 24, no. 7, pp. 971–987, 2002.

[12] H. J. Huisman, M. R. Engelbrecht, and J. O. Barentsz, "Accurate estimation of pharmacokinetic contrast-enhanced dynamic MRI parameters of the prostate," J Magn Reson Imaging, vol. 13, no. 4, pp. 607–614, Apr 2001.

[13] G. Brix, W. Semmler, R. Port, L. R. Schad, G. Layer, and W. J. Lorenz, "Pharmacokinetic parameters in cns gd-dtpa enhanced mr imaging." Journal of computer assisted tomography, vol. 15, no. 4, pp. 621–628,

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org