# THE DEVELOPMENT OF A MULTI-STAGE LEARNING SCHEME USING NEW TISSUE DESCRIPTORS FOR AUTOMATIC GRADING OF PROSTATIC CARCINOMA

*Clara Mosquera-Lopez, Sos Agaian**

The University of Texas at San Antonio
Dept. of Electrical and Computer Engineering
San Antonio, TX, USA

*Alejandro Velez-Hoyos*

Pablo Tobon Uribe Hospital
Dept. of Pathology
Medellin, Colombia

## ABSTRACT

This paper introduces a new system for the automated classification of prostatic carcinomas from biopsy images. The important components of the proposed system are (1) the new features for tissue description based on hyper-complex wavelet analysis, quaternion color ratios, and modified local binary patterns; and (2) a new framework for multi-stage learning that integrates both multi-class and binary classifiers. The system performance is estimated by employing Hold-out cross-validation in a dataset of 71 prostate cancer biopsy images with different Gleason grades. Simulation results show that the presented technique is able to correctly classify images in 98.89% of the test cases. Furthermore, the system is robust in terms of sensitivity (0.9833) and specificity (0.9917). We have demonstrated the efficacy of our system in distinguishing between Gleason grades 3, 4 and 5.

***Index Terms—*** Automated Gleason grading, multi-classifier systems, histopathology image analysis, quaternion features

## 1. INTRODUCTION

Prostate cancer (PCa) is the second most diagnosed cancer among American men. It is estimated that about 233,000 new cases will be diagnosed and 29,480 men will die of prostate cancer in 2014 [1]. The gold standard method for the PCa diagnosis is the analysis of needle biopsy tissue sections [2], which consists of the visual evaluation of histopathology images and the assignment of Gleason grades. The Gleason score, composed by the most predominant Gleason grades, is the most important prognostic indicator for prostate cancer [3]. The Gleason grade characterizes the degree to which the tumor resembles healthy tissue. Examples of histopathology images of different Gleason patterns are shown in Figure 1.

The process of assigning Gleason grades to a histopathology image is subjective and time-consuming due to its heavy reliance on physician interpretation and tissue complexity, which often leads to high levels of intra- and interobserver variability [4]. Intending to increase the reproducibility of the grading process and to save pathologists time, several computer-aided diagnosis (CAD) systems have been proposed [5–13]. Each method has its own advantages and shortcomings. In general, researchers have used a variety of features obtained from textural, architectural, and morphometric analysis to classify histopathology images. For instance, Diamond et al. [5] employed glandular and nuclear area as well as Haralick features. Tabesh et al. [14] used a combination color channel histograms, fractal analysis, and wavelets along with color, texture, and morphometric properties of histological objects in order to detect cancerous regions and discriminate between low- and high-Gleason grades. Naik et al. [6] developed a CAD system that extracts morphological and Voronoi, Delaunay, and minimum spanning tree graphs features in order to solve binary classification problems: benign epithelium vs grade 3, benign epithelium vs grade 4, and grade 3 vs grade 4. Huang et al. [7] proposed a set of fractal-based features to classify cancerous patterns belonging to Gleason 1 to 5. Nguyen et al. [8] employed statistics of size, density and shape of tissue structures to classify benign tissue and Gleason grades 3 and 4. Almuntashri et al. [9] presented a method that combines features from wavelet transform and wavelet-based fractal analysis.

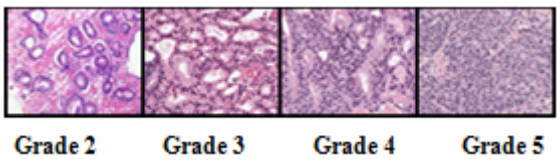


**Fig. 1**. Examples of Gleason patterns 2-5

Although the aforementioned and other techniques have been proposed so far, the development of new algorithms for CAD of PCa is an open problem. The performance of grading/scoring systems has to be further improved, especially when classifying intermediate Gleason patterns 3 and 4. Developed systems should clearly demonstrate that the accuracy

*This research has been partially supported by the NSF Grant (RD-0932339)

of interpretation of screening biopsy images with CAD is better than the one without using CAD. Therefore, there is still a long way to go before CAD systems for PCa become available commercially and widely used in clinics and screening centers.

The goal of this paper is to introduce a new system for the automated classification of prostatic carcinomas from biopsy images in a reproducible manner using a multi-classifier learning scheme. The rest of this paper is organized as follows: Section 2 introduces the new Gleason grading system including discussion on feature extraction methods and classification scheme; Section 3 presents experimental results, and Section 4 concludes the paper.

## 2. OVERVIEW OF THE PROPOSED PROSTATE CANCER GRADING SYSTEM

In the literature, multi-class classification problems such as PCa grading are tackled using two common approaches: (1) one-shot classification (OSC), and (2) decomposition of the problem into a set of binary problems. OSC is especially limited and may produce large classification errors when dealing with multiple similar classes [12]. The second approach is to build a system of multiple binary classifiers using one-vs-all (OVA), one-vs-one (OVO), and simultaneous classification [13]. Most of the published works on PCa grading use one of the above mentioned approaches.

Another strategy to grade prostatic carcinomas was developed by Doyle et al. [12]. In their cascaded approach, successive classications are performed, beginning with the most broad (i.e. cancer detection) and proceeding to increasingly granular separations (Gleason grading). However, the reported simulation results show that the cascaded approach only outperforms the OSC and OVA schemes in terms of the positive predictive value (PPV).

In this work we present a new system for the assignment of Gleason grades to prostate cancer histopathology images. This is an extension of our previous works in [10,11] and provides an integrated framework for multi-class classification that can be used in other applications. The general architecture of the system with the specific components used for the application discussed in this paper is illustrated in Figure 2. The proposed system consists of a two-stage classifier. First, we use OSC with a group of features to assign an initial label to an input pattern. Second, the classification is refined by using several pairwise SVM classifiers, which are trained using different feature vectors. The second level of the system is activated only if two or more classes are close to each other, producing large probability for classification error.

### 2.1. Feature extraction

In this work, we use six different feature extraction methods. Three new feature sets used with the Bayesian classifier are computed from the quaternion wavelet transform (QWT) coefficients, quaternion ratios, and the histograms of multi-resolution local binary patterns. The remaining feature vectors used with the two-class SVM classifiers are based upon the discrete Haar wavelet transform, color fractal dimension, and morphometric characteristics of the tissue, which were previously developed by the authors in [10, 11].

#### *2.1.1. Quaternion wavelet features*

The QWT is a hyper-complex extension of the real wavelet transform. It allows us to analyze the textural complexity of images since the magnitude and local phase information contains rich structural information. Although the QWT has been applied to texture classification problems, it has not been applied to PCa grading before. More details about the computation of the QWT are given in [15]. As in the case of real wavelet coefficients, the resulting QWT coefficients are separated into 4 sub-bands at each decomposition level, but they are quaternion numbers of the form $q = a + bi + cj + dk$, whose magnitude is given by $|q| = \sqrt{a^2 + b^2 + c^2 + d^2}$. We construct the QWT feature vector with statistics (mean and standard deviation) of the coefficients magnitude and phase along with the energy of coefficients at each decomposition level. The energy is computed as indicated below:

$$E_{s,j} = \frac{1}{I_{0E}} \sum_{x,y} |q_{x,y}|^p \tag{1}$$

$E_{s,j}$ is the energy of the sub-band $s = \{1, 2, 3, 4\}$ corresponding to approximation, horizontal, vertical, and diagonal sub-bands; and $j = 1, 2, ...J$ is the QWT decomposition level. $I_{0E}$ is a normalization term proportional to the energy of the original input image; and $p \geq 2$ is a real tuning parameter.

#### *2.1.2. Quaternion ratio features*

Quaternion representation is used in this section for representing a color triplet in the RGB color model. An RGB image pixel is represented as a pure quaternion as $q = Ri+Gj+Bk$. Then, each image is converted into a matrix of pure quaternion numbers of size $M \times N$. For each pixel in the quaternion input image, a color descriptor is computed using a $n \times n$ neighborhood of surrounded pixels. The value of $n$ should be an odd number and its minimum value is 5. The square neighborhood is then divided into four overlapping sub-regions $R_i$ (as shown in Figure 3) that do not include the central pixel. Next, the expected value of the quaternion numbers in each sub-region is computed using the following expression:

$$E\left[q_{R_i}\right] = E\left[R_{R_i}\right] + E\left[G_{R_i}\right] + E\left[B_{R_i}\right] \tag{2}$$

The local descriptor at a central pixel $g_c$, located at the position $x, y$, is computed as follows:

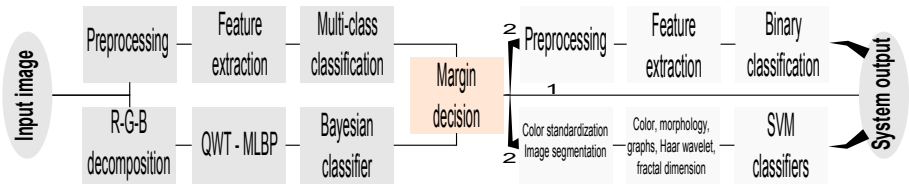


**Fig. 2**. Flow diagram of the proposed multi-classifier system

$$QRBP\left(x,y\right)=\sum_{p=0}^{7}w_{p}f\left(q_{c},q_{p}\right)\cdot 2^{p} \tag{3}$$

In equation (3),

$$q_{p}=\begin{cases} E\left[q_{R_{p}}\right] & if\ \ p=0,1,2,3 \\ q\left(x-1,y\right) & if\ \ p=4 \\ q\left(x,y+1\right) & if\ \ p=5 \\ q\left(x+1,y\right) & if\ \ p=6 \\ q\left(x,y-1\right) & if\ \ p=7 \end{cases} \tag{4}$$

The function $f\left(\right)$ is a thresholding function defined as:

$$f\left(q_{c},q_{p}\right)=\begin{cases} 1 & if\ \sum_{i=1}^{3}q_{p,i}-q_{c,i}\geq T \\ 0 & otherwise \end{cases} \tag{5}$$

Finally, each $w_{p}=\frac{q_{p}}{q_{c}}$ is a quaternion ratio. The resulting quaternion weights $w_{p}^{'}$ incorporate interactions among the different color channels. For instance, the dot product between color vectors represents intra-channel interactions and the cross product represent inter-channel relationships.

After building the quaternion matrix representing the pixel ratio and binary patterns, the singular value decomposition (SVD) of the resulting matrix QRBP is computed to generate the feature vector for image classification from the singular values of the quaternion matrix. In this paper, we use the complex adjoint matrix representation of a quaternion matrix for quaterion SVD computation.

#### *2.1.3. Multi-resolution local binary pattern features*

A number of features are extracted based on a proposed modification of the local binary pattern (LBP) operator [16]. We compute the feature vector by defining multi-resolution LBP (MLBP) using low-resolution images resulting from applying the discrete wavelet transform decomposition instead of varying the neighborhood radio. The value of the LBP code of a center pixel $g_{c}$ in a $3\times 3$ neighborhood at the $j^{th}$ decomposition level is given by:

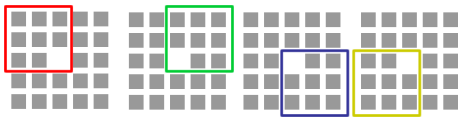

**Fig. 3**. Overlapping sub-regions

$$MLBP_{8,j}\left(x,y\right)=\sum_{p=0}^{7}f\left(g_{p}-g_{c}\right)\cdot 2^{p} \tag{6}$$

The function $f\left(x\right)$ is defined as $f\left(x\right)=1$ if $x\geq T$, and $f\left(x\right)=0$ otherwise. The described operation is performed for values of $T\in\left[0,30\right]$. In this work, we compute the normalized histogram of uniform patterns for each color channel independently and then we concatenate them. The final $MLBP$ feature vector is composed by the low-frequency components of the resulting 1-D histogram signals.

#### *2.1.4. Features for binary SVM classifiers*

The features utilized with the SVM classifiers were developed by the authors in [10, 11]. The first set [11] integrates color, gland morphology and architectural characteristics of tissue structures obtained after image segmentation. These vectors are more effective when distinguishing Gleason grade 3 vs 4. The second set of features [10] is based on textural analysis. We extended the definition of color fractal dimension to increase the separation between intermediate Gleason patterns by generating a new color model and weighting functions for the computation of the fractal dimension. Also, new features were proposed based on the joint probability of color channel wavelet coefficients in order to exploit inter-channel dependencies. These feature vectors perform very well recognizing Gleason grade 4 vs 5. In this paper, we use the best-performing features for each specific classification refinement task.

### 2.2. Classification

For the multi-class problem we use a Bayes classifier [17]. However, any multi-class-capable classifier may be used. Bayesian classifiers employ the Bayes theorem to estimate the probability of an object represented by a feature vector $x$ of dimension $D$ being in class $c_{j}$ by assuming independency among features. If $\left(x,c_{j}\right)$, is a random variable with joint probability $p\left(x,c_{j}\right)$, then the Bayes classifier select the class

for an observation $x$ as $f(x) = \arg\max p(c_j|x)$. The a posteriori probabilities $p(c_j|x)$ are used to determine if the input image should be processed by the SVM sub-systems. We refine the classification outcome from the Bayesian classifier if two or more classes have similar probabilities or if the most probable classes are not contiguous Gleason patterns. We use a softmax function to map the a posteriori class probabilities to a finite interval $[0, 1]$ and construct a discrete probability distribution. From this distribution, we define whether or not an image has to be reprocessed by the SVM stage, by setting the minimum acceptable probability of the predicted class to 65% and the minimum difference between the probability of two or more classes to 10%. Furthermore, the two more probable classes define the set of features that should be used for classification refinement. For instance, the differentiation between intermediate Gleason grades 3 and 4 is more accurate using morphological features, whereas classification tasks between Gleason grades 4 and 5 are more accurate using textural features because some important tissue structures are not present or occluded. It is important to note that the pool of classifiers in the refinement stage mostly perform a single binary classification tasks between the most probable classes.

## 3. EXPERIMENTAL RESULTS AND DISCUSSION

Simulations were carried out on a database of 71 images of Hematoxylin and Eosin (H&E)-stained prostate tissue. The dataset contains 30 image regions of Gleason 3, 30 of Gleason 4, and 11 of Gleason grade 5. Since our image set is small in size and unbalanced, for bayesian classification we adjusted the prior probabilities according to the distribution of classes in the database and selected the best performing classification threshold using cross-validation. For all SVM classifiers, we used linear kernel given the high dimension of the feature spaces compared to the number of training samples. Hold-out cross-validation was employed to estimate the performance of the system. The average correct classification rate (CCR) and other indicators of system performance at each classification stage including sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) [18] are summarized in Table 1.

From Table 1, it can be observed that the accuracy of the multiclass classifier is improved after the refinement step. The average accuracy increases from 83.78% to 98.89%. In general, the performance indicators from the OSC step to the following refinement step were improved or remained the same in specific cases. The sensitivity value for the positive group Grade 5, which is the worst performing indicator of the Bayesian classifier reaches a remarkable 0.97 at the output of the whole system, after classification refinement.

For comparison purposes, Figure 4 shows the reported accuracy of various published Gleason grading systems. Note that only the experiments reported by the authors of this paper in [9–11] are directly comparable since the experiments were carried out using the same database and experimental protocols.

| | Grade 3 | Grade 4 | Grade 5 |
|---|---|---|---|
| Stage 1: OSC multi-class classification (Bayes classifier) | | | |
| CCR | 0.8000 | 0.8800 | 0.8333 |
| Sensitivity | 0.9800 | 0.7700 | 0.5200 |
| Specificity | 0.7100 | 0.9350 | 0.9051 |
| PPV | 0.7279 | 0.9051 | 0.9200 |
| NPV | 0.9928 | 0.9200 | 0.8400 |
| Stage 2: Classification refinement (2-class SVM) | | | |
| CCR | 0.9833 | 0.9933 | 0.9900 |
| Sensitivity | 0.9800 | 1.0000 | 0.9700 |
| Specificity | 0.9850 | 0.9900 | 1.0000 |
| PPV | 0.9847 | 0.9900 | 1.0000 |
| NPV | 0.9933 | 1.0000 | 0.9900 |

**Table 1**. Performance of the proposed system, averaged over 100 simulation runs

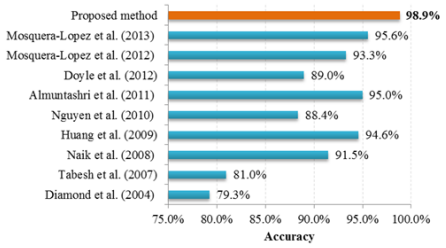


**Fig. 4**. Comparison of accuracy of Gleason grading systems

## 4. CONCLUSIONS

In this paper, we have developed a two-stage multi-classifier system for the assignment of Gleason grades (3,4, and 5) to H&E biopsy images. The system uses novel tissue descriptors to represent textural characteristics of the histopathology images. In this work, new quaternion features are introduced in order to represent color images and capture intra- and interchannel information that is often lost when color channels are processed independently. The features used in our system are combined such that each classification sub-system achieves the best possible performance. The system correct classification rate reaches 98.89% across all the considered Gleason grades. This performance indicator outperforms the accuracy of previously developed CAD systems for Prostate cancer and demonstrates the efficacy of the developed feature vectors. The preliminary results of this research must be validated using a larger cohort of images. We will focus our future work on integrating the concept of semi-supervised learning into the framework, which will allow us to train and test the system using sets of available unlabeled data.

## 5. REFERENCES


[1] American Cancer Society, "What are the key statistics about prostate cancer?," 2013.

[2] B. Djavan and M. Margreiter, "Biopsy standards for detection of prostate cancer," *World J. of Urology*, vol. 25, no. 1, pp. 11–17, 2007.

[3] G. T. Mellinger, D. Gleason, and J. Bailar, "The histology and prognosis of prostatic cancer," *J. Urology*, vol. 97, no. 2, pp. 331–337, 1967.

[4] C. Mosquera-Lopez, S. Agaian, A. Velez-Hoyos, and I. Thompson, *Computer-Aided Prostate Cancer Diagnosis: Principles, Recent Advances, and Future Prospective*, chapter 9, pp. 229–269, SPIE Press, Bellingham, WA, 2013.

[5] J. Diamond and et al., "The use of morphological characteristics and texture analysis in the identification of tissue composition in prostatic neoplasia," *Human Pathology*, vol. 35, pp. 1121–1131, 2004.

[6] S. Naik and et al., "Automated gland and nuclei segmentation for grading of prostate and breast cancer histopathology," in *5th IEEE Int. Symp. Biomed. Imag.: From Nano to Macro (ISBI)*, 2008, pp. 284–287.

[7] P. Huang and C. Lee, "Automatic classification for pathological prostate images based on fractal analysis," vol. 28, no. 7, pp. IEEE Trans. Med. Imag., 2009.

[8] K. Nguyen, A. Jain, and R. Allen, "Automated gland segmentation and classification for gleason grading of prostate tissue images," in *Int. Conf. on Pattern Recognition (ICPR)*, 2010, pp. 1497–1500.

[9] A. Almuntashri and et al., "Gleason grade-based automatic classification of prostate cancer pathological images," in *IEEE Int. Conf. on Systems, Man, and Cybernetics (SMC)*, 2011, pp. 2696–2701.

[10] C. Mosquera-Lopez and S. Agaian, "A new set of wavelet- and fractals-based features for gleason grading of prostate cancer histopathology images," in *SPIE Imag. Process.: Algorithms and Syst*, 2013, vol. 8655.

[11] C. Mosquera-Lopez and et al., "Exploration of efficacy of gland morphology and architectural features in prostate cancer gleason grading," in *IEEE Int. Conf. on Systems, Man, and Cybernetics (SMC)*, 2012, pp. 2849–2854.

[12] S. Doyle and et al., "Cascaded discrimination of normal, abnormal, and confounder classes in histopathology: Gleason grading of prostate cancer," *BMC Bioinformatics*, vol. 10, no. 282, 2012.

[13] A. Greenblatt, C. Mosquera-Lopez, and S. Agaian, "Quaternion neural networks applied to prostate cancer gleason grading," in *IEEE Int. Conf. on Systems, Man, and Cybernetics (SMC)*, 2013.

[14] A. Tabesh and et al., "Multifeature prostate cancer diagnosis and gleason grading of histological images," *IEEE Trans. Med. Imag.*, vol. 26, no. 10, pp. 1366–1378, 2007.

[15] R. Soulard and P. Carr, "Quaternionic wavelets for texture classification," *Pattern Recognition Letters*, vol. 32, pp. 1669–1678, 2011.

[16] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. and Mach. Intell.*, vol. 24, no. 7, pp. 971–987, 2002.

[17] F. Teng, Y. Chen, and X. Dang, "Multiclass classification with potential function rules:margin distribution and generalization," *Pattern Recognition*, vol. 45, pp. 540–551, 2012.

[18] A. G. Lalkhen and A. McCluskey, "Clinical tests: Sensitivity and specificity," *Continuing Education in Anaesthesia, Critical Care and Pain*, vol. 8, no. 6, pp. 221–223, 2008.