Automated Classification for Pathological Prostate Images using AdaBoost-based Ensemble Learning

Chao-Hui HUANG and Emarene Mationg KALAW
MSD International GmbH (Singapore Branch)
Agency for Science, Technology and Research, Singapore
Email: huangch.tw@gmail.com

***Abstract*—We present an AdaBoost-based Ensemble Learning for supporting automated Gleason grading of prostate adenocarcinoma (PRCA). The method is able to differentiate Gleason patterns 4–5 from patterns 1–3 as the patterns 4-5 are correlated to more aggressive disease while patterns 1-3 tend to reflect more favorable patient outcome. This method is based on various feature descriptors and classifiers for multiple color channels, including color channels of red, green and blue, as well as the optical intensity of hematoxylin and eosin stainings. The AdaBoost-based Ensemble Learning method integrates the color channels, feature descriptors and classifiers, and finally constructs a strong classifier.**

We tested our method on the histopathological images and the corresponding medical reports obtained from The Cancer Genome Atlas (TCGA) using 10-fold cross validation, the accuracy achieved 97.8%. As a result, this method can be used to support the diagnosis on prostate cancer.

I. Introduction

Prostate cancer remains the most prevalent form of of cancer among matured men and it is the second leading cause of cancer death in some countries. In 2015, only in the U.S., there are about 220,800 new cases of prostate cancer, and approximately 27,540 prostate cancer related deaths [1].

Blinded needle sextant biopsy is the current gold standard for prostate cancer diagnosis. During biopsy, 10 to 20 needle cores are obtained and slides are analyzed under a microscope by a pathologist [2]. Gleason patterns of prostate cancer are characterized by the morphological and architectural features of Hematoxylin and Eosin (H&E) stained histopathological tissue. A correlation between Gleason score and cancer aggressiveness has been found. In general, Gleason patterns 4 and 5 are highly correlated to more aggressive cancer while patterns 1, 2 and 3 are tend to reflect better patient outcome.

Under a microscope, the pathologist will determine the primary and secondary Gleason patterns. The final Gleason score is decided by summing up the primary and the secondly Gleason patterns. *E.g.*, In a slide where the primary Gleason pattern is 4 and the second dominant pattern is 3, the final Gleason score would be 7.

In the images of low Gleason patterns, there are relatively distinct glands and lumens. For high Gleason patterns, the structures of glands and lumens break down and the boundaries are no longer distinguishable. As the Gleason grading is performed by a human expert based on visually examining the pathological slides, manually distinguishing the difference between Gleason pattern 3 and Gleason pattern 4 can be difficult, as such kind of medical diagnosis often involve subjective judgment.

As a result, inter-observer and intra-observer variability is not preventable. For the sake of tackling this problem, many reports have been published in last few decades. Many researchers focused on searching a proper texture descriptor. *E.g.*, Khouzani *et al.* suggested a method based on analyzing the energy and entropy of the images after multiwavelet transform [3]; Alexandratou *et al.* showed the results of computer-aided prostate cancer grading based on texture analysis using gray-level occurrence matrix [4]; and Huang *et al.* proposed an approach of integrating fractal dimension into the problems of automated Gleason grading [5].

Some researchers looked into a specific classification algorithm for prostate cancer grading. *E.g.*, Sparks *et al.* suggested utilizing manifold regularization via statistical shape model of manifolds [6]. Nguyen *et al.* proposed an algorithm, in which, segmentation is performed before classification [7]. Tabesh *et al.* first introduced feature selection for improving the performance of automated prostate grading [8]. On the other hand, Doyle *et al.* discovered the capability of using AdaBoost [9] for the problems of Gleason grading [10].

In this work, we propose an approach of integrating various types of feature descriptors, different color channels and classifiers. For the color channels, we used all red, green and blue channels. Other than these conventional color channels, we also include the staining channels by using color deconvolution, including both hematoxylin and eosin. For the feature descriptors, we used Factral Dimensions (FD), Entropy-based Fractal Dimension Estimation (EBFDE) and Gabor filtering. For the classifiers, we used Support Vector Classifier with linear and RBF kernels, AdaBoost Classifier, Decision Classifier, Random Forest Classifier, Linear Discriminant Analysis and Quadratic Discriminant Analysis. All of these color channels, feature descriptors and classifiers are integrated by AdaBoost-based Ensemble Learning algorithm. In the following, first we will introduce the details of the method in section II. Then, its experimental result will be presented in section III. Finally, the conclusions will be addressed in section IV.

II. METHOD

The proposed method is introduced as follows: first, color deconvolution and stain normalization for histopathology images is performed in order to extract the color channel of hematoxylin and eosin [11], [12]. From each image, 5 color channels are extracted, including red, green, blue, hematoxylin and eosin. Each of the channels is represented in as a 2-D gray-level image. Then, we compute various feature descriptors for all color channels. The feature descriptors include Fractal Dimension (FD), Entropy-based Fractal Dimension Estimation (EBFDE) and Gabor filter bank set. Thus, each input image is represented by 15 different types of features.

Third, we prepare a pool of classifiers. We use all features of all training images to train all classifiers. We wish to compose a strong classifier by properly selecting the feature descriptors and classifiers. For this we use AdaBoost-based Ensemble Learning [9]. During the training phase of AdaBoost, in order to prevent the over-fitting problem, we used k-fold cross validation to evaluate the performance for each combination of feature descriptors and classifiers. Finally a strong classifier is generated, which is a series of combination of feature descriptors and classifiers ans an weight is assigned for each combination. The further details will be introduced as follows:

A. Color Deconvolution

Inconsistencies in staining is almost not preventable in histopathological slide preparation. It is due to various causes, including the thickness variation of the specimen, condense of staining solution, *etc.* Inconsistencies in staining makes it difficult to perform quantitatively analysis on the slides [11], [12]. Thus, color deconvolution and stain normalization for histopathology images is required in order to perform quantitatively analysis. There are a few reports focusing on the topics of color deconvolution and stain normalization for histopathology image. In our experiments, we used the method proposed by Macenko *et al.*

B. Feature Descriptors

1) Features of Fractal Dimension by Box-Counting: Given a gray-scale image of size $M \times M$, assume we segment the image into a grid of smaller patches. Assume the size of each patch is $s \times s$, where $s \leq M/2$ and $M, s \in \mathbb{I}_{\leq 0}$, we can have a scale ratio of $r = s/M$. Assume the gray-scale image can be represented by the 3-D plot of a function, say $z = f(x, y)$, where (x, y) represents the location on the image and z is the gray-scale value at (x, y). The space of (x, y) at the least can be divided into a grid of $\lfloor M/s \rfloor^2$ patches (without concerning the residuals). Assume the image has G gray-level (*e.g.*, $G = 256$), then we can have $\lfloor G/h \rfloor = \lfloor M/s \rfloor$, where h represents a step size along the gray-level. Thus, each z falls into a box of which the size is $s \times s \times h$, and there is a stack of $\lfloor G/h \rfloor$ boxes for each patch on the grid.

Now, we give index (i, j) for the stack of each patch. In this patch, assume the maximal and minimal values of z fall into the k^{th} and l^{th} boxes of the stack. The contribution of N_r in the $(i, j)^{th}$ patch is calculated as:

$$n_r(i, j) = k - l + 1, \text{ and } N_r = \sum_{i,j} n_r(i, j). \quad (1)$$

N_r is computed over different scales of r. Then, the fractal dimension can be calculated from the slope of line approximated by least-square linear fitting for $\log(N_r)$ v.s. $\log(r)$.

2) Entropy-based Fractal Dimension Estimation (EBFDE): The EBFDE method is first introduced by Huang and Lee [5]. In EBFDE, first we set a parameter s which defines the size of the box in FD. Then, the given 2-D image, where the size of $M \times M$, is segmented into a grid of boxes, where the size of each box is $s \times s$. As a result, we can have a ratio $r = s/M$. Assume each box is identified by the index (i, j), we can compute the entropy for each box as follows:

$$e_r(i, j) = -\sum_{k=0}^{G} p_k log_2(p_k), \quad (2)$$

where G is the number of gray-levels (*e.g.*, $G = 256$). For the given image, we can summarize the entropy for all (i, j) as follows:

$$E_r = \sum_{i,j} e_r(i, j)^2. \quad (3)$$

Again, the entropy of fractal dimension can be calculated from the slope of line approximated by least-square linear fitting for $\log(E_r)$ v.s. $\log(r)$.

3) Gabor Filtering: A Gabor filter can be considered as a sinusoidal plane of given frequency and orientation modulated by a Gaussian envelop. A 2-D Gabor filter is represented as follows:

$$g(x, y; \lambda, \theta, \psi, \sigma, \gamma) = \exp\left(-\frac{x'^2 + \gamma 2y'^2}{2\sigma^2}\right) \exp\left(i\left(2\pi\frac{x'}{\lambda} + \psi\right)\right), \quad (4)$$

where

$$x' = x\cos\theta + y\sin\theta \text{ and } y' = -x\sin\theta + y\cos\theta. \quad (5)$$

In the above equation, λ is the sinusoidal wavelength, θ represents the orientation, ψ is the phase offset, σ is the standard deviation of the Gaussian envelope and γ specifies the ellipticity. In general, we operate only the parameters of λ and θ while keeping the rest parameters fixed.

Gabor filtering has been used as a feature extraction method for distinguishing homogeneous textures [13]. *E.g.*, Gabor filtering bank set is a method which frequently be used in the study of texture classification. In a Gabor filtering bank set, a group of 2-D Gabor filters spanning over various λ and θ is arranged. Thus, given an image, each Gabor filter can extract various features, *e.g.*, *magnitude*, *energy* and *entropy* [5]. As a result, a multi-dimensional feature descriptor is obtained.

C. Classifiers

The pool of weak learners is composed by the combination of the FD mentioned in the previous section and various types of existing classifiers, including: k-nearest neighbors, support vector classifiers, decision trees, random forests, adaptive boosting, Gaussian Naïve Bayes, linear discriminant analysis, quadratic discriminant analysis, as well as these classifiers with different parameters.

D. AdaBoost-based Ensemble Learning

Adaptive Boosting Algorithm (*a.k.a.* AdaBoost) is a machine learning algorithm first proposed by Frund *et al.*. AdaBoost sometimes has been classified as a meta-algorithm for machine learning as it integrate many other types of classification algorithms in order to improve the performance. These algorithms are called *weak learners* in AdaBoost. During the training phase, a weak learner is assigned an weight according to its accuracy on the given classification task. The portions that the training patterns are wrongly classified will be corrected by other weak learners. As a result, the performance is "boosted".

We propose AdaBoost-based Ensemble Learning by integrating various feature descriptors and classifiers. In the experiments, we defined 4 types of feature descriptors and 5 classifiers. The feature descriptors include fractal dimension on red channel, fractal dimension on eosin staining channel, Gabor filtering on red channel and Gabor filtering on eosin staining channel. The classifiers include Support Vector Classifier with linear kernel, Support Vector Classifier with RBF kernel, AdaBoost classifier, Decision Tree classifier and Random Forest classifier. All of the classifiers are able to accept giving weights for training patterns. As a result, in total there were 100 weak classifiers.

Algorithm 1 The proposed algorithm

procedure
Input:
  Training patterns:
$$\mathbf{X} = \{(\mathbf{x}_i, l_i) | 1 \leq i \leq n, \mathbf{x}_i \in \mathbf{R}^{M \times N}, l_i \in \{1, -1\}\}$$
  Pattern weights: $w_1, \cdots, w_n \in \mathbf{R}_+$
  Feature Descriptors (see text):
$$d_1(\cdot), \cdots, d_m(\cdot) \text{ and each } d_i(\cdot) \in \mathbf{R}^{D_i}$$
  Classifiers: $h_1(\cdot), \cdots, h_p(\cdot) \in \{1, -1\}$
Initial:
  Each $w_i = 1/n$
  Weak Classifier Pool:
$$\mathcal{H} = \{h_1(d_1(\cdot)), \cdots, h_p(d_m(\cdot))\}$$
Begin:
  for $t = 1$ to T **do**
    Divide training patterns into K parts:
$$\mathbf{X} = \{X_1, \cdots, X_K\}$$
    for $k = 1$ to K **do** K-fold cross validation
      Train $h_i(d_j(\mathbf{x}_q)), \forall i, j$ and each $\mathbf{x}_q \notin X_k$
      Compute
$$e_{i,j,k} = \sum_q w_q l_q h_i(d_j(\mathbf{x}_q))$$
$$\forall i, j \text{ and each } \mathbf{x}_q \in X_k$$
    end for
    $E_{i,j} = \sum_{k=1}^{K} e_{i,j,k}, \forall i, j$
    if $\min_{i,j} E_{i,j} \geq 1/2$ **then**
      Stop
    end if
    $i_t^*, j_t^* = \arg\min_{i,j} E_{i,j}$
    $\alpha_t = \frac{1}{2} \log \frac{1 + E_{i_t^*, j_t^*}}{1 - E_{i_t^*, j_t^*}}$
    Remove $h_{i_t^*}(d_{j_t^*}(\cdot))$ from $\mathcal{H}$
  end for
Output:
  Strong Classifier:
$$\text{sign}\left(\sum_{t=1}^{T} \alpha_t h_{i_t^*}(d_{j_t^*}(\mathbf{v}))\right) \text{ for input } \mathbf{v}$$
end procedure

III. Results

In our experiment, the digital slides and medical reports were obtained from the dataset of PRAD of TCGA. We selected 28 high grade (both gleason pattern primary and gleason pattern secondary above or equal to 4) cases and 22 low grade (both gleason_pattern_primary and gleason_pattern_secondary below or equal to 3) cases. The magnification were $10\times$. An expert selected cropped the regions of interest for each digital slides. Each image was a 419×448 RGB image. In total, there were 333 high grade and 349 low grade images.

First, we validated the all classifiers using 10-fold cross validation. The accuracy across all color channels, feature descriptors and classifiers are shown in Table I. Then, we validated proposed algorithm (see Algorithm 1). The performance achieved 97.8%.

IV. Conclusion

In this paper, we presented AdaBoost-based Ensemble Learning for prostate adenocarcinoma cancer (PRAD) grading. The proposed method integrates various feature descriptors and color channels, by using PRAD histopathological images obtained from The Cancer Genome Atlas (TCGA), the accuracy achieved 97.8% in 10-fold cross validation.

The results suggested an application of discovering the critical regions on the digital slides. Thus, the pathologist will be able to exam these critical areas before screening over the whole slide image. As a result, the application will be able to reduce the workload from the pathologist.

TABLE I

THE ACCURACY (%) ACROSS DIFFERENT COLOR CHANNELS AND FEATURE DESCRIPTORS AND CLASSIFIERS. NOTE THAT F+E MEANS THE FEATURES COMBINE FD AND EBFDE.

KNeighbors

	F+E	Gabor
Red	77.2	91.7
Green	82	92.2
Blue	76	89.8
Luminance	80.9	92.8
Hematoxlyn	78.1	90.9
Eosin	89.3	92.5

SVC (linear)

	F+E	Gabor
Red	50	50
Green	50	50
Blue	50	50
Luminance	50	50
Hematoxlyn	50	50
Eosin	50	50

SVC (rbf)

	F+E	Gabor
Red	50	53.5
Green	50	51.5
Blue	52.4	58.7
Luminance	52.4	51.7
Hematoxlyn	43.2	50
Eosin	57	54.8

Decision Tree

	F+E	Gabor
Red	68.9	87.1
Green	83.2	90.7
Blue	70.1	84.2
Luminance	74.8	88.1
Hematoxlyn	80.1	85.2
Eosin	84.6	87.6

Random Forest

	F+E	Gabor
Red	76.1	85.3
Green	82	87.8
Blue	71.5	84.4
Luminance	74.9	87.8
Hematoxlyn	81.7	82.4
Eosin	84.8	86.7

AdaBoost

	F+E	Gabor
Red	74.4	90.3
Green	84.6	89.5
Blue	70.1	85.2
Luminance	77.7	87.9
Hematoxlyn	82.2	85.7
Eosin	81.7	91.7

Gaußian Naïve Bayes

	F+E	Gabor
Red	70.6	73.8
Green	79.3	78.2
Blue	69.4	78.9
Luminance	72	78.2
Hematoxlyn	75.4	77.7
Eosin	71.4	76.2

LDA

	F+E	Gabor
Red	64.7	86.7
Green	73.5	84.8
Blue	69.4	86.3
Luminance	72.6	87.5
Hematoxlyn	72.5	88.1
Eosin	79.1	87.2

QDA

	F+E	Gabor
Red	78.3	91
Green	81.5	90.6
Blue	78	89.1
Luminance	81.7	90.1
Hematoxlyn	80.6	90.4
Eosin	82.5	91.8

REFERENCES

[1] "Cancer facts and figures 2015," America Cancer Socity, Report, 2015 2015.

[2] D. F. Gleason, *The veteran's administration cooperative urologic research group: histologic grading and clinical staging of prostatic carcinoma*. Philadelphia: Lea and Febiger, 1977, pp. 171–198.

[3] K. J. Khouzani *et al.*, "Automatic grading of pathological images of prostate using multiwavelet transform," in *23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 3, 2001, Conference Proceedings.

[4] E. Alexandratou *et al.*, "Computer aided grading of prostate cancer based on texture analysis," in *6th European Symposium in Biomedical Engineering*, 2008, Conference Proceedings.

[5] P.-W. Huang *et al.*, "Automatic classification for pathological prostate images based on fractal analysis," *IEEE Trans. Med. Im.*, vol. 28, no. 7, pp. 1037–1050, 2009.

[6] R. Sparks *et al.*, "Gleason grading of prostate histology utilizing manifold regularization via statistical shape model of manifolds," in *Proc. SPIE 8315, Medical Imaging 2012: Computer-Aided Diagnosis, 83151J*, 2012, Conference Proceedings.

[7] K. Nguyen *et al.*, "Automated gland segmentation and classification for gleason grading of prostate tissue images," in *International Conference on Pattern Recognition*, 2010, Conference Proceedings.

[8] A. Tabesh *et al.*, "Automated prostate cancer diagnosis and gleason grading of tissue microarrays," in *SPIE Medical Imaging*, vol. 5747, 2005, Conference Proceedings, pp. 58–70.

[9] Y. Freund *et al.*, "A decision-theoretic generalization of on-line learning and an application to boosting," in *EuroCOLT '95 Proceedings of the Second European Conference on Computational Learning Theory*, 1995, Conference Proceedings, pp. 23–27.

[10] S. Doyle *et al.*, "Automated grading of prostate cancer using architecture and textural image features," in *4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, 2007, Conference Proceedings.

[11] M. Macenko *et al.*, "A method for normalizing histology slides for quantitative analysis," in *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, 2009, Conference Proceedings, pp. 1107–1110.

[12] A. M. Khan *et al.*, "A nonlinear mapping approach to stain normalization in digital histopathology images using image-specific color deconvolution," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 6, pp. 1729–1738, 2014.

[13] A. K. Jain *et al.*, "Feature selection: evaluation, application, and small sample performance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 2, pp. 153–158, 1997.