

Computer Aided Histopathological Classification of Cancer Subtypes

S. Waheed^{1,2,3}, R. A. Moffitt^{2,3}, Q. Chaudry¹, A. N. Young³, and M.D. Wang^{1,2,*}.

¹Electrical and Computer Engineering, Georgia Institute of Technology

²Biomedical Engineering, Georgia Institute of Technology and Emory University

³Pathology & Laboratory Medicine, Emory University

* denotes equal contribution to this work

*maywang@bme.gatech.edu

Abstract— In this paper we present the results of our effort to develop a computer aided diagnosis system for pathological imaging data using Renal Cell Carcinoma as a case study. Traditionally, cancer diagnosis is performed by an expert pathologist studying biopsy tissue under a microscope. Due to the complex nature of the task and the heterogeneity of patient tissue, these methods are not only time consuming but also suffer from subjective variability. To improve the repeatability and accuracy of the diagnosis process, a computational diagnosis system is proposed here. In this paper we report that with our novel knowledge-based methodology, we are able to achieve high level of classification accuracy (98%) when trying to classify 64 images (n=64) using a simple Bayesian classifier based on 8 extracted features and complete-leave-one-out cross-validation. This methodology is implemented in MATLAB and is expected to aid pathologists in the clinical setting to diagnose Renal Cell Carcinoma as well as other types of cancer.

Keywords-cancer; image processing; pattern recognition; pathology

I. INTRODUCTION

Cancer is the second leading cause of death after heart disease in America. Although there has been a steady decrease in the incidence of death due to heart disease, no such trend can be observed in cancer. In addition to that, the risk of getting cancer is increasing due to major environmental, habitual and behavioral trends [12].

Even with impressive strides made in treating and curing cancer, further improvement of survival rate relies heavily on early diagnosis. Also important to clinical success is knowledge of the behavior of a particular cancer and its treatment, which depend on correct identification of cancer stage and/or subtype. It is therefore imperative that we not only diagnose cancer early, but also differentiate between its various subtypes quickly and accurately. To achieve relatively high differential accuracy, extensive training is usually required by a pathologist. Unfortunately, the current diagnostic paradigm of manual assessment of histology slides is slow and often irreproducible [8, 7, 1]. By leveraging the power of computational systems, we speed up the process of diagnosis and reduce the subjectivity in pathology.

In this paper we present the design, development and results of a novel computer-aided diagnosis system which seeks interaction with expert users throughout the diagnosis process. With the system, users can lend their expertise to the validation of feature extraction and quantification, and they can also select from a list of features they deem most important and appropriate for the classification at hand. This same tool is designed to be used by a variety of pathologists with different cancer classification goals as long as the system has had sufficient training.

II. BACKGROUND

A. Computer Aided Diagnosis

In the past, much effort has been devoted to development of automated diagnosis systems for various maladies. Although the application of computational methods represents a significant step forward in the fight against cancer and its early detection, the task is anything but trivial. Traditionally, most automated cancer diagnosis research has been targeted at differentiating between cancerous and normal tissue images [17, 19].

Since pathologists use deviations in cellular structure as a means to make a diagnosis, many previous methods have used the statistical variation of various image properties to help make a diagnosis. The use of morphological features, for example, was reported by Jiang et al. in their study of breast cancer classification and by Roula et al. for the grading of prostate cancer [13, 15]. The diagnosis system developed by Diamond et al. used a combination of structural and textural features to achieve an accuracy of 79.3% for the classification of Prostatic Neoplasia [4]. Esgiar et al. studied the classification of colonic mucosa using six different textural features and optical density and reported an overall accuracy of 90.2 % [6].

The choice of features in these studies was generally motivated by the hypothesis that the human eye uses these features for discrimination and so should automated systems. Recent attempts at segmentation have moved beyond this paradigm and instead include less intuitive features, such as fractal dimension, which are not easily detected or described by humans. In a follow-up study, Esgiar et al. reported an increase in accuracy of their system when fractal analysis was employed

along with previously identified textural features. Furthermore, they suggested that knowledge incorporation is needed for further increase in accuracy [5]. Hamilton et al. used such knowledge-guided segmentation to calculate features like the co-occurrence matrix and optical density to achieve 83% correct classification of colorectal dysplasia [10].

We agree with Esgair's assertion, and contend that classification accuracy can be increased by incorporating knowledge from an expert pathologist at every step of the system: image processing, feature extraction, and classification. Involving the user in the decision making process and allowing the user to bias the system will lead to a more accurate prediction. The belief is that having user interaction through every step of the process helps to encompass the vast heterogeneity found in tissue imaging data, thus making the system more robust to intra-class variation.

B. Renal Cell Carcinoma

We chose Renal Cell Carcinoma (or RCC) as a case study for the development of this tool primarily because little research has been done for the automated classification of RCC. Moreover, this problem is more complicated than simple normal versus cancerous tissue classification as RCC has four commonly observed clinical subtypes.

RCC is the most common form of kidney cancer arising from the renal tubule in adults [2]. More than 90% of clinically significant lesions can be diagnosed as one of the common subtypes of renal tumor: clear cell RCC (CC) (70-75%), papillary RCC (PAP) (10-15%), chromophobe RCC (CHR) (2-5%), and renal oncocytoma (ONC) (5%). Renal tumor subtypes exhibit several common morphological characteristics, making diagnosis difficult and subjective in many cases. Histopathologic classification is critical for the treatment of RCC because each of its subtypes is associated with a distinct clinical behavior. Development of a diagnosis technique with a quantitative approach to renal tumor classification is therefore critical and in high demand.

Expert knowledge of RCC features was incorporated into our system by letting a pathologist select the features most relevant to him for the diagnosis of RCC. This was coupled with the prior knowledge about the presence and/or absence of specific histological features and structures, such as red blood cells, blood vessels, lipid structures, and papillary bodies, in various subtypes of RCC. The outline of the design of this novel diagnostic tool targeted for clinical practice and translational research is given in the next section.

III. METHODS

A. System Design

A diagram of the system workflow is given in Fig. 1. It should be noted in particular that the user can interact through the modification (if necessary) of various image processing modules and through the selection of features to be used in the classification step. Interfaces for user interaction have been designed to be straightforward and simple, consisting of slide bars or checkboxes. Because of this, the tool should be accessible to the target user, the pathologist, independent of

any experience with image processing or data classification algorithms. As mentioned earlier, the motivation for providing this user interactivity is to keep the tool as general as possible by allowing the user to define his or her own inputs and classification goals.

The system begins with acquisition of standard photomicrographs of H&E stained biopsy tissue sections. The photomicrographs, once acquired, are processed to improve image quality and enhance objects of interest. Regions of interest are then segmented and passed onto the next stage for feature extraction and quantification. Features are extracted based on the expert knowledge built into the system capturing various expected RCC features. The images are then classified based on the extracted features and results are provided to the pathologist for evaluation and feedback. The user interface which manages the system is shown in Fig. 2. Finally, the RCC images, the extracted features, the classification results, and the clinician's diagnosis are stored in a database. The database is used to further train the classification system with the goal of having a practical and useful clinical tool for cancer diagnosis. The following sections describe in detail the methods used for image processing, feature extraction and classification.

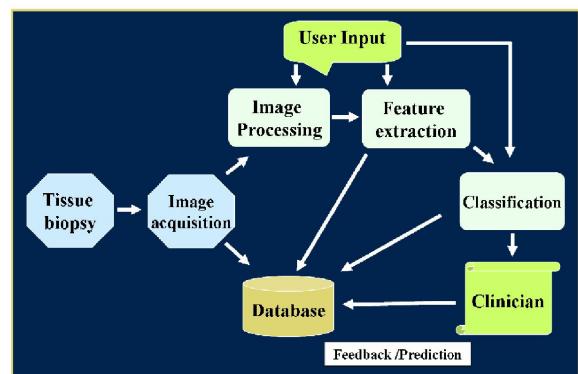


Figure 1. Workflow of the proposed system.

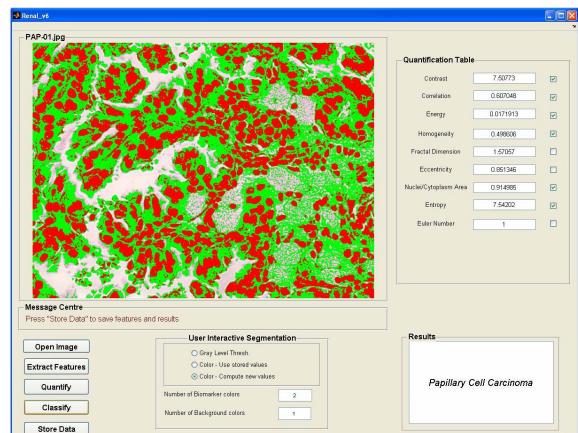


Figure 2. Screenshot of the interactive GUI for the diagnosis system.

B. Tissue Samples and Image Collection

All tissues in this study were derived from renal tumors resected by total nephrectomy. Tumors were fixed, processed,

sectioned and stained according to standard pathological procedures. Nephrectomy specimens were fixed for at least one hour in several volumes of 10% neutral buffered formalin, after which representative histologic samples (3-millimeter thickness) were obtained and fixed overnight in > 10 volumes of 10% neutral buffered formalin. Histologic samples were embedded in paraffin and microscopic sections (5-micrometer thickness) were prepared with a microtome and stained with hematoxylin & eosin (H&E). Representative photomicrographs of renal tumor sections were obtained at 200x total magnification and four images of 600 x 800 pixels were extracted from the original 1200 x 1600 field of view. Sixteen representative tumors, four from each of the four RCC subtypes, were used in this study for a total of 64 images.

C. Image Preprocessing

One of the main problems faced during segmentation of cellular features is the variation in staining within cellular features which leads to rough margins. In order to reduce this variation, images are smoothed with a Gaussian filter before segmentation is performed. Gaussian filter size is based on the size of the smallest feature of interest, the cell nucleus. The user retains the ability to adjust the size of the filter if necessary. Although the Gaussian filter tends to blur edges, an appropriately sized filter enhances segmentation results as compared to unfiltered counterparts, thereby validating the use of the filter.

Because Gaussian smoothing alters the texture of an image, smoothed images are not used for any texture or intensity-based feature extraction. Smoothed images are instead used only for the segmentation process. Once the object regions within an image are recognized and segmented, the smoothed image data is discarded, and original image data is used.

D. Segmentation

Both intensity-based and edge-based segmentation techniques were attempted for differentiating between nuclear and cytoplasmic regions of the image. Edge-based segmentation methods tended to fail, however, because of a lack of significant intensity variation between the nucleus and its surroundings.

Since acquired images are colored based on nuclear and cytoplasmic characteristic (a property of H&E staining), color information in the images is used as a criterion for segmentation. Segmentation is accomplished using the K-means algorithm, previously shown by Gunduz et al. to provide efficient pixel classification based on color information [9]. In the K-means algorithm, an objective function, given below as J , is minimized.

$$J = \sum_{j=1}^K \sum_{i \in S_j} \|x_i - \mu_j\|_2^2$$

In this case, J is the squared Euclidean distance of the n data points from their respective cluster centers, where x_i is a data point in cluster j , and μ_j is the j^{th} cluster center. For color segmentation, x_i is a three dimensional vector of [red, blue,

green], but this can be generalized to any number of dimensions.

The result of the K-means algorithm is that each pixel in the image is classified as being part of a cluster. For the case of RCC subtyping, $K=3$ clusters are used, corresponding to the nucleus, cytoplasm and background. Since the result of a K-means algorithm depends on the initial values of cluster centers, we allow the centers to be chosen based on user input. If necessary, the user may specify a number of clusters other than 3 to aid in segmentation. This flexibility is provided in anticipation of new imaging modalities such as multiplexed immunohistochemistry or fluorescence images. If the user is still not satisfied with the results of segmentation, they may also bias the results of the K-Means classification by moving simple slider controls, shown in Fig. 3 (top).

The raw segmentation results generated by K-means are then further refined as described below to reduce misclassification error and solve the problem of overlapping nuclear features. First, unwanted segmentation artifacts, such as small spots or holes, are removed by complementary erosion and dilation operations. A watershed algorithm based on original image intensity is then used to detect and separate overlapping nuclei as previously described [3, 14]. During all these processes, the user can choose to step in and bias the results using simple slider controls which function to grow or shrink each segmentation result. After segmentation is complete, feature extraction follows.

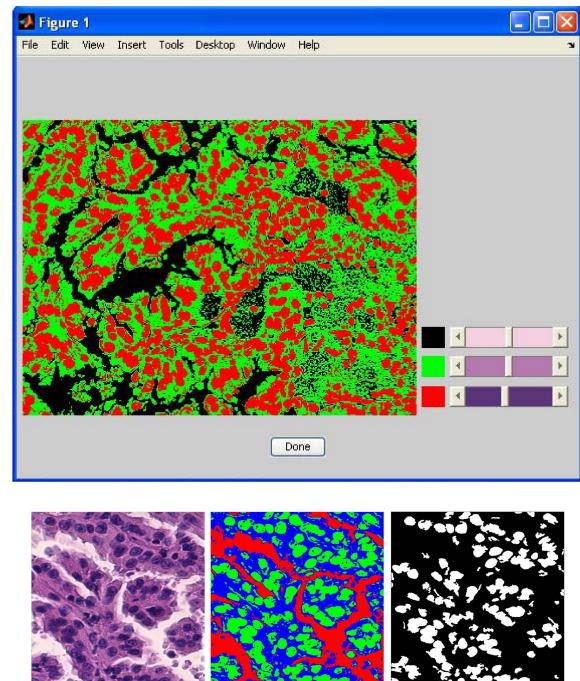


Figure 3. (Top) Interface for interaction with K-means clustering tool via slider bars. (Bottom) Left → Right: Original image of Papillary carcinoma, K-means segmented image, Segmented nuclei after post processing.

E. Feature Extraction

After regions of interest have been segmented, they are used to quantify cellular properties in a process called feature

extraction. Instead of calculating a wide variety of features and hoping to find hidden patterns, our system incorporates expert pathology knowledge by allowing the user to choose features that are already known to be important in differentiating between subclasses.

An example of this knowledge-modeling is the quantification of the papillary or finger-like features present extensively in the PAP subtype. Fractal dimension is used to model this finger-like structure due to its ability to detect the self-similarity of these papillary features. Fractal dimension has previously been used in image processing to quantify self similarity in images and is usually calculated over the whole image with a thresholding step. As an example of knowledge-modeling, the fractal dimension in this case is calculated only over the cellular regions of the image.

Schepers et al. showed that among the various algorithms proposed for the calculation of fractal dimension, spectral analysis provides results with highest fidelity [16]. Fractal dimension was tested using both spectral analysis and a box-counting method, and although the latter is a less complex algorithm to implement, the spectral analysis method was ultimately used because of its better performance during classification.

A second example of knowledge-modeling can be found in the way nuclear morphology is regarded. The observation that different subtypes of RCC display variation in nuclear density and nuclear shape is captured by directly quantifying these morphological features. PAP, for example, tends to show a higher nuclei to cytoplasm ratio while nuclei in CHR images are known to be more eccentric than others. Features like nuclear area, nuclear eccentricity, and the ratio of the area of the nuclei to cytoplasm are calculated to quantify these differences.

Knowledge about the presence of erythrocytes and blood vessels (CHR), the presence of fibrous and vascular cores (PAP), and the growth of nuclei in nest like structures (ONC) was quantified by calculating textural features like contrast, homogeneity, correlation, and energy. These features are calculated as previously described based off a pixel co-occurrence matrix [11]. These textural features are calculated for not only the entire image, but also the individual cellular structures where applicable.

Because of the modularity of the system, we fully expect that this tool will be able to quantify similar features from other forms of cancer imaging data. At the current stage, the user can customize which features will be used for classification. Once the panel of features is calculated for a given image, it is then passed on to the classification module which determines the subtype of cancer.

F. Classification

With properties of the biopsy images extracted and quantified, their distributions are used to classify samples into different subtypes. The classification scheme used here is a simple, multi-class Bayesian decision rule which assumes multivariate Gaussian distributions for the feature vector describing each image. The decision of our system is the

maximum-likelihood classification, given all of the hand-classified data in the training set. To attempt the most robust classifier possible, eight knowledge-based features were selected by a pathologist as inputs to the classifier, but any number of features are available for selection by the user.

IV. RESULTS

Table 1 shows the distributions of the eight features used for our final classification across each subtype of RCC. Examination of the table shows that a single feature alone is insufficient to differentiate between each of the different subtypes. This follows from the concept that that Renal tumor subtypes exhibit several common morphological characteristics. Adding additional features to the classifier helped separate the classes from each other, leading to a more accurate result. Fig. 4 is a scatterplot of this tabular data showing how the subtypes can be separated fairly easily with as few as three features.

TABLE I. FEATURE EXTRACTED FOR PAPILLARY (PAP), CLEAR CELL (CC), CHROMOPHOBIC (CHR) AND RENAL ONCOCYTOMA (ONC) WITH MEAN AND STANDARD DEVIATIONS

| | PAP | CC | CHR | ONC |
|-------------------|--------------|--------------|--------------|--------------|
| Correlation | 8.23 ±0.906 | 10.85 ±0.971 | 2.505 ±0.320 | 5.885 ±1.01 |
| Contrast | 0.648 ±0.077 | 0.43 ±0.038 | 0.532 ±0.031 | 0.454 ±0.050 |
| Energy | 0.017 ±0.005 | 0.018 ±0.002 | 0.076 ±0.016 | 0.029 ±0.006 |
| Homogeneity | 0.504 ±0.028 | 0.474 ±0.011 | 0.663 ±0.018 | 0.531 ±0.025 |
| Entropy | 7.766 ±0.047 | 7.446 ±0.033 | 6.716 ±0.071 | 7.288 ±0.080 |
| Fractal Dimension | 1.787 ±0.08 | 1.934 ±0.007 | 1.842 ±0.031 | 1.93 ±0.022 |
| Ratio of Area | 0.796 ±0.147 | 0.260 ±0.035 | 0.105 ±0.036 | 0.212 ±0.073 |
| Eccentricity | 0.812 ±0.025 | 0.827 ±0.021 | 0.792 ±0.052 | 0.766 ±0.033 |

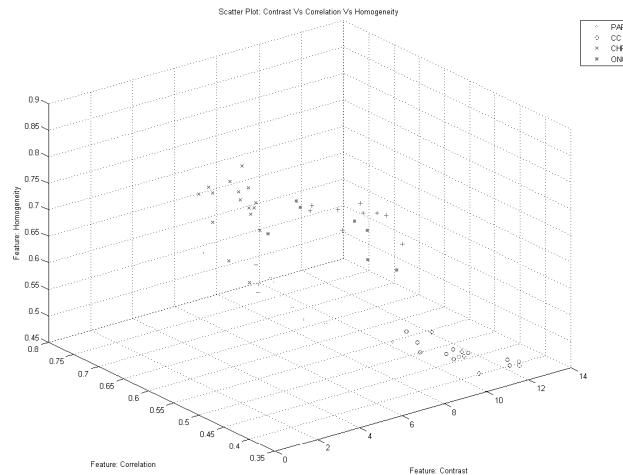


Figure 4. Scatter plot showing distribution of images for three co-occurrence features: Contrast, Correlation and Homogeneity (+, Papillary; o, Clear Cell; x, Chromophobe; *, Oncocytoma)

To estimate the ability of the classification rule to correctly predict the class of a new, unknown sample, complete leave-one-out (CLOO) cross-validation was performed, each time withholding one of the 64 images to act as the test sample.

Results showed that the classifier correctly predicted the class of the test sample 98.4%, or 63/64 of the time. Good results were also achievable with fewer features (Fig. 4), but in order to create a more robust classifier, eight features were eventually chosen. It should be noted here that a feature size of eight is appropriate considering the number of samples (sixteen) per class available for training. Increasing the number of features further, however would likely lead to over fitting or render the problem ill-posed.

V. DISCUSSION

The favorable classification results (98.4%) suggest that the strategy of feature extraction based on expert pathology knowledge is reasonable. This high accuracy result is not surprising, considering the visually compelling separation evident in Fig. 4. Still, this and all classification results must be met with at least some scrutiny.

The fact that this classification accuracy is high may be influenced by the fact that the user oversees the whole process from start to finish. For this study, the features selected for use in the classifier were chosen using results from all available data. Only after features were selected was any data held out for the cross-validation step. In theory however, leave-one-out cross validation should correct for any bias introduced here, because it is known to be an unbiased estimator of how the classifier will perform on completely new samples.

Where this type of cross-validation fails, however, is when new data are dissimilar from training data. For example, we would not expect our system to classify biopsy tissues with a different staining protocol at 98% accuracy. To provide a true estimator of how our system might perform in a clinical setting, we would require a large pool of training images from multiple sites, prepared by multiple personnel, and from a wide variety of patients. The results presented in this paper, in contrast, are based on images acquired through a single acquisition system from a limited pool of patients. Due to a lack of diverse RCC images, we are unable to satisfy this question completely at this time. Still, this system shows a marked improvement from previous reported classification results and is definitely a step in the right direction.

This successful case study of RCC subtype classification has shown promise for this tool's use as a more general pathology system. With the ability to process differently stained images and the flexibility to customize features already built-in, the system can readily be used for the classification of non-renal cancer imaging data.

In order to ensure our tool's universal clinical use, more work needs to be done. First, the system will be ported from its current Matlab implementation to a format more compatible with a clinical setting. In addition to the features already discussed here, a whole array of features have been added to the system including gabor filters, phase congruence analysis and fractal vectors. All these features will be available for the user to select from depending on his or her preference and the perceived importance during the diagnosis.

Aside from adding new feature extraction functionality, work is underway to implement a model which employs a

weighting function for the different features. With such a model, the pathologist will not only be able to select relevant features but will also be able to assign a unique importance to each of them for the classification problem.

It should be kept in mind, however that a clinical implementation will require much more training and validation. In particular, additional effort is required to adjust for the variation among image acquisition systems, tissue collection methods, and staining protocols.

VI. FUTURE TRENDS

Although computer aided diagnostic tools can never replace the expertise of a trained pathologist, they can assist pathologists and improve upon the status quo by increasing accuracy and reducing subjectivity. Furthermore, the power of pattern recognition can also be leveraged to find and extract features from these tissue images which are beyond human visual perception.

It is critical to note that the paradigm of cancer diagnostics based mainly on histopathology will soon have to evolve. True and Gao have contended that since histological patterns of cancer are not directly correlated with the underlying molecular profile that is responsible for cancer progression, new optical image technologies like Quantum dots should be used to provide further insight [18]. They state that "With new molecular profiling technologies, it should be possible to read the molecular signatures of an individual patient's tumor and correlate a panel of tissue biomarkers with clinical outcome and personalized therapy." Although our current diagnostic classification is based on H&E histopathology, all analysis techniques were developed and implemented with the vision to scale to advanced optical imaging techniques using multiplexed Quantum Dots or Surface Enhanced Raman Scattering probes. Our system can readily analyze and quantify such images and their properties and can integrate them with underlying histopathological information. The advent of these molecular profiling techniques promises the ability to correlate the histological patterns with the specific biomarker profile, thereby leading to better clinical diagnosis and representing a new horizon for the field of molecular pathology [20].

ACKNOWLEDGMENT

This research has been supported by grants from National Institutes of Health (Bioengineering Research Partnership R01CA108468, P20GM072069, Center for Cancer Nanotechnology Excellence U54CA119338), Georgia Cancer Coalition (Distinguished Cancer Scholar Award to MDW), National Science Foundation (Fellowship to RAM), Hewlett Packard, and Microsoft Research.

REFERENCES

- [1] A. B. Ackerman, Discordance among expert pathologists in diagnosis of melanocytic neoplasms, *Human Pathology*, 27 (1996), pp. 1115-1116.
- [2] M. B. Amin, M. B. Amin, P. Tamboli, J. Javidan, H. Stricker, M. de-Peralta Venturina, A. Deshpande and M. Menon, Prognostic impact of histologic subtyping of adult renal epithelial neoplasms: an experience of 405 cases, *Am J Surg Pathol*, 26 (2002), pp. 281-291.

- [3] P. Bamford and B. C. Lovell, A Water Immersion Algorithm for Cytological Image Segmentation, APRS Image Segmentation Workshop (1996), pp. 75-79.
- [4] J. Diamond, N. H. Anderson, P. H. Bartels, R. Montironi and P. W. Hamilton, The use of morphological characteristics and texture analysis in the identification of tissue composition in prostatic neoplasia, *Human Pathology*, 35 (2004), pp. 1121-1131.
- [5] A. N. Esgiar, R. N. G. Naguib, B. S. Sharif, M. K. Bennett and A. Murray, Fractal analysis in the detection of colonic cancer images, *Information Technology in Biomedicine, IEEE Transactions on*, 6 (2002), pp. 54-58.
- [6] A. N. Esgiar, R. N. G. Naguib, B. S. Sharif, M. K. Bennett and A. Murray, Microscopic image analysis for quantitative measurement and featureidentification of normal and cancerous colonic mucosa, *Information Technology in Biomedicine, IEEE Transactions on*, 2 (1998), pp. 197-203.
- [7] E. R. Farmer, R. Gonin and M. P. Hanna, Discordance in the histopathologic diagnosis of melanoma and melanocytic nevi between expert pathologists, *Human Pathology*, 27 (1996), pp. 528-531.
- [8] K. A. Fleming, Evidence-based pathology, *J Pathol*, 179 (1996), pp. 127-8.
- [9] C. Gunduz, B. Yener and S. H. Gultekin, The cell graphs of cancer, *Bioinformatics*, 20 (2004), pp. 145-151.
- [10] P. W. Hamilton, P. H. Bartels, D. Thompson, N. H. Anderson, R. Montironi and J. M. Sloan, Automated Location of Dysplastic Fields in Colorectal Histology Using Image Texture Analysis, *The Journal of Pathology*, 182 (1997), pp. 68-75.
- [11] M. Hauta and R. Lenzt, Generalized Co-Occurrence Matrix for Multispectral Texture Analysis, *Proceedings of ICPR*, 2 (1996), pp. 785.
- [12] A. Jemal, R. Siegel, E. Ward, T. Murray, J. Xu and M. J. Thun, *Cancer Statistics, 2007*, CA: A Cancer Journal for Clinicians, 57 (2007), pp. 43-66.
- [13] Y. Jiang, R. M. Nishikawa, R. A. Schmidt, C. E. Metz, M. L. Giger and K. Doi, Improving breast cancer diagnosis with computer-aided diagnosis, *Acad Radiol*, 6 (1999), pp. 22-33.
- [14] N. Malpica, C. O. de Solorzano, J. J. Vaquero, A. Santos, I. Vallcorba, J. M. Garcia-Sagredo and F. del Pozo, Applying watershed algorithms to the segmentation of clustered nuclei, *Cytometry*, 28 (1997), pp. 289-297.
- [15] M. Roula, J. Diamond, A. Bouridane, P. Miller and A. Amira, A multispectral computer vision system for automatic grading of prostatic neoplasia, *Biomedical Imaging, 2002. Proceedings. 2002 IEEE International Symposium on* (2002), pp. 193-196.
- [16] H. E. Schepers, J. van Beek and J. B. Bassingthwaigte, Four methods to estimate the fractal dimension from self-affinesignals [medical application], *Engineering in Medicine and Biology Magazine, IEEE*, 11 (1992), pp. 57-64.
- [17] J. P. Thiran and B. Macq, Morphological feature extraction for the classification of digitalimages of cancerous tissues, *Biomedical Engineering, IEEE Transactions on*, 43 (1996), pp. 1011-1020.
- [18] L. D. True and X. Gao, Quantum Dots for Molecular Pathology: Their Time Has Arrived, *Journal of Molecular Diagnostics*, 9 (2007), pp. 7.
- [19] R. F. Walker, P. Jackway, B. Lovell and I. D. Longstaff, Classification of cervical cell nuclei using morphologicalsegmentation and textural feature extraction, *Proceedings of the 1994 Australian and New Zealand Conference on Intelligent Information Systems*. (1994), pp. 297-301.
- [20] A. N. Young, M. B. Petros, J. A. Natan and M. Wang, Nanomolecular Histopathology for Renal Tumor Classification, *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the* (2005), pp. 723-726.