

Malignancy Detection in Fine-Needle Aspiration Biopsy of the Thyroid using Multispectral Microscopy and Bag-of-Features

[anonymized]

Abstract—Fine needle aspiration (FNA) is widely accepted as the most direct, accurate and cost-effective diagnostic procedure in the management of nodular thyroid disease. However, the false positive rate of cytological grading of FNA smears suffers from the cytopathologic difficulties in differentiating benign and malignant follicular tumors. In this paper, we propose a computer-aided detection tool to help improve the confidence of the diagnosis with quantitative measurements. Bag-of-features classification strategy is implemented, leveraging its benefits in using patch-like features, thus achieving independence from the result of image segmentation or cellular level cytological analysis. In addition, we demonstrate in the test results that multispectral imaging affords more discriminative power than data in the color space.

I. INTRODUCTION

The presence of palpable thyroid nodules is a common clinical problem and epidemiologic studies show that up to 7% of the US adult population has single or multiple nodules within the thyroid gland [1]. Fine needle aspiration (FNA) is widely accepted as the most direct and accurate diagnostic procedure in the management of nodular thyroid disease. It is also the most cost-effective tool for the initial screening and triage of thyroid nodule cases [2]. However, FNA has two major limitations: non-diagnostic results and suspicious results. Non-diagnostic results are usually caused by unsatisfactory specimens acquired through FNA. Repeated FNA procedure may be needed in this case. Suspicious results are related to cytopathologic difficulties in differentiating between benign and malignant follicular tumors since the differences are subtle and often unrealizable on visual examination. The standard procedure in such cases is to classify them as “follicular neoplasm” without specifying their benign or malignant character. For those cases that are suspicious, Arda et al. [3] reported that 54% patients underwent surgery for further histological examinations. The rest of patients are exempt from surgery because they respond positively to a three-months hormonal treatment. Because of these limitations, the result of statistical analysis on the performance of the FNA examination varies according to how the unsatisfied and suspicious cases are handled during the analysis. Unsatisfactory specimens are excluded from most analyses. Reported specificity values of FNA cytological studies vary in the range of 72% to 100% [1]. The lower bound of this range means a high false positive rate which could lead to unnecessary surgery and thyroidectomy. The variation of the of the specificity is due to several factors [1]: first, the quality of FNA is related to the skill of the aspirator; second, the statistical result may depend on the inclusion or exclusion of unsatisfactory and suspicious cases during the analysis; third, the expertise of the cytologist could

be a decisive factor that affects the accuracy of the result. In conclusion, it is hard to maintain a consistent result if the diagnostic procedure is tightly coupled with the experience and knowledge of people. Computer-aided diagnosis, combining elements of artificial intelligence and digital image processing, can be utilized to minimize the variation. In this paper, we exploit automatic classification methods on multispectral data with an intention to reduce the false positive rate of cytological interpretations of the FNA examination. The purpose of this research is not to replace the role of cytologist, but to find a method that could provide supplemental information for the cytologist and help improve the confidence of the diagnosis with quantitative measurements.

Computer aided detection of malignancy has been widely used as a screening tool and shown to be useful in many cases [4], [5], [6]. However, reports of automatic diagnosis related to the FNA biopsy of the thyroid are few. Several factors may have contributed to this shortage of applications. First, cytological classification of FNA samples is complicated. Camargo et al. [7] described a four-grade classification system based on the shape, distribution, arrangement of the cells and other elements including chromatin, cytoplasm and intercellular material, etc. The four grades are listed as benign pattern grade I, indeterminate grade II, suspicious pattern grade III, and malignant pattern grade IV. The malignant grade can be further classified into papillary pattern, medullar pattern, anaplastic pattern, and malignant lymphoma. Differences between some grades are subtle and the diagnostic confidence is low for the suspicious pattern grade. Second, samples abstracted through FNA procedures are usually contaminated by blood cells. Under the microscope, these blood cells are observed to have similar shape and size as the target cells. In addition, many imaging results of FNA samples show that cells tend to be overlapped under cytological preparations. This overlapping pattern poses a great challenge to some image pre-processing procedures, e.g., image segmentation.

Majority of algorithms developed for malignancy detection of fine needle aspiration cytology (FNAC) images use either the cytological features or features derived from image processing techniques, e.g., color or texture. Delibasis et al. [8] proposed a supervised classification algorithm using the Artificial Immune Systems (AIS). In order to train the system, 61 cytological features (cellularity, cohesiveness, colloid, monolayers, follicular formations, microfollicles, phagocytes, multinucleated giant cells, oxyphilic cells, lymphocytes, intranuclear inclusions, etc.) are distracted from each smear. 12 cytological features are selected by using sequential float forward search algorithm [9], and the final diagnosis is based on the evaluation of the 12 features by the clinician. They reported higher than average accuracy scores by properly tuning the

AIS. However, their method depends solely on the cytological features and depends heavily on the experienced cytologists' ability to evaluate all 12 cytological features precisely for each smear to be tested. It is a delicate piece of work that is not easy even for a skilled cytopathologist, since it requires the visual inspection and evaluation of subtle morphological and textural differences. Another approach for the malignancy detection depends heavily on the image processing techniques. For example, the method described by Daskalakisa et al. [10] requires accurate segmentation of nuclei as the segmentation result is of crucial importance to guarantee correct feature extractions. Such a requirement can be hardly achieved in images with overlapping cells and dense cellular distribution. An example of a typical case with overlapping cellular pattern is shown in Figure 1. Figure 2 shows the typical processing

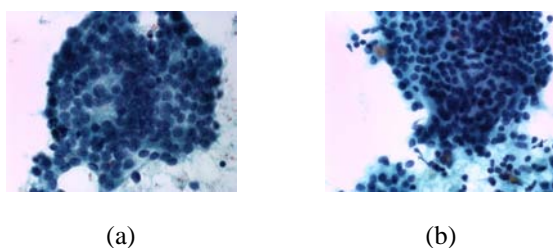


Fig. 1. Typical Images of Papanicolau Stained Fine Needle Aspiration Cytological Smears

pipeline for the above computerized analyses.

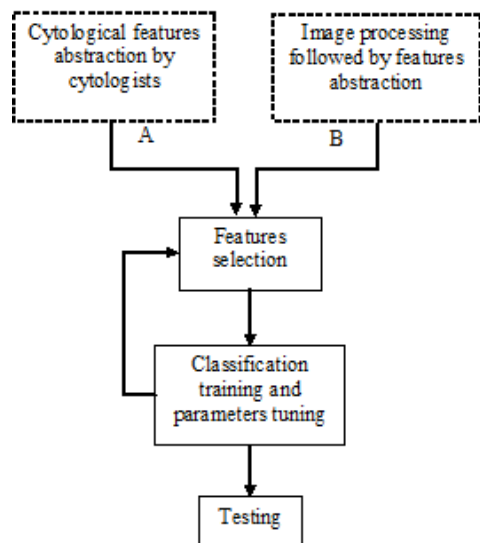


Fig. 2. Computer-aided diagnosis procedure: features can be abstracted through cytological analysis (A) or imaging processing techniques (B) or both

Both cytological features and features related to cells' morphological or statistical properties are local descriptors. The workload and technical difficulties (e.g. image segmentation) related to this kind of descriptors make it hard to be applied extensively in practice. In this paper, we propose to quantify FNAC specimens by using bag-of-features model on multispectral data. One of the concerns of using global descriptors is the possibility of losing detailed information

held by local descriptors. However, spectral data compensate for a potential loss by extending the feature space. According to [11], standard color or gray images only roughly map the true spectral content of an image. Color is not an intrinsic physical property either. Spectral imaging, on the other hand, captures images with accurate spectral content correlated with spatial information and reveals the chemical or anatomic features of the target. Spectral microscopy has been well utilized in the analysis of abnormalities in FNA cytological samples. Multispectral imaging retrieves spectrally resolved information of an imaged scene. It allows for the simultaneous measurement of spectral and spatial information of a sample. The unique transmission spectra of biological tissue provides additional information that is potentially useful for better classification. Molecular interactions between tissues and histological dyes revealed by spectral images, however, can be hardly retrieved through monochrome or color-based sensors [11]. Figure 3 shows spectral images obtained with different wavelengths.

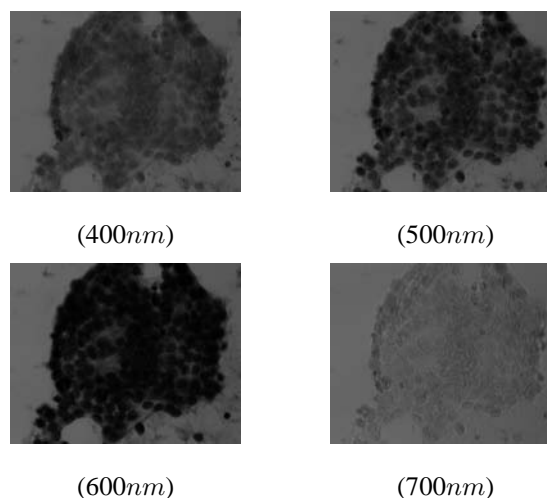


Fig. 3. Spectral images of FNA smear obtained with different wavelengths

As a model using global descriptors, bag-of-features representations have become popular for content based image classification. It can recognize the context of a scene without having to recognize the objects that are present [12][13][14]. It treats images as collections of orderless patches. Each patch is represented by a visual descriptor vector. By sampling a representative set of patches from the image, the image is characterized by the resulting distribution of samples in their corresponding descriptor space. It implies that the image can be classified without segmentation or cytological interpretation. Figure 4 shows the processing pipeline of the bag-of-features model. In this paper, we apply bag-of-features method on both FNAC color and multispectral images. We demonstrate that it has the potential to help improve the diagnostic confidence for detecting malignancy of thyroid.

The rest of the paper is organized as follows. In section II, we describe the procedure for sampling patches and characterizing the result distribution. Section III studies the parameter of the codebook. The choice of classifier is discussed in section IV. Section V summarizes the algorithm. Section VI

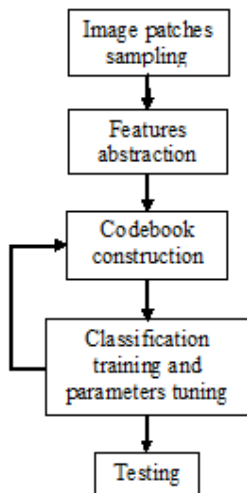


Fig. 4. Processing diagram of bag-of-features classification model: each image is represented by a histogram descriptor derived from the codebook

analyzes the experiment results and the paper is concluded in section VII.

II. PATCHES AND FEATURE DESCRIPTOR

Bag-of-features evolved from texton, a basic element that characterizes the texture by repetition. The identity of the textons plays a more important role than the spatial arrangement for stochastic textures. Construction of bag-of-features starts by sampling patches from the image. Patches can be densely sampled, randomly sampled or sampled by keypoint detectors. Keypoint detectors have been the center of the research of computer vision in the last decade and has many applications [15][16][17]. One of the prominent properties of modern keypoint detectors is that they can be both scale and affine invariant. This feature is important for object recognition in computer vision since images are usually taken from different angles and different distances. However, images of FNA smears are obtained in a controlled environment with a fixed scale and cells are approximately round-shaped. So the invariant property is not a major concern here. On the other hand, keypoint detectors are more sensitive to corners and edges and tend to ignore areas with a uniform distributed pattern. Areas occupied by the cytoplasm in FNAC images could be missed by keypoint detectors although the subtle differences of the appearance of the cytoplasm plays an important role in determining the discriminative power [7]. In this paper, we combine dense and random sampling methods for selecting patches from the region of interest (ROI). Figure 5(b) shows the ROI obtained by thresholding the FNAC image. A square bounding box of fixed size is densely sampled from the ROI. A patch will be discarded if more than 10% of its area is not overlapped with the ROI. The size of the bounding box will be decided by cross validation during the training process.

Once patches are selected, the next question is how to describe them. One way is to use the raw intensity value or corresponding histogram. However, it is less efficient for the

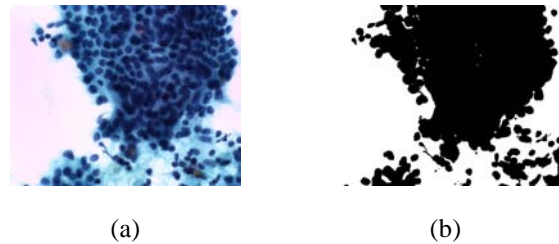


Fig. 5. (a) FNAC image (b) Black area indicates ROI distracted by thresholding

purpose of classification [18]. Studies of the human visual system support a multiscale texture analysis approach, since the visual cortex can be modeled as a set of independent channels, each with a particular orientation and spatial frequency tuning. Multi-resolution analysis method such as wavelet or wavelet packet decompositions is one of the most popular texture extraction methods that has better representation for intuitive properties like roughness, granulation and regularity [19]. A 2-D wavelet transform is obtained by bandpass filtering in a specific direction (vertical, horizontal and diagonal) and contains detailed directional information at a scale. The original image can be represented by a set of statistical descriptors (e.g. mean and variance) at several scales and spatial information is retained within each subimage. In this paper, a 2-D wavelet filter is applied to each patch and leads to a decomposition of 10 subimages. Mean and variance of coefficients of each subimage are distracted as feature descriptors. A total of 20 statistical features are distracted from the filtering result.

Three parameters of the CIE L*a*b* (CIELAB) color space are added for the color patches. The CIELAB standard defines an approximately uniform color space, which means that a change of the same amount in a color value should produce a change of about the same visual importance [20].

III. CODEBOOK

Statistical features collected from each image need to be quantized. In particular, each image can be encoded by the vector quantization in the appearance space. Vector quantization of bag-of-features is similar to the bag-of-words model. Each patch is a “visual word” and patches come form “visual vocabulary” just like words coming from the dictionary. The image is represented by frequencies of these “visual words” as shown in Figure 6. Owing to its simplicity, K-means is a popular algorithm for learning the vocabulary. K-means minimizes sum of squared Euclidean distances between points x_i and their nearest cluster centers m_k :

$$D(x, m) = \sum_{clusterk} \sum_{x_i \in clusterk} (x_i - m_k)^2 \quad (1)$$

By clustering collected statistical features through K-means, each cluster center produced by K-means becomes a code-word. The vector quantization is performed by taking a feature vector and adding it to the index of the nearest codeword (K-means center) in a codebook. Thus, each image is represented by a histogram descriptor that records the frequency of each codeword. The problem remains unsolved for the this strategy

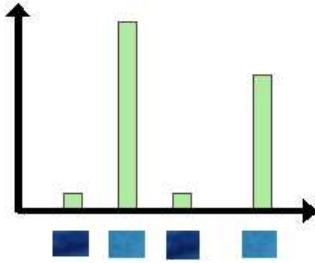


Fig. 6. Represent images by frequencies of “visual words”

is the size of the codebook. Having too few number of clusters will lead to under representative of some patches. However, large size of codebook can easily overfit the data. In this paper, we have built the codebook with various sizes and the size parameter is decided by cross validation during the training process. Figure 7 shows the codeword representation of images for the malignancy case and benign case with a codebook of size 10.

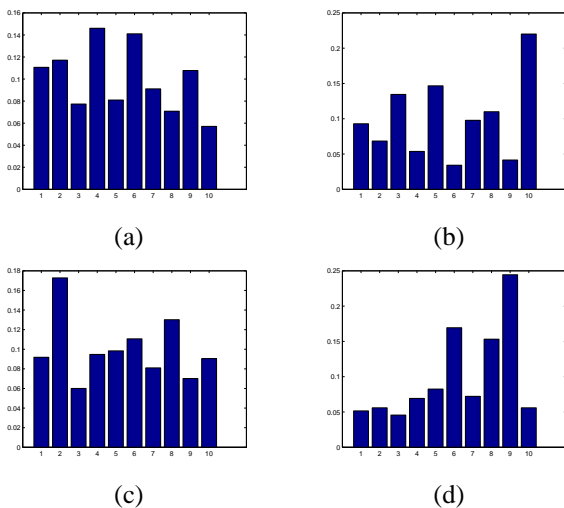


Fig. 7. (a) codeword for the malignant case of a color image (b) codeword for the benign case of a color image (c) codeword for the malignant case of a spectral image (d) codeword for the benign case of a spectral image

IV. CLASSIFICATION

Kernel-based support vector machine (SVM) is used for the classification. The non-separable case can be formulated as [21]:

$$\begin{aligned} \operatorname{argmin} & \frac{1}{2} \|w\|^2 + C \sum_i \xi_i \\ \text{s.t.} & y_i(w^T \cdot \phi(x_i) + b) \geq 1 - \xi_i; \quad \xi_i \geq 0 \end{aligned} \quad (2)$$

where C is a constraint on the Lagrange multipliers and set by the user, ξ is the slack variable that measures the degree of misclassification of the data and ϕ is the data mapping function. So $\sum_i \xi_i$ is an upper bound on the number of training

errors. Kernel-based SVM can be used in the case where the decision function is not a linear function of data. One of the popular kernel functions is radial basis function (RBF) defined as follows:

$$k(x_1, x_2) = \exp(-\gamma \|x_1 - x_2\|^2) \text{ for } \gamma > 0 \quad (3)$$

For the histogram descriptor, the implicit assumption made by the Euclidean distance computation in the RBF kernel is that histograms are aligned, so that only corresponding bins from both histograms are compared to each other. The disadvantage of this method lies in the fact that alignment assumption makes the distance measure sensitive to possible distortions in histogram descriptors. In addition, perturbations in bin positions due to quantization effect is another concern for *bin-to-bin* comparisons. The earth-mover distance (EMD) [22] alleviates the quantization concern by allowing bins at different locations to be partially matched. The EMD is defined as the minimal cost to transform one histogram into the other. Given two histograms p and q , let I be a set of suppliers, J a set of consumers, the EMD can be formally described as:

$$EMD(p, q) = \operatorname{argmin}_{f_{ij}} \frac{\sum_{i,j} f_{ij} d_{ij}}{\sum_{i,j} f_{ij}} \quad (4)$$

where f_{ij} denotes flows, d_{ij} is the ground distance between bin i and bin j of two histograms. The computation of EMD is based on the solution of bipartite network flow problem. In this paper, both Euclidean distance and EMD distance are tested for their discriminative performance.

V. SUMMARY OF THE ALGORITHM

Algorithm 1 Bag-of-features model for the malignancy detection of FNA biopsy of the thyroid

```

for each image do
  sampling patches with different sizes
  for each patch do
    distract wavelet features and color features if applicable
  end for
end for
cluster collected features by K-means
construct codebook with different sizes by vector quantization
represent each image by a histogram descriptor that records the frequency of each codeword in the codebook
for each patch size do
  for each codebook size do
    find the best parameters for the kernel-based SVM through 10-fold validation and save the parameters
  end for
end for
find the best accuracy rate of training and its corresponding patch size and codebook size
classify the testing data by selected parameters

```

VI. EXPERIMENTS

Images to be classified are obtained from slides of cytological smears that are Papanicolau stained for the purpose of differentiating pathologies in fine needle aspirated cells from thyroid nodules. A total of 66 smears are used in the test. Each smear has one color image and 31 spectral images with a wavelength between $400nm$ and $700nm$. All smears has been graded as malignant or benign after the histologic review. 33 out of 66 smears are benign and the other 33 cases are malignant. Among the 66 smears, 12 smears are classified as follicular neoplasm (suspicious cases) and 3 out of 12 are graded as follicular carcinoma (malignant cases) and the other 9 smears are graded as follicular adenoma (benign cases). Both malignant and benign cases are randomly split into 70% for training and 30% for testing. In either training set or testing set, the number of benign cases and the number of magligant cases are balanced. Results are reported as the average of 100 repeated experiments. A 10-fold cross validation is employed for the parameter selection. The parameters include patch size, codebook size, constant C of the SVM and the constant γ of the SVM. The search range for these parameters is pre-defined as follows:

$$\text{patch size} \in \{32 \times 32, 64 \times 64, 128 \times 128\}$$

$$\text{codebook size} \in \{10, 50, 100\}$$

$$C \in \{2^{-5}, 2^{-3}, 2^{-1}, 1, 2^1, 2^3, 2^5, 2^7, 2^9, 2^{11}\}$$

$$\gamma \in \{2^{-5}, 2^{-4}, 2^{-3}, 2^{-2}, 1/2, 1, 2, 2^2, 2^3, 2^4, 2^6\}$$

To acquire multispectral images, we use the Olympus BX51 optical microscope and a grating based spectral light source. 2D images are acquired by using a high resolution CCD camera. The Czerny-Turner type monochromator from PTI can provide a tunable light emission spectrum at $10nm$ resolution. A wavelength range from $400nm$ - $700nm$ is used in this study. A total number of 31 pictures are taken with wavelength separation of $10nm$. The images are acquired by using the Photometric SenSysTMCCD camera having 768×512 pixels ($9 \times 9 \mu m$) at 8-bit digitization. The condenser, aperture diaphragm, and the field stop were kept constant during measurements.

The results for a classifier is evaluated by the classification accuracy. Table I shows the accuracy of the classification using bag-of-features with color image on various combinations of different patch and codebook sizes. False positive rates for

TABLE I

CLASSIFICATION ACCURACY USING COLOR IMAGE. THE FIRST COLUMN IS THE PATCH SIZE AND THE FIRST ROW IS THE CODEBOOK SIZE

size	10	50	100
32x32	73.3%	74.1%	74.0%
64x64	72.8%	69.5%	71.7%
128x128	65.2%	68.0%	70.9%

the classification using color image on various combinations of different patch and codebook sizes are shown in table II. In the test setup, there are total 12 smears of follicular neoplasm and 3 out of 12 are follicular carcinoma. Because of the limitation

TABLE II

CLASSIFICATION FALSE POSITIVE RATE USING COLOR IMAGE. THE FIRST COLUMN IS THE PATCH SIZE AND THE FIRST ROW IS THE CODEBOOK SIZE

size	10	50	100
32x32	25.2%	22.3%	24.4%
64x64	21.9%	23.4%	22.0%
128x128	28.6%	21.9%	25.5%

of FNA indicated in section I, the possibilities for these patients to undergo surgery for the histological examinations are high. Even with perfect diagnosis result on all non-follicular cases, the potential false positive rate, in this case, could be as high as 25%. Table II shows that the false positive rate using bag-of-features classification on color image could be as low as 21.9%, which make it potentially useful to reduce the false positive rate of the diagnosis that involves suspicious cases. However, its corresponding accuracy is below expectation.

After running the same test on all 31 multispectral images, we rank the classification ability of data of each band by the classification accuracy and specificity. Spectral images with wavelength of $600nm$ have the best classification performance. Table III and table IV show their classification accuracy and false positive rate. Tests on multispectral data show that

TABLE III

CLASSIFICATION ACCURACY USING SPECTRAL IMAGES WITH WAVELENGTH OF $600nm$. THE FIRST COLUMN IS THE PATCH SIZE AND THE FIRST ROW IS THE CODEBOOK SIZE

size	10	50	100
32x32	73.0%	75.6%	73.7%
64x64	73.1%	76.6%	80.7%
128x128	74.5%	73.4%	78.0%

TABLE IV

CLASSIFICATION FALSE POSITIVE RATE USING SPECTRAL IMAGES WITH WAVELENGTH OF $600nm$. THE FIRST COLUMN IS THE PATCH SIZE AND THE FIRST ROW IS THE CODEBOOK SIZE

size	10	50	100
32x32	22.0%	24.4%	24.4%
64x64	16.8%	19.6%	12.8%
128x128	14.5%	20.3%	14.7%

TABLE V

CLASSIFICATION ACCURACY AND FALSE POSITIVE RATE (FP) FOR SUSPICIOUS CASES USING SPECTRAL IMAGES WITH WAVELENGTH OF $600nm$. THE PATCH SIZE IS 64×64 AND THE CODEBOOK SIZE IS 100

Accuracy	False Positive
82.0%	11.1%

the classification accuracy can be as high as 80.7% with a corresponding false positive rate of 12.8%. Compared to the result of color image, the improvement is significant. The ROC curves for the classification using spectral data at $600nm$ and data from color images are shown in figure 9. For those smears classified as follicular neoplasm, detection accuracy is at 82% with a false positive rate of 11.1% if classified with spectral data with wavelength of $600nm$ (shown in table V).

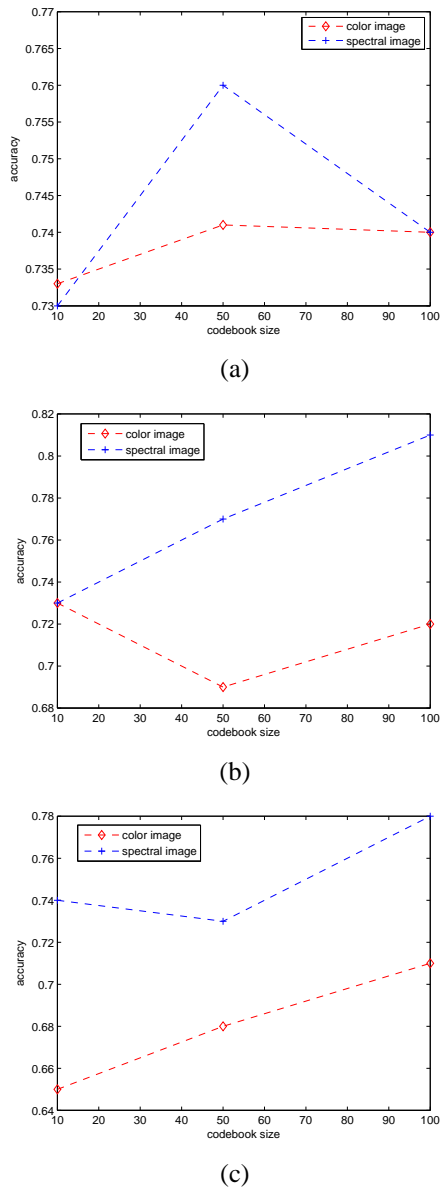


Fig. 8. Classification accuracy using color images (red) and spectral images (blue) with different patch sizes and codebook sizes; (a) patch size: 32x32; (b) patch size: 64x64; (c) patch size: 128x128

The classification result of follicular smears indicates that the proposed classification strategy has the same discriminative power over the follicular neoplasm as it does over other cytological structures since our method is independent of cytological features.

If we replace the Euclidean distance with the EMD distance in the SVM kernel, the best classification accuracy is at 81.8% with a false positive rate of 12.9%. So in our test, there is no significant classification improvement by using EMD kernel. However, since we have limited searching ranges of parameters, it is possible that the optimal value for the EMD kernel is out of the pre-defined scope and we will leave it for the future research. Figure 10 shows some sample images classified successfully by this bag-of-features method.

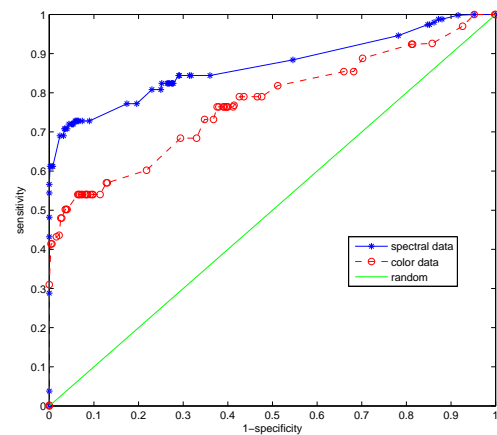


Fig. 9. ROC curves of classification results using spectral data with a wavelength of 600nm (blue) and data of color images (red)

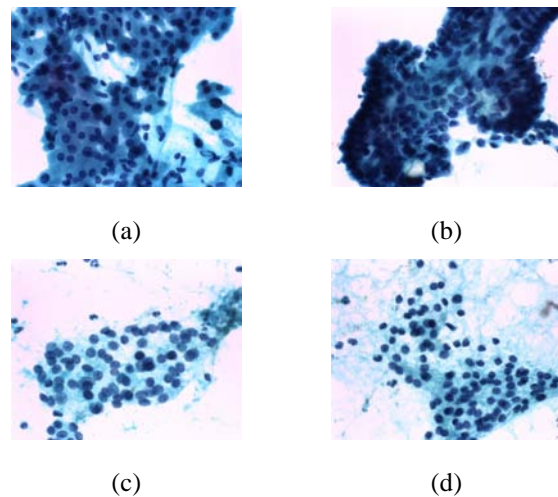


Fig. 10. (a) benign case (b) malignant case (c) follicular adenoma (d) follicular carcinoma

VII. CONCLUSION

In this paper, we have proposed to use bag-of-features classification to detect malignant cases in FNAC smears. We also demonstrated discriminative power of multispectral data over data in the color space. Although our tests show that the false positive rate is reduced with this computer-aided method, the accuracy of this method does not exceed the manual procedure. Part of the reason is probably due to the fact we were using all areas occupied by cells as ROI. However, for an experienced cytologists, particular locations on the smear are more valuable than anywhere else in term of cytological features or certain image patterns. In the future research, we will improve the method by incorporating cytopathologic knowledge into the bag-of-features model in order to increase its classification accuracy.

REFERENCES

- [1] H. Gharib and J.R. Goellner, "Fine-needle aspiration biopsy of the thyroid: An appraisal," *Annals of Internal Medicine*, vol. 118, pp. 282–289, 1993.

- [2] M. Amrikachi, I. Ramzy, S. Rubinfeld, and T. Wheeler, "Accuracy of fine-needle aspiration of thyroid," *Arch Pathol Lab Med*, vol. 125, pp. 484488, 2001.
- [3] I.S. Arda, S. Yildirim, and S. Firat, "Fine needle aspiration biopsy of thyroid nodules," *Arch Dis Child*, vol. 85, pp. 313–317, 2001.
- [4] N. Situ, X. Yuan, J. Chen, and G. Zouridakis, "Malignant melanoma detection by bag-of-features classification," in *EMBS*, 2008, pp. 3110–3113.
- [5] C. Lu, A. Devos, J. Suykens, C. Afus, and S. Huffel, "Bagging linear sparse bayesian learning models for variable selection in cancer diagnosis," *IEEE Transactions on Information Technology in Biomedicine*, vol. 11, pp. 338–347, 2007.
- [6] R. Llobet, J. Pérez-Cortés, A. Toselli, and A. Juan, "Computer-aided detection of prostate cancer," *International Journal of Medical Informatics*, vol. 76, pp. 547–556, 2007.
- [7] R. Camargo, E.K. Tomimori, M. Knobel, and G. Medeiros-Neto, "Pre-operative assessment of thyroid nodules: Role of ultrasonography and fine needle aspiration biopsy followed by cytology," *CLINICS*, vol. 62, pp. 411–418, 2007.
- [8] K.K.K. Delibasis, P.P.A. Asvestas, G.G.K. Matsopoulos, E.E. Zoulias, and S.S. Tseleni-Balafouta, "Computer aided diagnosis of thyroid malignancy using an artificial immune system classification algorithm," *IEEE Transactions on Information Technology in Biomedicine*, vol. PP, 2008.
- [9] P. Pudil, J. Nonovocova, and J. Kittler, "floating search methods in feature selection," *Pattern Recognition Letters*, vol. 15, pp. 1119–1125, 1994.
- [10] A. Daskalakisa, S. Kostopoulou, P. Spyridonou, D. Glotsos, P. Ravazoulab, M. Kardarib, I. Kardarib, D. Cavourasc, and G. Nikiforidisa, "Design of a multi-classifier system for discriminating benign from malignant thyroid nodules using routinely H&E-stained cytological images," *Computers in Biology and Medicine*, vol. 38, pp. 196–203, 2008.
- [11] R.M. Levenson, "Spectral imaging perspective on cytomics," *Cytometry*, vol. 69A, pp. 592–600, 2006.
- [12] J. Zhang, M. Marszalek, S. Lazebnik, and C. Schmid, "Local features and kernels for classification of texture and object categories: a comprehensive study," *International Journal of Computer Vision*, vol. 73, pp. 213–238, 2007.
- [13] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *CVPR*, 2006, pp. 2169–2178.
- [14] L. Fei-Fei and P. Perona, "A bayesian hierarchical model for learning natural scene categories," in *CVPR*, 2005, pp. 524–531.
- [15] S.R. Gunn, "On the discrete representation of the laplacian of gaussian," *Pattern Recognition*, vol. 32, pp. 1463–1472, 1999.
- [16] D.G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, pp. 91110, 2004.
- [17] K. Mikolajczyk and C. Schmid, "Scale & affine invariant interest point detectors," *International Journal of Computer Vision*, vol. 60, pp. 6386, 2004.
- [18] E. Nowak, F. Jurie, and B. Triggs, "Sampling strategies for bag-of-features image classification," in *ECCV*, 2006, pp. 490–503.
- [19] G.V. Wouwer, P. Scheunders, and D.V. Dyck, "Statistical texture characterization from discrete wavelet representations," *IEEE Transactions on Image Processing*, vol. 8, pp. 592–598, 1999.
- [20] M. Nischik and C. Forster, "Analysis of skin erythema using true-color images," *IEEE Transactions on Medical Imaging*, vol. 16, pp. 711–716, 1997.
- [21] C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery*, vol. 2, pp. 121–167, 1998.
- [22] Y. Rubner, C. Tomasi, and L.J. Guibas, "A metric for distributions with applications to image databases," in *IEEE International Conference on Computer Vision*, 1998.