The 2010 IEEE/RSJ International Conference on
Intelligent Robots and Systems
October 18-22, 2010, Taipei, Taiwan

Tumor CE Image Classification Using SVM-Based Feature Selection

Baopu Li and Max Q.-H. Meng, *Fellow IEEE*

Abstract— **In this paper, we propose a new scheme aimed for gastrointestinal (GI) tumor capsule endoscopy (CE) images classification, which utilizes sequential forward floating selection (SFFS) together with support vector machine (SVM). To achieve this goal, candidate features related to texture characteristics of CE images are extracted. With these candidate features, SFFS based on SVM is applied to select the most discriminative features that can separate normal CE images from tumor CE images. Comprehensive experiments on our present CE image data verify that it is promising to employ the proposed scheme to recognize tumor CE images.**

I. Introduction

Capsule endoscopy is a novel technique used to detect GI tract related diseases since it enables direct visualization of the small intestine in human body for the first time. It has been reported by Given Imaging, the manufacturer of the first generation CE, that over 800,000 patients have been diagnosed with this new technology [1].

As shown in Fig.1, capsule endoscopy (CE), measuring 26mm in length and 11mm in diameter, is a pill-shaped device which consists of a short-focal-length complementary metal oxide semiconductor (CMOS) camera, light source, battery and radio transmitter. After a CE is swallowed by a patient who has a diet for about 12 hours, this small device propelled by peristalsis begins to record images while moving forward along the digestive tract. At the same time, the images are transmitted wirelessly to a special recorder attached to the waist. This process continues for about eight hours until CE battery is used up. Finally, all the image data are downloaded into a computer, and clinicians can examine the video and diagnose.

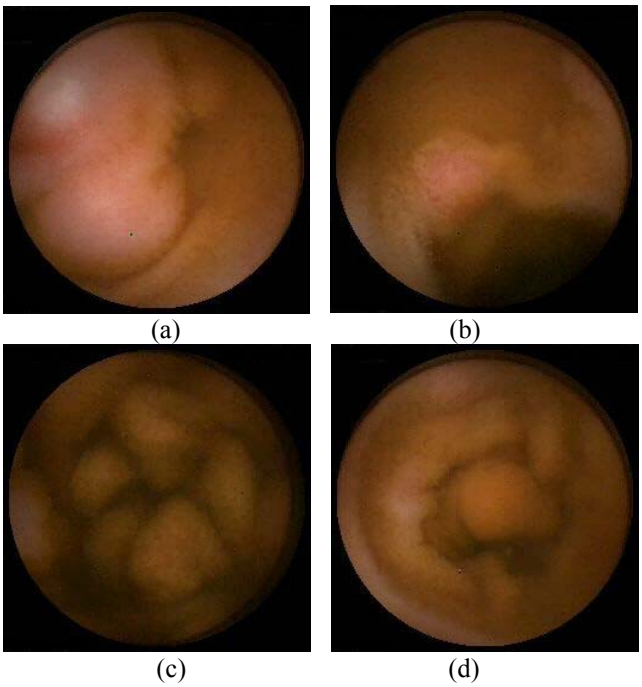
Fig.1 Capsule endoscopy

The authors are with the Department of Electronic Engineering at The Chinese University of Hong Kong, Hong Kong, SAR, China, email: {bpli & max}@ee.cuhk.edu.hk. This research is supported by SHIAE project #8115021 of the Shun Hing Institute of Advanced Engineering of The Chinese University of Hong Kong, awarded to Max Meng.

However, there exists a major weakness associated with CE. There are about 50,000 images in total per examination for one patient, and it costs an experienced clinician about two hours on average to review and analyze all the video data [2]. There is also some other limitations about CE such as no control of the CE direction, limited camera field of view and incapacity to handle occlusion, which are beyond the scope of the presented work at present.

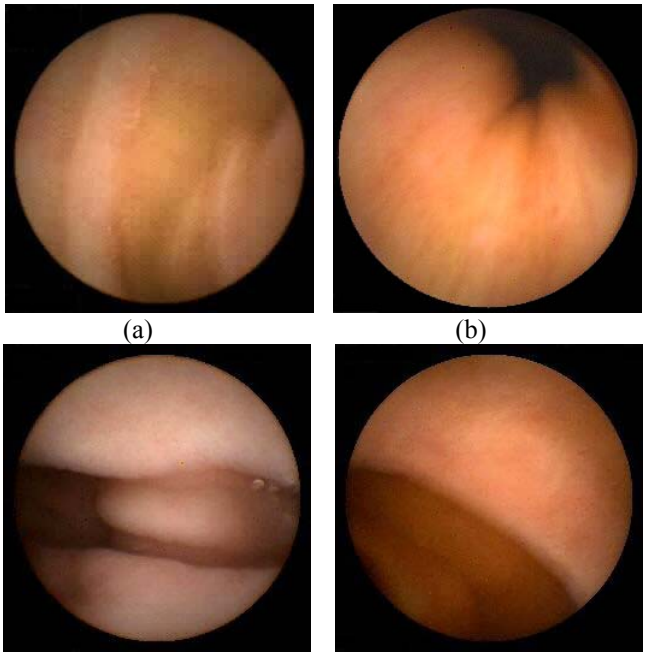
Because of its gradually wide application, some studies have been investigated towards the direction of partially automating inspection of CE images to reduce the burden of clinicians. The authors in [3] proposed a method using color distribution to discriminate stomach, intestine and colon tissue. An interesting approach of choosing MPEG-7 visual descriptors as the feature extractor to detect several diseases such as ulcers and bleeding in the GI tract was advanced in [4]. We have studied bleeding region detection and ulcer region detection for CE images in [5-7], the preliminary experimental results validate that the proposed schemes work fine for CE images.

We continue to delve into tumor CE image recognition in this work since tumor is another common disease in GI tract. Tumor is a swelling or lesion formed by an abnormal growth of cells, it is characterized by its color, bleeding tendency and so on. Fig.2 and Fig.3 illustrate some tumor CE images and some normal CE images, respectively.



(a) (b) (c) (d)

Fig. 2 Some tumor CE images



(a) (b) (c) (d)

Fig.3 Some normal CE images

From these illustrations, it can be noticed that tumor in CE images show great variations in color and shape. Moreover, the size of tumor region also varies greatly in each image. Even for normal CE images, they still show different color and texture pattern. This situation poses a great challenge to feature design since it may be difficult to design a single powerful feature to discriminate normal CE images from tumor CE images. Moreover, the image resolution of CE images used in our experiments is only 256×256, while traditional commercial endoscopy images have a higher resolution such as 720×480 or even higher. We attempt to design a computer aided detection (CAD) system for tumor CE image classification in this paper. To achieve this goal, textural features of CE images are extracted as candidate image features. The extracted features cannot guarantee a good detection performance and the computational cost of classification grows drastically with feature dimension size. Therefore, we apply sequential forward floating feature selection (SFFS) method to reduce the dimension of image features and improve the classification accuracy.

The remainder of this paper is designed as follows. The texture features for CE images using wavelet transform based local binary pattern is discussed in the following section, followed by feature selection scheme and the classifier used in our work. Section IV gives the experimental results, and we draw some conclusions at the end of this paper.

II. Color Texture Extracted

Since clinicians mainly use color and textural features to judge the status of CE images, we will investigate color textural features to discriminate normal CE images from tumor ones. As has been demonstrated in our preliminary work [8], wavelet based local binary pattern may exhibit a promising performance for tumor detection in small intestine. In this study, we choose this approach as the candidate features to describe GI tract tumor. It should be pointed out that the work of this paper is different from our previous work mainly in the following three aspects: First, the location of tumor is extended to the whole GI tract rather than only the small intestine. Moreover, the features used in this paper are refined by SVM based feature selection. In addition, the experiments in this paper are more comprehensive.

A. Local Binary Pattern

Local binary pattern (LBP) texture operator, which was first proposed by Ojala et al. [9], is invariant against any monotonic grey scale transformation and is computationally simple. The image pixels are first labeled by thresholding the difference between the central pixel and its neighbors using the step function. Then values of the pixels in the thresholded neighborhood are multiplied by binomial weights given to the corresponding pixels. Finally, values of the products are summed up to obtain an LBP number of this neighborhood. A simple yet efficient multi-resolution approach to gray-scale and rotation invariant texture based on LBP was further proposed by them in [10]. In this paper, they found that some local binary patterns are fundamental, and these fundamental patterns are called 'uniform'. The uniform patterns have circular structures that contain few transitions from 0 to 1. In order to formally define the 'uniform' patterns, a uniformity measure U was introduced, which corresponds to the number of spatial transitions (bitwise $0/1$ changes) in the pattern. For instance, both pattern 00000000 and pattern 11111111 have U values of 0, and patterns that have U values of at most 2 are designated as 'uniform'. Based on the above discussions, a uniform operator is defined:

$$LBP^{u2}_{P,R} = \begin{cases} \sum_{p=0}^{P-1} s(g_p - g_c) & if\ U(LBP_{P,R}) \le 2 \\ P+1 & otherwise \end{cases} \quad (1)$$

where $U(LBP_{P,R}) = |s(g_{P-1} - g_c) - s(g_0 - g_c)| + \sum_{p=1}^{P-1} |s(g_p - g_c) - s(g_{p-1} - g_c)|$

and $s(x)$ is the sign function. This operator is a useful measure of local image texture, and it is robust to monotonic transformation of gray scale and also simple to implement. CE images suffer from illumination variation due to the specific imaging circumstances such as limited range of the illumination in the digestive tract. Therefore, it is necessary to consider illumination variation effects on textures of CE images since texture features are not constant to illumination variation. Uniform local binary pattern demonstrates rather robust performance to illumination change.

B. Discrete Wavelet Transform

Wavelet is a useful tool for multi-resolution analysis of an image. As for CE images, wavelet can provide zooming ability and local characterization to better analyze the mucosa of the inner GI tract, leading to more capable image analysis using information at different scales. Because wavelet transform for an image can be implemented with discrete wavelet transform (DWT), we briefly review it as follows.

DWT for an image is implemented with a separable filter-bank to an image [11], and the image is convoluted with a low-pass filter L and high-pass filter H recursively. Due to the decomposition of an image using DWT, the image is transformed to four sub-images which are generally denoted as LL, LH HL and HH. The LL sub-image comes from low pass filtering in both directions and it is the most like original picture, so it is called the approximation component. The remaining sub-images are called detailed components. The HL is derived from high pass filtering along the horizontal direction and low pass filtering along the vertical direction and so has the label HL. The other two sub-images LH and HH have similar explanations. In this study, we apply three levels DWT to each channel of a CE image.

Because textural features are better encoded in the middle wavelet detailed sub-images [12], we choose the middle level sub-images, i.e., HL2, LH2 and HH2, as the basis for textural feature analysis. For color images, three channels result in nine such sub-images. The aforementioned uniform LBP is further applied to each sub-image to describe the color textural features for CE images. Using the uniform LBP histogram of each sub-image in each color channel, we may obtain six statistical measurement of the histogram as features of the texture [13] in order to reduce the number of features. Hence, each CE image can be characterized with a feature vector with 54 ($6\times3\times3$) elements.

To provide more features for further analysis, we will investigate the proposed feature extraction methods in RGB space, HSI space and YCbCr space. RGB color space is the most convenient color space, and HSI color space and YCbCr space are representative spaces that separate color information into chromaticity and intensity. We will combine all the features from these three color spaces as candidate textural features to represent CE images, so a CE image will be represented with a length of 162 (54×3) features. Although features from different color spaces may be uncorrelated, they may contain some supplementary information which may provide better discrimination ability for tumor CE image classification. And experiments which will be presented later validate such an effect that is helpful to improve the detection accuracy.

III. Feature selection and classification

Based on the previous section, a total of 162 texture features are extracted for each CE image. We hope to know which ones are useful in discriminating tumor CE images from normal CE images since there may exist some redundancy between the proposed color textural features. Principal component analysis has been a useful tool to reduce the feature dimension. However, it cannot guarantee a good classification performance using the principal component of the features. As the main goal of our present study is to find those features that show best discrimination ability for tumor CE images and normal images, we turn to feature selection approach. In this section, we will use the classification performance of SVM as the criterion to guide feature selection process.

A. Feature selection

Feature selection (FS) is a problem of finding an optimal subset of features based on some selecting algorithms. Not only the efficiency of recognition can be improved by reducing the number of features, but the performance of classification may be improved by the optimal subset features. Traditionally, feature selection is implemented independently of the classifier used. This strategy is efficient since its computational burden is not much. However, separation of feature selection from classification might lead to possible loss of information related to classification. To solve this problem, another strategy, wrapper methods, were proposed. Wrapper approach consists of searching for the optimal subset features that can maximize the generalization ability of a classifier. In this work, since our target is to find the best color textural features that separate normal CE images from tumor ones, we will adopt the wrapper method to implement feature selection. It should be pointed out that better generalization ability of the obtained feature subsets is at the cost of a larger computational burden caused by classifiers.

Another important issue in feature selection is feature subset search, which has also been actively investigated [14]. Complete search performs a thorough search for the optimal subset features. Exhaustive search is a typical example of this kind of method, whose time complexity grows exponentially with the number of features, preventing its practical usage in experiments. Branch-and-bound feature selection algorithm was advanced for searching an optimal subset of features. However, it requires a monotonic feature selection criterion, which may not hold true in reality [15]. Suboptimal search methods were further proposed to overcome the above shortcomings. Typical examples were sequential backward selection (SBS) and its counterpart, sequential forward selection (SFS). SBS sequentially removes features from a full candidate set until the removal of further features decrease the criterion. For SFS, features are sequentially added to an empty candidate set until the addition of further features does not increase the criterion. Both of these two methods work fast due to the comparative less search burden. However, they are shown to have nesting problems [16], i.e., SFS approach does not re-select the discarded features while SBS approach will not give up features once they have been chosen. To overcome such a shortcoming, sequential floating forward search (SFFS) algorithm was proposed by Pudil et. al. [16], and it is claimed that the produced features come close to those of branch-and-bound without its drawback. In this approach, some backwards search steps are also tested to backwardly select the subset feature for each forward step. SFFS is employed in our proposed scheme to select the most desired features due to advantages mentioned above.

Assume Y is the candidate image features with size n and X are the selected features with size m. The detailed procedures of SFFS can be described in the following steps [16]:

Input: $Y = \{y_i \mid i = 1, 2...n\}$

Output: $X_k = \{x_i \mid i = 1, 2...k, x_i \in Y\} \quad k = 0, 1...m$

Initialization:

$$X_0 = NULL \quad k = 0$$

Step1 (Inclusion)

$$x_{add} = \arg\max J(X_k + x), x \in Y - X_k$$

$$X_{k+1} = X_k + x_{add} \; k = k + 1$$

Step2 (Exclusion)

$$x_{delete} = \arg\max J(X_k - x), x \in X_k$$

$$J(X_k - \{x_{delete}\}) > J(X_{k-1}), then$$

$$X_{k-1} = X_k - x_{delete}, k = k - 1$$

Go to step 2

Else go to step 1

Till termination

The function J used here is the accuracy of SVM, the larger it is, the better the features selected. After using SFFS on the features, X_m will be the feature subsets that perform best classification on training data sets in terms of the criterion function, i.e., accuracy of SVM. Meanwhile, the dimension of features will be reduced in this process.

B. Support vector machine

Support vector machine is a state-of-the-art technique used for classification, and it has been shown to demonstrate better accuracy and computational advantages over some other traditional classification approaches [17].

An SVM constructs a binary classifier from a set of labeled patterns called training set. Let $(x_i, y_i) \in R^N \times \{-1, +1\}, i = 1...n$ be such a set. The purpose here is to select a hyper-plane function $f: \; R^N \to \{\pm 1\}$ from a given class of functions such that it can maximize the margin separating the two classes of samples in a high-dimensional space. The hyper-plane of a liner classification could be written as $\omega x + b = 0$, where ω and b are parameters. By rescaling such two parameters, the margin mentioned above can be denoted as

$$d = 2 \Big/ \|w\|^2 \tag{2}$$

The learning task for a SVM can be expressed as an optimization problem

$$\min_w \left\{ \frac{\|w\|^2}{2} \right\} \text{ with } y_i(\omega \cdot x_i + b) \geq 1, \quad i = 1, 2, ...n \tag{3}$$

The dual version of the above problem corresponds to the solution of the following quadratic program:

$$\arg\max J = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1, j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j \tag{4}$$

Subject to $\alpha_i \geq 0, \sum_{i=1}^{n} \alpha_i y_i = 0$, where α_i is a real number.

The decision boundary can then be built from the solutions of (4) and the final SVM can be written in the following form:

$$f(x) = \sum \alpha_i y_i x_i \cdot x + b \tag{5}$$

IV. Experiment results

A. Experimental Data

A data set composed of 600 representative tumor CE images and 600 normal CE images from 12 patients' video data was built by two experts of GI tract with about 5 years of experiences with CE images. These images are obtained from M2A CE, the first generation product of Given imaging company, and the resolution of these images is 256×256. 50 normal images and 50 tumor images was chosen by these experts from each patient's video segments. Concerning normal images, we mainly mean those images that have healthy mucosal appearance. However, to make the normal image data more characteristic, some CE images with a few bubbles and some indigested food or feces are also chosen by these experts. To avoid the possible overlap between continuous frames, these experts chose the frames at some interval, which is greater than or equal to 4 frames since the M2A CE records images at a frame rate of 2 frames per second. The original images are manually labeled to provide the ground truth.

B. Experimental Results

In our experiments, we divided our data into two subsets with equal size, and we used one subset to choose the optimal feature sets. After selection of the features, we made use of the entire data set to evaluate the performance of the selected features. In the process of evaluation using selected features, we exploited 2-fold cross-validation for all our classification experiments in order to prevent over-fitting of the classification results. Moreover, since the image data from the same patient may have similar appearances sometimes, all the images from one patient are put in either the training set or the test set. As such, each time 300 normal samples and 300 tumor samples from six patients are used as the training sets, while the left 300 normal samples and 300 tumor samples from the left six patients are used as the test sets, and we repeat this procedure 2 times. Finally, the average recognition rates are used to assess the performance of classification. It should be pointed out that the data used in the process of feature selection is also chosen from six patients' image data.

Since we use classification accuracy of SVM as the feature selection criterion, we employ LIBSVM [18] for the implementation of classification. Meanwhile, we also use Gaussian radial basis function (RBF) as the kernel function in our experiments. The optimal values of two parameters, i.e.,

penalty parameter and the kernel parameter of RBF, for the classification were found using a grid search approach [19] in each folder validation experiment.

The recognition results using all the features concatenated in three color spaces and the result using features selected by SFFS are demonstrated in Table 1. It can be noticed that accuracy of the selected features with SFFS is 88.3%, showing an improvement margin of 5.7% compared to the original features. It should be noticed that 5.7% improvement in the field of medical decision is encouraging in fact. Such a result demonstrates that the proposed scheme using SFFS can improve the discrimination ability of the proposed color texture features.

TABLE1 CLASSIFICATION RESULTS USING ALL FEATURES WITHOUT FS AND WITH FS (%)

	Features	Features after SFFS
Specificity	80.7±2.04	83.5±2.16
Sensitivity	84.6±1.92	93.0±1.87
Accuracy	82.6±1.96	88.3±2.08

TABLE2 CLASSIFICATION RESULTS USING ORIGINAL FEATURES EXTRACTED IN DIFFERENT COLOR SPACE (%)

	RGB	YCbCr	HSI
Specificity	71.8±3.16	70.0±3.62	50.0±3.18
Sensitivity	54.7±3.24	56.2±4.26	80.1±4.08
Accuracy	63.3±3.22	63.1±3.84	65.4±3.64

We further illustrate performances of the original separate features extracted in different color spaces in Table 2. It can be noticed that the best performance of the original features in individual color space is only 65.4%, which is obtained in HSI color space. Moreover, the performance difference among different color spaces in terms of average accuracy is not much. However, it can be observed that HSI provides a best sensitivity of 80.1%, while the other two color spaces produce rather good performance of specificity. By comparing Table 2 to that of Table 1, we can see clearly that after combing the features extracted from three different color space, the features demonstrate significantly superior discrimination ability for tumor CE images classification. This is due to the fact that features from different color spaces may contain complementary information and integration of features from different color spaces may have a positive influence on detection. After SFFS, the selected features show further classification accuracy improvement. In addition, the features extracted with SFFS contain only 10 features, which reduce greatly the number of the original features integrated from three color spaces. The increase in accuracy may be ascribed that using classification accuracy of SVM to guide the feature selection, the most representative features are extracted. In addition, feature redundancy is also reduced to some extent after using SFFS, resulting in better performance of classification.

To further illustrate the performance of the proposed scheme, we compared it with a recent method also aimed for tumor detection for endoscopic images, i.e., the color wavelet covariance (CWC) features [20]. CWC feature is a recently new technique to describe color textural features that are built upon the covariance of second-order textural measures in wavelet domain of color channels of an image. We also used 2-folder cross-validation for CWC to test its discrimination ability for WCE images. The performance of CWC on our collected data using SVM is 75.83% in specificity and 82.83% in sensitivity, which is inferior to the best performance obtained with the features selected by SFFS. It should be noted that here in CWC experiments, the parameters were also optimized by grid search.

V. CONCLUSIONS

Tumors are a common disease in GI tract and CE is a new mini-invasive technology to examine the GI tract. It is necessary to develop a computer aided diagnosis system for tumor detection for CE images, which is a very challenging problem in fact due to the fact that tumors exhibit great variations in GI tract. In this paper, wavelet based uniform local binary pattern is first proposed to discriminate tumor CE images from normal ones. To make the features more discriminative, SVM based SFFS feature selection scheme is employed to refine the proposed features. Comprehensive experimental results show that for tumor CE image classification, features obtained with SFFS achieves a best accuracy of 88.3%, resulting in an improvement margin of 2.8% in specificity and 8.4% in sensitivity compared to features without selection. Moreover, the dimensionality of the features were drastically reduced to 10.The problem of further increasing specificity for tumor recognition remains to be solved in the near future since there is still much room for its improvement.

ACKNOWLEDGMENT

We would like to show our sincere thanks to James Lau, a professor in Prince of Wales Hospital in Hong Kong, for providing us the CE image data mentioned in this paper. We also thank the anonymous reviewers whose comments greatly improved this manuscript.

REFERENCES

[1] Given Imaging Ltd., Internet site address: http://www.givenimaging.com
[2] D. G. Adeler, C. J.Gostout, "Wireless capsule endoscopy," Hospital Physician, pp.14-22, May, 2003.
[3] J. Berens, M. Mackiewicz, and D. Bell. "Stomach, intestine and colon tissue discriminators for wireless capsule endoscopy images," Proceedings of SPIE. Conference on Medical Imaging, volume 5747, Bellingham, WA, 2005, pp.283-290
[4] M.T. Coimbra. J.P.S. Cunha, "MPEG-7 visual descriptors—contributions for automated feature extraction in capsule endoscopy," IEEE Transactions on Circuits and Systems for Video Technology, Volume 16, Issue 5, pp.628- 637, 2006
[5] B. Li, Max Q.-H. Meng, "Computer aided detection of bleeding regions in capsule endoscopy images," IEEE Transactions on Biomedical Engineering, Vol. 56, No. 4, April, 2009, pp.1032-1039;

[6] B. Li, Max Q.-H. Meng, “Texture analysis for ulcer detection in capsule endoscopy images,” Image and Vision Computing, Vol. 27, No. 9, August, 2009, pp.1336-1342;

[7] B. Li, Max Q.-H. Meng, “Computer-based detection of bleeding and ulcer in wireless capsule endoscopy images by chromaticity moments,” Computers in Biology and Medicine, Vol. 39, No. 2, February, 2009, pp.141-147;

[8] B. Li, Max Q.-H. Meng, "Small bowel tumor detection for wireless capsule endoscopy images using textural features and support vector machine," IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2009), St. Louis, MO, USA, October, 2009

[9] T. Ojala, M. Pietikainen, D.Harwood “A comparative study of texture measures with classification based on feature distributions,” pattern recognition, vol.29, pp.51-59.

[10]T. Ojala, M. Pietikainen and T. Maenpaa “Multi-resolution gray-scale and rotation invariant texture classification with local binary pattern” IEEE Trans. on PAMI, vol.24, No.7, pp971-987,2002

[11]S. G. Mallat, “A theory for multiresolution signal decomposition: The wavelet representation,” IEEE Trans. Pattern Anal. Machine Intell., vol.11, pp.674–693, 1989.

[12]K. W. Abyoto, S. J. Wirdjosoedirdjo, and T. Watanabe, “Unsupervised texture segmentation using multiresolution analysis for feature extraction,” J. Tokyo Univ. Inform. Sci., vol. 2, pp. 49–61, Jan. 1998

[13]M. Boulougoura, E. Wadge, “Intelligent systems for computer-assisted clinical endoscopic images analysis,” proceedings of Second International Conference on Biomedical Engineering, Austria, Feb.16-18, 2004, pp.405- 408.

[14]M. Dash and H. Liu, “ Feature selection for classification,” Intelligent Data Analysis, Vol.1, 1997, pp.131-156

[15] A.K. Jain, Douglas Zongker, “Feature selection, evaluation, application, and small sample performance” IEEE Trans. on PAMI, vol.19, no.2, 1997,pp.153-158

[16]P. Pudil, J. Novovicova, J.Kittler, “Floating search methods in feature selection,” Pattern Recognition Letter, vol.15, no.1, 1994, pp. 1119-1125,

[17V. Vapnik, The Nature of Statistical Learning Theory. Springer Verlag, New York, 1995.

[18] C.-C. Chang, C.-J. Lin. LIBSVM: a library for support vector machines, 2001. Software available at http://www.csie.ntu.edu.tw/cjlin/libsvm.

[19] C. W Hsu, C.-C. Chang, C.-J. Lin, “A practical guide to support vector classification,” Technical Report, Department of Computer Science & Information Engineering, National Taiwan University, Taiwan. 2003 http://www.csie.ntu.edu.tw/cjlin/papers/ guide/guide.pdf

[20] S.A. Karkanis, Dimitris K. Iakovidis, Dimitris E. Maroulis, Dimitris A. Karras, “Computer-Aided tumor detection in endoscopic video using color wavelet features,” IEEE Trans. Info. Tech. in Biomedicine, vol. 7, no.3, pp.141-152, Sep. 2003.