

# Multi-Label Classification for Colon Cancer Using Histopathological Images

YAN XU,<sup>1,2</sup> LIPING JIAO,<sup>1</sup> SIYU WANG,<sup>1</sup> JUNSHENG WEI,<sup>1</sup> YUBO FAN,<sup>1</sup> MAODE LAI,<sup>3</sup> AND ERIC I-CHAO CHANG<sup>2\*</sup>

<sup>1</sup>State Key Laboratory of Software Development Environment, Key Laboratory of Biomechanics and Mechanobiology of Ministry of Education, Beihang University, Beijing 100191, China

<sup>2</sup>Microsoft Research, Beijing, China

<sup>3</sup>Department of Pathology, School of Medicine, Zhejiang University, China

**KEY WORDS** colon cancer; histopathological image; multi-label; multi-SVM

**ABSTRACT** Colon cancer classification has a significant guidance value in clinical diagnoses and medical prognoses. The classification of colon cancers with high accuracy is the premise of efficient treatment. Our task is to build a system for colon cancer detection and classification based on slide histopathological images. Some former researches focus on single label classification. Through analyzing large amount of colon cancer images, we found that one image may contain cancer regions of multiple types. Therefore, we reformulated the task as multi-label problem. Four kinds of features (Color Histogram, Gray-Level Co-occurrence Matrix, Histogram of Oriented Gradients and Euler number) were introduced to compose our discriminative feature set, extracted from our dataset that includes six single categories and four multi-label categories. In order to evaluate the performance and make comparison with our multi-label model, three commonly used multi-classification methods were designed in our experiment including one-against-all SVM (OAA), one-against-one SVM (OAO) and multi-structure SVM. Four indicators (Precision, Recall, F-measure, and Accuracy) under 3-fold cross-validation were used to validate the performance of our approach. Experiment results show that the precision, recall and F-measure of multi-label method as 73.7%, 68.2%, and 70.8% with all features, which are higher than the other three classifiers. These results demonstrate the effectiveness and efficiency of our method on colon histopathological images analysis. *Microsc. Res. Tech.* 76:1266–1277, 2013.

© 2013 Wiley Periodicals, Inc.

## INTRODUCTION

Colon cancer is one of the many cancers that lead to people's death. World Health Organization statistics show that: new global cancer prevalence was more than twelve million, and death more than 7.5 million in 2008. The high incidence and mortality of colon cancer are causing increasing attention all over the world.

Accurate colon cancer type determination is significantly important for cancer treatment. According to WHO histological classification of tumors of the colon and rectum (World Health Organization Classification of Tumours, 2000), common types of colon cancer include well-differentiated adenocarcinoma, moderately differentiated adenocarcinoma, poorly differentiated adenocarcinoma, mucinous adenocarcinoma, and signet-ring cell carcinoma. Figure 1 shows the different cancer types. For adenocarcinoma, the percentage of the tumour showing conformation of gland-like structures is used to define the grade. Well-differentiated adenocarcinoma shows glandular structures in more than 95% of the tumors; a moderately differentiated lesion has 50–95% glandular structures; poorly differentiated adenocarcinoma has 5–50%. Mucinous adenocarcinoma, this definition is used when more than 50% of the diseased tissue is

composed of mucin. Signet-ring cell carcinoma is defined by the existence of greater than 50% of tumor cells with prominent intracytoplasmic mucin. Other types that are rarely appearing in clinic are not considered here including medullary carcinoma, undifferentiated carcinoma, spindle cell carcinoma or sarcomatoid carcinoma, choriocarcinoma and paneth cell-rich (crypt cell carcinoma), etc. (World Health Organization Classification of Tumours, 2000).

With the digitalization of pathology slices becoming an inevitable tendency, colon cancer diagnosis highly depends on pathology image analysis. However, considering the property of colon cancer type definition, traditional manual analysis method is time-consuming

\*Correspondence to: Eric Chang, Microsoft Research T2-14463, No 5 Danling Street, Haidian District, Beijing 100080, People's Republic of China. E-mail: eric.chang@microsoft.com

REVIEW EDITOR: Prof. George Perry

Received 3 August 2013; accepted in revised form 5 September 2013

Contract grant sponsors: Microsoft Research, MSR eHealth grant; Contract grant sponsor: National Science Foundation of China; Contract grant number: 61073077; Contract grant sponsor: State Key Laboratory of Software Development Environment in Beihang University in China; Contract grant number: SKLSDE-2011ZX-13.

DOI 10.1002/jemt.22294

Published online 5 October 2013 in Wiley Online Library (wileyonlinelibrary.com).

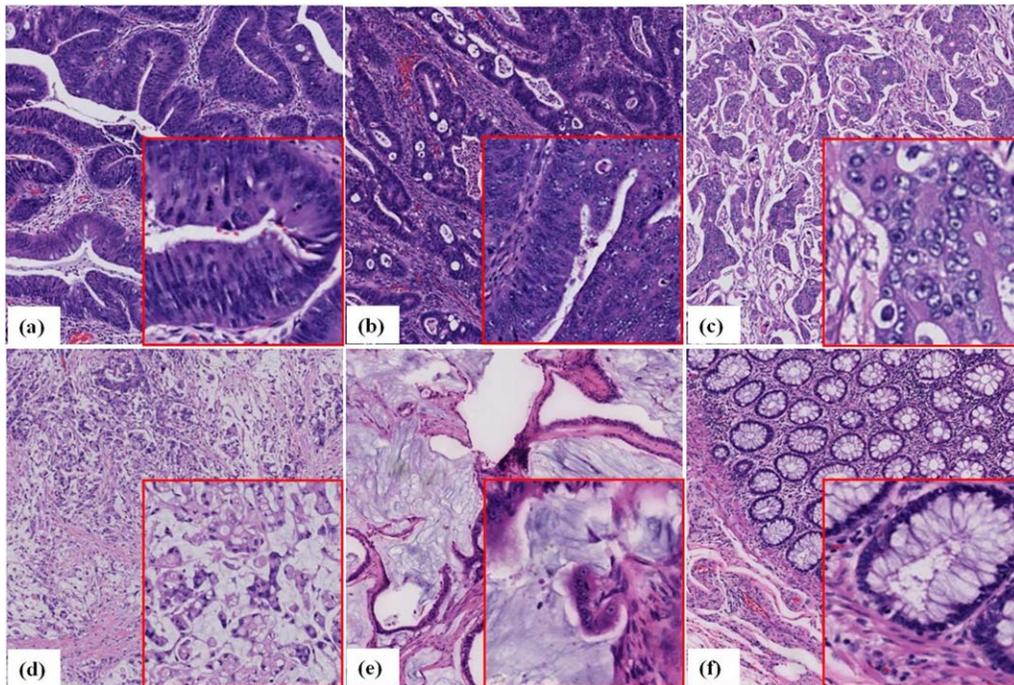


Fig. 1. Six images with different types: (a) well differentiated adenocarcinoma; (b) moderately differentiated adenocarcinoma; (c) poorly differentiated adenocarcinoma; (d) signet-ring cell carcinoma; (e) mucinous adenocarcinoma; (f) normal image. [Color figure can be viewed in the online issue, which is available at [wileyonlinelibrary.com](http://wileyonlinelibrary.com).]

and labor-costing. Hence, it is very valuable to design a set of automatic detecting system based on electronic pathology slices to reduce labor amount and increase diagnosing accuracy. An effective and efficient way is to build a classifier to implement colon cancer type classification automatically. Three commonly used multi-classification methods, one-against-all SVM (Kij-sirikul and Ussivakul, 2002), one-against-one SVM (Friedman, 1996), and multi-structure SVM (Crammer and Singer, 2000), were selected to build different classifiers for the multi-classification task of colon cancer image analysis.

However, according to clinical data annotation given by experts, one colon cancer image may contain different types of cancer regions. In this case, it is inaccurate to classify this image to only one type. For this reason, we proposed to consider the classification of such images as a matter of multi-label task. In traditional single-label classification method, each example is associated with a single label from a set of disjoint labels  $L$ ,  $|L| > 1$  (Tsoumakas and Katakis, 2007). While multi-label classification is concerned with a set of examples where each example is associated with a set of labels  $Y$  in  $L$ . Nowadays, multi-label classification which is mainly motivated by the tasks of text categorization, is applied to protein function classification, music categorization, semantic scene classification and many other similar modern applications successfully (Tsoumakas and Katakis, 2007). In our task, we concentrated on this issue and designed a classifiers based on multi-label for colon cancer images classification to enhance the veracity of analyzing for the colon cancer pathologists, the other three methods

were based on single label classification, so as to facilitate more effective diagnosis and treatment of colon cancer.

On the other hand, feature designation is well considered since feature is of significance for accurate classification result. Color is the first selected feature since the pathology images are dyed by means of Hematoxylin and Eosin technique (Allen, 1992; Huang et al., 2009). This technique uses two separate dyes, one staining the nucleus while the other staining the cytoplasm and connective tissue. Hematoxylin is a dark purplish dye which can stain the chromatin (nuclear material) within the nucleus, leaving it a deep purplish-blue color. Eosin is an orangish-pink to red dye that stains the cytoplasmic material including connective tissue and collagen, and leaves an orange-pink counterstain. This counterstain acts as a sharp contrast to the purplish-blue nuclear stain of the nucleus, and helps identify other entities in the tissues such as cell membrane (border), red blood cells, and fluid (Allen, 1992). Thus, color feature is one of the important features of pathology images (e.g., Fig. 1). Another two features we selected for classification are texture and shape. The cell structures are obviously different between common tissue and cancer ones, i.e., arrangement patterns (texture structure) and shapes of cells. To verify the contribution of each feature set, we repeated the classification process four times by adding a new feature set to the last feature set and utilizing each feature set individually to produce a new classifier, followed by comparison of the classification results.

Two contributions were described in this article: (1) Reformulating the issue of cancer classification from

another perspective, and realizing multi-classification of colon cancer by multi-classification based on multi-label method. (2) Extracting four kinds of features to build our feature set, demonstrating how these features are suitable for colon cancer pathological image classification.

The article followed introduces with related works, describes in details the method of classifier building based on multi-label classification for colon cancer image classification along with other three single label multi-classification methods, designs an experiment to compare the classification results of different classifiers, and elaborates debates with discussion and conclusions according to the results.

### RELATED WORK

We divided related work into two broad categories: (1) research on histopathological images, more specifically, tissue histopathology slides can be digitized and stored in digital image form, and (2) classification based on multi-label SVM.

Histopathology is a very hot topic, and histopathological images are gradually becoming more common. With the advent and cost-effectiveness of whole-slide digital scanners, tissue histopathology slides can now be digitized and stored in digital image form (Madabhushi, 2009). Hence, research on pathological section image classification is booming these years. At first, such operations merely concentrate on binary classification (normal and cancer), while explorations in multi-classification becomes more various in recent years. Esgiar et al. (1998) introduced some research of early days, in which data used for the experiment was acquired by micrography and multivariate analysis method was taken. The disadvantage reflected in the low resolution of Esgiar's examples, is restricting the information content to a low grade. Shuttleworth et al. (2002) combined color and texture information of both high and low resolution images based on discriminate analysis method. Amin et al. (2003) realized multi-classification of colon cancer images by means of genetic algorithm, but the experiment result contained some large diversities. The author believed such shortage is due to the distinct differentiation degrees of two examples. However, we prefer to take this for multiple reasons, first, features reflecting image differentiation degrees are not extracted; second, the property of classifiers has the possibility to improve, etc. Nwoye et al. (2006) got a result with high accuracy through building classifiers to process the images by using the co-occurrence matrices feature based on fast fuzzy back-propagation algorithm. Altunbay et al. (2010) took advantage of structural features from color graphs and realized the grading to colon tissues according to SVM, which achieved a big breakthrough in the field of pathological image based cancer grading. With the development of computer technology, algorithm at this realm has been studied extensively in recent years. Xu et al. (2012) compared multiple clustering instance learning (MCIL) algorithm with multiple kernels learning (Vedaldi et al., 2009), multiple instance learning (Dundar et al., 2010) and other relative algorithms, which were able to accomplish the classification as well as the segmentation and cluster of image at the same time. The above list shows some representative

articles related to colon cancer pathological image classification.

Multi-label is a ubiquity of debate within nature, especially in the fields of texts (McCallum, 1999) and figures (Li et al., 2004). To minimize the efforts of labeling text and without sacrifice the classification accuracy, a novel multi-label active learning approach was produced (Yang et al., 2009). Huang et al. (2011) applied multi-label categorical K-nearest neighbor (ML-CKNN), which empirically showed to outperform ML-KNN (Zhang and Zhou, 2007) and other multi-label algorithms.

As we all know, in the natural scene images, an image may contain many different objects, that is multiple instances, so multi-label approach was used widely on classifying multiple instances of natural scene images. Furthermore Grady and Funka-Lea (2004) proposed a semi-automated image segmentation method based on multi-label by giving a small number of pixels with user-defined labels which can obtain high-quality image segmentation. Another phenomenon was found that not only multiple instances but also multiple class labels existed, so the multi-instance multi-label learning BOOST (MIMLBOOST) and multi-instance multi-label learning SVM (MIMLSVM) algorithms were proposed which can achieve good performance in an application to scene classification (Zhou and Zhang, 2006). To solve the same problem, an integrated multi-label multi-instance learning (Zha et al., 2008) approach based on hidden conditional random fields was proposed, which meanwhile captured both the correlations among the labels in a single formulation and the connections between semantic labels and regions.

In reality, one or more keywords are invariably found within a text or a sentence. Also most images contain more than a single scene or object. Therefore, such examples are propitious to be marked with several labels. All kinds of statistic methods and machine learning methods, as is constantly deepened and extended especially in the machine learning field, have been applied for data mining as well as image labeling. Elisseeff and Weston (2002) presented a Support Vector Machine like learning system to deal with multi-label issues. Other methods were capable to solve problems concerning multi-label includes: k-nearest-neighbor-based ranking approach (Chiang et al., 2012; Oliveira et al., 2008; Zhang, 2010), probabilistic neural network algorithm (Chiang et al., 2012), fuzzy similarity measure and k-nearest neighbors (FSKNN) (Jiang et al., 2012), etc.

Even though large amount of colon cancer classification relative researches are now being implemented (Altunbay et al., 2010; Amin et al., 2003; Esgiar et al., 1998; Nwoye et al., 2006; Shuttleworth et al., 2002; Xu et al., 2012), they are mostly single-label. Our problem is extraordinarily similar to those that multi-label is appropriate for, therefore we adopt this method into application.

## METHODS

### Methods Overview

The key point in our task focused on the feature and classifier. In the following sections, we extract rational feature set based on image analyzing; besides, we

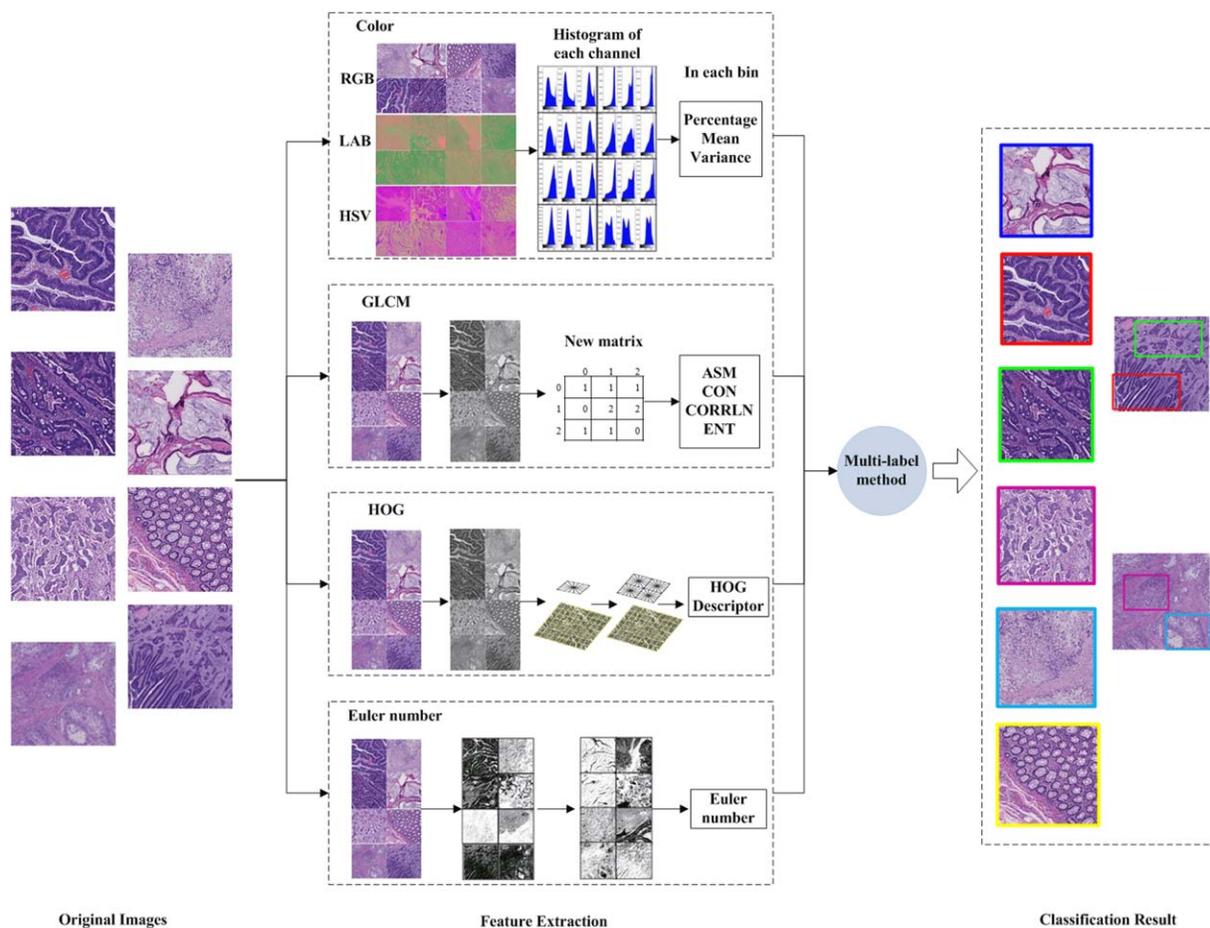


Fig. 2. Multi-label method flow chart. Blue rectangle: mucinous adenocarcinoma; Red rectangle: well differentiated adenocarcinoma; Green rectangle: moderately differentiated adenocarcinoma; Purple rectangle: poorly differentiated adenocarcinoma; Cyan rectangle: signet-ring cell carcinoma; Yellow rectangle: normal image. [Color figure can be viewed in the online issue, which is available at [wileyonlinelibrary.com](http://wileyonlinelibrary.com).]

introduce three commonly classifications and combined with classifiers build by Multi-label SVM method (Elisseff and Weston, 2002). The flow chart of the method is in Figure 2.

**Feature Extraction**

Feature extraction is a key step to build a classifier. Most commonly used features in pathological image classifying include: color, texture, and shape. Since there are large gaps between the colors of various colon cancer images dyed by Hematoxylin and Eosi technique, as well as the texture and shape among different kinds of image organizations, features extracted in this paper including color features based on histogram (Wang et al., 2010), texture features based on Gray-Level Co-occurrence Matrix (GLCM) (Jiao et al., 2013; Kong et al., 2009) and two kinds of shape features based on Histogram of Oriented Gradients (HOG) (Dalal and Triggs, 2005) and Euler number (Sleigh, 2001). From Figure 1 we can observe the differences of color, texture, and shape among different types of colon cancer images.

**Color: Color Histogram**

Cancerous cells and tissues are differentiating and secreting mucus all the time, leading to different species distribution within cells and tissues from the normal ones. After stained by Hematoxylin and Eosin, these differences among different cancer types are prominent, reflected in color diversity of the image. In this situation color feature will contribute to better classification results.

Feature descriptor is formed by the statistical data of color histogram that represents the distribution of colors in an image (Wang et al., 2010). Specifically, nine color channels' statistical data were obtained by concerning with three different color spaces (RGB, LAB, and HSV) while each space consists of three color channels. Each color channel was then further equally divided into five bins. For each bin, the pixels with the color values span within the bin were used to obtain the data we require, including the percentage of the pixels with respect to the whole pixel number in the image, the mean and variance of the pixel color values.

**Texture: GLCM Features**

Texture features derived from gray scale changes in space with a certain form. For example, normal cells are regularly arranged in the image, while cancerous cells are arranged irregularly. To extract the texture information of an image, Haralick proposed a second-order statistical method, named as Gray-Level Co-occurrence Matrix (GLCM) (Jiao et al., 2013; Kong et al., 2009). Here are some commonly used feature values showed in (1), (2), (3), and (4). Angular second moment (ASM): Always called energy, it is the sum of the GLCM elements' square, which indicates the distribution of gray level and texture coarseness.

$$ASM = \sum_i \sum_j p(i,j)^2 \tag{1}$$

Contrast (CON): It is the GLCM's contrast near the main diagonal moment of inertia. It reflects the clarity of the images and texture depth.

$$CON = \sum_i \sum_j (i-j)^2 p(i,j) \tag{2}$$

Correlation (CORRLN): It measures the similarity degree of GLCM elements in row or column directions, therefore the correlation values reflect the relevance of local gray images.

$$CORRLN = \left[ \sum_i \sum_j ((i,j)p(i,j)) - \mu_x \mu_y \right] / \sigma_x \sigma_y \tag{3}$$

Entropy (ENT): It reflects the randomness of image texture. When the GLCM's all values are equal, it obtains the maximum value. On the contrary, if the value is uneven, this value is small.

$$ENT = - \sum_i \sum_j p(i,j) \log p(i,j) \tag{4}$$

The GLCM method for feature extraction is described as follows. Consider an image of size  $N \times N$  pixels that has  $G$  kinds of Gray levels. Then, we can get a size  $G \times G$  co-occurrence matrix  $p_d$ . For the element  $(i, j)$  of  $p_d$  where gray levels are  $i$  and  $j$  respectively, the number of pixel pairs is determined by the formula (5).

$$p_d(i,j) = |\{(r,s), (t,v) : I(r,s)=i, I(t,v)=j\}| \tag{5}$$

In the formula  $(r, s), (t, v) \in N \times N, (t, v) \in (r + dx, s + dy)$ , “| |” stands for element number which often called set potential. Displacement vector is  $d=d(dx, dy)$  (Atlamazoglu et al., 2001),  $dx$  and  $dy$  are the displacements of  $r$  and  $s$ .

A simple example is showed in Figure 3. In it, Table (a) is the original image, Table (b) is the image transformed by  $d = d(1, 1)$  orientations' GLCM, and Table (c) is the image transformed by  $d = d(1, 0)$  orientations' GLCM. So 16 texture features were obtained by using the four GLCM features mentioned above. Each of the four GLCM features contains four orientations ( $d = d(1, 0), d = d(1, 1), d = d(0, 1),$  and  $d = d(-1, 1)$ ), so 16 features which represent the image texture features were obtained.

**Shape: HOG Features**

As we mentioned before, the pathology images are dyed by means of Hematoxylin and Eosin technique.

Table.(a) 4x4 original image

0	1	2	2
0	0	1	1
1	1	2	2
1	2	1	0

Table.(b) GLCM  $d=(1,1)$

	0	1	2
0	1	1	1
1	0	2	2
2	1	1	0

Table.(c) GLCM  $d=(1,0)$

	0	1	2
0	1	2	0
1	1	2	3
2	0	1	2

Fig. 3. An example of GLCM computation from the original image.

Hematoxylin stains the chromatin within the nucleus, leaving it a deep purplish-blue color. Meanwhile, Eosin stains the cytoplasmic material including connective tissue and collagen, and leaves an orange-pink counterstain. This technique generates a sharp contrast of cells and structures with apparent edges. To extract the edge features, histogram of oriented gradients (HOG) (Dalal and Triggs, 2005; Zhu et al. 2006) were used as the edge feature descriptors. HOG was proposed based on the idea that the distribution of pixel gradient values or edge directions can be used to describe the local object appearance and shape in an image. The HOG descriptors are implemented by four steps. (1) Graying and dividing the image into a quantity of small connected regions, called cells. (2) Computing a histogram of gradient directions or edge orientations for the pixels within each of the cells. (3) Local histogram contrast normalization that is able to result in better invariance to illumination or shadowing changes. Contrast normalization is attained by first defining blocks that across regions larger than cells in the image, then calculating the histogram intensity with each block, and finally normalizing all cells within the block. (4) Combining all the block histograms to form the HOG descriptor. The framework of HOG implementation process is showed in Figure 4.

**Shape: Euler Number Features**

Different types of cancer vary widely in the shape of the cells and tissues. From the red rectangle in Figure 1, we can simply find the connectivity structure's differences. The measurement method of most common spatial integrity (e.g., the number of cavities within cavity area) is called Euler function, which describes these functions with only one parameter, that is, Euler number (Hirzebruch, 1990).

To achieve the method, images needs binarization and negation afterwards. Euler number = (number of

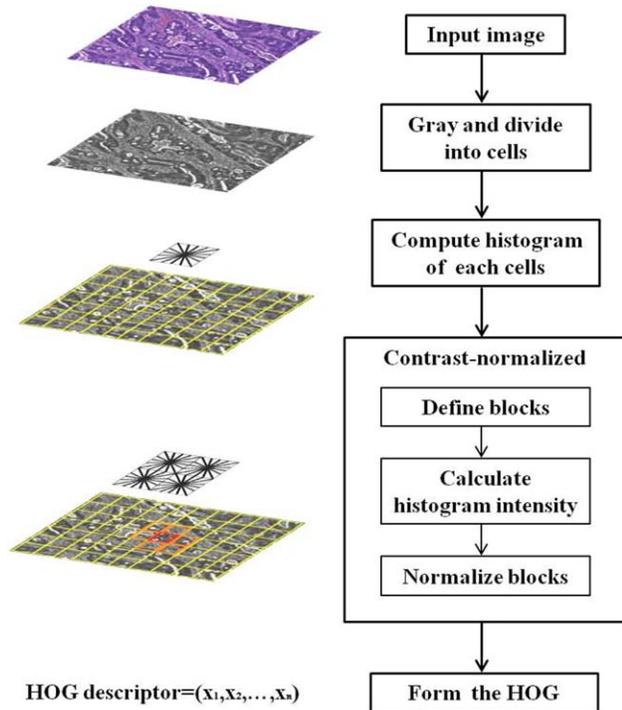


Fig. 4. Framework of HOG. [Color figure can be viewed in the online issue, which is available at [wileyonlinelibrary.com](http://wileyonlinelibrary.com).]

cavities) – (number of fragments – 1), the number of cavities here refer to the number of polygon cavities contained by external polygons themselves, and the number of fragments means the number of polygons contained by fragment area.

### Classifiers

SVM proposed in computer and statistics science is an excellent method of machine learning and has been extensively applied in classification and regression analysis. Histopathology images multi-classification based on SVM has gained lots of progress, such as one-against-all SVM (Kijirikul and Ussivakul, 2002), one-against-one SVM (Friedman, 1996) and multi-structure SVM (Crammer and Singer, 2000).

On the other hand, in text mining or bioinformatics area, literature for multi-labeled classification existed. A multi-label algorithm for labeling risk factors was applied to text classification (Huang et al., 2011), which could automatically identify 25 types of risk factors. Zhang and Zhou (2005) proposed a new method to take advantage of k-Nearest neighbor based algorithm for protein function classification, which was highly competitive to other existing multi-label learners. To achieve semantic scene classification, the scene can be described by multiple class labels (Boutell et al., 2004).

However, all these methods are proposed for text mining and bioinformatics multi-labeled problems, instead of focusing on colon cancer classification. To compare the capability for colon cancer image analysis, these four methods are stated briefly in this section.

For the three state of art methods, one-al one-against-all SVM (OAA), one-against-one BSVM (OAO), and structure SVM, the data is predefined as: given k

number of classes and l number of training examples, the training data set can be formed as  $(x_1, y_1), \dots, (x_l, y_l)$ , where  $x_i \in R^n$ ,  $i=1, \dots, l$  is the  $i_{th}$  example and  $y_i \in \{1, \dots, k\}$  is the class label of the  $i_{th}$  example.

### One-Against-All

In this approach, there is one binary SVM for each class to separate members of that class from members of other classes.

In the one-against-all approach, we build as many binary classifiers as there are classes, each trained to separate one class from the rest.

As we mentioned before, the idea is to train k SVM models each one separating one class from the rest. Once we have those binary classifiers, we use the probability outputs to predict new instances by picking the class with the highest probability.

This method constructs k number of SVM models. The  $i_{th}$  SVM model is trained by all the examples where the examples of the  $i_{th}$  class are set to positive label and all the other examples are set to negative label. The model is computed by solving k number of l-variable quadratic programming problems:

$$\min_{\omega^i, b^i, \xi_j^i} \frac{1}{2} (\omega^i)^T \omega^i + C \sum_{j=1}^l \xi_j^i$$

$$(\omega^i)^T \phi(x_j) + b^i \geq 1 - \xi_j^i, \text{ if } y_j = i, \quad (6)$$

$$(\omega^i)^T \phi(x_j) + b^i \leq -1 + \xi_j^i, \text{ if } y_j \neq i,$$

$$\xi_j^i \geq 0, \quad j=1, \dots, l,$$

where the function  $\phi$  is used to map the example  $x_i$  to a higher dimensional space and C is the penalty parameter. This equation leads to K decision functions:

$$\begin{aligned} & (\omega^1)^T \phi(x) + b^1, \\ & \vdots \\ & (\omega^k)^T \phi(x) + b^k. \end{aligned} \quad (7)$$

To predict a new instance x, we choose the classifier with the largest decision function value. Such that the class label of example x is computed as:

$$\arg \max_{i=1, \dots, k} ((\omega^i)^T \phi(x) + b^i) \quad (8)$$

### One-Against-One BSVM

One-against-one BSVM (Friedman, 1996) is also named Pairwise classification, in which there is one binary SVM for each pair of classes to separate members of one class from members of the other. For instance, if k is the number of classes, then  $k(k-1)/2$  classifiers are constructed and each one trains data from two classes. For instance, when classify example of i class label and j class label, the classification problem is formed as:

$$\begin{aligned}
& \min_{\omega^{ij}, b^{ij}, \xi_t^{ij}} \frac{1}{2} (\omega^{ij})^T \omega^{ij} + C \sum_t \xi_t^{ij} \\
& (\omega^{ij})^T \phi(x_i) + b^{ij} \geq 1 - \xi_t^{ij}, \quad \text{if } y_i = i, \\
& (\omega^{ij})^T \phi(x_t) + b^{ij} \leq -1 + \xi_t^{ij}, \quad \text{if } y_t = j, \\
& \xi_t^{ij} \geq 0.
\end{aligned} \tag{9}$$

After all the  $k(k-1)/2$  classifiers are constructed, prediction can be achieved. In classification we use a voting strategy: each binary classification is considered to be a voting where votes can be cast for all data points  $x$ . If  $\text{sign}((\omega^{ij})^T \phi(x) + b^{ij})$  says  $x$  is in the  $i_{th}$  class, then the vote for the  $i_{th}$  class is added by one; otherwise, the  $j_{th}$  class is added by one. The class label of  $x$  is set as the maximum number of votes.

While multiclass learning using output codes provides a simple and powerful framework it cannot capture correlations between the different classes since it breaks a multiclass problem into multiple independent binary problem.

### Structure SVM

Structure SVM is aimed to deal with the problem of learning a mapping from inputs to interdependent and structured output spaces. For the training example of input-output pairs  $(x_1, y_1), \dots, (x_l, y_l) \in \mathbf{X} \times \mathbf{Y}$ , structure-SVM generates structured output spaces  $\mathcal{Y}$ , such that output may be sequences, labeled trees or graphs. The main idea is to generalize large margin methods to the broader problem of learning structured responses by specifying a discriminate function  $F: \mathbf{X} \times \mathbf{Y} \rightarrow \mathbf{R}$  that exploit the structure and dependencies within output spaces. A prediction can be obtained for a given input  $x$  by maximizing  $F$ , formulated under a parameter vector as following:

$$\begin{aligned}
f(x; w) &= \arg \max_{y \in \mathbf{Y}} F(x, y; w) \\
F(x, y; w) &= \langle w, \Psi(x, y) \rangle
\end{aligned} \tag{10}$$

where  $w$  is a parameter vector, and  $\Psi$  extracts the features jointly from inputs and outputs, depending on the nature of the problem and special cases.

Given a classifier  $f(x; w)$  and an example  $(\bar{x}, \bar{y})$ , we say that  $f(x; w)$  misclassifies an example  $x$  if  $f(\bar{x}; w) \neq \bar{y}$ . A loss function  $\Delta(y, \bar{y})$  quantifies the loss associated with the prediction  $y = f(\bar{x}; w)$  and  $\bar{y}$ . Thus the risk or empirical error for this problem is given as:

$$\mathfrak{R}_P^\Delta(f) = \int_{\mathbf{X} \times \mathbf{Y}} \Delta(y, f(x)) dP(x, y), \tag{11}$$

where the  $P(x, y)$  denotes the data generating distribution. The goal is to find a classifier  $f$  that can achieve a small risk and also generalize well. In order to avoid the expensive computation, the notion of margin is used to recast the problem as a quadratic optimization problem. Four cost functions are set up to make the problem more feasible to be solved.

The condition of zero training error can be written as a set of linear constraints:

$$\forall i, \forall y \in \mathbf{Y} \setminus y_i : \langle w, \delta\psi_i(y) \rangle > 0, \tag{12}$$

where we have defined the shorthand

$$\delta\psi_i(y) \equiv \psi(x_i, y_i) - \psi(x_i, y) \tag{13}$$

To generalize large margin methods to the broader problem of learning structured responses,  $w$  that satisfies  $\|w\| \leq 1$  is selected. Moreover two approaches are used to deal with arbitrary loss functions. The first approach is to rescale the slack variables according to the loss incurred in each of the linear constraints. With introduced slack variable to allow errors in the training set, the resulting margin quadratic optimization problem is: (Slack variable  $\frac{C}{n} \sum_{i=1}^n \xi_i$  for SVM1,  $\frac{C}{2n} \sum_i \xi_i^2$  for SVM2)

$$\begin{aligned}
& \min_{w, \xi} \frac{1}{2} \|w\|^2 + \frac{C}{n} \sum_{i=1}^n \xi_i, \quad \text{s.t. } \forall i, \xi_i \geq 0 \\
& \text{SVM}_1^{\text{As}} : \forall i, \forall y \in \mathbf{Y} \setminus y_i : \langle w, \delta\psi_i(y) \rangle \geq 1 - \frac{\xi_i}{\Delta(y_i, y)} \\
& \text{SVM}_2^{\text{As}} : \forall i, \forall y \in \mathbf{Y} \setminus y_i : \langle w, \delta\psi_i(y) \rangle \geq 1 - \frac{\xi_i}{\sqrt{\Delta(y_i, y)}}
\end{aligned} \tag{14}$$

---

**Algorithm 1** Algorithm for solving SVM<sub>0</sub> and the loss re-scaling formulations SVM<sub>1</sub><sup>As</sup> and SVM<sub>2</sub><sup>As</sup>

---

- 1: Input:  $(x_1, y_1), \dots, (x_n, y_n), C, \varepsilon$
- 2:  $S_i \leftarrow \phi$  for all  $i=1, \dots, n$
- 3: repeat
- 4:   **for**  $i=1, \dots, n$  **do**
- 5:     set up cost function

$$\text{SVM}_1^{\text{As}} : H(y) \equiv (1 - \langle \delta\psi_i(y), w \rangle) \Delta(y_i, y)$$

$$\text{SVM}_2^{\text{As}} : H(y) \equiv (1 - \langle \delta\psi_i(y), w \rangle) \sqrt{\Delta(y_i, y)}$$

$$\text{SVM}_1^{\Delta m} : H(y) \equiv \Delta(y_i, y) - \langle \delta\psi_i(y), w \rangle$$

$$\text{SVM}_2^{\Delta m} : H(y) \equiv \sqrt{\Delta(y_i, y)} - \langle \delta\psi_i(y), w \rangle$$

$$\text{where } w \equiv \sum_j \sum_{y' \in S_j} \alpha_{jy'} \delta\psi_j(y').$$

- 6:   compute  $\hat{y} = \arg \max_{y \in \mathbf{Y}} H(y)$
  - 7:   compute  $\xi_i = \max \{0, \max_{y \in S_i} H(y)\}$
  - 8:   **if**  $H(\hat{y}) > \xi_i + \varepsilon$  **then**
  - 9:      $S_i \leftarrow S_i \cup \{\hat{y}\}$
  - 10:     $\alpha_S \leftarrow$  optimize dual over  $S, S = \cup_i S_i$ .
  - 11:    **end if**
  - 12:    **end for**
  - 13: **until** no  $S_i$  has changed during iteration
- 

where  $C > 0$  is the usual regularization constant parameter a constant that controls the tradeoff between training error minimization and margin maximization. The second approach is to rescale the margin for the special case of the Hamming loss, which result to the margin constraints as:

$$\begin{aligned} \text{SVM}_1^{\Delta m} : \forall i, \forall y \in Y \setminus y_i : \langle w, \delta \psi_i(y) \rangle &\geq \Delta(y_i, y) - \xi_i \\ \text{SVM}_2^{\Delta m} : \forall i, \forall y \in Y \setminus y_i : \langle w, \delta \psi_i(y) \rangle &\geq \sqrt{\Delta(y_i, y)} - \xi_i \end{aligned} \quad (15)$$

The optimization problem is solved then solved efficiently by a cutting plane method that exploits the sparseness and structural decomposition of the problem. The algorithm pseudo code is showed in Algorithm 1.

### Multi-Label (Rank-SVM)

When the colon cancer images are analyzed, it is often found that one image contains several regions that are belonging to different cancer types. In this case, it is inaccurate to classify this image to only one type. For more accurate colon cancer image analysis, we reformulate this colon cancer image classification as a multi-label problem, which means each point in a training set is associated to more than one label while the size of this label set is unknown. One effective approach for multi-label classification is Rank-SVM, proposed by Elisseeff and Weston (2002). A brief introduction is given as following.

We are given a training data set, the training example  $X=R^d$  is a  $d$ -dimensional input space  $\{x_1, \dots, x_l\}$  and the output space  $Y$  contains  $2^Q$  elements formed by the labels  $\{1, \dots, Q\}$ . One output element corresponds to one set of labels, which means a vector. Thus each training example is associated with one element from the output space. The goal of the training process is to find a learning set:

$$S = \{(x_1, y_1), \dots, (x_l, y_l)\} \subset (X, Y)^l, \quad (16)$$

drawn identically and independently from an unknown distribution  $D$  and a function  $f$  such that the following generalization error is as low as possible:

$$R(f) = E_{(x,y) \sim D} [c(f, x, y)] \quad (17)$$

where  $c$  is a real-value loss function. For linear models we considered here, a ranking based system is proposed to minimizing the empirical error.

Given  $Q$  vectors  $w_1, \dots, w_Q$  and  $k$  bias  $b_1, \dots, b_Q$ , assume the size of label set of the training example  $x$  is known as  $s(x)$ , a ranking value is defined as:

$$r_q(x) = \langle w_q, x \rangle + b_q. \quad (18)$$

If and only if  $r_q(x)$  is among the first  $s(x)$  elements  $(r_1(x), \dots, r_Q(x))$ , the label  $q$  is considered to be in the label set of  $x$ .

As a ranking system, the Ranking Loss (Schapire and Singer, 2000) is defined as following to represent the average fraction of pairs that are not correctly ordered.

$$RL(f, x, y) = \frac{1}{|y||\bar{y}|} |(i, j) \in y \times \bar{y}, r_i(x) \leq r_j(x)|, \quad (19)$$

where  $\bar{y}$  denotes the complementary set of  $y$  in  $\{1, \dots, Q\}$ . The empirical error is measured by the appropriate cost function, which is Ranking Loss in

this case. A good system has a high precision and a low Ranking Loss.

To obtain a linear model that minimizes the Ranking Loss and at the same time having a low complexity, margin is introduced to represent the complexity. The margin of  $(x, y)$  is the signed distance of  $x$  to the decision boundary, expressed as:

$$\min_{q \in y, p \in \bar{y}} \frac{\langle w_q - w_p, x \rangle + b_q - b_p}{\|w_q - w_p\|}. \quad (20)$$

when decision boundary of  $x$  is  $\langle w_q - w_p, x \rangle + b_q - b_p = 0$ , where  $q$  belongs to the label set of  $x$  and  $l$  does not. For the well ranked data in the learning set, parameters  $w_q$  can be normalized such that:

$$\langle w_q - w_p, x \rangle + b_q - b_p \geq 1. \quad (21)$$

Therefore, maximizing the margin on the whole learning set is done by solving the problem:

$$\begin{aligned} \max_{w, j=1, \dots, Q} \min_{(x,y) \in S} \min_{q \in y, p \in \bar{y}} \frac{1}{\|w_q - w_p\|^2}, \\ \langle w_q - w_p, x \rangle + b_q - b_p \geq 1, (q, p) \in y \times \bar{y} \end{aligned} \quad (22)$$

For learning set that is not ranked, margin is maximized by firstly redefining the Ranking Loss as following when the condition  $\langle w_q - w_p, x_i \rangle + b_q - b_p \geq 1 - \xi_{iqp}$  is satisfied for  $(q, p) \in y_i \times \bar{y}_i$ :

$$\frac{1}{l} \sum_{i=1}^l \frac{1}{|y_i||\bar{y}_i|} \sum_{(q,p) \in y_i \times \bar{y}_i} \theta(-1 + \xi_{iqp}), \quad (23)$$

where  $\theta$  is the Heaviside function; then after some calculations, the final quadratic optimization problem is further reformulated as:

$$\begin{aligned} \max_{w, j=1, \dots, Q} \sum_{q=1}^Q \|w_q\|^2 + C \sum_{i=1}^l \frac{1}{|y_i||\bar{y}_i|} \sum_{(q,p) \in y_i \times \bar{y}_i} \xi_{iqp}, \\ \langle w_q - w_p, x_i \rangle + b_q - b_p \geq 1 - \xi_{iqp}, (q, p) \in y_i \times \bar{y}_i, \xi_{iqp} \geq 0 \end{aligned} \quad (24)$$

After ranking process, each example is corresponded to a ranking set that the possibility of this example belongs to each of the labels is given. To determine exactly what labels an example  $x$  is associated with, the size of the label set  $s(x)$  should be predicted, which is obtained by a threshold based method:

$$s(x) = |\{f_q(x) > t(x)\}|, \quad (25)$$

where  $t(x)$  is the threshold function defined as:

$$t(x_i) = \arg \min_t |\{q \in y, f_q(x_i) \leq t\}| + |\{k \in \bar{y}, f_k(x_i) \geq t\}|. \quad (26)$$

Then the  $s(x)$  labels that rank in the top  $s(x)$  places in the sorted ranking set are set to this example  $x$ .

## EXPERIMENTS AND RESULTS

To demonstrate the effectiveness of classify colon cancer image based on multi-label method, three

commonly used multi-classification methods (one-against-all SVM, one-against-one SVM, and multi-structure SVM) are selected as the baseline for comparison.

### Data

The images are collected in Department of Pathology of Zhejiang University from January to September 2010. We obtain all images from the Nano Zoomer 2.0HT digital slice scanner produced by Hamamatsu Photonics with a magnification factor of 40. Our data include 138 patients with colon pathology images. The size of the whole slice is about  $200,000 \times 200,000$  pixels, which is hardly to process. In this experiment, considering of the computation expense, we cut each slice into small images with size of  $10,000 \times 10,000$  pixels. To balance the quantities of the images with different cancer types, 230 images are randomly chosen as the dataset.

### Annotations

To ensure the quality of the ground truth, images are carefully studied and labeled by well-trained experts. Each image is independently annotated by two pathologists; the third pathologist moderates their discussion until they reach the final agreement on the result. All images are labeled as abnormal or normal image; and for abnormal images, the cancer type is also labeled.

### Dataset

Our data contain the six most common types of colon cancer pathological images: well differentiated tubular adenocarcinoma (H), moderately differentiated tubular adenocarcinoma (M), poorly differentiated tubular adenocarcinoma (L), mucinous adenocarcinoma (Mu), Signet-ring carcinoma (R) and Non-cancer (N). And also four multi-label types: well and moderately differentiated tubular adenocarcinoma (HM), moderately and poorly differentiated tubular adenocarcinoma (ML), poorly differentiated tubular adenocarcinoma and Signet-ring carcinoma (LR), Signet-ring carcinoma and mucinous adenocarcinoma (RMu). We use the same abbreviations for each type in the following sections. Words in brackets are short for the title of each type. Table 1 shows the components of dataset. Number of images in the dataset.

### Experiments

To make comparison with multi-label classification model, three other methods are designed in our experiment. Also, for the purpose of verifying the contribution of each feature set, we first test the feature set individually, then repeat the classification process four times by adding a new feature set to the last feature set and utilizing each feature set individually to produce a new classifier. The time for feature extraction is 16.8 s. The seconds for training and classification are 6.7 and 6.2, 6.9 and 6.6, 7.1 and 5.7, 270.3 and 213.5 for OAA, OAO, structure SVM and multi-label SVM, respectively. The experiments are con-

ducted on Matlab R2010b, based on computer hardware configuration of (Processor: Intel(R) Core(TM)2 Duo CPU E7500 @2.93GHz 2.94GHz; RAM: 2.00GB).

Cross-validation (Whitehall and Lu, 1991) is a method used to estimate the generalization ability of a statistical analysis. K-fold cross validation (Kong et al., 2009) is one of the most commonly used method that can eliminate the bias of random selection for proving the performance of classifier by dividing dataset into two groups: one used as training model and the other used as test model. In  $k$ -fold cross-validation whole dataset first should be divided into  $k$  equal subsets. Then,  $k-1$  subsets are randomly selected as training set for testing the rest examples. Finally, repeat this process for all subsets and calculate the final result which is the average of the  $K$  results. In this article,  $K$  is chosen as 3.

### RESULTS

The performance of the classifiers is evaluated by precision, recall and F-measure. We use TP and FP to denote the number of the instances that truly have type  $i$  among all those correctly classified as type  $i$  and the other types respectively, and FN to denote the number of the instances that do not have type  $i$  among all those misclassified as type  $i$  (Dong et al., 2012).

It should be noticed that multi-label images may be classified to single-label or multi-label by the prediction. A result analysis rule is made here for this situation. For instance, given an image belongs to both H and M types, namely an HM image, the evaluation result is made according to the classification result Y. For the single label result condition, if  $Y=H$  (or M), then the TP value of H (or M) will be increased by 1. If  $Y=L$  (or Mu, R, N), then the FN value of L (or Mu, R, N) will be increased by 1. This image may also be classified as a multi label image. In this condition, if  $Y=HM$ , then both the TP values of H and M are increased by 1. If  $Y=H$  (or M) and L (or HMu, HR, HN), then the TP value of H (or M) increased by 1 and the FN value of L (or Mu, R, N) increased by 1. Otherwise if Y is classified as the two types from L, Mu, R and N, then both the FN values of these two types are increased by 1.

For each type the calculating methods are as follows:

$$\text{precision} = \frac{TP_i}{TP_i + FP_i} \quad (27)$$

$$\text{recall} = \frac{TP_i}{TP_i + FN_i} \quad (28)$$

For all types:

$$\text{Overall precision} = \frac{\sum_{i=1}^k TP_i}{\sum_{i=1}^k (TP_i + FP_i)} \quad (29)$$

$$\text{Overall recall} = \frac{\sum_{i=1}^k TP_i}{\sum_{i=1}^k (TP_i + FN_i)} \quad (30)$$

$$F\text{-measure} = \frac{2 \times \text{Overall precision} \cdot \text{Overall recall}}{\text{Overall precision} + \text{Overall recall}} \quad (31)$$

TABLE 1. Number of images of our dataset

H	M	L	R	Mu	N	HM	ML	LR	RMu
30	30	30	20	30	30	15	15	15	15

Accuracy is used as the evaluation criterion for natural scene images classification based on multi-label method. Here for cancer type classification, this criterion is also used. The accuracy is formulated as below (Ladicky et al., 2009).  $N_{ii}$  is the number of the type  $i$  images classified as  $i$ .  $N_{ij}$  is the number of the type  $i$  images classified as  $j$ .

$$\text{Accuracy} = \frac{N_{ii}}{\sum_j N_{ij}} \quad (32)$$

According to the definitions of precision and recall, the two values are equal in multi-classification.

In the first experiment, four feature sets (HOG, GLCM, Euler number and color) are tested individually under four classifiers, to compare how the feature set and classifier contribute to the classification result. Table 2 shows the performances of the four feature sets individually used. Color feature outperforms other three feature sets no matter which classifier is used. This further proves that the Hematoxylin and Eosin technique is a mature method for colon cancer image achievement. The performance of GLCM feature set is second to color feature set. GLCM feature describe the essential characteristics of cell structure.

To further validate the contribution of the four feature sets and compare the capability of the four classifiers, another experiment is designed. In this experiment, new feature sets are added to the original feature set one by one. The result is showed in Table 3. According to the result, all the four methods have better performance when all the four feature sets are used. Moreover, when all the four features are added, multi-label has the best performance than the other three methods. It had a higher  $F$ -measure than OAA

with  $\Delta_{4-1}=14.1\%$ , a higher  $F$ -measure than OAO with  $\Delta_{4-2}=8.6\%$ , and also a higher  $F$ -measure than structure SVM with  $\Delta_{4-3}=4\%$ .

Table 4 shows the performances of the four methods from the cancer type's view. The four methods are all able to achieve a highest score for normal images, and a lowest  $F$ -score for signet-ring cancer type detection.

### DISCUSSION

In this article, we proposed a new method for colon cancer diagnoses using pathological images based on machine learning technique. Some former researches focused on single label, but not consistent with the fact that colon cancer image classification is a multi-label problem. In our research, we reformulated this issue as multi-label problem and implement multi-classification using multi-label SVM. Also, four different kinds of features were extracted to build feature set.

Experimental results show that, multi-label approach has a much significantly performance than the other three methods which is in accordance with our hypothesis of regarding colon cancer as multi-label problem. Given 60 multi-label images in the experiment, multi-label SVM classified 37 images correctly, OAA SVM 21 images, OAO SVM 23 images, and structure SVM 28 images. The experimental results demonstrate that multi-label method has a powerful ability to classify multi-label colon pathology images.

According to the experiments in which different feature sets are applied to the classification process, the results show the effectiveness and efficiency of color feature set for histopathology colon cancer image analysis. In addition, we validate the classification results from the view of each cancer type. For the images belong to normal type, the  $F$  values are all higher than

TABLE 2. Performance of classifiers using individual feature set

Feature	OAA			OAO			Structure SVM			Multi-label SVM		
	P(%)	R(%)	F(%)	P(%)	R(%)	F(%)	P(%)	R(%)	F(%)	P(%)	R(%)	F(%)
1.HOG	20.3	63.2	30.7	37.2	26.7	31.1	38.8	34.1	36.3	41.5	43.0	42.3
2.GLCM	59.4	26.4	36.6	32.9	59.8	42.4	60.7	51.0	55.5	62.3	55.9	58.9
3.Euler	37.2	26.7	31.1	41.6	38.3	39.9	50.9	51.5	51.2	51.7	50.0	50.9
4.Color	31.0	49.3	38.3	56.8	54.6	55.7	55.8	62.7	59.0	50.5	81.9	62.4

TABLE 3. Performance of classifiers using different feature sets

Feature	OAA				OAO				Structure SVM				Multi-label SVM		
	P(%)	R(%)	F(%)	$\Delta_{2-1}(\%)$	P(%)	R(%)	F(%)	$\Delta_{3-2}(\%)$	P(%)	R(%)	F(%)	$\Delta_{4-3}(\%)$	P(%)	R(%)	F(%)
HOG	20.3	63.2	30.7	0.4	37.2	26.7	31.1	5.2	38.8	34.1	36.3	6.0	41.5	43.0	42.3
+GLCM	28.2	58.5	38.0	0.6	60.0	28.5	38.6	14.9	54.9	52.2	53.5	4.6	54.6	62.0	58.1
+Euler	31.6	57.3	40.7	1.1	55.1	33.7	41.8	12.4	59.6	49.7	54.2	8.4	59.1	66.5	62.6
+Color	56.2	57.2	56.7	5.5	65.5	59.3	62.2	4.6	69.8	64.1	66.8	4.0	73.7	68.2	70.8

TABLE 4. Performance of classifiers for different types

Feature	OAA			OAO			Structure SVM			Multi-label SVM		
	P(%)	R(%)	F(%)	P(%)	R(%)	F(%)	P(%)	R(%)	F(%)	P(%)	R(%)	F(%)
H	66.7	47.6	55.6	75.9	52.4	62.0	77.4	57.1	65.8	78.8	61.9	69.3
L	55.4	69.2	61.5	54.1	80.0	64.5	61.5	78.4	69.0	64.6	80.8	71.8
M	44.6	61.7	51.8	59.6	54.9	57.1	57.7	57.7	57.7	66.0	62.3	64.1
R	87.5	17.9	29.8	87.5	23.3	36.8	65.0	43.3	52.0	84.2	53.3	65.3
Mu	50.0	77.1	60.7	50.9	77.1	61.4	61.4	75.0	67.5	60.9	77.8	68.3
N	87.0	66.7	75.5	88.9	80.0	84.2	93.1	90.0	91.5	90.0	90.0	90.0
Macro-average	56.7	65.2	60.7	69.5	61.3	65.1	69.4	66.9	68.1	74.1	71.0	72.5
Micro-average	56.2	57.2	56.7	65.5	59.3	62.2	69.8	64.1	66.8	73.7	68.2	70.8

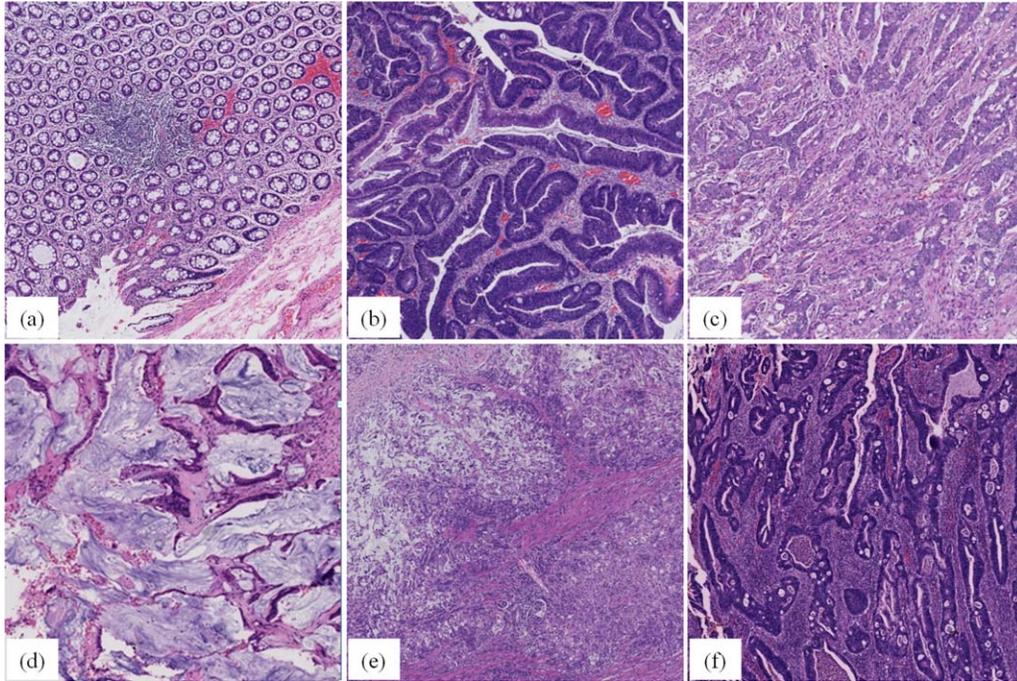


Fig. 5. Typical success and failure cases of classification. Success cases: (a) normal image; (b) well differentiated adenocarcinoma; (c) poorly differentiated adenocarcinoma; (d) mucinous adenocarcinoma. Failure cases: (e) signet-ring cell carcinoma and poorly differentiated adenocarcinoma classified as poorly differentiated adenocarcinoma; (f) well differentiated adenocarcinoma and moderately differentiated adenocarcinoma classified as moderately differentiated adenocarcinoma. [Color figure can be viewed in the online issue, which is available at [wileyonlinelibrary.com](http://wileyonlinelibrary.com).]

other types of images, which is due to the significance color and texture difference of normal cancer type images compared with other types of images. The images belonging to moderately differentiated adenocarcinoma are of higher detection possibility. The multi-label HM (one image belonging to well and moderately differentiated adenocarcinoma) type of images (e.g., Fig. 5) is of the possibility to be classified as moderately differentiated adenocarcinoma. Similarly, HM images are hardly to be recognized clinically. At the same time, LR (one image belonging to poorly differentiated adenocarcinoma and signet-ring cell carcinoma) images tend to be classified as poorly differentiated adenocarcinoma. This is because the signet-ring cell carcinoma area is too small in the whole LR images to be recognized correctly. Mucinous adenocarcinoma images tend to be classified as adenocarcinoma by OAA or OAO method due to the property of mucinous adenocarcinoma that mucus lake contains ingredients of adenocarcinoma. Poorly differentiated images are sometimes classified as moderately differentiated adenocarcinoma especially when poorly images are very close to moderately images according to the definition of adenocarcinoma. The classification results of LMu (one image belonging to signet-ring cell carcinoma and mucinous adenocarcinoma) images generated by structure SVM and multi-label SVM are better than the others, since these two methods take the structure information of an image into consideration.

The reason why multi-label method is better can be explained from the angle of algorithm. The loss functions of all these classifiers should be minimized during

the learning process. However, for the three single-label methods that many classifiers are built, though the loss function of each classifier is minimized, the sum loss function is not insured to be the least. On the contrary, it is the overall information that multi-label SVM takes advantage of to build minimum loss function.

Even though the result is satisfying, it should be admitted that, the system can be improved in the following two aspects in the future. First, special features of pathology colon cancer images should be introduced. Such as the appearance of signet-ring cell carcinoma looks like a ring; mucinous adenocarcinoma with a lot of mucus lake. Extracting features like these have an effective resolution to improve image classification result. The second improvement is to define an application window for pathology image analysis.

During the process of extraction from multi-label images, if the region belongs to one label is in a relatively small proportion of the whole image, its features may be overwhelmed by the whole image' features. To solve this problem, window can be introduced to scan the image. The image is analyzed in each rectangle window.

## CONCLUSIONS

In this paper, a new method for colon cancer diagnoses is proposed. To obtain classifiers multi-label SVM was used to train our data. Compared with other three traditional multi-SVM methods, this method took multi-label phenomenon into consideration. In multi-label phenomenon one image could contain more than one single label. Every image was extracted with four

feature sets. The best results were obtained when combining all features together with precision of 73.7%, recall of 68.2% and F-measure of 70.8%. This result is very satisfying, convincing us that multi-label SVM combining pathology images have strong potentials for colon cancer diagnoses. By using this method, pathologist can obtain a supporting role. Thus, colon cancer detection based on machine learning is very promising.

### ACKNOWLEDGMENT

The authors would like to thank Lab of Pathology and Pathophysiology, Zhejiang University in China to provide data and help.

### REFERENCES

- Allen TC. Hematoxylin and eosin. 1992. Laboratory methods in histotechnology. American Registry of Pathology, Washington, DC, pp. 53–58.
- Altunbay D, Cigir C, Sokmensuer C, Gunduz-Demir C. 2010. Color graphs for automated cancer diagnosis and grading. *IEEE Trans Biomed Eng* 57:665–674.
- Amin SA, Filippas J, Naguib RNG, Bennett MK. 2003. A parallel system for performing colonic tissue classification by means of a genetic algorithm. Recent Advances in Parallel Virtual Machine and Message Passing Interface: 10th European Parallel Virtual Machine and Message Passing Interface Users Group Meeting (Euro PVM/MPI 2003), 2840:560–564.
- Atlamazoglu V, Yova D, Kavantzias N, Loukas S. 2001. Texture analysis of fluorescence microscopic images of colonic tissue sections. *IFMBE, World Congress Med Biol Eng Comp* 39:145–151.
- Boutell MR, Luo J, Shen XP, Brown, CM. 2004. Learning multi-label scene classification. *Pattern Recognition*, 37(9), 1757–1771.
- Chiang TH, Lo HY, Lin SD. 2012. A ranking-based KNN approach for multi-label classification. *Journal of Machine Learning Research, Workshop and Conference Proceedings*, 25:81–96.
- Crammer K, Singer Y. 2000. On the algorithmic implementation of multiclass kernel-based vector machines. *J Machine Learn Res* 2: 265–292.
- Dalal N, Triggs B. 2005. Histograms of Oriented Gradients for Human Detection. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005*, 1:886–893.
- Dong S, Zhou DD, Ding W. 2012. Traffic Classification Model Based on Integration of Multiple Classifiers. *Journal of Computational Information Systems* 8(24):10429–10437.
- Dundar MM, Badve S, Raykar V, Jain RK, Sertel O, Gurcan MN. 2010. A multiple instance learning approach toward optimal classification of pathology slides. 20th International Conference on Pattern Recognition (ICPR), 2010.
- Elisseff A, Weston J. 2002. A kernel method for multi-labelled classification. *Advances in Neural Information Processing Systems*, 14: 681–687.
- Esgiar AN, Naguib RN, Sharif BS, Bennett MK, Murray A. 1988. Microscopic image analysis for quantitative measurement and feature identification of normal and cancerous colonic mucosa. *IEEE Trans Inf Technol Biomed* 2:197–203.
- Friedman J. 1996. Another approach to polychotomous classification. Technical report. Department of Statistics, Stanford University.
- Grady L, Funka-Lea G. 2004. Multi-label image segmentation for medical applications based on graph-theoretic electrical potentials. *ECCV 2004 Workshop on Computer Vision Approaches to Medical Image Analysis and Mathematical Methods in Biomedical Image Analysis*, 3117:230–245.
- Hamilton SR, Aaltonen LA (Eds.). 2000. *World Health Organization Classification of Tumours. Pathology and Genetics of Tumours of the Digestive System*. Lyon: IARC Press. pp. 104–111.
- Hirzebruch F and Höfer T. 1990. On the Euler number of an orbifold. *Mathematische Annalen*, 286(1–3):255–260.
- Huang KW, Li ZL. 2011. A Multilabel Text Classification Algorithm for Labeling Risk Factors in SEC Form 10-K. *Journal ACM Transactions on Management Information Systems (TMIS)*, 2(3), Article 18.
- Huang PW, Lee Cheng-Hsiung, Lin Phen-Lan. 2009. Support vector classification for pathological prostate images based on texture features of Multi-Categories. *IEEE, International Conference on Systems, Man, and Cybernetics, TX*, pp. 912–916.
- Jiang JY, Tsai SC, Lee SJ. 2012. FSKNN: Multi-label text categorization based on fuzzy similarity and k-nearest neighbors. *Expert Systems Appl* 39:2813–282.
- Jiao LP, Chen Q, Li SY, Xu Y. 2013. Colon Cancer Detection Using Whole Slide Histopathological Images. *World Congress on Medical Physics and Biomedical Engineering, 2012, Beijing, China IFMBE Proceedings*, 39:1283–1286.
- Kijisirikul B, Ussivakul N. 2002. Multiclass support vector machines using adaptive directed acyclic graph. *Proceedings of International Joint Conference on Neural Networks* 1:980–985.
- Konga J, Sertela O, Shimada H, Boyer KL, Saltz JH, Gurcan MN. 2009. Computer-aided evaluation of neuroblastoma on whole-slide histology images: Classifying grade of neuroblastic differentiation. *Pattern Recognition* 42(6):1080–1092.
- Kong J, Sertel O, Shimada H, Boyer KL, Saltz JH, Gurcan MN. 2009. Computer-aided evaluation of neuroblastoma on whole-slide histology images: Classifying grade of neuroblastic differentiation. *Pattern Recognition* 42(6):1080–1092.
- Ladicky L, Russell C, Kohli P, Torr, PH. 2009. Associative Hierarchical CRFs for Object Class Image Segmentation. *International Conference on Computer Vision (ICCV)*, 2009.
- Li XC, Wang L, Sung E. 2004. Multi-label SVM active learning for image classification. *IEEE International Conference on Image Processing (ICIP)*, pp. 2207–2210.
- Madabhushi A. 2009. Digital pathology image analysis: Opportunities and challenges. *Imaging Med* 1:7–10.
- McCallum KA. 1999. Multi-label text classification with a mixture model trained by EM. *AAAI 99 Workshop on Text Learning*.
- Nwoye E, Khor LC, Satnam S, Woo WL. 2006. A novel fast fuzzy neural network backpropagation algorithm for colon cancer cell image discrimination. *Lecture Notes Comp Sci* 3973:760–769.
- Oliveira E, Ciarelli PM, Badue C, De Souza AF. 2008. A comparison between a KNN based approach and a PNN algorithm for a multi-label classification problem. *8th International Conference on Intelligent Systems Design and Applications*, 2:628–633.
- Schapire RE, Singer Y. 2000. BoosTexter: A boosting-based system for text categorization. *Machine learning*, 39(2/3):135–168.
- Sleigh A. 2001. Notes For the First Year Lecture Course: An Introduction to Fluid Mechanics. School of Civil Engineering, University of Leeds, pp 118–120.
- Shuttleworth JK, Todman AG, Naguib RNG, Newman BM, Bennett MK. 2002. Multi-resolution color texture analysis for classifying colon cancer images. *Engineering in Medicine and Biology, 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society EMBS/BMES Conference* 2:1118–1119.
- Tsoumakas G, Katakis I. 2007. Multi-label classification: An overview. *Int J Data Warehousing Mining* 3:1–13.
- Vedaldi A, Gulshan V, Varma M, Zisserman A. 2009. Multiple kernels for object detection. *IEEE 12th International Conference on Computer Vision*, 2009.
- Wang XY, Wu JF, Yang HY. 2010. Robust image retrieval based on color histogram of local feature regions. *Multimedia Tools Appl* 49: 323–345.
- Whitehall BL, Lu SC. 1991. Machine learning in engineering automation: The present and the future. *Comp Industry* 17(2-3):91–100.
- Xu Y, Zhu JY, Chang E, Lai MD, Tu ZW. 2012. Multiple clustered instance learning for histopathology cancer image classification, segmentation and clustering. *Computer Vision and Pattern Recognition, Providence, Rhode Island, CVPR*. 2012, 1:964–971.
- Yang BS, Sun JT, Wangy TJ, Chen Z. 2009. Effective Multi-Label Active Learning for Text Classification. *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, New York, USA*, pp 917–926.
- Zhang ML. 2010. A k-nearest neighbor based multi-instance multi-label learning algorithm. *Int Conference Tools Artificial Intelligence* 2:207–212.
- Zhang M L, Zhou ZH. 2005. A k-nearest neighbor based algorithm for multi-label classification. *IEE Int Conf Granular Comp* 2:718–721.
- Zhang ML, Zhou ZH. 2007. ML-KNN: A lazy learning approach to multi-label learning. *Pattern Recogn* 40:2038–2048.
- Zhou ZH, Zhang ML. 2006. Multi-Instance Multi-Label Learning with Application to Scene Classification. In *Proceedings of NIPS*, pp 1609–1616.
- Zhu Q, Yeh MC, Cheng KT, Avidan S. 2006. Fast Human Detection Using a Cascade of Histograms of Oriented Gradients. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2006*, 2:1491–1498.
- Zha ZJ, Hua XS, Mei T, Wang JD, Qi GJ, Wang ZF. 2008. Joint multi-label multiinstance learning for image classification. *IEEE Conference on Computer Vision and Pattern Recognition*, 2008. CVPR.