Computer-aided estimation for the risk of development of gastric cancer by image processing

Yoshihiro Sasaki[1], Ryukichi Hada[1], Tetsuro Yoshimura[2], Norihiro Hanabata[2], Tatsuya Mikami[2] and Shinsaku Fukuda[2],

[1] Department of Medical Informatics and Gastroenterology[2], Hirosaki University Hospital, Honcho-53, Hirosaki 036-8563, Japan.

{ Yoshihiro Sasaki, Ryukichi Hada, Tetsuro Yoshimura, Norihiro Hanabata, Tatsuya Mikami and Shinsaku Fukuda} gahiro@cc.hirosaki-u.ac.jp

Abstract. The aim of this study was to establish a computer-aided estimating system for determining the risk of development of gastric cancer, achieved by image processing on an ordinary endoscopic picture. Digital endoscopic pictures of the background gastric mucosa in 26 Helicobacter pylori (*H. pylori*) positive patients with early intestinal type gastric cancer and age-gender-matched *H. pylori* positive subjects without cancer were used. The pictures were processed for 15 pictorial parameters. Out of the 15 pictorial parameters, 3 parameters were found to characterize the background gastric mucosa with gastric cancer against that without. Based on the Bayes decision theory, the computer-aided estimating system has been established. Sensitivity, specificity, positive predictive value and negative predictive value of the Bayes classifier were found to be 0.64, 0.64, 0.65 and 0.63, respectively. This method may permit an effective selection of the high risk population of gastric cancer needing follow-up endoscopy.

Keywords: Helicobacter pylori; Gastric cancer; Endoscopy; Image processing; Computer-aided diagnosis

1 Introduction

Most studies have found Helicobacter pylori (*H. pylori*) to be closely associated with the development of gastric cancer [1], [2], [3], [4], [5]. Among the patients with *H. pylori* infection, histologic findings at higher risk for gastric cancer have included severe gastric atrophy, corpus-predominant gastritis and intestinal metaplasia [1]. In Japan, early detection of gastric cancer has been a crucial clinical problem for the mortality control because the incidence of gastric cancer has been about 9-fold higher compared with that in the USA [6]. When gastric atrophy advances, the submucosal vasculature becomes to varying extent traceable on endoscopy. Although it is quite easy for endoscopists to recognize and describe the endoscopic features of atrophy as a risk factor for gastric cancer in empiric terms, pictorial information has been extremely refractory to precise definition and to analysis by computers.

Recently, computer aided diagnosis (CAD) methods have been developed for image interpretation in large-scale breast or lung cancer screening studies. CAD methods are important for: (a) improving the sensitivity of cancer detection, (b)

reducing interobserver variation in image interpretation, (c) improving efficiency of screening by identifying suspect lesions [7]. We have developed a quantitative method to characterize endoscopic features in a processed endoscopic picture [8]. The aims of this study were to characterize background gastric mucosa at higher risk for gastric cancer and to establish a CAD system for determining the risk of development of gastric cancer in a population with positive *H. pylori*.

For this purpose, digital endoscopic pictures of the background gastric mucosa in subjects with or without gastric cancer were processed for 15 pictorial parameters characterizing endoscopic features. Out of the 15 pictorial parameters, 3 parameters were found to characterize the background gastric mucosa with gastric cancer against that without. Sensitivity, specificity, positive predictive value and negative predictive value of this CAD system were found to be 0.64, 0.64, 0.65 and 0.63, respectively.

2 Materials and Methods

In this section, we present an image processing method to extract endoscopic features of the background gastric mucosa in subjects with or without gastric cancer for CAD system determining the risk of development of gastric cancer, and evaluate the performance of this system.

2.1 Subjects

A total of 2584 subjects underwent gastroscopy between January 2002 and February 2003 at Hirosaki University Hospital. Out of the subjects, we enrolled 26 consecutive patients with early well differentiated gastric carcinoma (19 males and 7 females; mean age, 65.1 years; range, 34 to 77). Out of 185 consecutive patients with non-ulcer dyspepsia, 26 age-gender-matched patients served as controls (19 males and 7 females; mean age, 65.1 years; range, 35 to 77). A total of 52 subjects were enrolled in this study. All subjects gave written informed consent. The study protocol was approved by the Ethics Committees of Hirosaki University Hospital and was reviewed annually.

2.2 Endoscopy

Endoscopy was performed with local anaesthesia (lidocaine). Endoscopic pictures were taken under the neutral chromatic parameters of an endoscopy system (EVIS 240 for a control unit and XK240 for an endoscope; Olympus, Tokyo, Japan), and saved in a digital filing system (EVIS-F; Olympus, Tokyo, Japan). Two biopsy specimens were taken from the greater curvature of the antrum and the middle body of the stomach for a rapid urease test (Helicocheck, Ohtsuka Pharmaceutical, Tokyo, Japan). In patients with suspected gastric cancer, additional biopsies were performed for histological diagnosis.

2.3 Pictorial features of the endoscopic images

The endoscopic picture of the lesser curvature of the gastric body distended by a sufficient amount of air (Figure 1) was used. In electronic endoscopy, red (wavelength > 600 nm), green (500 to 600 nm) and blue light (<500 nm) are sequentially irradiated on the surface of the mucosa. Part of the lights incident on the mucosa, after suffering scattering and absorption, is re-emitted in a nearly diffuse

state [9]. A single endoscopic picture is composed of the red, green and blue diffuse reflectance images. The color of any object in the digitized picture can be quantified by the intensity of red (R), green (G) and blue (B) reflectance. As hemoglobin is the predominant chromophore in the gastrointestinal mucosa, a colorant endoscopic picture is equivalent to a grey scale picture with index of hemoglobin (IHB=32log2[R/G]) assigned to each pixel (Figure 2) [8]. The grey scale picture with blackout or halation excluded was processed for 4 basic statistical parameters (mean IHB, SD of IHB, skewness of IHB, kurtosis of IHB) and 11 textural features that are related with spatial arrangements of the grey scale intensities (Table 1) [10]. Criteria of parameters for a reliable CAD system can be stated as follows: 1) Significant difference is demonstrated between the two subject groups and 2) the parameters are mutually independent to the other ones in a set of parameters with a significant difference. Out of the 15 pictorial parameters, SD of IHB, skewness of IHB and f3 (correlation feature) have been found to meet the criteria, thereby uniquely characterizing the background mucosa with cancer against that without cancer.

Table 1. Textural features.

Number	Name	Characteristic
f1	Angular second moment	Homogeneity
f2	Contrast	Contrast
f3	Correlation	Linearity
f4	Sum of square	Heterogeneity
f5	Inverse difference moment	*
f6	Sum average	*
f7	Sum variance	*
f8	Sum entropy	*
f9	Entropy	*
f10	Difference variance	*
f11	Difference entropy	*

*See detail in the reference 10.

2.4 Bayes classifier for the computer aided diagnosis system

A computer-aided estimating system for determining the risk of development of gastric cancer was established on the basis of Bayes decision theory [11]. Let ω_1 and ω_2 be the state of the background gastric mucosa with or without gastric cancer, respectively. Let the feature vector $\boldsymbol{x}$ be a 3-component column vector, where the component x_1, x_2 and x_3 designate SD of IHB, skewness of IHB and f3, respectively, and let $p(\boldsymbol{x}|\omega_i)$ be the state-conditional probability density function for $\boldsymbol{x}$. Then, the likelihood for the background gastric mucosa with gastric cancer $f(\boldsymbol{x})$ can be computed from $p(\boldsymbol{x}|\omega_i)$ by Bayes rule:

$$f(\boldsymbol{x})= p(\boldsymbol{x}|\omega_1) / (p(\boldsymbol{x}|\omega_1) + p(\boldsymbol{x}|\omega_2)).$$

$p(\boldsymbol{x}|\omega_i)$ (i=1, 2) is calculated as

$$p(\boldsymbol{x}|\omega_i)= \exp[-1/2(\boldsymbol{x}-\mu)^t\Sigma^{-1}(\boldsymbol{x}-\mu)]/ [(2\pi)^2|\Sigma|^{1/2}],$$

where μ is the 3-component mean vector in ω_i, Σ is the 3 by 3 covariance matrix, $(\boldsymbol{x}-\mu)^t$ is the transpose of $x-\mu$, Σ^{-1} is the inverse of Σ, and $|\Sigma|$ is the determinant of Σ. When a given gastric mucosa with positive *H. pylori* has a feature vector $\boldsymbol{x}$, the risk

for the development of gastric cancer is given by f(***x***), which increases from 0 to 1 as the risk increases.

2.5 Assessment for the performance of CAD system

Five-fold cross-validation method was employed. A total of 26 data samples were divided into 5 test samples and 21 training samples. The training samples were utilized to obtain f(***x***). The test samples were used to evaluate the prediction performance (including sensitivity, specificity, positive predictive value and negative predictive value) of f(***x***) at varying cut-off values from 0.1 to 0.9. The averaged sensitivity and specificity over a total of 5 test sets were plotted against the cut-off values (Figure 3). The optimal cut-off, which can be defined as the crossing point of the two curves, was found to be 0.42.

2.6 Statistical analysis

Pictorial parameters were expressed as mean ± SD. Statistical significance was evaluated using Student's or Welch's unpaired t-test. The individual p-values were corrected for multiplicity effects due to multiple testing [12]. Differences between means were considered significant if the corrected-p value < 0.05.

3 Experimental results

The pictorial parameters in the background mucosa with or without gastric cancer were shown in Table 2. The background mucosa with cancer had a significantly larger SD of IHB (6.8±1.1, p=0.013), skewness of IHB (-0.17±0.63, p=0.021) and f3 (4.6±1.0, p=0.015) than that without cancer (6.0±0.9 for SD of IHB; -0.69±0.51 for skewness of IHB; 3.7±1.2 for f3). In a set of the three parameters, anyone was independent to the other ones. The other pictorial parameters f1, f4, f7, f8 and f9 have also significantly differed between the two categories. But, they have not been independent to anyone of the three parameters (SD of IHB, skewness of IHB and f3).

An example of the background mucosa with gastric cancer was 6.4 in SD of IHB, -0.5 in skewness of IHB, 5.0 in f3 and estimated at high risk (f(***x***)=0.85) (Figure 4), while that without cancer was 4.5 in SD of IHB, 0.1 in skewness of IHB, 3.6 in f3 and estimated at low risk (f(***x***)=0.24) (Figure 5). Under the cut-off value of 0.42, sensitivity, specificity, positive predictive value and negative predictive value in estimating whether at high risk or not were found to be 0.64, 0.64, 0.65 and 0.63, respectively.

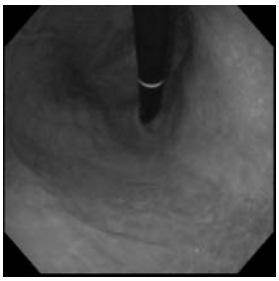

Fig. 1. The endoscopic picture of the lesser curvature of the gastric body distended by a sufficient amount of air.

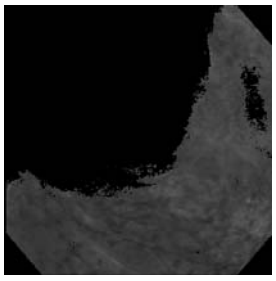

Fig. 2. The grey scale picture composed of pixel arrays with the index of hemoglobin assigned to each pixel for pictorial parameters. The black (0 intensity) area designates blackout or halation area excluded from analysis.

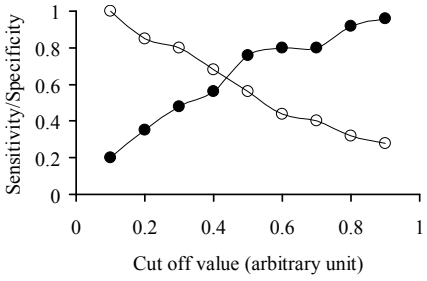

Fig.3. Sensitivity (open circle) and Specificity (filled circle) against cut-off values.

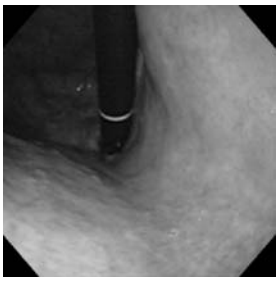

Fig. 4. An example of the background mucosa with gastric cancer (SD of IHB=6.4, skewness of IHB= -0.5, f3=5.0 and estimated at high risk or f(***x***) =0.85).

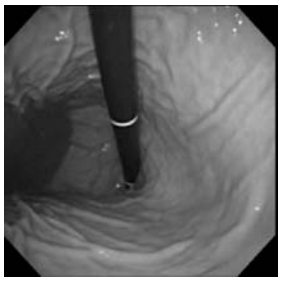

Fig. 5. An example of the background mucosa without gastric cancer (SD of IHB=4.5, skewness of IHB=0.1, f3=3.6 and estimated at low risk or f(***x***) =0.24).

Table 2. Basic statistical parameters and textural features in the background mucosa with or without gastric cancer.

	With cancer	Without cancer	p value*
mean*	56.4±7.6	58.6±8.5	0.350
SD*	6.8±1.1	6.0±0.9	0.013
Skewness	-0.2±0.6	-0.7±0.5	0.021
Kurtosis	5.9±3.4	7.3±3.8	0.263
f1	1.2±0.3	1.4±0.4	0.026
f2	2.0±0.6	1.8±0.4	0.270
f3	4.6±1.0	3.7±1.2	0.015
f4	1.9±0.6	1.4±0.4	0.014
f5	4.7±0.5	4.8±0.4	0.337
f6	2.6±0.4	2.7±0.5	0.339
f7	2.2±0.8	1.6±0.5	0.015
f8	2.6±0.2	2.4±0.2	0.028
f9	3.9±0.3	3.7±0.2	0.025
f10	1.7±0.5	1.6±0.4	0.290
f11	3.1±0.3	3.0±0.2	0.316

mean*, the mean of gray scale intensity of index of haemoglobin; SD*, the standard deviation of gray scale intensity of index of haemoglobin; p value* (corrected according to Benjamini & Hochberg method, see detail in the reference 12), the background gastric mucosa with cancer vs. without cancer; the parameters were expressed as mean±SD.

4 Discussion

We have developed a reliable method to quantify endoscopic features in a processed grey scale picture with IHB assigned, which has led to a computer-aided grading system for endoscopic severity in ulcerative colitis [7]. In the present study, the background gastric mucosa with gastric cancer has been demonstrated to be

uniquely characterized against that without cancer by three pictorial parameters. With the use of the parameters, a computer-aided estimation system for the risk of development of gastric cancer has been established.

In Japan, endoscopy is performed frequently for gastric cancer screening even in subjects without any symptom because of the high incidence of the disease [6]. Clinical interest has been particularly focussed on quantitative endoscopic features for the risk of development of gastric cancer among subjects with *H. pylori* infection to establish a screening or follow-up protocol with a higher cost-performance.

Among the patients with *H. pylori* infection, histologic findings at higher risk for gastric cancer have included severe gastric atrophy, corpus-predominant gastritis and intestinal metaplasia [1]. Out of the three histologic findings, gastric atrophy is characterized as the presence of characteristic patterns or mesh-like or river-like submucosal vasculature. The degree of mucosal atrophy (the clarity or complexity of the patterns) or the area of the atrophic mucosa have been considered to correlate with the three pictorial parameters (SD of IHB, skewness of IHB and f3).

Pictorial information has been extremely refractory to precise definition and to analysis by computers, although it is quite easy for human observers to recognize and describe in empirical terms. With the advent of the electronic endoscope, digital filing system and image processor, it has become possible to quantify any element composing a digitized endoscopic picture by mathematical processes. We have noticed that endoscopic features for diagnosis are concerned with spatial arrangements of IHB. The spatial arrangements expressed by empirical terms as fine, coarse, smooth, rippled, irregular or lineated are quantified by the basic statistical parameters or textural features. The parameters of diagnostic value may vary with diseases or organs studied and are not necessarily independent to each other [10]. For a reliable CAD system, it is essential to determine the parameters of a diagnostic value on the basis of above stated criteria.

In gastrointestinal endoscopy, image interpretation is inherently associated with interobserver variation. We have previously established CAD for grading endoscopic severity in patients with ulcerative colitis [8] or CAD for detecting flat type early esophageal cancer [13]. In the present study, it has been demonstrated that CAD methods can be utilized as a means for an effective screening of gastric cancer by identifying the high risk population needing follow-up endoscopy. The challenge is now to construct a database of the endoscopic images for the optimization of CAD software, leading to the creation of a more scientific basis for evaluation and comparison of image processing methods [7].

References

1. Uemura, N., Okamoto, S., Yamamoto, S., Matsumura, N., Yamaguchi, S., Yamakido, M., Taniyama, K., Sasaki, N., Schlemper, R.J.: Helicobacter pylori infection and the development of gastric cancer. N. Eng. J. Med. 354, 784--789 (2001)
2. Forman, D., Newel,l D.G., Fullerton, F., Yarnell, J.W., Stacey, A.R., Wald, N., Sitas, F.: Association between infection with Helicobacter pylori and risk of gastric cancer: evidence from a prospective investigation. Brit. Med. J. 302, 1302--1305 (1991)

3. Huang, J.Q., Sridhar, S., Chen, Y., Hunt, R.H.: Meta-analysis of the relationship between Helicobacter pylori seropositivity and gastric cancer. Gastroenterol. 114, 1169--1179 (1998)
4. Danesh, J.: Helicobacter pylori infection and gastric cancer: systemic review of the epidemiological studies. Aliment. Pharm. Therap. 13, 851--856 (1999)
5. Tabata, H., Fuchigami, T., Kobayashi, H., Sakai, Y., Nakanishi, M., Tomioka, K., Nakamura, S., Fujishima, M.: Helicobacter pylori and mucosal atrophy in patients with gastric cancer: a special study regarding the methods for detecting Helicobacter pylori. Dig. Dis. Sci. 44, 2027--2034 (1999)
6. Lambert, R., Guilloux, A., Oshima, A., Pompe-Kirn, V., Bray, F., Parkin, M., Ajiki, W., Tsukuma, H.: Incidence and mortality from stomach cancer in Japan, Slovenia and the USA. Int. J. Cancer 97, 811--818 (2002)
7. New NCI Initiatives in Computer Aided Diagnosis, http://imaging.cancer.gov/reportsandpublications/publications/NewNCIInitiativesinComputerAidedDiagnosis
8. Sasaki, Y., Hada, R., Munakata, A.: Computer-aided grading system for endoscopic severity in patients with ulcerative colitis. Dig. Endosc. 15, 206--209 (2003)
9. Wyszecki, G., Stiles, W.S.: Color science: concepts and methods, quantitative data and formulae. John Wiley & Sons, New York (1982)
10. Haralick, R.M., Shanmugam, K., Dinstein, I.: Textural features for image classification. IEEE Trans. Syst. Man. Cybern. 6, 610-- 621 (1973)
11. Duda, R.O., Hart, P.E.: Bayes Decision Theory: In Pattern Classification and Scene Analysis. John Wiley & Sons, New York (1973)
12. Benjamini, Y., Hochberg, Y.: Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Statist. Soc. Ser. B 57, 298--300 (1995)
13. Mikami, T., Sasaki, Y., Fukuda, S., Hada, R., Munakata, A.: Computer-aided diagnosis of superficial type early oesophageal cancer by image processing on ordinary endoscopic pictures. In: Digestive Disease Week, Gastrointest. Endosc. 59, pp258 (2004)