Classification of Gastric Cancer Risk From X-ray Images Based on Efficient Image Features Related to Serum Hp Antibody Level and Serum PG Levels

Kenta Ishihara (student member)†, Takahiro Ogawa (member)†, Miki Haseyama (member)†

Abstract In this paper, a fully automatic gastric cancer risk classification method with the aim of constructing a computer-aided diagnosis (CAD) system is presented. Two-stage classification is used in the proposed method for determining gastric cancer risk. In the first stage, the proposed method detects *H. pylori*-infected patients, i.e., detection of patients who have gastric cancer risk, and the proposed method classifies the level of gastric cancer risk, i.e., high or low, from *H. pylori*-infected patients in the second stage. In each stage, we derive new image features that are closely related to values of blood examination via kernel canonical correlation analysis. The introduction of these new image features provides classification improvement in each stage, and it is the main contribution of this paper. Consequently, accurate classification becomes feasible by the proposed method. Experimental results obtained by applying the proposed method to real X-ray images show that our method outperforms several comparative methods.

Key words: Gastric cancer risk, *H. pylori* infection, blood examination, kernel canonical correlation analysis.

1. Introduction

It is now known that chronic atrophic gastritis triggered by *Helicobacter pylori* (*H. pylori*) infection causes the development of gastric cancer\(^1\)\(^2\)\(^3\), and it is known that gastric cancer rarely occurs in patients not infected with *H. pylori*\(^3\). Therefore, gastric cancer screening focusing on *H. pylori* infection and gastric atrophy has been actively executed for reducing the number of gastric cancer patients.

There are various types of examination including blood examination, fluoroscopy, and endoscopy. A blood examination (ABC classification\(^4\)) focusing on *H. pylori* infection and degree of atrophic gastritis has been introduced in each municipality and in hospitals in Japan. *H. pylori* infection and degree of atrophic gastritis are assessed by serum Hp antibody level and serum pepsinogen (PG) levels including PG I and PG II levels respectively. As shown in Table 1, the ABC classification categorizes gastric cancer risk into four groups, i.e., groups A, B, C and D, based on blood samples. It has been reported that the lowest group of gastric cancer risk is group A, followed in order by groups B, C and D\(^3\). The number of patients in group D is small since the Hp antibody level has high sensitivity, and groups C and D are therefore sometimes merged into one group. In this way, the ABC classification sometimes categorizes gastric cancer risk into three groups, i.e., group A, group B, and group C and D. In other words, it derives three conditions of patients, namely, non-gastric cancer risk, low gastric cancer risk and high gastric cancer risk.

Although the ABC classification can be used to assess the presence of gastric cancer risk, it can not be used to identify gastric cancer patients. For identification of gastric cancer patients, monitoring of stomach by fluoroscopy or endoscopy is needed. However, these examinations have a problem. Although endoscopy is better than fluoroscopy for detecting early gastric cancer, there is only a small number of facilities in which endoscopy can be performed for all patients. Annual Report of Gastrointestinal Cancer Screening reported that the number of patients inspected by fluoroscopy was 6,776,770, whereas the number of patients inspected by

<table>
<thead>
<tr>
<th>H.p antibody level</th>
<th>Negative</th>
<th>Positive</th>
</tr>
</thead>
<tbody>
<tr>
<td>PG levels**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>Positive</td>
<td>D</td>
<td>C</td>
</tr>
</tbody>
</table>

* Generally, “Hp antibody level ≥ 10U/ml” is positive.
** Generally, “PG I ≤ 70ng/ml and PG I/PG II ≤ 3” is positive.

Received February 24, 2016; Revised June 13, 2016; Accepted August 9, 2016
† Graduate School of Information Science and Technology Hokkaido University.(Sapporo, Japan)
endoscopy was 485,429 in Japan in 2013\(^9\). Therefore, it is difficult to perform endoscopy for all patients. On the other hand, although there are many facilities with X-ray equipment, the number of doctors specializing in fluoroscopy has decreased in Japan\(^6\). Furthermore, the number of X-ray images that doctors must examine is large since X-ray images are generally taken at multiple positions per patient.

The utilization of computer-aided diagnosis (CAD) systems for X-ray images is useful for diagnosing many cases. CAD systems that automatically analyze images have been attracting attention\(^7\), and CAD systems have therefore become important for supporting doctors in diangose of X-ray images.

We have proposed methods that enable detection of \textit{H. pylori} infection from gastric X-ray images\(^8\)\(^9\). Although it has been confirmed that these methods enable accurate detection of \textit{H. pylori} infection, a CAD system that can not only detect \textit{H. pylori} infection but also classify gastric cancer risk is needed for supporting doctors. By classifying gastric cancer risk, doctors can decide the interval of re-examination for each gastric cancer risk, i.e., the interval of re-examination for patients with a high risk is short, whereas that for patients with a low risk is long.

The relationship between the values of blood examination, i.e., “Hp antibody level and PG levels” and “symptoms of \textit{H. pylori} infection that appear on gastric X-ray images” has been reported\(^10\). This relationship was investigated by comparing the results of fluoroscopy and the results of blood examination. In particular, the stomach progressed atrophy has the same tendency of symptom. Therefore, we expect that gastric cancer risk can be classified accurately by focusing on the relationship between values of blood examination and symptoms of \textit{H. pylori} infection.

In this paper, we propose a new classification method for gastric cancer risk, i.e., group A, group B and group C and D, from gastric X-ray images. In the proposed method, we try to classify three conditions of patients, namely, non-gastric cancer risk, low gastric cancer risk and high gastric cancer risk. A two-stage classification, as shown in Fig. 1, is used in the proposed method. In the first stage, the proposed method detects \textit{H. pylori}-infected patients, i.e., detection of patients who have gastric cancer risk. In the second stage, the proposed method classifies gastric cancer risk, i.e., high or low, for \textit{H. pylori}-infected patients detected in the first stage. Particularly, the symptoms between “group A” and “groups B, C, and D” are significantly different from those between “group B” and “group C and D.” The former focuses on the presence of gastric cancer risk, whereas the later focuses on the degree of gastric cancer risk, i.e., high or low. Therefore, we can regard the two stages as “the detection of the patients having gastric cancer risk” and “the classification of the degree of gastric cancer risk”, respectively. From this point of view, the two-stage classification approach is suitable for the classification of the gastric cancer risk.

In each stage, the proposed method focuses on the relationship between values of blood examination, i.e., “Hp antibody level and PG levels” and “symptoms of \textit{H. pylori} infection that appear on gastric X-ray images”. In order to effectively use the above relationship in our classification method, we derive new image features by projecting original image features, which are directly extracted from X-ray images, to a space that provides the best correlation with values of blood examination via Kernel Canonical Correlation Analysis (KCCA)\(^11\). In image processing and computer vision communities, various studies have been carried out for effectively and collaboratively using heterogeneous features\(^12\). KCCA has been used in many studies for obtaining a subspace that provides the optimal relationship between heterogeneous features\(^13\)\(^14\). KCCA can also consider a non-linear correlation between two input heterogeneous
features. Thus, we derive new KCCA-based image features that effectively reflect the correlation with values of blood examination since these values complexly change with the progress of atrophy. In this way, the proposed method can capture the details of symptoms that appear on X-ray images as new image features. The introduction of these features is the greatest contribution of this paper, and it is the first trial in gastric cancer risk classification. By using the new image features, we classify gastric cancer risk by a classifier in each stage.

In our method, multiple classification results per patient are obtained by applying the classification to each position at which the X-ray image is taken, and these classification results are integrated via an accuracy-based voting method. Consequently, the proposed method can achieve successful classification of gastric cancer risk based on derivation of the new image features focusing on the relationship between “values of blood examination” and “symptoms of H. pylori infection that appear on gastric X-ray images”.

This paper is organized as follows. First, new image feature extraction using values of blood examination via KCCA is explained in Section 2. Next, the classification and integration of gastric cancer risk are presented in Section 3. In Section 4, experimental results obtained by applying our method to actual X-ray images are presented, and the effectiveness of our method is shown. Finally, conclusions are presented in Section 5.

2. KCCA-based Image Feature Extraction Using Values of Blood Examination

In this section, a method for calculating the KCCA-based image features using values of blood examination is explained. In 2.1, we briefly explain the relationship between values of blood examination and symptoms of H. pylori infection that appear on gastric X-ray images. Note that this relationship has been investigated by comparing the results of fluoroscopy and the values of blood examination\(^{10}\). In 2.2, we present a method for calculating the KCCA-based image features.

2.1 Relationship Between Values of Blood Examination and Symptoms of H. Pylori Infection

In this subsection, we explain the relationship between values of blood examination and symptoms of H. pylori infection that appear on gastric X-ray images. Infection of the stomach with H. pylori causes inflammation, and the stomach walls become thin with atrophy. The above symptoms are called “chronic atrophic gastritis”. The gastric cancer risk becomes higher with the development of chronic atrophic gastritis, and thus blood values of blood examination complexly change as shown below.

**Hp antibody level (U/ml)**

Positive for Hp antibody level is defined by optical density values according to the manufacturer’s protocol\(^{3}\). In this paper, the patients with Hp antibody level \(\geq 10\) U/ml were classified as H. pylori-infected. Although Hp antibody level naturally increases when the stomach is infected with H. pylori, this level decreases due to natural shedding caused by the severe atrophy.

**PG levels**

Serum pepsinogens originating from the stomach mucosa are classified into two groups, namely PG I and PG II\(^{15}\). In this paper, the patients with both PG I \(\leq 70\) ng/ml and PG I/PG II ratio \(\leq 3\) were classified as atrophic gastritis. The proposed method uses the following blood values.

**PG I (ng/ml)**

PG I is secreted only from the fundic gland mucosa\(^{15}\). Although PG I level increases with slight atrophy, its level immediately decreases with progress of the atrophy.

**PG II (ng/ml)**

PG II is secreted from the fundic, the antral and the duodenal mucosa\(^{15}\). PG II level does not decrease until the atrophy in the stomach becomes severe.

**PG I/PG II ratio**

PG I/PG II ratio is obtained by dividing PG I level by PG II level. This value decreases on initially H. pylori infection since PG II level increases more than PG I level. It further decreases with progress of the atrophy since PG I immediately decrease.

Furthermore, variations of the mucosal surface and gastric folds appear on X-ray images of the stomach infected with H. pylori\(^{10}\) as shown in Fig. 2. The details of these variations are shown below.

**Mucosal surface pattern**

The mucosal surface of Hp antibody-positive patients generally has a rough pattern such as a valvate pattern.

**Fold distribution**
Folds on the gastric surface are classified into the following four groups:

(1) Folds that exist on the whole bodies of the stomach.

(2) Folds that exist on the whole bodies of the stomach except the lower side of the lesser curvature.

(3) Folds that exist only on the side of the greater curvature.

(4) Folds that exist only on a part of the side of the greater curvature or no existence of folds.

Patients who are PG-negative are principally classified into groups (1) and (2), whereas patients who are PG-positive are principally classified into groups (3) and (4). Most patients in whom atrophy has progressed in the stomach are classified into group (4).

**Fold shape**

The fold shape of Hp antibody-negative patients is slim and straight. On the other hand, the fold shape of Hp antibody-positive patients is wide and not straight.

Since the aforementioned symptoms on X-ray images are closely related to serum Hp antibody level and serum PG levels, calculation of image features focusing on the relationship with the blood data is desirable. Therefore, it is necessary to introduce a new approach that considers this relationship between “values of blood examination” and “symptoms of *H. pylori* infection that appear on gastric X-ray images” into CAD systems.

### 2.2 Calculation of KCCA-based Image Features

In this subsection, we present a method for calculating KCCA-based image features that have a relationship with values of blood examination. The values of blood examination, i.e., Hp antibody level and PG levels, complexly change with the development of the atrophy. Naturally, Hp antibody level increases when the stomach is infected with *H. pylori*. However, Hp antibody level decreases when the atrophy becomes severe since *H. pylori* cannot live in the stomach. On the other hand, PG I and PG II levels both increase with slight atrophy. Although PG I level immediately decreases with progress of atrophy, PG II level does not decrease until the atrophy in the stomach becomes severe.

As described above, the relationship between “values of blood examination” and “symptoms of *H. pylori* infection that appear on gastric X-ray images” is complex. Therefore, the proposed method calculates new image features using values of blood examination based on KCCA, which can consider a non-linear correlation.

Given an X-ray image taken at the pth ($p = 1, 2, \cdots, P; P$ being the number of positions) position from the ith ($i = 1, 2, \cdots, N; N$ being the number of patients) patient, the proposed method extracts image features shown in **Table 2** from this X-ray image.
Note that the details of each image feature are denoted in Appendix. Then we define an image feature vector \( x_p^i \in \mathbb{R}^M \) from these image features, where \( M \) is the dimension of the image feature vector. Specifically, the features other than HOG features are calculated for each of the blocks that are obtained by dividing the target image into \( 8 \times 8 \) regions. This means that \( 101 \times 8 \times 8 = 6464 \) features are obtained. On the other hand, HOG features are calculated by dividing the image into \( 6 \times 6 \) cells, and the block size is \( 3 \times 3 \) cells. Furthermore, a 9-bin gradient histogram is calculated from each cell. Therefore, \( 3 \times 3 \times 9 = 81 \) features are obtained from each block. In summary, \( 6464 + 1296 = 7760 \) features are obtained for each target gastric X-ray image, i.e., \( M = 7760 \).

Considering trade-off relationship between “the discriminative power of these features for the classification of \( H. pylori \) infection” and “its computation cost”, we divide the images into \( 8 \times 8 \) blocks. Specifically, the bigger the size of blocks becomes, the more difficult the representation of the detailed symptoms is, whereas the smaller the size of blocks becomes, the higher the calculation cost is since the dimension of the calculated image features becomes higher. Thus, from this trade-off relationship, we experimentally determine the size of blocks. Nevertheless, we should find the optimal size as a future work. Furthermore, we include the outside black area in this area division. Although the outside black area affects texture features, these meaningless features are eliminated by the calculation of correlation between the image features and the blood features via KCCA. The reason why we calculate image features from the whole regions of images is shown as follows. Since an imaging equipment which can take X-ray images not having the black area has been diffused, we calculated the visual features from all areas of images considering the use of these images.

Furthermore, \( H. pylori \) antibody level, PG I level, PG II level, and PG I/PG II of the \( i \)th patient are denoted as a blood feature vector \( y^i \in \mathbb{R}^L \), where \( L (= 4) \) is the dimension of the blood feature vector.

We obtain two sets of \( N \) feature vectors for position \( p \), \( X_p = [x_p^1, x_p^2, \ldots, x_p^N] \) and \( Y = [y^1, y^2, \ldots, y^N] \). To discover a non-linear correlation that does not generally appear in the original feature space, feature vectors of each set, i.e., \( x_p^i \) and \( y^i \), are mapped into non-linear feature spaces via \( \phi_{x_p} (\cdot) \in \mathbb{R}^{\phi_{x_p}} \) and \( \phi_{y_p} (\cdot) \in \mathbb{R}^{\phi_{y_p}} \), respectively. From the mapped results \( \Phi_{x_p} = [\phi_{x_p} (x_p^1), \phi_{x_p} (x_p^2), \ldots, \phi_{x_p} (x_p^N)] \) and \( \Phi_{y_p} = [\phi_{y_p} (y^1), \phi_{y_p} (y^2), \ldots, \phi_{y_p} (y^N)] \), KCCA seeks to maximize the correlation

\[
\rho_p = \frac{s_{x_p}^\top s_{y_p}}{\sqrt{s_{x_p}^\top s_{x_p}} \sqrt{s_{y_p}^\top s_{y_p}}} \tag{1}
\]

between

\[
s_{x_p} = H \Phi_{x_p}^\top w_{x_p} \tag{2}
\]

and

\[
s_{y_p} = H \Phi_{y_p}^\top w_{y_p} \tag{3}
\]

over the projection directions \( w_{x_p} \in \mathbb{R}^{\phi_{x_p}} \) and \( w_{y_p} \in \mathbb{R}^{\phi_{y_p}} \), where \( H \) is a centering matrix. This means that KCCA finds the directions \( w_{x_p} \) and \( w_{y_p} \) that maximize the correlation \( s_{x_p}^\top s_{y_p} \) of corresponding projections subject to \( s_{x_p}^\top s_{x_p} = 1 \) and \( s_{y_p}^\top s_{y_p} = 1 \). However, finding \( w_{x_p} \) and \( w_{y_p} \) by directly maximizing Eq. (1) is generally difficult since the dimensions of the projected spaces are very high. Hence, we employ the kernel trick, where kernel functions \( k_{x_p} (x_p^i, x_p^j) = \phi_{x_p} (x_p^i)^\top \phi_{x_p} (x_p^j) \) and \( k_{y_p} (y^i, y^j) = \phi_{y_p} (y^i)^\top \phi_{y_p} (y^j) \) are respectively used for the following calculation. Furthermore, \( w_{x_p} \) and \( w_{y_p} \) can be rewritten as

\[
w_{x_p} = \Phi_{x_p} H \alpha_{x_p}, \tag{4}
\]

\[
w_{y_p} = \Phi_{y_p} H \alpha_{y_p}. \tag{5}
\]

Then Eq. (1) is rewritten as follows:

\[
\rho_p = \frac{\alpha_{x_p}^\top \hat{K}_{x_p,y_p} \alpha_{y_p}}{\sqrt{\alpha_{x_p}^\top \hat{K}_{x_p} \alpha_{x_p} \alpha_{y_p}^\top \hat{K}_{y_p} \alpha_{y_p}}}, \tag{6}
\]

\[
\hat{K}_{x_p,y_p} = H K_{x_p} H H K_{y_p} H, \tag{7}
\]

\[
\hat{K}_{x_p} = H K_{x_p} H H K_{x_p} H + \zeta_{x_p} K_{x_p}, \tag{8}
\]

\[
\hat{K}_{y_p} = H K_{y_p} H H K_{y_p} H + \zeta_{y_p} K_{y_p}, \tag{9}
\]

where \( K_{x_p} \) and \( K_{y_p} \) are the gram matrices whose \( (j,k) \)th elements are \( k_{x_p} (x_p^j, x_p^k) \) and \( k_{y_p} (y^j, y^k) \), respectively. Furthermore, \( \zeta_{x_p} \) and \( \zeta_{y_p} \) are regularization parameters. As shown in Eqs. (8) and (9), the regularization terms \( \zeta_{x_p} K_{x_p} \) and \( \zeta_{y_p} K_{y_p} \) are newly introduced to obtain a unique solution.

The optimal values of \( \alpha_{x_p} \) and \( \alpha_{y_p} \) can be found by solving the following Lagrangean multiplier approach:

\[
\mathcal{L} = \alpha_{x_p}^\top \hat{K}_{x_p,y_p} \alpha_{y_p} - \frac{\lambda_{x_p}}{2} (\alpha_{x_p}^\top \hat{K}_{x_p} \alpha_{x_p}) - \frac{\lambda_{y_p}}{2} (\alpha_{y_p}^\top \hat{K}_{y_p} \alpha_{y_p}), \tag{10}
\]

where \( \lambda_{x_p} \) and \( \lambda_{y_p} \) are Lagrange multipliers. Consequently, the optimal directions \( \alpha_{x_p} \) and \( \alpha_{y_p} \) can be obtained by solving the following eigenvalue problem:
where $\lambda_p = \lambda_{x_p} = \lambda_{y_p} (= \rho_p)$ and $\rho_p$ corresponds to the canonical correlation. As solutions of Eq. (11), the proposed method obtains some eigenvectors as $\alpha^j_x (j = 1, 2, \cdots, d; d$ being the number of eigenvectors) and $\alpha^j_y$, respectively. Then we derive the following projection matrix

$$
\begin{bmatrix}
0 & K_{x_p,y_p} \\
\hat{K}_{x_p,y_p} & 0
\end{bmatrix}
\begin{bmatrix}
\alpha_x \\
\alpha_y
\end{bmatrix}
= \lambda_p
\begin{bmatrix}
K_{x_p} & 0 \\
0 & K_{y_p}
\end{bmatrix}
\begin{bmatrix}
\alpha_x \\
\alpha_y
\end{bmatrix},
$$

(11)

by calculating $d$ direction vectors whose corresponding eigenvalues are larger than the other ones.

By using the projection $W_{x_p} \in \mathbb{R}^{d \times d}$, the proposed method can map the image feature vector $\phi_{x_p}(x_p^i)$ into the subspace having the maximum correlation with the blood feature vector $\phi_{y_p}(y^i)$. Therefore, accurate classification of gastric cancer risk is expected by using the new feature vector $v^i_p \left( = W_{x_p}^T (\phi_{x_p}(x_p^i) - \hat{\phi}_{x_p}) \right) \in \mathbb{R}^d$ which is obtained by applying the projection $W_{x_p}$ to the original image feature vector $\phi_{x_p}(x_p^i)$, namely $v^i_p$ is a KCCA-based image feature vector in our method. Note that $\hat{\phi}_{x_p}$ is the mean vector of $\phi_{x_p}(x_p^i)$, i.e.,

$$\frac{1}{N} \sum_{i=1}^{N} \phi_{x_p}(x_p^i).$$

In the proposed method, the object of KCCA is to obtain the projection $W_{x_p}$, which can map the image feature vectors into the subspace having the maximum correlation with the blood feature vectors. In order to obtain the projection, we use only the training image feature vectors and their blood feature vectors. In the testing phase, the proposed method can map the testing image feature vector into the above subspace without using the blood feature vectors.

### 3. Classification and Integration of Gastric Cancer Risk

In this section, the classification method for X-ray images taken at each position and the integration method for all positions’ classification results for each stage are presented. As shown in Fig. 1, the proposed method performs two-class classification of gastric X-ray images for each stage. A patient detected as an $H.\ pylori$-infected patient in the first stage is classified again in the second stage based on the same manner as that in the first stage. As shown in Fig. 3, the proposed method integrates multiple classification results for obtaining the final result in each stage since the proposed method constructs a Support Vector Machine (SVM)\(^{(16)}\) for each position at which the X-ray image is taken.

Given a KCCA-based test image feature vector $v_p$ whose class label is unknown, this feature vector is classified by an SVM trained by using the training vectors $v^i_t$. Note that $v_p$ of the test data which do not have blood data is obtained by applying the projection $W_{x_p}$ to test image feature vector $x_p$, i.e., $v_p \left( = W_{x_p}^T (\phi_{x_p}(x_p) - \hat{\phi}_{x_p}) \right) \in \mathbb{R}^d$. Then we can obtain the $p$th classification result $l_p \in \{1, 0\}$, where 1 and 0 correspond to positive and negative, respectively.

Next, we calculate each stage’s final classification result by integrating all positions’ results $l_p (p = 1, 2, \cdots, P)$ via an accuracy-based voting method. We newly add threshold processing to Weighted Majority Voting\(^{(17)}\) since doctors can change classification accuracy of each class by modifying the threshold accordingly. For example, when doctors want to accurately detect patients with $H.\ pylori$ infection, the setting threshold satisfy these requests. Our voting method assigns
greater powers, i.e., higher weights, to competent classifiers. Specifically, the decision function to obtain each stage’s final result \( \hat{l} \) is defined as follows:

\[
\hat{l} = \begin{cases} 
1 & \text{if } \mu > T \\
0 & \text{otherwise}
\end{cases},
\]

(13)

\[
\mu = \frac{\sum_{p=1}^{P} \beta_{p,1} l_{p}}{\sum_{p=1}^{P} (\beta_{p,1} l_{p} + \beta_{p,0} (1 - l_{p}))},
\]

(14)

where \( \beta_{p,1} \) and \( \beta_{p,0} \) are respectively the \( p \)th classifier’s accuracy for results of 1 (positive) and 0 (negative), which are approximately calculated by applying cross-validation to training data. Moreover, \( T \) is a predefined threshold. As shown in the above procedures, we can perform the classification of gastric cancer risk for the X-ray image taken from each position and the integration of all positions’ classification results.

4. Experimental Results

The effectiveness of the proposed method was evaluated by performing experiments using gastric X-ray images taken from individuals having gastric atrophy of different degrees of development and from individuals \( H. \) pylori not infected with, i.e., patients having no gastric cancer risk. An outline of the experiment is given in 4.1, and the obtained results are presented and discussed in 4.2.

4.1 Experimental Conditions

A total of 2100 samples was used in this experiment. Each sample consisted of ABC classification results, values of blood examination and gastric X-ray images taken at eight positions (\( P = 8 \)), namely there were 2100×8 gastric X-ray images. Each image was of 8-bit gray scale and had a resolution of 1024 × 1024 pixels. The ground truth for this image set was obtained from the ABC classification results, which strongly correlate to gastric cancer risk as shown in Table 1. As described above, there are medical facilities in which groups C and D are assessed as the same gastric cancer risk since the number of patients classified as group D is generally small because of the high sensitivity of Hp antibody level. Therefore, we performed three-class classification, i.e., group A, group B and group C and D, of gastric X-ray images. In this experiment, 462 samples in group C and D were regarded as high gastric cancer risk, and 508 samples in group B were regarded as low gastric cancer risk. Furthermore, 1130 samples in group A were considered to be non-gastric cancer risk.

In order to confirm the effectiveness of considering the relationship between “values of blood examination” and “symptoms of \( H. \) pylori infection” and considering a non-linear correlation based on KCCA, the following comparative methods were also evaluated.

1. Comparative method 1 (CM1) uses the minimal-Redundancy-Maximal-Relevance (mRMR)\(^\text{15}\) feature selection algorithm, which can consider the redundancy among image features and relevance between image features and class label based on mutual information. Hence, the mRMR algorithm does not use values of blood examination.

2. Comparative method 2 (CM2) applies Kernel Principal Component Analysis (KPCA)\(^\text{19}\), which is a common dimensionality reduction method not using class label, i.e., an unsupervised method, to image feature vectors \( \mathbf{x}_p \).

3. Comparative method 3 (CM3) uses Canonical Correlation Analysis (CCA)\(^\text{20}\), which can consider only a linear correlation between two input heterogeneous features. Note that the dimension of the new image feature vector after applying CCA was set to four since the dimension of the blood feature vector was four. This value was theoretically the maximum value in CCA.

4. Comparative method 4 (CM4) directly uses all image features.

The verification method was 10-fold cross-validation. For evaluation of each stage’s classification, we utilized the following sensitivity (Sen) and specificity (Spe), Youden Index (YI) which are commonly used to evaluate reliability of clinical examinations in the field of medicine:

\[
\text{Sen} = \frac{TP}{TP + FN},
\]

(15)

\[
\text{Spe} = \frac{TN}{TN + FP},
\]

(16)

\[
\text{YI} = \text{Sen} + \text{Spe} - 1.
\]

(17)

\( TP, TN, FP \) and \( FN \) represent the numbers of true positive, true negative, false positive and false negative, respectively. In the first stage, negative means the samples in group A, and in the second stage, negative means the samples in group B. We also utilized the Receiver Operating Characteristic curve (ROC curve) which displays the sensitivity on the vertical axis and the false positive rate (equal to 1-specificity) on the horizontal axis and the confusion matrix known in the field of machine learning. Each column of the matrix represents the number of samples in a predicted class, while each row represents the number of samples in a true class.
Furthermore, we applied Welch’s t-test as a significant difference test to the Youden Index calculated by 10-fold cross-validation for claiming the effectiveness of the proposed method.

In each method, we changed the dimension of the image feature in mRMR, KPCA, CCA and KCCA by 50 from 50 to 300 to determine its best value. The Gaussian kernel was used for the kernel function in KPCA and KCCA, with the best kernel parameter. The parameters, such as the dimension, the kernel parameter and the regularization parameters, were set to the values which output the highest Youden Index. To evaluate the effectiveness of the KCCA-based new image feature, the linear kernel was used for the kernel function in SVM whose regularization parameter was set to the same value in our method and the comparative methods.

4.2 Results and Discussion

The results of experiments are presented and discussed in this subsection. Table 3 shows the first stage’s sensitivity, specificity and Youden Index of the proposed method (PM) and the comparative methods (CM1, CM2, CM3 and CM4).

Table 3  First stage’s results of sensitivity, specificity and Youden Index in the proposed method (PM) and the comparative methods (CM1, CM2, CM3 and CM4).

<table>
<thead>
<tr>
<th></th>
<th>Spec (group A)</th>
<th>Sen (groups B, C and D)</th>
<th>Youden Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>PM</td>
<td>0.866</td>
<td>0.892</td>
<td>0.758</td>
</tr>
<tr>
<td>CM1</td>
<td>0.854</td>
<td>0.874</td>
<td>0.723</td>
</tr>
<tr>
<td>CM2</td>
<td>0.825</td>
<td>0.833</td>
<td>0.658</td>
</tr>
<tr>
<td>CM3</td>
<td>0.859</td>
<td>0.847</td>
<td>0.707</td>
</tr>
<tr>
<td>CM4</td>
<td>0.819</td>
<td>0.856</td>
<td>0.675</td>
</tr>
</tbody>
</table>

Table 4  Second stage’s results of sensitivity, specificity and Youden Index in the proposed method (PM) and the comparative methods (CM1, CM2, CM3 and CM4).

<table>
<thead>
<tr>
<th></th>
<th>Spec (group B)</th>
<th>Sen (group C and D)</th>
<th>Youden Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>PM</td>
<td>0.724</td>
<td>0.827</td>
<td>0.551</td>
</tr>
<tr>
<td>CM1</td>
<td>0.657</td>
<td>0.729</td>
<td>0.386</td>
</tr>
<tr>
<td>CM2</td>
<td>0.628</td>
<td>0.695</td>
<td>0.322</td>
</tr>
<tr>
<td>CM3</td>
<td>0.631</td>
<td>0.740</td>
<td>0.371</td>
</tr>
<tr>
<td>CM4</td>
<td>0.588</td>
<td>0.697</td>
<td>0.285</td>
</tr>
</tbody>
</table>

Table 5  Results of Welch’s t-test whose significant level is set as $\alpha = 0.01$.

<table>
<thead>
<tr>
<th></th>
<th>First stage</th>
<th>Second Stage</th>
</tr>
</thead>
<tbody>
<tr>
<td>CM1</td>
<td>$p = 0.054$</td>
<td>$p &lt; 0.01$</td>
</tr>
<tr>
<td>CM2</td>
<td>$p &lt; 0.01$</td>
<td>$p &lt; 0.01$</td>
</tr>
<tr>
<td>CM3</td>
<td>$p &lt; 0.01$</td>
<td>$p &lt; 0.01$</td>
</tr>
<tr>
<td>CM4</td>
<td>$p &lt; 0.01$</td>
<td>$p &lt; 0.01$</td>
</tr>
</tbody>
</table>

As shown in Tables 3 and 4, sensitivity, specificity and Youden Index of our method are higher than those of the comparative methods. Furthermore, comparative method 3 outperforms comparative methods 2 and 4 from Tables 3 and 4. We confirmed that the Youden Index of the proposed method was statistically higher than those of comparative methods 2, 3 and 4 at the 1 percent level in the first stage, and was statistically higher than those of all comparative methods at the 1 percent level in the second stage as shown in Table 5. From the aforementioned results, we can confirm the effectiveness of considering the relationship between “values of blood examination” and “symptoms of H. pylori infection”. Since the proposed method’s sensitivity of the first stage is 0.892, patients with gastric cancer risk or not was determined in such a way that Youden Index was the highest. This parameter setting is commonly used in the field of medicine. Therefore, the proposed method can increase the sensitivity in the first stage by setting the threshold $T$ to a lower value than the value used in this experiment. Furthermore, Figures 4 and 5 show the ROC curves of the first stage and the second stage, respectively. From these ROC curves, the performance of the proposed method was higher than those of the comparative methods since the curve of the proposed method was drawn on the uppermost part.

Table 6 shows the confusion matrices of the proposed method and comparative methods. The confusion matrices show the number of samples classified into each class. Furthermore, Fig 6 shows the gastric X-ray images classified correctly and incorrectly by the proposed method for each gastric cancer risk. From Table 6, the number of samples classified correctly in the proposed method is the largest in this experiment. A comparison of the proposed method and comparative method 3 show the effectiveness of considering a non-linear correlation based on KCCA. From the aforementioned results, the consideration of a non-linear correlation between “Hp antibody level and PG levels” and “symptoms of H. pylori infection that appear on gastric X-ray images” is important for accurate classification of gastric cancer risk.

***Note that when focusing on Hp antibodies as ground truth, the classification results by specialists are as follows: sensitivity of 0.932 and specificity of 0.867.***

4.2 Results and Discussion

The results of experiments are presented and discussed in this subsection. Table 3 shows the first stage’s sensitivity, specificity and Youden Index of the proposed method (PM) and the comparative methods, and Table 4 shows the second stage’s sensitivity, specificity and Youden Index of the proposed method and the comparative methods. Furthermore, Table 5 shows the results of Welch’s t-test. This statistical test was performed between our method and each comparative method through the 10-fold cross-validation.

Furthermore, we applied Welch’s t-test as a significant difference test to the Youden Index calculated by 10-fold cross-validation for claiming the effectiveness of the proposed method.

In each method, we changed the dimension of the image feature in mRMR, KPCA, CCA and KCCA by 50 from 50 to 300 to determine its best value. The Gaussian kernel was used for the kernel function in KPCA and KCCA, with the best kernel parameter. The parameters, such as the dimension, the kernel parameter and the regularization parameters, were set to the values which output the highest Youden Index. To evaluate the effectiveness of the KCCA-based new image feature, the linear kernel was used for the kernel function in SVM whose regularization parameter was set to the same value in our method and the comparative methods.

4.2 Results and Discussion

The results of experiments are presented and discussed in this subsection. Table 3 shows the first stage’s sensitivity, specificity and Youden Index of the proposed method (PM) and the comparative methods, and Table 4 shows the second stage’s sensitivity, specificity and Youden Index of the proposed method and the comparative methods. Furthermore, Table 5 shows the results of Welch’s t-test. This statistical test was performed between our method and each comparative method through the 10-fold cross-validation.

As shown in Tables 3 and 4, sensitivity, specificity and Youden Index of our method are higher than those of the comparative methods. Furthermore, comparative method 3 outperforms comparative methods 2 and 4 from Tables 3 and 4. We confirmed that the Youden Index of the proposed method was statistically higher than those of comparative methods 2, 3 and 4 at the 1 percent level in the first stage, and was statistically higher than those of all comparative methods at the 1 percent level in the second stage as shown in Table 5. From the aforementioned results, we can confirm the effectiveness of considering the relationship between “values of blood examination” and “symptoms of H. pylori infection”. Since the proposed method’s sensitivity of the first stage is 0.892, patients with gastric cancer risk or not was determined in such a way that Youden Index was the highest. This parameter setting is commonly used in the field of medicine. Therefore, the proposed method can increase the sensitivity in the first stage by setting the threshold $T$ to a lower value than the value used in this experiment. Furthermore, Figures 4 and 5 show the ROC curves of the first stage and the second stage, respectively. From these ROC curves, the performance of the proposed method was higher than those of the comparative methods since the curve of the proposed method was drawn on the uppermost part.

Table 6 shows the confusion matrices of the proposed method and comparative methods. The confusion matrices show the number of samples classified into each class. Furthermore, Fig 6 shows the gastric X-ray images classified correctly and incorrectly by the proposed method for each gastric cancer risk. From Table 6, the number of samples classified correctly in the proposed method is the largest in this experiment. A comparison of the proposed method and comparative method 3 show the effectiveness of considering a non-linear correlation based on KCCA. From the aforementioned results, the consideration of a non-linear correlation between “Hp antibody level and PG levels” and “symptoms of H. pylori infection that appear on gastric X-ray images” is important for accurate classification of gastric cancer risk.

***Note that when focusing on Hp antibodies as ground truth, the classification results by specialists are as follows: sensitivity of 0.932 and specificity of 0.867.***
5. Conclusions

A fully automatic method for classification of gastric cancer risk from gastric X-ray images has been presented in this paper. The proposed method first detects \textit{H. pylori}-infected patients, and then the proposed method classifies gastric cancer risk, \textit{i.e.}, high or low, from the previously detected \textit{H. pylori}-infected patients. In these classification procedures, the proposed method can consider the relationship between “values of blood examination” and “symptoms of \textit{H. pylori} infection” by calculating a new KCCA-based image feature in each stage. Finally, the effectiveness of the proposed method was evaluated by performing experiments using actual gastric X-ray images.

In future work, we plan to propose a method that can estimate the results of endoscopy from X-ray images since endoscopy is the most reliable examination. In this method, we can use features collaboratively using the blood features and the image features since we did not directly use the values of blood examination in this paper.

Acknowledgement

In this research, we utilized image data and results that were kindly provided by the Medical Examination Center of Yamagata City Medical Association. We would also like to sincerely thank Katsuhiro Mabe of Hakodate National Hospital, Shigemi Nakajima of Social Insurance Shiga Hospital, Harufumi Oizumi of Yamagata City Medical Association, Kazuya Yoshizawa of the Faculty of Medicine, Yamagata University, and many others for providing data, images and results as well as invaluable advice. This work was partly supported by Grant-in-Aid for Scientific Research (B) JP25280036 from JSPS.

References

12) C. Xu, D. Tao, and C. Xu, “A survey on multi-view learning,”

Fig. 4 First stage’s ROC curves obtained by changing the threshold $T$.

Fig. 5 Second stage’s ROC curves obtained by changing the threshold $T$. 
Table 6  Confusion matrices of the proposed method and the comparative methods. Each value is the number of samples.

<table>
<thead>
<tr>
<th></th>
<th>True class</th>
<th></th>
<th>Predicted class</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Group A</td>
<td>Group B</td>
<td>Group C and D</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Group A</td>
<td>979</td>
<td>84</td>
<td>67</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Group B</td>
<td>88</td>
<td>304</td>
<td>116</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Group C and D</td>
<td>17</td>
<td>77</td>
<td>368</td>
<td></td>
</tr>
</tbody>
</table>

(a) Proposed method

<table>
<thead>
<tr>
<th></th>
<th>True class</th>
<th>Predicted class</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Group A</td>
<td>Group B</td>
<td>Group C and D</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Group A</td>
<td>965</td>
<td>78</td>
<td>87</td>
</tr>
<tr>
<td></td>
<td>Group B</td>
<td>94</td>
<td>271</td>
<td>143</td>
</tr>
<tr>
<td></td>
<td>Group C and D</td>
<td>28</td>
<td>118</td>
<td>316</td>
</tr>
</tbody>
</table>

(b) Comparative method 1

<table>
<thead>
<tr>
<th></th>
<th>True class</th>
<th>Predicted class</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Group A</td>
<td>Group B</td>
<td>Group C and D</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Group A</td>
<td>932</td>
<td>111</td>
<td>87</td>
</tr>
<tr>
<td></td>
<td>Group B</td>
<td>123</td>
<td>242</td>
<td>143</td>
</tr>
<tr>
<td></td>
<td>Group C and D</td>
<td>39</td>
<td>129</td>
<td>294</td>
</tr>
</tbody>
</table>

(c) Comparative method 2

<table>
<thead>
<tr>
<th></th>
<th>True class</th>
<th>Predicted class</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Group A</td>
<td>Group B</td>
<td>Group C and D</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Group A</td>
<td>925</td>
<td>117</td>
<td>88</td>
</tr>
<tr>
<td></td>
<td>Group B</td>
<td>106</td>
<td>236</td>
<td>116</td>
</tr>
<tr>
<td></td>
<td>Group C and D</td>
<td>34</td>
<td>130</td>
<td>298</td>
</tr>
</tbody>
</table>

(d) Comparative method 3

<table>
<thead>
<tr>
<th></th>
<th>True class</th>
<th>Predicted class</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Group A</td>
<td>Group B</td>
<td>Group C and D</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Group A</td>
<td>971</td>
<td>75</td>
<td>84</td>
</tr>
<tr>
<td></td>
<td>Group B</td>
<td>117</td>
<td>247</td>
<td>144</td>
</tr>
<tr>
<td></td>
<td>Group C and D</td>
<td>31</td>
<td>112</td>
<td>319</td>
</tr>
</tbody>
</table>

(e) Comparative method 4

Fig. 6  Examples of gastric X-ray images with their classification results obtained by the proposed method. “GT” and “Result” mean the ground truth and the classification result, respectively. Furthermore, “A”, “B” and “C and D” mean group A, group B and group C and D, respectively.
Appendix: Calculation of image features

- **Intensity histogram-related features (4 dimensions per block)**
  An intensity histogram which has 256 bins is calculated. From the normalized intensity histogram such that its sum is equal to one, the mean, variance, skewness and kurtosis are calculated.

- **Co-occurrence matrix-related features** (9 dimensions per block)
  From each of the four direction co-occurrence matrices whose size is 256×256, the proposed method calculates the following 9 features: contrast, correlation, second angular moment, uniformity, variance, inverse difference moment, entropy and correlation information measure. Then the average for each feature is obtained.

- **Gabor Wavelet-based features (56 dimensions per block)**
  Gabor Wavelet transform is applied to the gastric X-ray image in 8 individual directions. For each obtained Wavelet coefficient matrix, we calculate the maximum value and the minimum value as well as the histogram mode, mean, variance, skewness and kurtosis.

- **Adaptive Local Binary Patterns (ALBP)** (22 dimensions per block)
  ALBP is based on texture features for Local Binary Pattern and consists of two elements, namely Adaptive Local Binary Patterns Histogram (ALBPH) and Gradient Adaptive Local Binary Patterns (GALBP). In the propose method, ALBPH shows a histogram of the 11 texture patterns calculated from the differences between the center pixel and the neighborhood pixel values. GALBP calculates the gradient for each pixel and the sum of the gradient for each texture pattern. Finally, $11 + 1 = 22$ ALBP features can be obtained.

- **Hu-moment invariant** (7 dimensions per block)
  The proposed method calculates the scale- and rotation-invariant Hu-moment features from the binary image obtained by applying Canny filter.

- **Hough transform-based features (2 dimensions per block)**
  The proposed method applies the Hough transform to the binary image after applying Canny filter and obtains the peak point in Hough space.

- **Histograms of Oriented Gradients (HOG) features** (1296 dimensions per image)
  Each X-ray image is divided into $6 \times 6$ cells, and a block is composed of $3 \times 3$ cells. For each block, a 9-bin gradient histogram is calculated for each cell. Therefore, $3 \times 3 \times 9 = 81$ features are obtained from each block. Finally, $16 \times 81 = 1296$ HOG features can be obtained.

- **Sobel filter-based edge features (1 dimension per block)**
  The proposed method calculates the sum of the gradients obtained by applying the Sobel filter.

Kenta Ishihara received his B.S. and M.S. degrees in Electronics and Information Engineering from Hokkaido University, Japan in 2014 and 2016, respectively. He is currently pursuing a Ph.D. degree at the Graduate School of Information Science and Technology, Hokkaido University. His research interests are medical image processing and its applications. He is a student member of the IEEE and IEICE.
Takahiro Ogawa received his B.S., M.S. and Ph.D. degrees in Electronics and Information Engineering from Hokkaido University, Japan in 2003, 2005 and 2007, respectively. He is currently an assistant professor in the Graduate School of Information Science and Technology, Hokkaido University. His research interests are multimedia signal processing and its applications. He has been an Associate Editor of ITE Transactions on Media Technology and Applications. He is a member of the IEEE, EURASIP, IEICE, and Institute of Image Information and Television Engineers (ITE).

Miki Haseyama received her B.S., M.S. and Ph.D. degrees in Electronics from Hokkaido University, Japan in 1986, 1988 and 1993, respectively. She joined the Graduate School of Information Science and Technology, Hokkaido University as an associate professor in 1994. She was a visiting associate professor of Washington University, USA from 1995 to 1996. She is currently a professor in the Graduate School of Information Science and Technology, Hokkaido University. Her research interests include image and video processing and its development into semantic analysis. She has been a Vice-President of the Institute of Image Information and Television Engineers, Japan (ITE), an Editor-in-Chief of ITE Transactions on Media Technology and Applications, a Director, International Coordination and Publicity of The Institute of Electronics, Information and Communication Engineers (IEICE). She is a member of the IEEE, IEICE, Institute of Image Information and Television Engineers (ITE) and Acoustical Society of Japan (ASJ).