



Title	Helicobacter Pylori infection detection from gastric X-ray images based on feature fusion and decision fusion
Author(s)	Ishihara, Kenta; Ogawa, Takahiro; Haseyama, Miki
Citation	Computers in biology and medicine, 84, 69-78 <a href="https://doi.org/10.1016/j.combiomed.2017.03.007">https://doi.org/10.1016/j.combiomed.2017.03.007</a>
Issue Date	2017-05-01
Doc URL	<a href="http://hdl.handle.net/2115/70005">http://hdl.handle.net/2115/70005</a>
Rights	© 2017. This manuscript version is made available under the CC-BY-NC-ND 4.0 license <a href="http://creativecommons.org/licenses/by-nc-nd/4.0/">http://creativecommons.org/licenses/by-nc-nd/4.0/</a>
Rights(URL)	<a href="http://creativecommons.org/licenses/by-nc-nd/4.0/">http://creativecommons.org/licenses/by-nc-nd/4.0/</a>
Type	article (author version)
File Information	CBM_ishihara.pdf



[Instructions for use](#)

# Helicobacter Pylori Infection Detection from Gastric X-ray Images Based on Feature Fusion and Decision Fusion

Kenta Ishihara, Takahiro Ogawa, Miki Haseyama

Graduate School of Information Science and Technology, Hokkaido University, Kita-14, Nishi-9, Sapporo-shi, 060-0814 Japan  
TEL: +81-11-706-6078, FAX: +81-11-706-7369

E-mail: {ishihara,ogawa}@lmd.ist.hokudai.ac.jp, miki@ist.hokudai.ac.jp

## Abstract

In this paper, a fully automatic method for detection of *Helicobacter pylori* (*H. pylori*) infection is presented with the aim of constructing a computer-aided diagnosis (CAD) system. In order to realize a CAD system with good performance for detection of *H. pylori* infection, we focus on the following characteristic of stomach X-ray examination. The accuracy of X-ray examination differs depending on the symptom of *H. pylori* infection that is focused on and the position from which X-ray images are taken. Therefore, doctors have to comprehensively assess the symptoms and positions. In order to introduce the idea of doctors' assessment into the CAD system, we newly propose a method for detection of *H. pylori* infection based on the combined use of feature fusion and decision fusion. As a feature fusion scheme, we adopt Multiple Kernel Learning (MKL). Since MKL can combine several features with determination of their weights, it can represent the differences in symptoms. By constructing an MKL classifier for each position, we can obtain several detection results. Furthermore, we introduce confidence-based decision fusion, which can consider the relationship between the classifier's performance and the detection results. Consequently, accurate detection of *H. pylori* infection becomes possible by the proposed method. Experimental results obtained by applying the proposed method to real X-ray images show that our method has good performance, close to the results of detection by specialists, and indicate that the realization of a CAD system for determining the risk of *H. pylori* infection is possible.

**Keywords:**

*H. pylori* infection detection, gastric X-ray images, feature fusion, decision fusion.

## 1. Introduction

The results of many studies have shown that *Helicobacter pylori* (*H. pylori*) infection causes the development of gastric cancer [1, 2, 3]. *H. pylori* infection triggers a series of inflammatory reactions [4, 5], and among these reactions, chronic atrophic gastritis is considered to be the first step of a sequence of mucosal changes in the stomach leading to cancer [2]. Early detection of *H. pylori* infection will enable subsequent elimination of *H. pylori* and a reduction in gastric cancer risk [6, 7]. Since *H. pylori* infection is an important factor in alteration of acid secretion, development of ulcers and carcinogenic potential [8], eradication of *H. pylori* has become an important issue for curing gastroduodenal diseases [9].

A gastric cancer risk classification scheme (ABCD classification [10]) based on *H. pylori* infection and degree of atrophic gastritis has been introduced in each municipality and in hospitals in Japan. The ABCD classification is a blood examination based on the combination serum *H. pylori* antibody level (Hp antibody level) and serum pepsinogen (PG) levels. Hp antibody level is used for estimating *H. pylori* infection, and PG levels are the serum markers of atrophic gastritis, which is a precancerous change in the stomach [11]. In the ABCD classification, screened patients are classified into four categories,

Table 1: ABCD classification for gastric cancer risk. Note that the Hp antibody of category D is (-), but this category is regarded as *H. pylori* infection. Thus, category A represents the absence of *H. pylori* infection, and categories B, C and D represent the presence of *H. pylori* infection.

		Hp antibody level	
		(-)	(+)
PG levels	(-)	A	B
	(+)	D	C

*i.e.*, categories A, B, C and D, based on results of examinations of blood samples as shown in Table 1. Category A is regarded as the lowest category of gastric cancer risk, and category D is regarded as the highest category of gastric cancer risk even if Hp antibody level is low, *i.e.*, no detectable *H. pylori* presence. Patients identified as category D have severe mucosal atrophy associated with intestinal metaplasia [12]. There is the potential for the reduction of anti-Hp antibody production due to severe atrophy since *H. pylori* hardly survives in the intestinal metaplasia mucosa [13]. Therefore, the gastric cancer risk is high despite shedding of *H. pylori*. For efficient diagnosis, the low-risk category, *i.e.*, category A, can be excluded from further detailed examination. Although the ABCD classifica-

tion is useful for screening patients, the results generally tend to include some false negative results [10, 14]. The false negative results are results for patients who have been incorrectly regarded as not having gastric cancer risk even though they have the risk. Hp antibody level may decrease due to severe atrophy as described above. Furthermore, PG levels after *H. pylori* eradication are changed to normal levels even though gastric acid secretion may not have ceased [15], *i.e.*, remaining gastric cancer risk. The ABCD classification tends to categorize patients after eradication treatment into category A, thus giving rise to false negative results.

Therefore, doctors must directly monitor the stomachs of patients to avoid the above problem. Diagnosis is generally made by endoscopy or photofluorography. For endoscopic images, a classification method of *H. pylori*-related gastric histology has been reported [16]. Although endoscopic examination is better than X-ray examination for detecting early gastric cancer [17], the usefulness of endoscopic examination for mass screening is limited [18]. Since many patients need to be diagnosed every year, diagnosis by endoscopy for all candidates with a risk of *H. pylori* infection is impossible. Therefore, stomach X-ray examination has become more significant. After screening patients by stomach photofluorography, performing endoscopy as an additional examination enables more accurate and efficient gastric cancer examination. Thus, stomach X-ray examination is effective for identifying patients with a risk of gastric cancer.

It has also been reported that *H. pylori* infection can be visually detected from gastric X-ray images and that highly reliable diagnosis by specialists is possible [19, 20]. However, in X-ray diagnosis, the workload for doctors increases due to the necessity of a high degree of experience and specialist knowledge. Computer-aided detection (CAD) systems that automatically analyze images have been attracting attention since they can decrease the workload on doctors [21]. The objective for the use of a CAD system is first screening for selecting patients who require follow-up examination, *i.e.*, X-ray examination and endoscopy, by doctors for diagnosis. Therefore, if a CAD system has good performance for detection of *H. pylori* infection, it can reduce the workload on doctors. A CAD system can also be used to support a diagnosis by an inexperienced doctor as a second opinion.

Recently, in the field of medical imaging, many methods have been developed for detecting medical abnormalities [22, 21, 23]. The idea of a combination scheme, *i.e.*, decision fusion, for detecting an abnormality has been introduced in reports [21, 23]. These methods combine the results obtained from exiting detection methods. However, as far as we know, there have been only a few recent studies on detection of gastric abnormalities by initially using X-ray images. Abe *et al.* [24] proposed new features that can be used to discriminate gastric cancer in X-ray images. Since gastric X-ray images are usually taken from several positions, the combined use of images may yield better performance.

In fields other than medical imaging, many methods based on combination schemes [25], *i.e.*, feature fusion and decision fusion, have been proposed. For feature fusion, multiple kernel learning (MKL) has been successfully applied to a number

of tasks in computer vision [26]. Furthermore, recent studies have shown promising performance of MKL for object recognition [27]. There have also been many studies on decision fusion. Much interest has been shown in confidence-based decision fusion schemes [28, 29]. Alam *et al.* [28] developed a fusion framework for audio-visual biometric identification. The framework works well when input samples presented are contaminated by noise, *e.g.*, detector noise, bit error and additive noise. Topcu *et al.* [29] classified each block that is obtained by dividing face images into rectangular regions and applied a linear combination to each block's classification results. Then the coefficients of linear combinations are obtained from some schemes, *e.g.*, accuracy-based weighting and Fisher discriminant analysis-based weighting. These fusion schemes can be used to determine whether patients are infected with *H. pylori* nor not.

A method for detection of *H. pylori* infection that is based on feature fusion and decision fusion is presented in this paper. For accurate detection of *H. pylori* infection, we focus on two unique characteristics of gastric X-ray examination. First, the accuracy of stomach X-ray examination differs depending on the symptoms of *H. pylori* infection that are focused on [30], *e.g.*, mucosal surface, fold distribution and fold shape. Secondly, gastric X-ray images are usually taken from several positions, and the accuracy of stomach X-ray examination differs depending on the position. Due to these characteristics, feature fusion and decision fusion must both be used. Feature fusion can integrate each visual feature by considering its importance, and decision fusion can integrate multiple detection results by considering which position from which X-ray image was taken will lead to good detection results from the relationship between the classifier's performance and the classifier's confidence. Therefore, we first introduce MKL as feature fusion since MKL can combine visual features with consideration of the importance of each visual feature by linear combinations of visual features. For realizing effective feature fusion based on MKL, the proposed method refines each visual feature based on the minimal-Redundancy-Maximal-Relevance (mRMR) algorithm [31] and Kernel Local Fisher Discriminant Analysis (KLFDA) [32]. Note that the mRMR algorithm and KLFDA are a feature selection method and a dimensionality reduction method, respectively. By applying these methods to original visual features, we can obtain visual features that have more discriminative power. We also use confidence-based decision fusion developed in [33], which can integrate multiple detection results by considering not only the classifier's performance but also the relationship between feature vectors and detection results unlike simple decision fusion. This enables identification of classifiers that output better results for target features to determine the final optimal detection results. We can also consider how confidently each classifier obtains detection results by using confidence scores obtained in MKL-based classification as the feature vector. By considering the characteristics of a doctor's inspection based on the combined use of feature fusion and decision fusion, accurate detection of *H. pylori* infection is achieved by the proposed method.

The contribution of this paper is the combined use of feature

Table 2: Overview of the symptoms of *Helicobacter pylori* infection in gastric X-ray images.

			Corresponding blood test
<b>Mucosal Surface</b>	Smooth type	A uniform, fine mucosal pattern	Hp-negative
	Coarse type	A rough mucosal pattern with fine bumps	Hp-positive
<b>Fold Distribution</b>	1)	Folds can be seen in all areas of the gastric body.	Hp-negative
	2)	Folds can be seen in all areas of the gastric body other than the lower side of the lesser	
	3)	Folds can be seen only in the side of the greater curvature.	Hp-positive
	4)	Folds can be seen only in parts of the side of the greater curvature or are completely	
<b>Fold Shape</b>	Normal	Slim, straight, smooth, small (height not large), slow (rise and fall not rapid), and soft	Hp-negative
	Abnormal	Large, non-straight, with a rough surface (surface appearing bumpy upon magnification), with a large height and rapid rise and fall	Hp-positive
	Absent	Folds are missing.	

fusion and decision fusion for considering the characteristics of a stomach X-ray examination. Although we have published some previous papers [34, 35, 36], their contribution is different from this paper’s one. In a previous report [34], we introduced decision fusion without the use of feature fusion, *i.e.*, not considering the importance of each symptom. We have also introduced the idea of double reading, which is examination of gastric X-ray images, *i.e.*, photofluorography, by two readers in a method based on two kinds of classifiers [35]. Although we used MKL as a kind of classifier, the theoretical methodology of the previous method was insufficient since the previous method was used in a pre-study for demonstrating the idea of double reading in machine learning. Particularly, MKL in the previous method were not methodologically introduced the characteristic that the accuracy of X-ray examination differed depending on the symptom of *H. pylori* infection focused on. For additional improvement in performance using the framework in our previous paper, we have to improve the performance of MKL based on the refinement of visual features. In the previous method, all of the visual features were aggregated into three kinds of feature types, texture feature, shape feature and gradient-based feature, and only the mRMR algorithm was applied for feature selection to each type of feature. After having been selected by the mRMR algorithm, the three types of features were combined on the basis of MKL. However, we have revealed that the symptoms of *H. pylori* infection focused on by each kind of visual feature are different. Furthermore, we considered that the accurate description of the symptom that doctors were focused on enabled by increasing the discriminative power of the visual features. Therefore, the proposed method enables accurate detection of *H. pylori* by improving the discriminative power of each visual feature based on KLFDA and combining all of these features based on MKL. For the reason mentioned above, we consider the method presented in this paper is a refined method considering both medical knowledge and knowledge of information science. On the other hand, we have calculated effective visual features related to measured values in a blood examination [36]. Although X-ray images and measured values in blood examination, *i.e.*, Hp antibody level and PG levels, were used in the previous method [36], only X-ray images are used in the

method presented here. Therefore, the motivation of investigation for the present method and that for the previous method are different [36].

This paper is organized as follows. First, the extraction and selection of features from gastric X-ray images are explained in Section 2. The feature fusion-based method for detection of *H. pylori* infection is presented in Section 3. Confidence-based decision fusion is explained in Section 4. In Section 5, experimental results obtained by applying our method to actual X-ray images are presented, and the effectiveness of the proposed method is shown. Finally, conclusions are presented in Section 6.

## 2. Feature Extraction and Selection

In this section, the extraction and selection of visual features used in our method are explained.

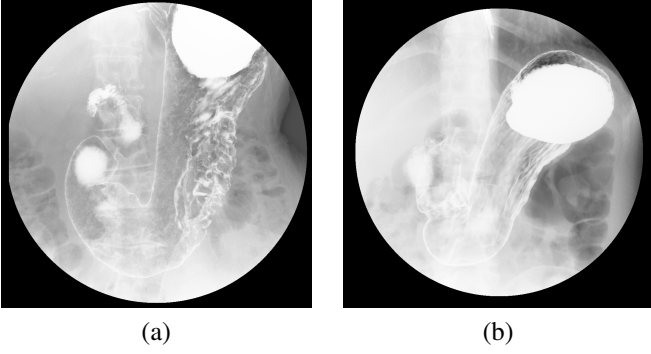
### 2.1. Feature Extraction

A summary of the symptoms of *H. pylori* infection in gastric X-ray images is shown in Table 2. Figure 1 shows an example of gastric X-ray images used in this study. Figure 1 (a) shows an image with *H. pylori* infection (hereafter, positive image), and Fig. 1 (b) shows an image with no *H. pylori* infection (hereafter, negative image). These images can be examined visually by looking for characteristics such as overall shape of the stomach and irregularities and folds in the gastric mucosal surface. It has been reported that high values for both sensitivity and specificity can be obtained when specialists manually perform diagnosis of *H. pylori* infection according to the above symptoms [20]. As shown in Table 2, *H. pylori* infection has the following main symptoms.

- (i) Gastric mucosal surface in cases of *H. pylori* infection usually has a coarse mucosal surface pattern.
- (ii) Gastric folds in cases of *H. pylori* infection usually exist only on the side of the greater curvature or do not exist.
- (iii) Folds in cases of *H. pylori* infection are usually defined as wide, non-straight, rough and large.

Table 3: Features extracted in the proposed method.

Feature type	Detail of features	Dimension
Texture features	Intensity histogram-related features	$4 \times 8 \times 8$
	Co-occurrence matrix-related features [37]	$9 \times 8 \times 8$
	Gabor Wavelet-based features	$56 \times 8 \times 8$
	ALBP[38]	$22 \times 8 \times 8$
Shape features	Hu-moment invariant [39]	$7 \times 8 \times 8$
	Hough transform[40]-based features	$2 \times 8 \times 8$
Gradient-based Features	HOG features [41]	1296
	Sobel filter[42]-based edge features	$1 \times 8 \times 8$
Total dimension		7760

Figure 1: Examples of gastric X-ray images used in this study: (a) an image with *H. pylori* infection and (b) an image with no *H. pylori* infection. These two images show typical features of images with/without *H. pylori* infection.

In order to represent the above symptoms, the proposed method extracts texture features, shape features and gradient-based features from each image as shown in Table 3. Note that each image ( $X \times Y$  pixels) is a gray-scale image, not including color information. The details of the features used in the proposed method are shown below.

### Texture Features

- Intensity histogram-related features (4 dimensions)  
An intensity histogram is calculated from the gastric X-ray image. The histogram is normalized such that its sum is equal to one. From the obtained histogram, the mean, variance, skewness and kurtosis are calculated.
- Co-occurrence matrix-related features [37] (9 dimensions)  
The proposed method calculates a co-occurrence matrix that represents inter-pixel relationships for each of the four directions  $\theta = \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$  from the gastric X-ray image. From each of the calculated co-occurrence matrices, 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)  
The proposed method applies a Gabor Wavelet transform to the gastric X-ray image in 8 individual directions ( $\theta =$

$\{0^\circ, 45^\circ, 90^\circ, 135^\circ, 180^\circ, 225^\circ, 270^\circ, 315^\circ\}$ ). For each obtained wavelet coefficient matrix, the proposed method calculates the maximum value and minimum value as well as the histogram mode, mean, variance, skewness and kurtosis.

- Adaptive Local Binary Pattern (ALBP) [38] (22 dimensions)  
This feature is proposed for improving the Local Binary Pattern (LBP) [43]. The ALBP has robustness of image rotation, and it can represent the difference between a smooth region and an unsmooth region.

The target X-ray images do not have any color information, and it is important to monitor the brightness of each part. Therefore, intensity histogram-related features are adopted. For each local area, it is also important to monitor the direction of folds. Co-occurrence matrix-related features and Gabor Wavelet-based features are representative features for representing their characteristics. Since the mucosal surfaces are smooth in patients not infected with *H. pylori* and coarse in *H. pylori*-infected patients, the difference is extracted by using ALBP.

### Shape Features

- Hu-moment invariants [39] (7 dimensions)  
The proposed method first extracts edges from gastric X-ray images by using the Canny filter [44]. Then the proposed method calculates the scale- and rotation-invariant Hu-moment features from the obtained binary image.
- Hough transform-based features (2 dimensions)  
Similar to the Hu-moment invariants, the proposed method first extracts edges using the Canny filter and obtains a binary image. The proposed method then applies the Hough transform [40] to the binary image and obtains the peak values for  $\rho$  and  $\theta$  in the  $\rho$ - $\theta$  space.

For monitoring the fold distributions in both the lesser curvature and greater curvature, features representing the global direction of folds within the stomach are necessary. Therefore, in the proposed method, we introduce Hu-moment invariants and Hough transform-based features for finding such global directions.

### Gradient-based Features

- Histogram of Oriented Gradient (HOG) features [41] (1296 dimensions)

The proposed method extracts HOG features that are robust to local geometric and brightness changes, as described in [41]. The proposed method uses a block size of  $\frac{X}{2} \times \frac{Y}{2}$  pixels and cell size of  $\frac{X}{6} \times \frac{Y}{6}$  pixels, *i.e.*, each block is composed of  $3 \times 3$  cells. The blocks are obtained at intervals of  $\frac{X}{6} \times \frac{Y}{6}$  pixels (same as the cell size), and 16 blocks are totally extracted from the image. Furthermore, for each block, a 9-bin gradient histogram is calculated from 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 [42]-based edge features (1 dimension)

The proposed method calculates the sum of the gradients obtained by applying the Sobel filter to the gastric X-ray image.

Since HOG features can detect an object's shape by using a gradient histogram, we expect that it can represent the gastric fold shape. Furthermore, for monitoring the mucosal surface, it is important to represent intensity differences between neighboring pixels. Therefore, Sobel filter-based edge features are useful.

The above features other than HOG features are calculated for each block, the blocks having been obtained by dividing the target image into  $8 \times 8$  regions. This means that  $101 \times 8 \times 8 = 6464$  features are obtained. Totally,  $6464 + 1296$  (HOG features) = 7760 features are obtained for each target gastric X-ray image. Then we denote each feature calculated from the X-ray images taken at the  $p$ -th ( $p = 1, 2, \dots, P_{max}$ ;  $P_{max}$  being the number of positions) position as a feature vector  $\mathbf{v}_{p,f} \in \mathbb{R}^{M_f}$  ( $f = 1, 2, \dots, F$ ;  $F$  being the number of visual features, *i.e.*,  $F = 8$ ). Note that  $M_f$  is the dimension of the  $f$ -th feature, *i.e.*,  $\sum_{f=1}^F M_f = 7760$ , and we use the X-ray images taken in eight positions ( $P_{max} = 8$ ). As shown in Fig. 1, we cannot obtain any image information from the surrounding black regions or any effective features for detecting *H. pylori* infection from outside the stomach region. Thus, we have to select features that can represent the symptoms of *H. pylori* infection and exclude features that do not have any image information.

## 2.2. Feature Selection and Discriminative Feature Extraction

In the previous subsection, we showed how to obtain feature vectors  $\mathbf{v}_{p,f}$  that have  $M_f$  dimensions from each X-ray image. In the training stage, we assume the label  $y \in \{1, -1\}$  that represents the presence or absence of *H. pylori* infection is given for each patient. Then, from the  $M_f$  features, the proposed method performs feature selection based on the mRMR algorithm [31] to select  $m_{p,f} (< M_f)$  features. This feature selection algorithm considers the redundancy of the features and relevance of the features to the label to select  $m_{p,f}$  features. It should be noted that the optimal dimension of each visual feature shown in Table 3 is different depending on the position. The redundancy and the relevance are calculated on the basis of mutual information, and the mRMR algorithm selects the feature for which the value obtained by subtracting the redundancy from relevance is high. Thus, the proposed method eliminates redundant features

that cause performance degradation of the following classification procedures in *H. pylori* detection and enables the use of features that are strongly connected to *H. pylori* infection.

Unlike the classification method using endoscopy images, a region of interest (ROI) is generally not provided in X-ray images. In addition, the important region is generally different for each feature since symptoms of *H. pylori* infection can be found in several parts of the stomach. From this point, we must not only calculate suitable features but also find which parts should be used for the classification. Therefore, in the proposed method, we calculate the features that have location information, *i.e.* blocks, and perform optimal feature selection based on the mRMR algorithm to obtain optimal feature values in the optimal locations.

We also use KLFDA [32] for the selected features to obtain the features that have more discriminative power for *H. pylori* infection. KLFDA introduces the idea of Locality-Preserving Projection (LPP) [45] to Kernel Discriminant Analysis (KDA) [46]. KDA and LPP both calculate embedding transformation such that the between-class variance is maximized, and nearby data pairs in the original space are close in the embedding space. Therefore, we can obtain more discriminative features based on KLFDA. After performing KLFDA, we define a new feature vector  $\mathbf{x}_{p,f} \in \mathbb{R}^{t_{p,f}}$  that contains  $t_{p,f}$  projected features. Then, by using this non-conventional approach, we can solve the classification problem of X-ray images that does not include an ROI, and it is more difficult than the classification of endoscopy images [16].

## 3. Feature Fusion-based Detection of *H. pylori* Infection for Each Position

In this section, the classification method for X-ray images taken from each position is presented. In the proposed method, two-class classification of gastric X-ray images, *i.e.*, infected or uninfected, is performed on the basis of feature fusion. As described in Section 2, symptoms of *H. pylori* infection in X-ray images appear in various parts of the stomach. On the other hand, the accuracy of X-ray examination differs depending on the symptom focused on [30]. Therefore, we expect that feature fusion can combine the visual features with consideration of the above differences in accuracy.

In this paper, we combine each visual feature vector based on MKL, which linearly combines kernel functions calculated from the feature vectors  $\mathbf{x}_{p,f}$ . Given a training dataset consisting of  $N$  vectors  $\mathbf{x}_{p,f}^i \in \mathbb{R}^{t_{p,f}}$  ( $i = 1, 2, \dots, N$ ) and their corresponding labels  $y^i \in \{1, -1\}$ , the feature vectors  $\mathbf{x}_{p,f}^i$  are mapped into a non-linear feature space by  $\phi_{p,f}(\mathbf{x}_{p,f}^i) \in \mathbb{R}^{d_{p,f}}$  ( $d_{p,f} \gg t_{p,f}$ ). Then separation of the two classes becomes feasible through an optimal hyperplane defined by a weight factor  $\mathbf{w}_{p,f} \in \mathbb{R}^{d_{p,f}}$  and a bias  $b_p \in \mathbb{R}$ . Specifically, when we denote  $X_p$  as a set of input vectors  $\mathbf{x}_{p,f}$  ( $f = 1, 2, \dots, F$ ) whose class label is unknown, the optimal hyperplane for the  $p$ -th position is defined as

$$f_p^{\text{mkl}}(X_p) = \sum_{f=1}^F \beta_{p,f} \mathbf{w}_{p,f}^\top \phi_{p,f}(\mathbf{x}_{p,f}) + b_p, \quad (1)$$

where  $\beta_{p,f}$  corresponds to the coefficient of the kernel linear combination. MKL obtains the optimal values of  $\mathbf{w}_{p,f}$ ,  $b_p$ , and  $\beta_{p,f}$  by solving the following primal problem:

$$\begin{aligned} \min_{\mathbf{w}_{p,f}, b_p, \beta_{p,f}, \xi_{p,i}} \quad & \frac{1}{2} \sum_{f=1}^F \beta_{p,f}^2 \|\mathbf{w}_{p,f}\|^2 + C_p \sum_{i=1}^N \xi_{p,i}, \\ \text{s.t.} \quad & \forall i: y^i \left( \sum_{f=1}^F \beta_{p,f} \mathbf{w}_{p,f}^\top \phi_{p,f}(\mathbf{x}_{p,f}^i) + b_p \right) \geq 1 - \xi_{p,i}, \\ & \xi_i \geq 0, \beta_{p,f} \geq 0, \sum_{f=1}^F \beta_{p,f} = 1, \end{aligned} \quad (2)$$

where  $C_p$  is a constant that represents a trade-off between the number of misclassified samples in the training data and separation of the remaining samples with a maximum margin. Furthermore,  $\xi_{p,i}$  is the  $p$ -th position's slack variable for sample  $i$ . The slack variable represents the magnitude of error, i.e., the distance between the vector of misclassified samples and a hyperplane.

The aim of MKL is to find the optimal hyperplane that minimizes the cost function in Eq. (2) consisting of two criteria: margin maximization and error minimization. We obtain a dual problem of the primal problem in Eq. (2) by applying the Lagrange multiplier method as follows:

$$\begin{aligned} \max_{\alpha_p, \beta_{p,f}, \lambda_p} \quad & \sum_{i=1}^N \alpha_p^i - \lambda_p, \\ \text{s.t.} \quad & \sum_{i=1}^N \alpha_p^i y^i = 0, \\ & 0 \leq \alpha_p^i \leq C_p, (i = 1, 2, \dots, N), \\ & \frac{1}{2} \sum_{i,j=1}^N \alpha_p^i \alpha_p^j y^i y^j \sum_{f=1}^F \beta_{p,f} \kappa_{p,f}(\mathbf{x}_{p,f}^i, \mathbf{x}_{p,f}^j) \leq \lambda_p, \forall f, \end{aligned} \quad (3)$$

where  $\alpha_p^i$  is the Lagrange multipliers of the constraint related to the usual Support Vector Machine (SVM) [47] problem, whereas  $\lambda_p$  is associated with the constraint on  $\beta_{p,f}$ . Furthermore,  $\kappa_{p,f}(\cdot, \cdot)$  is a kernel function representing the inner product of  $\phi_{p,f}(\mathbf{x}_{p,f}^i)$  and  $\phi_{p,f}(\mathbf{x}_{p,f}^j)$ , i.e.,  $\kappa_{p,f}(\mathbf{x}_{p,f}^i, \mathbf{x}_{p,f}^j) = \langle \phi_{p,f}(\mathbf{x}_{p,f}^i), \phi_{p,f}(\mathbf{x}_{p,f}^j) \rangle$ . Using the Lagrangian formulation, Eq. (1) is rewritten as follows:

$$f_p^{\text{mkl}}(X_p) = \sum_{i=1}^N \alpha_p^i y^i \sum_{f=1}^F \beta_{p,f} \kappa_{p,f}(\mathbf{x}_{p,f}^i, \mathbf{x}_{p,f}) + b_p. \quad (4)$$

Finally, the class label  $y_p$  of  $X_p$  is determined according to  $y_p = \text{sign}[f_p^{\text{mkl}}(X_p)]$ . If  $y_p$  is positive,  $X_p$  belongs to the positive class whose label is 1. Otherwise, it belongs to the negative class of -1.

In the above procedures, we can detect *H. pylori* infection from gastric X-ray images taken from each position. MKL can detect *H. pylori* infection accurately since it considers the importance of each symptom of *H. pylori* infection. Specifically, as the values of  $\beta_{p,f}$  in Eq. (1) become larger, the

importance of the corresponding visual features also becomes greater, and the difference in the importance is thus represented by the difference in the values of  $\beta_{p,f}$ . It should be noted that multiple gastric X-ray images are generally taken from several ( $=P_{\max}$ ) positions when performing diagnosis of gastric X-ray images. Therefore, we can obtain  $P_{\max}$  results by the classifiers  $f_p^{\text{mkl}}(X_p)$  ( $p = 1, 2, \dots, P_{\max}$ ) shown in Eq. (4).

#### 4. Confidence-based Decision Fusion

In this section, we show the procedures for integration of the multiple detection results obtained from the classifiers shown in the previous subsection. In gastric X-ray examination, doctors usually focus on a specific position in which symptoms of *H. pylori* tend to appear. In other words, the performance of gastric X-ray examination is different depending on the positions of X-ray images. However, doctors cannot always perform examination from X-ray images taken from specific positions due to the barium reservoir and duodenum overlapping. In such cases, doctors examine X-ray images taken from other positions, and these images are evaluated comprehensively.

In order to realize evaluation as performed by doctors, we have to introduce confidence-based decision fusion, which can consider how confidently each classifier obtains detection results. Therefore, we newly introduce the decision fusion in [33], which can combine detection results using the relationship between the classifier's performance and the confidence score.

First, the classifier's performance  $P_p^{\text{se}}$  (sensitivity) and  $P_p^{\text{sp}}$  (specificity) of the  $p$ -th classifier are defined as follows:

$$P_p^{\text{se}} = P[y_p = 1 | y = 1], \quad (5)$$

$$P_p^{\text{sp}} = P[y_p = 0 | y = 0], \quad (6)$$

where the classification result  $y_p \in \{1, 0\}$  is obtained by rewriting the class label  $\{1, -1\}$  in the previous section, and  $y \in \{1, 0\}$  is the true class label. Secondly, a classification model, i.e., a linear discriminating function  $f_{\mathbf{w}}$ , is defined as follows:

$$f_{\mathbf{w}}(\mathbf{z}) = \mathbf{w}^\top \mathbf{z}, \quad (7)$$

where  $\mathbf{z} \in \mathbb{R}^{P_{\max}+1}$  represents the integrating  $P_{\max}$  confidence scores  $f_p^{\text{mkl}}(X_p)$  and a bias, i.e.,  $\mathbf{z} = [1, f_1^{\text{mkl}}(X_1), f_2^{\text{mkl}}(X_2), \dots, f_{P_{\max}}^{\text{mkl}}(X_{P_{\max}})]^\top$ , and  $\mathbf{w} \in \mathbb{R}^{P_{\max}+1}$  is a weight vector that is calculated from the training dataset. Furthermore, we define the posterior probability for a positive class using  $f_{\mathbf{w}}(\mathbf{z})$  as follows:

$$\begin{aligned} P[y = 1 | \mathbf{z}, \mathbf{w}] &= \sigma(f_{\mathbf{w}}(\mathbf{z})) \\ &= \frac{1}{1 + \exp(-\mathbf{w}^\top \mathbf{z})}. \end{aligned} \quad (8)$$

Next, we define a training dataset as  $\mathcal{D} = \{\mathbf{z}^i, y_1^i, y_2^i, \dots, y_{P_{\max}}^i\}_{i=1}^N$  consisting of  $N$  samples' feature vectors  $\mathbf{z}^i$  and  $P_{\max}$  classification results  $(y_1^i, y_2^i, \dots, y_{P_{\max}}^i)$ , where  $\mathbf{z}^i \in \mathbb{R}^{P_{\max}+1}$  corresponds to  $\mathbf{z}$ . Note that the feature vector  $\mathbf{z}^i$ , the detection results of training data  $y_p^i \in \{1, 0\}$ , and

the performance, *i.e.*,  $P_p^{\text{se}}$  and  $P_p^{\text{sp}}$ , are obtained by 10-fold cross-validation of training data. The aim of this decision fusion is calculation of the optimal weight vector  $\hat{\mathbf{w}}$  in Eq. (7), and it is found by maximizing the following log-likelihood using the training dataset  $\mathcal{D}$ ,  $P_p^{\text{se}}$  and  $P_p^{\text{sp}}$ .

$$\hat{\mathbf{w}} = \arg \max_{\mathbf{w}} \{\ln P[\mathcal{D}|\mathbf{w}]\}, \quad (9)$$

where

$$P[\mathcal{D}|\mathbf{w}] = \prod_{i=1}^N [A^i \sigma(\mathbf{w}^\top \mathbf{z}^i) + B^i \{1 - \sigma(\mathbf{w}^\top \mathbf{z}^i)\}], \quad (10)$$

$$A^i = \prod_{p=1}^{P_{\max}} (P_p^{\text{se}})^{y_p^i} (1 - P_p^{\text{se}})^{(1-y_p^i)}, \quad (11)$$

$$B^i = \prod_{p=1}^{P_{\max}} (1 - P_p^{\text{sp}})^{y_p^i} (P_p^{\text{sp}})^{(1-y_p^i)}. \quad (12)$$

In [33], the log-likelihood is maximized by an unsupervised learning approach. On the other hand, since the ground truth is available in the proposed method, we perform reformulation of the target problem. Given the true labels of the training data  $\mathbf{y} = [y^1, y^2, \dots, y^N]^\top$ , the complete data log-likelihood can be written as follows:

$$\ln P[\mathcal{D}, \mathbf{y}|\mathbf{w}] = \sum_{i=1}^N [y^i \ln \{\sigma(\mathbf{w}^\top \mathbf{z}^i) A^i\} + (1 - y^i) \ln \{(1 - \sigma(\mathbf{w}^\top \mathbf{z}^i)) B^i\}]. \quad (13)$$

In order to find the optimal weight vector  $\hat{\mathbf{w}}$ , we maximize this likelihood by using the Newton-Raphson method [48]. Let  $\mathbf{g}(\mathbf{w})$  and  $\mathbf{H}(\mathbf{w})$  be the gradient vector and the Hessian matrix, respectively. The weight vector  $\mathbf{w}$  is calculated by iterating the following equations until convergence:

$$\mathbf{w} \leftarrow \mathbf{w} - \eta \mathbf{H}^{-1}(\mathbf{w}) \mathbf{g}(\mathbf{w}), \quad (14)$$

$$\mathbf{g}(\mathbf{w}) = \sum_{i=1}^N \{y^i - \sigma(\mathbf{w}^\top \mathbf{z}^i)\} \mathbf{z}^i, \quad (15)$$

$$\mathbf{H}(\mathbf{w}) = - \sum_{i=1}^N \sigma(\mathbf{w}^\top \mathbf{z}^i) \{1 - \sigma(\mathbf{w}^\top \mathbf{z}^i)\} \mathbf{z}^i (\mathbf{z}^i)^\top, \quad (16)$$

where  $\eta$  is a step length.

Consequently, we can obtain the optimal weight vector  $\hat{\mathbf{w}}$ . Then the final detection label  $\hat{y}$  of test data is obtained by the following equations:

$$\hat{y} = \begin{cases} 1 & \text{if } \mu > T \\ 0 & \text{otherwise,} \end{cases} \quad (17)$$

where

$$\mu = \frac{A \sigma(\hat{\mathbf{w}}^\top \mathbf{z})}{A \sigma(\hat{\mathbf{w}}^\top \mathbf{z}) + B \{1 - \sigma(\hat{\mathbf{w}}^\top \mathbf{z})\}}, \quad (18)$$

$$A = \prod_{p=1}^{P_{\max}} (P_p^{\text{se}})^{y_p} (1 - P_p^{\text{se}})^{(1-y_p)}, \quad (19)$$

$$B = \prod_{p=1}^{P_{\max}} (1 - P_p^{\text{sp}})^{y_p} (P_p^{\text{sp}})^{(1-y_p)}, \quad (20)$$

where  $T$  is a predefined threshold.

As shown in the above procedures, we can perform decision fusion based on the collaborative use of multiple detection results. Decision fusion can consider not only the classifiers' performance but also the relationship between confidence scores and detection results. Specifically,  $A$  in Eq. (19) and  $B$  in Eq. (20) represent the classifiers' performance. Furthermore, since the weight vector  $\mathbf{w}$  in Eq. (7) is calculated by using the confidence vector  $\mathbf{z}^i$  and the performance  $A^i$  and  $B^i$ , we can consider how confidently each classifier obtains the detection results. Consequently, the proposed method can accurately detect *H. pylori* infection even if the classification results of some images are not correct.

## 5. Experimental Results

The effectiveness of the proposed method was evaluated by performing experiments using gastric X-ray images taken from individuals infected with *H. pylori* and uninfected individuals. An outline of the experiment is given in Section 5.1, and the obtained results are presented and discussed in Section 5.2.

### 5.1. Experimental Conditions

In the experiment, we used 2100 samples from composed of 1056 male samples and 1044 female samples. Each sample consisted of the ABCD classification result and gastric X-ray images taken in eight positions ( $P_{\max} = 8$ ). Therefore, we used 2100×8 gastric X-ray images in the experiment. The numbers of samples in age brackets were as follows: twenties, 3; thirties, 329; forties, 631; fifties, 733; sixties, 374; and seventies, 30. The ground truth for this experiment was obtained from the ABCD classification results, which strongly correlate to gastric cancer risk as shown in Table 1. Consequently, a sample was considered to be uninfected, *i.e.*, category A, if all of the following three conditions were satisfied: Hp antibody level < 10 U/ml, PGI > 70 ng/ml, and PGI/II ratio  $\geq 3$ . Otherwise, the target sample was considered to be infected, *i.e.*, category B, C or D.

In this experiment, we regarded 1129 samples in category A as uninfected samples (negative samples) and 971 samples in categories B, C and D as infected samples (positive samples). To avoid the degradation of the detection performance due to misclassification of ABCD classification, we prepared the samples by removing samples suspected as being false negative in the ABCD classification. We regarded samples in category A with PGI levels  $\leq 30$  ng/ml, PGII levels  $\geq 15$  ng/ml or PGI/PGII < 4 as suspected false negative samples. In order to confirm the effectiveness of feature extraction based on KLFDA by the mRMR algorithm, MKL-based feature fusion, and decision fusion, the following comparative methods were also evaluated. We denoted the experiments for confirming their effectiveness as experiments 1, 2 and 3.

#### • Experiment 1 (feature selection and dimensionality reduction)



Table 4: Overview of the comparative methods and the proposed method. The tick marks indicate the approaches introduced in each method.

Comparative methods	Feature Selection and Dimensionality Reduction					Classifier			Decision Fusion	
	mRMR [31]	KPCA [49]	KLFDA [32]	mRMR+KPCA	mRMR+KLFDA	SVM [47]	SCW [50]	MKL [27]	Majority voting [33]	
1	×	×	×	×	×	×	×	✓	×	✓
2	✓	×	×	×	×	×	×	✓	×	✓
3	×	✓	×	×	×	×	×	✓	×	✓
4	×	×	✓	×	×	×	×	✓	×	✓
5	×	×	×	✓	×	×	×	✓	×	✓
6	×	×	×	×	✓	✓	×	×	×	✓
7	×	×	×	×	✓	×	✓	×	×	✓
8	×	×	×	×	✓	×	×	✓	×	×
9	×	×	×	×	✓	×	×	✓	✓	×
Proposed Method	×	×	×	×	✓	×	×	✓	×	✓

- (1) Comparative method 1 does not use the mRMR algorithm and KLFDA, *i.e.*, this method uses all of the features.
- (2) Comparative method 2 uses mRMR-based feature selection only.
- (3) Comparative method 3 uses only Kernel Principal Component Analysis (KPCA) [49], which is a common dimensionality reduction method that does not use class information, *i.e.*, an unsupervised method.
- (4) Comparative method 4 uses only KLFDA.
- (5) Comparative method 5 uses KPCA after exclusion of unnecessary or undesirable visual information by the mRMR algorithm.

#### • Experiment 2 (classifier)

- (6) Comparative method 6 trains the SVM classifier using feature vectors obtained by performing mRMR and KLFDA. Note that the feature vectors using in this method are obtained by concatenating  $\mathbf{v}_{p,1}, \mathbf{v}_{p,2}, \dots, \mathbf{v}_{p,F}$ . Furthermore, the feature vector  $\mathbf{z}$  is composed of SVM's discriminant function, which is similar to  $f_p^{\text{mkl}}(\cdot)$ .
- (7) Comparative method 7 uses the Soft Confidence-Weighted (SCW) learning scheme [50], which is an extension of Confidence-Weighted learning [51] for soft margin learning. In this method, feature vectors are obtained in the same manner as that for comparison method 6. Furthermore, the feature vector  $\mathbf{z}$  is composed of SCW's discriminant function.

#### • Experiment 3 (decision fusion)

- (8) Comparative method 8 performs detection of *H. pylori* infection using an image obtained from only one position, *i.e.*, decision fusion is not used. Note that we select the position for which the classifier has the best performance.
- (9) Comparative method 9 combines multiple detection results by using majority voting. The final detection result becomes *H. pylori* infection when more than  $T_{\text{mv}}$  classifiers detect *H. pylori* infection, where  $T_{\text{mv}}$  is a predefined threshold.

An overview of the above comparative methods and the proposed method is shown in Table 4.

The dimensions of the features after mRMR, KPCA and KLFDA, *i.e.*,  $m_{p,f}$  and  $t_{p,f}$ , were set to values that output the best results in our method and the comparative methods. The Gaussian kernel was used for the kernel function in KPCA and KLFDA, with the best kernel parameter. On the other hand, the linear kernel was used for MKL since the feature vectors that are obtained on the basis of KPCA and KLFDA have been calculated in a non-linear feature space. Comparative methods 1 and 2 employed the Gaussian kernel for MKL since the features of these methods are in a linear space before applying MKL.

We used 10-fold cross-validation for the model evaluation method. Then we used sensitivity (Sen), specificity (Spe), harmonic mean (HM) of sensitivity and specificity, and area under the ROC curve (area under the curve, AUC) as evaluation criteria. In the ROC curve, Sen and 1-Spe, *i.e.*, false positive rate, are plotted on the vertical axis and the horizontal axis, respectively. The accuracy of the ROC curve is determined on the basis of the AUC, and AUC= 1 represents the theoretical maximum accuracy. Furthermore, we applied Welch's t-test as a significant difference test to HM obtained by 10-fold cross-validation.

#### 5.2. Results and Discussion

Figure 2 shows the ROC curves obtained by changing the threshold  $T$  and  $T_{\text{mv}}$ . Table 5 shows the values of Sen, Spe, HM, and AUC in experiments 1, 2, and 3. Note that Se, Spe, and HM were obtained at the optimal threshold, which was set to the value that outputs the nearest point to (0, 1) on the ROC curve. The point is shown in Fig. 2 as +. Table 6 shows Se and Spe of each position's classifier.

First, we verify the results of experiment 1. As shown in Fig. 2 (a), the proposed method outperforms the comparative methods, and as shown in Table 5, sensitivity, harmonic mean, and AUC of the proposed method are higher than those of comparative methods 1-5. A comparison of comparative method 1 and the other methods showed the effectiveness of feature selection and dimensionally reduction. Furthermore, since the proposed method and comparative method 5 both outperform comparative methods 3 and 4, the elimination of unnecessary or

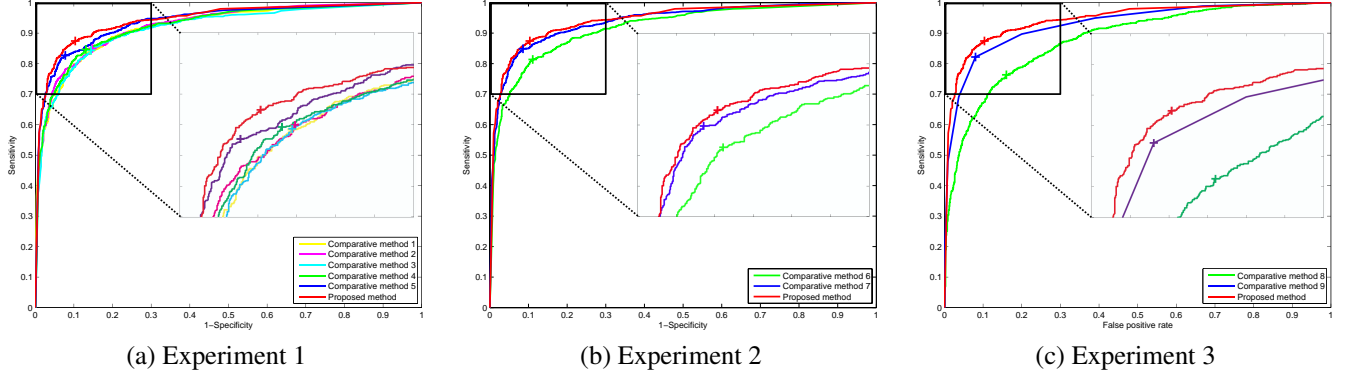


Figure 2: ROC curves for detection of *H. pylori* infection by the proposed method and the comparative methods: (a), (b), and (c) were obtained by experiments 1, 2 and 3, respectively. The point shows the best result at the optimal threshold.

Table 5: Results of sensitivity, specificity, harmonic mean of sensitivity and specificity, and AUC for the proposed method and the comparative methods.

	Methods	Sen	Spe	HM	AUC
	<b>Proposed method</b>	<b>0.874</b>	0.896	<b>0.885</b>	<b>0.944</b>
<b>Experiment 1</b>	Comparative method 1	0.843	0.857	0.850	0.921
	Comparative method 2	0.850	0.852	0.851	0.925
	Comparative method 3	0.844	0.856	0.850	0.916
	Comparative method 4	0.847	0.869	0.858	0.923
	Comparative method 5	0.823	<b>0.922</b>	0.870	0.937
<b>Experiment 2</b>	Comparative method 6	0.813	0.889	0.849	0.918
	Comparative method 7	0.849	0.914	0.880	0.939
<b>Experiment 3</b>	Comparative method 8	0.763	0.836	0.798	0.879
	Comparative method 9	0.822	0.920	0.868	0.930

Table 6: Results of sensitivity, specificity, and harmonic mean of sensitivity and specificity for each of the position's classifier in experiment 2.

Position $p$	Proposed method			Comparative method 6			Comparative method 7		
	Sen	Spe	HM	Sen	Spe	HM	Sen	Spe	HM
1	<b>0.763</b>	<b>0.835</b>	<b>0.797</b>	0.698	0.814	0.752	0.749	0.819	0.782
2	0.760	<b>0.834</b>	<b>0.795</b>	0.696	0.815	0.751	<b>0.779</b>	0.804	0.791
3	0.744	<b>0.825</b>	<b>0.782</b>	0.715	0.796	0.753	<b>0.771</b>	0.780	0.775
4	0.750	<b>0.839</b>	<b>0.792</b>	0.708	0.808	0.755	<b>0.784</b>	0.795	0.789
5	0.744	<b>0.803</b>	0.772	0.713	0.810	0.758	<b>0.792</b>	<b>0.773</b>	<b>0.782</b>
6	0.638	<b>0.770</b>	0.698	0.620	0.763	0.684	<b>0.704</b>	0.713	<b>0.708</b>
7	0.715	<b>0.799</b>	<b>0.755</b>	0.684	0.762	0.721	<b>0.740</b>	0.733	0.736
8	0.687	<b>0.744</b>	<b>0.714</b>	0.614	0.732	0.668	<b>0.743</b>	0.683	0.712
Average	0.725	<b>0.806</b>	<b>0.763</b>	0.681	0.788	0.730	<b>0.758</b>	0.763	0.760

undesirable visual features by the mRMR algorithm is effective before calculating discriminative features. As evidence that the proposed method can exclude such visual features, Fig. 3 shows the locations selected with feature values by the mRMR algorithm. Note that Fig. 3 only shows the results for one position from which the X-ray images in Fig. 1 are taken. From Figs. 1 and 3, it is obvious that the features selected by the mRMR algorithm are generally extracted from gastric regions. Therefore, we can obtain effective features for detection of *H. pylori* infection by the mRMR algorithm. As a similar consideration, the use of not only feature selection but also discriminative feature extraction is desirable for accurate detection of *H. pylori* infection since the proposed method and comparative method 5 outperform comparative method 2. Furthermore, the proposed method outperforms comparative method 5, and these methods respectively use KLFDA and KPCA. KLFDA can obtain more discriminative features since KLFDA considers class information, i.e.,  $y^i$ . For a similar reason, the results of comparative method 4 are better than those of comparative method 3. From the aforementioned results, the effectiveness of the mRMR algorithm and KLFDA is verified.

Next, we verify the results of experiment 2. From the results shown in Fig. 2 (b) and Table 5, the proposed method outperforms comparative methods 6 and 7. When we focus on the harmonic mean of each position's classifier in Table 6, there is

an average differences of 3.3% between the proposed method and comparative method 6. There is also an average difference of 0.3% between the proposed method and comparative method 7. Although MKL is a classifier similar to SVM, there are the above differences in performance since SVM does not consider integration of the features. Comparative method 7 is inferior to the proposed method for a reason similar to that for SVM. In order to confirm that MKL can consider the importance of each symptom, we consider the relationship between the value of  $\beta_{p,f}$  and the symptom focused on by each visual feature. For example, in one position from which the X-ray images in Fig. 1 are taken, the values of  $\beta_{p,f}$  assigned to ALBP and Gabor features are 0.2807 and 0.2556, respectively, and these values are higher than those of the other features. Figure 4 also shows the locations selected from ALBP and Gabor features by the mRMR algorithm in order to verify the symptom focused on by each visual feature. As shown in Fig. 4, the locations selected by the mRMR algorithm are different between ALBP and Gabor. When we collate Fig. 4 with Fig. 1, it is obvious that ALBP and Gabor features represent the mucosal surface pattern and the fold shapes, respectively. Therefore, since the symptoms of *H. pylori* infection that are focused on are different according to the visual feature, MKL can represent the importance of each symptom by the difference of  $\beta_{p,f}$ . Consequently, the proposed method improves the performance of each position's classifier,

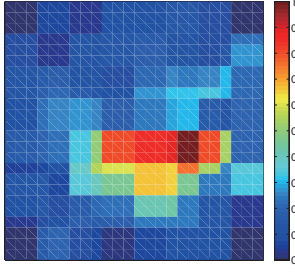


Figure 3: Locations selected with feature values by the mRMR algorithm from all features shown in Table 3. The frequencies with which the regions are selected by mRMR were scaled to the range [0, 1].

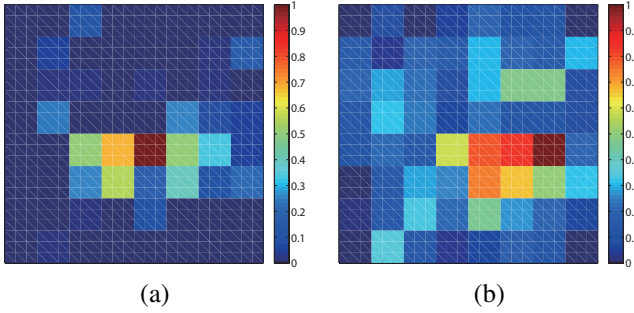


Figure 4: Locations selected with feature values by the mRMR algorithm: (a) and (b) were obtained from ALBP and Gabor features, respectively. The frequencies with which the regions are selected by mRMR were scaled to the range [0, 1].

and the final detection results are more accurate than those of the comparative methods.

We next verify the results of experiment 3. From the results shown in Fig. 2 (c) and Table 5, the proposed method outperforms comparative methods 8 and 9. From a comparison of the proposed method and comparative method 9, we can confirm the effectiveness of considering the differences in the classifiers' performance since majority voting treats all classifiers' performance as equal. Furthermore, since the results of comparative method 8, which uses images taken from only one position, *i.e.*, decision fusion is not used, are inferior to those of the other methods in experiment 3, comprehensive evaluation of X-ray images taken from several positions is important. The above-described results indicate that it is important to integrate multiple results for accurate detection of *H. pylori* infection and to consider each classifier's performance in decision fusion.

Finally, we show the results of Welch's t-test in Table 7. As shown in Table 7, we confirmed that the harmonic mean of the proposed method was statistically higher than those of comparative methods 1, 2, 3, 4, 6 and 8 at the 5 percent level. Although we could not confirm a significant difference between the proposed method and comparative methods 5, 7 and 9, these p-values did not show that the proposed method is inferior to these comparative methods. We should propose a new method that can show a significant difference from all comparative methods as a future work.

Table 7: Results of Welch's t-test whose significant level is set as  $\alpha = 0.05$ .

	Methods	p value
Experiment 1	Comparative method 1	$p < 0.05$
	Comparative method 2	$p < 0.05$
	Comparative method 3	$p < 0.05$
	Comparative method 4	$p < 0.05$
	Comparative method 5	$p = 0.16$
Experiment 2	Comparative method 6	$p < 0.05$
	Comparative method 7	$p = 0.34$
Experiment 3	Comparative method 8	$p < 0.05$
	Comparative method 9	$p = 0.061$

## 6. Conclusions

A fully automatic method for detection of *H. pylori* infection from gastric X-ray images is presented in this paper. We focus on the unique characteristic of gastric X-ray examination. The accuracy of X-ray examination is influenced by the symptoms of *H. pylori* infection focused on and the positions from which X-ray images are taken. From this characteristic, we employ MKL and decision fusion: MKL can weight each visual feature considering its importance, and decision fusion can consider which classifier leads to good detection results from the relationship between the classifier's performance and the confidence score obtained in MKL-based classification. Finally, the effectiveness of the proposed method was evaluated by performing experiments using actual gastric X-ray images from patients with and those without *H. pylori* infection. In a future work, we plan to investigate automatic extraction of the ROI and to improve the detection performance by re-examining the features.

## Acknowledgements

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, Shigemitsu Nakajima of Japan Community Health care Organization 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 JSPS KAKENHI Grant Number JP25280036.

## References

- [1] N. Uemura, S. Okamoto, S. Yamamoto, N. Matsumura, S. Yamaguchi, M. Yamakido, K. Taniyama, N. Sasaki, R. J. Schlemper, Helicobacter pylori infection and the development of gastric cancer, New England Journal of Medicine 345 (11) (2001) 784–789.
- [2] H. Ohata, S. Kitauchi, N. Yoshimura, K. Mugitani, M. Iwane, H. Nakamura, A. Yoshikawa, K. Yanaoka, K. Arai, H. Tamai, Y. Shimizu, T. Takeshita, O. Mohara, M. Ichinose, Progression of chronic atrophic gastritis associated with helicobacter pylori infection increases risk of gastric cancer, Int. Journal of Cancer 109 (1) (2004) 138–143.

- [3] H. Watabe, T. Mitsushima, Y. Yamaji, M. Okamoto, R. Wada, T. Kokubo, H. Doi, H. Yoshida, T. Kawabe, M. Omata, Predicting the development of gastric cancer from combining helicobacter pylori antibodies and serum pepsinogen status: a prospective endoscopic cohort study, *Gut* 54 (6) (2005) 764–768.
- [4] J. Parsonnet, G. D. Friedman, D. P. Vandersteen, Y. Chang, J. H. Vogelmann, N. Orentreich, R. K. Sibley, Helicobacter pylori infection and the risk of gastric carcinoma, *New England Journal of Medicine* 325 (16) (1991) 1127–1131.
- [5] J. H. Walsh, W. L. Peterson, The treatment of helicobacter pylori infection in the management of peptic ulcer disease, *New England Journal of Medicine* 333 (15) (1995) 984–991.
- [6] K. Fukase, M. Kato, S. Kikuchi, K. Inoue, N. Uemura, S. Okamoto, S. Terao, K. Amagai, S. Hayashi, M. Asaka, Effect of eradication of helicobacter pylori on incidence of metachronous gastric carcinoma after endoscopic resection of early gastric cancer: an open-label, randomized controlled trial, *The Lancet* 372 (9636) (2008) 392–397.
- [7] L. Fuccio, R. M. Zagari, L. H. Eusebi, L. Laterza, V. Cennamo, L. Ceroni, D. Grilli, F. Bazzoli, Meta-analysis: Can helicobacter pylori eradication treatment reduce the risk for gastric cancer?, *Annals of Internal Medicine* 151 (2) (2009) 121–128.
- [8] E. M. El-Omar, I. D. Penman, J. E. Ardi, R. S. Chittajallu, C. Howie, K. E. McColl, Helicobacter pylori infection and abnormalities of acid secretion in patients with duodenal ulcer disease, *Gastroenterology* 109 (3) (1995) 681–691.
- [9] K. E. McColl, A. El-Nujumi, L. Murray, E. El-Omar, D. Gillen, A. Dickson, A. Kelman, T. E. Hilditch, The helicobacter pylori breath test: A surrogate marker for peptic ulcer disease in dyspeptic patients, *Gut* 40 (3) (1997) 302–306.
- [10] T. Kudo, S. Kakizaki, N. Soharu, Y. Onozato, S. Okamura, Y. Inui, M. Mori, Analysis of abc (d) stratification for screening patients with gastric cancer, *World Journal of Gastroenterology* 17 (43) (2011) 4793–4798.
- [11] K. Miki, Gastric cancer screening by combined assay for serum anti-Helicobacter pylori IgG antibody and serum pepsinogen levels – “ABC method”, in: *Proc. the Japan Academy Series B, Physical and Biological Sciences*, 2011, pp. 405–414.
- [12] S. Mizuno, I. Miki, T. Ishida, M. Yoshida, M. Onoyama, T. Azuma, Y. Habu, H. Inokuchi, K. Ozasa, K. Miki, Y. Watanabe, Prescreening of a high-risk group for gastric cancer by serologically determined helicobacter pylori infection and atrophic gastritis, *Digestive Diseases and Sciences* 55 (11) (2010) 3132–3137.
- [13] M. Asaka, D. Graham, Strategy for eliminating gastric cancer in japan, *Helicobacter* 15 (6) (2010) 486–490.
- [14] M. Yamaoka, S. Nakajima, Barium X-ray examination is useful for gastric cancer screening and for identifying those at both high and low risk for gastric cancer, *Journal of gastroenterological cancer screening* 49 (1) (2011) 20–31, (in Japanese).
- [15] K. Iijima, T. Koike, Y. Abe, N. Ara, K. Uno, A. Imatani, S. Ohara, T. Shimosegawa, Alteration of correlation between serum pepsinogen concentrations and gastric acid secretion after h. pylori eradication, *Journal of Gastroenterology* 44 (8) (2009) 819–825.
- [16] C.-R. Huang, P.-C. Chung, B.-S. Sheu, H.-J. Kuo, P. Mikulas, Helicobacter pylori-related gastric histology classification using support-vector-machine-based feature selection, *IEEE Trans. Information Technology in Biomedicine* 12 (4) (2008) 523–531.
- [17] K. Sugano, Screening of gastric cancer in asia, *Best Practice & Research Clinical Gastroenterology* 29 (6) (2015) 895–905.
- [18] M. Kato, M. Asaka, Recent development of gastric cancer prevention, *Japanese Journal of Clinical Oncology* 42 (11) (2012) 987–994.
- [19] M. Horiguchi, D. Yamauchi, T. Tsukada, N. Takaya, J. Takaya, M. Takaya, The efficiency of gastric X-ray examination for gastric cancer screening : Gastric X-rays image diagnosis of normal gastric mucosa without helicobacter pylori infection, *Health evaluation and promotion* 33 (5) (2006) 510–516, (in Japanese).
- [20] S. Nakajima, M. Yamaoka, K. Doi, M. Nishimura, A. Matsui, J. Sato, M. Shimada, Y. Okumura, A new gastric cancer screening method by barium X-ray examination with diagnosing helicobacter pylori infection, *Journal of gastroenterological cancer screening* 46 (4) (2008) 461–471, (in Japanese).
- [21] M. Niemeijer, M. Loog, M. Abramoff, M. Viergever, M. Prokop, B. van Ginneken, On combining computer-aided detection systems, *IEEE Trans. Medical Imaging* 30 (2) (2011) 215–223.
- [22] X. Gao, Y. Wang, X. Li, D. Tao, On combining morphological component analysis and concentric morphology model for mammographic mass detection, *IEEE Trans. Information Technology in Biomedicine* 14 (2) (2010) 266–273.
- [23] L. Hogeweg, C. I. Sanchez, P. Maduskar, R. Philipsen, A. Story, R. Dawson, G. Theron, K. Dheda, L. Peters-Bax, B. van Ginneken, Automatic detection of tuberculosis in chest radiographs using a combination of textural, focal, and shape abnormality analysis, *IEEE Trans. Medical Imaging* 34 (12) (2015) 2429–2442.
- [24] K. Abe, H. Nakagawa, M. Minami, H. Tian, Features for discriminating normal cases in mass screening for gastric cancer with double contrast X-ray images of stomach, *Journal of Biomedical Engineering and Medical Imaging* 1 (6).
- [25] U. G. Mangai, S. Samanta, S. Das, P. R. Chowdhury, A survey of decision fusion and feature fusion strategies for pattern classification, *IETE Technical Review* 27 (4) (2010) 293–307.
- [26] S. Bucak, R. Jin, A. Jain, Multiple kernel learning for visual object recognition: A review, *IEEE Trans. Pattern Analysis and Machine Intelligence* 36 (7) (2014) 1354–1369.
- [27] A. Vedaldi, V. Gulshan, M. Varma, A. Zisserman, Multiple kernels for object detection, in: *Proc. IEEE Int. Conf. Computer Vision*, 2009, pp. 606–613.
- [28] M. R. Alam, M. Bennamoun, R. Togneri, F. Soheli, A confidence-based late fusion framework for audio-visual biometric identification, *Pattern Recognition Letters* 52 (0) (2015) 65–71.
- [29] B. Topcu, H. Erdogan, Decision fusion for patch-based face recognition, in: *Proc. Int. Conf. Pattern Recognition*, 2010, pp. 1348–1351.
- [30] S. Nakajima, M. Yamaoka, K. Doi, M. Nishimura, Barium X-ray characteristics of the stomach with or without h. pylori infection and the diagnostic value, *Japanese Journal of Helicobacter Research* 8 (2) (2007) 18–21, (in Japanese).
- [31] H. Peng, F. Long, C. Ding, Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy, *IEEE Trans. Pattern Analysis and Machine Intelligence* 27 (8) (2005) 1226–1238.
- [32] M. Sugiyama, Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis, *Journal of Machine Learning Research* 8 (2007) 1027–1061.
- [33] V. C. Raykar, S. Yu, L. H. Zhao, A. Jerebko, C. Florin, G. H. Valadez, L. Bogoni, L. Moy, Supervised learning from multiple experts: Whom to trust when everyone lies a bit, in: *Proc. the 26th Annual Int. Conf. Machine Learning*, 2009, pp. 889–896.
- [34] K. Ishihara, T. Ogawa, M. Haseyama, Helicobacter pylori infection detection from multiple X-ray images based on decision level fusion, in: *Proc. IEEE Int. Conf. Image Processing*, 2014, pp. 2769–2773.
- [35] K. Ishihara, T. Ogawa, M. Haseyama, Helicobacter pylori infection detection from multiple X-ray images based on combination use of support vector machine and multiple kernel learning, in: *Proc. IEEE Int. Conf. Image Processing*, 2015, pp. 4728–4732.
- [36] K. Ishihara, T. Ogawa, M. Haseyama, Classification of gastric cancer risk from X-ray images based on efficient image features related to serum Hp antibody level and serum PG levels, *ITE Trans. Media Technology and Applications* 4 (4) (2016) 337–348.
- [37] R. Haralick, K. Shanmugam, I. Dinstein, Textural features for image classification, *IEEE Trans. Systems, Man and Cybernetics* 3 (6) (1973) 610–621.
- [38] C.-H. Lin, C.-W. Liu, H.-Y. Chen, Image retrieval and classification using adaptive local binary patterns based on texture features, *IET Image Processing* 6 (7) (2012) 822–830.
- [39] M.-K. Hu, Visual pattern recognition by moment invariants, *IRE Trans. Information Theory* 8 (2) (1962) 179–187.
- [40] R. O. Duda, P. E. Hart, Use of the hough transformation to detect lines and curves in pictures, *Commun. ACM* 15 (1) (1972) 11–15.
- [41] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, in: *Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition*, Vol. 1, 2005, pp. 886–893.
- [42] W. Gao, X. Zhang, L. Yang, H. Liu, An improved sobel edge detection, in: *Proc. IEEE Int. Conf. Computer Science and Information Technology*, Vol. 5, 2010, pp. 67–71.

- [43] T. Ojala, M. Pietikäinen, D. Harwood, A comparative study of texture measures with classification based on featured distributions, *Pattern Recognition* 29 (1) (1996) 51–59.
- [44] J. Canny, A computational approach to edge detection, *IEEE Trans. Pattern Analysis and Machine Intelligence* 8 (6) (1986) 679–698.
- [45] X. He, P. Niyogi, Locality preserving projections, in: *Advances in Neural Information Processing Systems* 16, MIT Press, 2004, pp. 153–160.
- [46] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, K. Muller, Fisher discriminant analysis with kernels, in: *Proc. IEEE Signal Processing Society Workshop on Neural Networks for Signal Processing IX*, 1999, pp. 41–48.
- [47] C. Cortes, V. Vapnik, Support-vector networks, *Machine Learning* 20 (3) (1995) 273–297.
- [48] A. Ben-Israel, A newton-raphson method for the solution of systems of equations, *Journal of Mathematical Analysis and Applications* 15 (2) (1966) 243–252.
- [49] B. Schölkopf, A. Smola, K.-R. Müller, Nonlinear component analysis as a kernel eigenvalue problem, *Neural Computation* 10 (5).
- [50] J. Wang, P. Zhao, S. C. Hoi, Exact soft confidence-weighted learning, in: *Proc. the Int. Conf. Machine Learning*, 2012, pp. 121–128.
- [51] M. Dredze, K. Crammer, F. Pereira, Confidence-weighted linear classification, in: *Proc. the 25th Int. Conf. Machine Learning*, 2008, pp. 264–271.