A lesion-based convolutional neural network improves endoscopic detection and depth prediction of early gastric cancer

Hong Jin Yoon

Department of Medicine

The Graduate School, Yonsei University

A lesion-based convolutional neural network improves endoscopic detection and depth prediction of early gastric cancer

Hong Jin Yoon

Department of Medicine

The Graduate School, Yonsei University

A lesion-based convolutional neural network improves endoscopic detection and depth prediction of early gastric cancer

Directed by Professor Jie-Hyun Kim

The Master's Thesis
submitted to the Department of Medicine
the Graduate School of Yonsei University
in partial fulfillment of the requirements for the degree
of Master of Medial Science

Hong Jin Yoon

December 2019

This certifies that the Master's Thesis
of Hong Jin Yoon is approved.

Thesis Supervisor : Jie-Hyun Kim

Thesis Committee Member#1 : Eun-Suk Cho

Thesis Committee Member#2 : Hyunki Kim

The Graduate School
Yonsei University

December 2019

<TABLE OF CONTENTS>

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

A Lesion-Based Convolutional Neural Network Improves Endoscopic Detection and Depth Prediction of Early Gastric Cancer

Hong Jin Yoon

Department of Medicine
The Graduate School, Yonsei University

(Directed by Professor Jie-Hyun Kim)

In early gastric cancer (EGC), tumor invasion depth is an important factor for determining the treatment method. However, as endoscopic ultrasonography has limitations when measuring the exact depth in a clinical setting as endoscopists often depend on gross findings and personal experience. The present study aimed to develop a model optimized for EGC detection and depth prediction, and we investigated factors affecting artificial intelligence (AI) diagnosis. We employed a visual geometry group(VGG)-16 model for the classification of endoscopic images as EGC (T1a or T1b) or non-EGC. To induce the model to activate EGC regions during training, we proposed a novel loss function that simultaneously measured classification and localization errors. We experimented with 11,539 endoscopic images (896 T1a-EGC, 809 T1b-EGC, and 9834 non-EGC). The areas under the curves of receiver operating characteristic curves for EGC detection and depth prediction were 0.981 and 0.851, respectively. Among the factors affecting AI prediction of tumor depth, only histologic differentiation was significantly associated, where undifferentiated-type histology exhibited a lower AI accuracy. Thus, the lesion-based model is an appropriate training method for AI in EGC. However, further improvements and validation are required, especially for undifferentiated-type histology.

Key words : early gastric cancer; artificial intelligence; convolutional neural networks; endoscopy

A Lesion-Based Convolutional Neural Network Improves Endoscopic Detection and Depth Prediction of Early Gastric Cancer

Hong Jin Yoon

Department of Medicine
The Graduate School, Yonsei University

(Directed by Professor Jie-Hyun Kim)

I. INTRODUCTION

Accurate staging is the basis for determining an appropriate treatment plan for suspected early gastric cancer (EGC) based on endoscopy or biopsy findings. As the indications for endoscopic resection (ER) and minimally invasive surgery are usually decided by the T-stage, tumor invasion depth is crucial for determining the treatment modality [1-3].

EGC is categorized as tumor invasion of the mucosa (T1a) or that of the submucosa (T1b). Endoscopic ultrasonography (EUS) is useful for T-staging of gastric cancer because it can delineate each gastric wall layer [4,5]. However, EUS is not superior to conventional endoscopy for T-staging of EGC, having a low accuracy of approximately 70% [6,7]. Therefore, there has been increasing interest in the field of medical imaging regarding modalities for predicting EGC depth.

Recently, deep learning-based artificial intelligence (AI) has shown remarkable progress across multiple medical fields. Diagnostic imaging is currently the highest and most efficient application of AI-based analyses in medical fields [8,9]. AI using endoscopic images has been applied to diagnose neoplasms in the gastrointestinal tract [10,11]. Deep convolutional neural networks (CNNs) are a type of deep learning model that are widely used for image analysis [12]. However, they differ from general image classification as the difference in EGC depth in endoscopic images is subtler and more difficult to discern. Therefore, more sophisticated image classification methods are required.

Although conditions such as easily distinguishable visual features and large-scale datasets play key roles in the performance improvements of natural image classification models, these conditions are difficult to be applied to EGC detection and EGC depth prediction models. Although the definitions of invasion depth in EGC differ, features such as textures, shapes, and colors are visually similar. In addition, each degree of invasion depth may not have sufficient training images to cover all types of visual features, because of the fine-scale granularity. Therefore, models for EGC detection and depth prediction may be used to focus on other visually distinguishable patterns rather than EGC. For example, model weights may initially be tuned to find a tiny particle appearing on most images in a T1b-EGC training set rather than extracting features from homogeneous regions. Therefore, it is critical to guide the model to learn the visual features of EGC regions rather than those of other gastric textures.

The present study aims to develop a model and training method optimized for EGC depth prediction, evaluate its diagnostic performance, and investigate factors affecting AI diagnosis.

II. MATERIALS AND METHODS

1. Patients

This study included 800 patients (538 men and 262 women; age: 26-92 years; mean age: 62.6 years) with an endoscopic diagnosis of EGC at the Gangnam Severance Hospital, Yonsei University College of Medicine, Seoul, Korea, between January 2012 and March 2018. EGC was suspected based on endoscopy findings and all patients underwent a curative treatment by either operation or ER for gastric cancer. The invasion depth was confirmed pathologically through specimens obtained after the treatment. This study was approved by the Institutional Review Board of Gangnam Severance Hospital (no. 3-2017-0365).

2. Data preparation (endoscopic image collection)

Endoscopy was performed for screening or preoperative examinations. Images were captured using standard endoscopes (GIF-Q260J, GIF-H260, and GIF-H290; Olympus Medical Systems, Co., Ltd., Tokyo, Japan). The image of the lesion should have both close-up and a distant view so that the size and position of the lesion can be identified. Additionally, the amount of gas insufflation should be adjusted appropriately to reflect the condition of the lesion and its surrounding area.

We collected 11,686 endoscopic images, including 1,097 T1a-EGC, 1,005 T1b-EGC, and 9,834 non-EGC images. The non-EGC images were endoscopic images of the gastric mucosa that were not EGC, including chronic gastritis, chronic atrophic gastritis, intestinal metaplasia, and erosion. The images with poor quality were filtered out. The image inclusion criteria comprised white light images and images with whole lesions. However, images with motion-blurring, out of focus, halation, and poor air insufflation were excluded. Finally, 11,539 images (896 T1a-EGC, 809 T1b-EGC, and 9834 non-EGC) were selected. To prepare the image dataset for the models, the selected images were randomly organized into five different folds to assess how the trained model was generally applicable while avoiding overfitting or testset selection bias.[13] The five folds were used to train and evaluate the deep learning models. All the folds were independent, and the training:validation:testing dataset ratio at each fold was 3:1:1 (Table 1). The images extracted from one patient were assigned to a fold; therefore, the number of images between the folds differed slightly (Table 2). The validation set that was a totally independent fold than the training folds was used to observe the training status during the training. After training the model, the other independent fold was used to evaluate the model performace as a testing set. For example, cross validation-group 1 of Table 2 used the first three folds (A, B, and C) as the training set, the fourth fold (D) as the validation set, and the remaining folds (E) as the testing set (Table 1).

Table 1. Groups for 5-fold cross validation.

Group	Training	Validation	Test
1	A,B,C	D	E
2	B,C,D	E	A
3	C,D,E	A	B
4	D,E,A	B	C
5	E,A,B	C	D

Table 2. Composition of the five-fold cross-validation dataset.

		Normal		Lesion		Mucosal		Submucosal	
		Case	Image	Case	Image	Case	Image	Case	Image
CV-1	Train	405	7701	420	1028	236	534	186	494
	Validation	51	330	138	330	68	150	61	150
	Test	54	330	125	330	60	150	63	150
CV-2	Train	408	7712	420	1026	232	534	188	492
	Validation	53	330	124	330	57	150	63	150
	Test	63	330	139	330	76	150	61	150
CV-3	Train	410	7642	397	1029	213	546	184	483
	Validation	53	330	139	330	76	150	58	150
	Test	51	330	147	330	74	150	66	150
CV-4	Train	399	7659	403	1015	215	541	188	474
	Validation	52	330	146	330	78	150	64	150
	Test	66	330	131	330	73	150	57	150
CV-5	Train	384	7638	412	1017	221	533	191	484
	Validation	59	330	130	330	73	150	54	150
	Test	58	330	139	330	70	150	63	150

CV, *cross validation group*

3. Convolutional neural network and training

We used two networks on two training methods to evaluate which one allowed the CNN to be better oriented to EGC regions. The two network models were based on a transfer learning method with the VGG-16 network pre-trained on ImageNet, which is a large-scale dataset published for the image classification task to effectively initialize and train network weights [14,15]. The first model was a typical method that computed the loss between the real and predicted classes of input data. The second was a novel method that used the weighted sum of gradient-weighted class activation mapping (Grad-CAM) and cross-entropy losses. To let the model focus on the fine-grained features of EGC regions, we employed a novel loss function by adding Grad-CAM[16,17]. Although most existing visualization methods require an additional module to generate visual explanations, Grad-CAM can visualize activation statuses that are gradually changed over the training time as its initial architectures [18,19]. The gradually activated EGC regions of the input RGB image, which were produced by passing trained layers, are shown in Figure 1. The blue and red colors on Grad-CAM indicate lower and higher activation values, respectively.

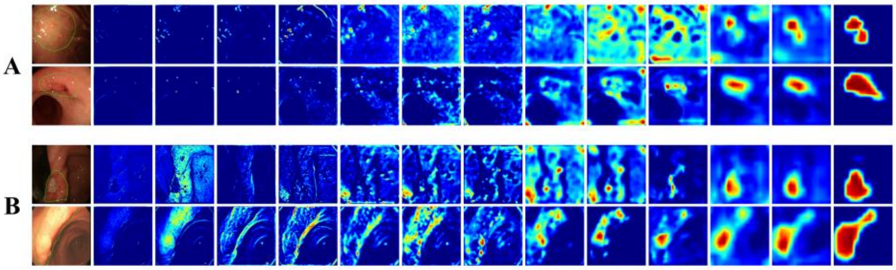

Figure 1. Examples of the Grad-CAM output extracted from each convolutional layer of the trained lesion-based VGG-16. The green lines on the first column indicate the actual EGC regions. The images from the second column (the first convolutional layer) to the last column (the last convolutional layer) are the

activated map extracted from each convolutional layer of the network. A, T1a-EGC. B, T1b-EGC.

The proposed novel method allows the training procedure to optimize an objective that simultaneously minimizes not only the classification error (real classes - predicted classes) but also the localization error (real lesion mask - activated Grad-CAM). The real lesion mask is part of an endoscopic image that the endoscopist identified as real EGC area. We named this novel method “lesion-based VGG-16.” An overview of the proposed algorithm for the computer-aided diagnosis (CAD) of EGC is shown in Figure 2.

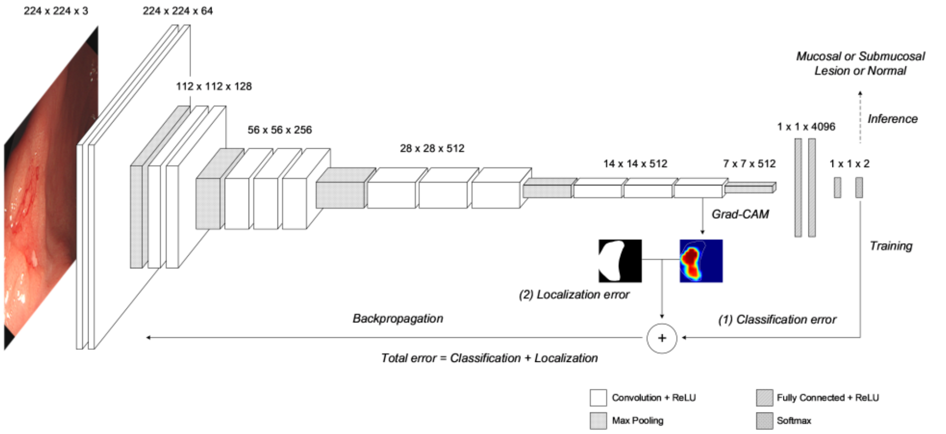

Figure 2. Overview of the VGG-16-based model. The solid lines indicate the paths for training the model, while the dotted line indicates the path to infer the class.

4. Lesion-based training

To detect EGCs and predict their depth, we used the transfer learning method with a VGG-16 network, which is a well-known model comprising 16 convolutional layers and a rectified linear unit. The last three convolutional layers are fully connected to their output. We adapted the last layer of VGG-16 to a two-dimensional, fully connected layer to classify the input endoscopy image into two classes (EGC *vs*. non-EGC and T1a-EGC *vs*. T1b-EGC). A two-way softmax layer was connected to the last fully connected layer to set the output classification probability to [0, 1]. First, to initialize the weights of the network, VGG-16 was trained on the ImageNet classification dataset. Subsequently, all layer weights were fine-tuned by learning our dataset.

Most existing classification models were trained to minimize only one loss of the classification results, such as the cross-entropy loss[14,20]. However, there are limitations in obtaining high classification performances by optimizing an objective function with the classification loss. To precisely detect EGCs and predict their depth, we proposed a loss function that jointly measures the classification loss of probability and localization loss of the activated regions. The trained model using Grad-CAM loss was defined as the lesion-based network in this study (lesion-based VGG-16). We used the cross-entropy loss $\mathcal{L}_{\mathrm{C}}$ for the classification loss and Grad-CAM loss $\mathcal{L}_{\mathrm{G}}$ based on Grad-CAM, which visualizes the activated regions of the convolutional layer, for the localization loss [21]. Cross-entropy loss has been used to measure the performance of a classification model that outputs probability values for each target class between 0 and 1. If the predicted probability diverges from the actual class of an input image, cross entropy loss will be larger according to the following equation:

$$\mathcal{L}_{\mathrm{C}} = -\sum_{c=1}^{M} y_c(x) \log p(c|x, \mathcal{W}),$$

where $y_c(x)$ and $p(c|x, \mathcal{W})$ are the actual label and predicted probability defined by model weights $\mathcal{W}$ of input EGC image x for class c. M denotes the number of target classes (e.g., $M = 2$ for EGC vs. non-EGC classification model). Grad-CAM of the last convolutional layer was used to measure Grad-CAM loss. The low activation value indicated does not mean complete exclusion from the features, as the activation values are relative in an image.

Activated regions with high value were more often used as visual features in the following fully connected layers. Grad-CAM loss was measured as the dice coefficient loss between the gold-standard EGC binary mask and Grad-CAM output from the last convolutional layer as follows:

$$\mathcal{L}_G = -\sum_{c=1}^{M} \frac{2 \times |P_c^G(x) \cap M_c(x)|}{|P_c^G(x) \cup M_c(x)|},$$

where $P_c^G(x)$ and $M_c(x)$ indicate Grad-CAM consisting of activation values for class c from the last convolutional layer and binary mask of actual EGC regions of the input endoscopy image x, respectively [22]. The final loss $\mathcal{L}$ was computed by a weighted summation of cross-entropy and Grad-CAM losses, as follows:

$$\mathcal{L} = \alpha\mathcal{L}_C + (1-\alpha)\mathcal{L}_G, \quad 0 \leq \alpha \leq 1.$$

In this study, we empirically set the weight α to 0.5. The network was optimized using the adaptive moment estimation (Adam) optimizer with an initial learning rate of 1e-5. The training batch size was set to 32 [23].

For training the model (solid arrows), the network takes a paired input consisting of an RGB image and a binary mask, where EGC regions are filled with a pixel value of 1. The proposed model takes an RGB image with a size of 224 x 224 x 3 as input. Before entering the input image into the network, the mean value (from ImageNet dataset) subtraction was performed on each color channel of the input RGB image to serve as the distribution of data points to the center (i.e., 0) [2].

We applied two types of data augmentation (image augmentation and mask augmentation) to our training dataset. First, image augmentations including random flip, random small rotation, and random elastic deformation were generated 10 times for each image and mask pair. Next, mask augmentations consisting of random kernel erosions, dilations, and random elastic transforms were applied to the binary masks generated by the image augmentation to reflect the diversity of labeling. Although image augmentation was equally performed on the RGB image and corresponding binary mask pair, mask augmentation was only applied to the binary mask image without transformation of the RGB image. As a result, the prepared training image was augmented 100 times.

We implemented the models on an Intel-Core i7-7700K 4.2 GHz processor with 32 GB RAM and a GeForce Titan XP graphic card with 12 GB CUDA memory. In our batch implementation environment, approximately 40 images were processed per second.

5. Invasion depth prediction

For training, the model requires a binary mask corresponding to the EGC region of an input RGB image to measure Grad-CAM loss; there is no need to feed a binary mask to the trained network for obtaining the classification probability of the test image (inference line of Figure 2). Since we set our network to focus on the fine-grained features of EGC regions during training, the network gradually activated cancer regions to the infer class while the input was consecutively propagated. By passing the RGB image to the trained network, we were able to obtain the probability for each class.

6. EGC localization

During the training, the model was trained to properly detect EGCs (or predict their depth) while activating EGC regions by simultaneously optimizing the cross-entropy loss and Grad-CAM loss. By utilizing this characteristic, we segmented the connected components of pixels with an activation value of more than 0.5 as a cancer region localization result. To show the effectiveness of localization of activated regions on the last convolutional layer, we computed the ratio of intersection between the actual EGC region and activated region to the actual EGC region, according to changes in the threshold overlap ratios, which ranged from 0.01 to 1.0 with a step size of 0.01. Among the correctly detected EGCs or predicted depth of EGCs, a correct lesion localization was defined as when the intersection regions over the real EGC region of an input RGB image exceeded the threshold overlap ratio. Consequently, a fraction of the number of images in which the lesions were precisely activated to the number of correctly classified images was computed as a metric.

7. Evaluation

To evaluate the performance of EGC detection and depth prediction models, we measured the sensitivity (%), specificity (%), positive predictive value (PPV) (%), negative predictive value (NPV) (%), and area under the curve (AUC) of receiver operating characteristic (ROC) curves by summing all cross-validation folds. Because the number of test images comprising each cross-validation fold was different, it was insufficient to evaluate their generalized performances. Therefore, we randomly selected a fixed number of images from each class of the test dataset. The test set of EGC depth prediction included 300 images, comprising 150 T1a-EGC and 150 T1b-EGC images. The EGC detection model test set included 660 images consisting of 330 EGC and 330 non-EGC images. Additionally, 90 EGC images not included in the cross-validation datasets were also tested. A total of 1,590 and 3,390 images were evaluated for predicting EGC depth and detecting EGC, respectively.

Since the network was trained by activating EGC regions, activated regions extracted by Grad-CAM at the last convolutional layer could be considered as suspected cancer regions when an endoscopy image was fed to the network. To demonstrate the utility of cases where activated regions can be localized EGC regions, we evaluated the EGC localization performances.

8. Statistical analysis

Chi-squared and Fisher’s exact tests were used to evaluate the associations among various categorical variables. Univariable and multivariable logistic regression analyses were performed to identify factors significantly affecting the AI accuracy. Odds ratios (ORs) and relevant 95% confidence intervals (CIs) were calculated. Analyses were performed using SAS version 9.4 (SAS Institute, Cary, NC, USA) or IBM SPSS Statistics for Windows, version 23.0 (IBM Co., Armonk, NY, USA) and *P*-values <0.05 indicated statistical significance.

III. RESULTS

1. Baseline clinicopathological characteristics of the subjects

The patients included 538 men and 262 women with a mean age of 62.6 years (range: 26–92 years). The mean lesion size (±SD) was 23.7±15.1 mm. There were 428 (53.5%) mucosal-depth lesions (T1a-EGC) and 372 (46.5%) submucosal-depth lesions (T1b-EGC). The histology types of lesions according to the World Health Organization (WHO) classification included well-differentiated (321/800, 40.1%), moderately-differentiated (268/800, 33.5%), and poorly-differentiated adenocarcinoma (103/800, 12.9%) and signet-ring cell carcinoma (108/800, 13.5%). The histologic types according to the Japanese classification included differentiated (589/800, 73.6%) and undifferentiated adenocarcinoma (211/800, 26.4%). The baseline clinicopathological characteristics of the lesions are summarized in Table 3.

Table 3. Baseline clinicopathological characteristics of all patients.

Characteristics	Value
Age (years, mean±SD)	62.6 ±12.2
Male (n, %)	536 (67.2)
Tumor size (mm, mean ± SD)	23.7 ±15.1
Location of lesion (n, %)	
Upper one-third	74 (9.3)
Middle one-third	118 (14.7)
Lower one-third	608 (76)
Gross type (n, %)	
Elevated	171 (21.4)
Flat	285 (35.6)
Depressed	344 (43)
Lymphovascular invasion (n, %)	82 (10.3)
Perineural invasion (n, %)	14 (1.8)
T-stage (n, %)	
Mucosa (T1a)	428 (53.5)
Submucosa (T1b)	372 (46.5)
WHO classification (n, %)	
Well-differentiated	321 (40.1)
Moderately-differentiated	268 (33.5)
Poorly-differentiated	103 (12.9)
Signet ring cell carcinoma	108 (13.5)
Japanese classification (n, %)	
Differentiated	589 (73.6)
Undifferentiated	211 (26.4)
Lauren classification (n, %)	
Intestinal	606 (77.3)
Diffuse	156 (19.9)
Mixed	22 (2.8)

2. Diagnostic performance using the VGG-16

We first tested the VGG-16 trained using the cross-entropy loss function on the selected test image set. The sensitivity, specificity, PPV, NPV, and overall AUC for EGC detection were 80.7%, 92.5%, 91.9%, 82.0%, and 0.938, respectively. The values for EGC depth prediction were 81.7%, 75.4%, 78.0%, 79.3%, and 0.844, respectively (Table 4).

Table 4. Diagnostic accuracy of VGG-16 and lesion-based VGG-16.

	VGG-16		Lesion-based VGG-16	
	Detecting EGC	Predicting depth	Detecting EGC	Predicting depth
AUC	0.938	0.844	0.981	0.851
Sensitivity (%)	80.7	81.7	91.0	79.2
Specificity (%)	92.5	75.4	97.6	77.8
PPV (%)	91.9	78.0	97.5	79.3
NPV (%)	82.0	79.3	91.1	77.7

EGC, early gastric cancer; *AUC*, area under the curve; *PPV*, positive predictive value; *NPV*, negative predictive value

Subsequently, we evaluated the lesion-based VGG-16 on the same test image set. The sensitivity and specificity for EGC detection were 91.0% and 97.6%, respectively, and the PPV and NPV were 97.5% and 91.1%, respectively. The overall AUC was 0.981. The sensitivity and specificity of the prediction of tumor depth in the lesion-based VGG-16 were 79.2% and 77.8%, respectively, and the PPV and NPV were 79.3% and 77.7%, respectively. The overall AUC was 0.851 (Table 4 and Figure 3).

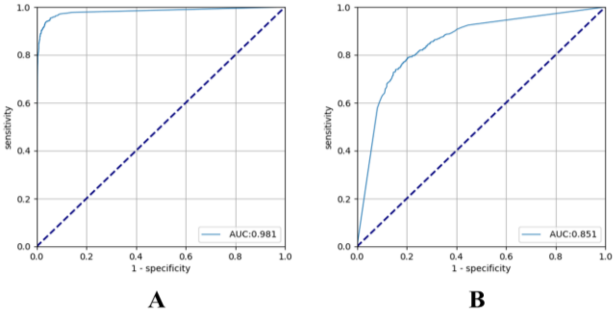

Figure 3. Receiver operating characteristic (ROC) curves of lesion-based VGG-16 for the test dataset with their areas under the curves (AUCs). A, EGC detection model. B, EGC depth prediction model.

3. Localization ability of the activated regions

We compared the localization ability of the activated regions on the last convolutional layer of the lesion-based VGG-16 to the that of VGG-16. Figures 4A and B show the localization performance of the activation results of the EGC detection and depth prediction models, respectively. The correct ratios at the 0.5 overlap ratio for detecting EGCs of the lesion-based VGG-16 and VGG-16 were 0.994 and 0.581, respectively. For predicting EGC depth, the lesion-based VGG-16 correctly activated the EGC regions with a ratio of 0.959 compared to 0.811 for VGG-16.

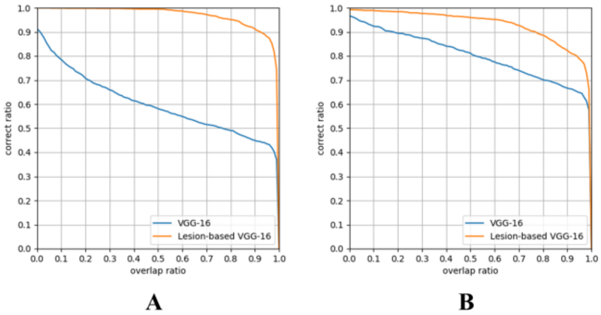

Figure 4. Localization performances of activated regions extracted using the Grad-CAM method. A, EGC detection models. B, EGC depth prediction model.

The activated regions extracted from the last convolutional layer of VGG-16 and lesion-based VGG-16 are shown in Figure 5. The activation regions of VGG-16 (the second row of Figure 5) did not precisely cover the actual EGC regions (first two columns of Figure 5), and in some cases, deviated from the EGC regions (last two columns of Figure 5). In contrast, the lesion-based VGG-16 (last row) attempted to completely activate and reach the EGC regions. In depth prediction, lesion-based VGG-16 reflected the actual EGC regions more accurately, as shown in Figure 5B. Figure 6 shows the correctly classified (first two rows) and misclassified (last row) images of the lesion-based VGG-16. Although the model misclassified the presence or depth of EGCs in some cases, the EGC region was accurately activated.

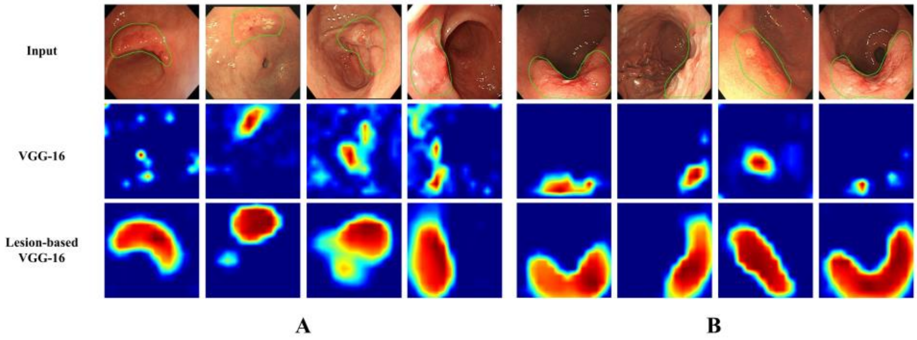

Figure 5. Example of RGB input images (the first row) and their Grad-CAM results extracted from the last convolutional layer of VGG-16 (the second row) and the lesion-based VGG-16 (the third row). The green lines of the input images indicate the actual EGC regions. In the Grad-CAM outputs, the red color is a higher activation value and the blue color is the opposite. A, EGC detection. B, EGC depth prediction.

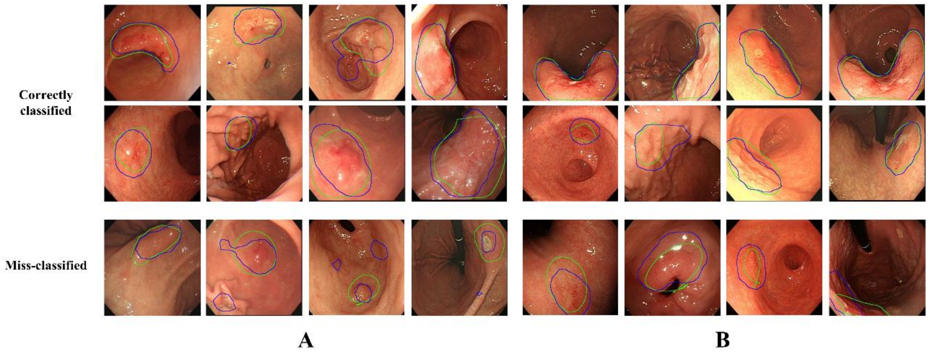

Figure 6. Classification results of lesion-based VGG-16. The green lines indicate the actual early gastric cancer (EGC) regions. The blue lines indicate the activated regions at testing. The first two rows are images precisely classified to their own classes, whereas the last row shows misclassified images. A, EGC detection. B, EGC depth prediction.

4. Factors associated with the accuracy of tumor detection by AI

EGCs with a flat morphology had a significantly lower accuracy for EGC detection than other gross types (*P* = 0.038) (Table 5). Relatively small size (1–13mm) (*P* = 0.002) and T1a-EGC (*P* = 0.001) were significantly associated with tumor detection. In multivariable analysis, small size (1–13 mm) (*P* = 0.006) and T1a-EGC (*P* = 0.019) showed statistically lower accuracies. The accuracies of EGC detection for tumors ≤5 and ≤10 mm were 88.4% and 89.4%, respectively. The EGC detection did not differ significantly according to the histologic differentiation and location.

Table 5. Factors affecting the accuracy of tumor detection.

	Accurate	**Inaccurate**	***P*-value**	**Odds ratio (95% CI)**	***P*-value**
Gross type (n, %)			0.038		
Elevated	169 (21.7)	2 (10.5)			
Flat	271 (34.9)	12 (63.2)			
Depressed	337 (43.4)	5 (26.3)			
T-stage (n, %)			0.001		0.019
Mucosa (T1a)	406 (52.3)	17 (89.5)		ref	
Submucosa (T1b)	371 (47.7)	2 (10.5)		5.891 (1.326-26.171)	
Size (n, %)			0.002		0.006
1–13 mm	162 (21.7)	11 (57.9)		ref	
≥ 14 mm	608 (78.3)	8 (42.1)		3.660 (1.427-9.384)	
Location (n, %)			0.780		
Upper one-third	72 (9.3)	2 (10.5)			
Mid one-third	115 (14.8)	3 (15.8)			
Lower one-third	590 (75.9)	14 (73.7)			
Japanese classification (n, %)			0.296		
Differentiated	575 (74)	12 (63.2)			
Undifferentiated	202 (26)	7 (36.8)			

5. Factors associated with the accuracy of T-staging by AI

Undifferentiated-type histology was the only factor significantly associated with a lower accuracy for T-stage prediction in both univariable and multivariable analyses (*P* = 0.001 and 0.033, respectively) (Table 6). The accuracy did not differ significantly according to the size.

Table 6. Factors affecting the accuracy of T staging.

	Accurate	Inaccurate	*P*-value	Odds ratio (95% CI)	*P*-value
Japanese classification (n, %)			0.001		0.033
Differentiated	446 (76.8)	132 (65.0)		ref	
Undifferentiated	135 (23.2)	71 (35.0)		0.491 (0.255-0.945)	
Gross type (n, %)			0.442		
Elevated	127 (21.9)	41 (20.2)			
Flat	212 (36.5)	67 (33.0)			
Depressed	242 (41.6)	95 (46.8)			
T-stage (n, %)			0.235		
Mucosa (T1a)	320 (55.1)	102 (50.3)			
Submucosa (T1b)	261 (44.9)	101 (49.7)			
Size (n, %)			0.329		
1–13 mm	137 (23.7)	44 (21.8)			
≥ 14 mm	442 (76.3)	158 (78.2)			

The factors associated with T-stage prediction were reanalyzed in undifferentiated-type histology. T1b was only significantly associated with a lower T-stage prediction accuracy (*P* = 0.015) (Table 7). Thus, factors associated with T-staging in undifferentiated-type histology were investigated. Relatively large size (≥14 mm) (*P* = 0.003) and poorly differentiated adenocarcinoma (*P* <

0.001) were significantly associated with T1b in undifferentiated-type histology (Table 8). Among undifferentiated-type EGCs, flat and elevated morphologies were more common in T1a and T1b, respectively.

Table 7. Factors affecting the accuracy of T staging in undifferentiated-type adenocarcinoma.

	Accurate	**Inaccurate**	***P*-value**	**Odds ratio (95% CI)**	***P*-value**
T-stage (n, %)			0.015		0.015
Mucosa (T1a)	97 (71.9)	39 (54.9)		ref	
Submucosa (T1b)	38 (28.1)	32 (45.1)		0.477 (0.262-0.869)	
Gross type (n, %)			0.152		
Elevated	15 (11.1)	7 (9.9)			
Flat	55 (40.7)	20 (28.1)			
Depressed	65 (48.2)	44 (62.0)			
Size (n, %)			0.444		
1–13 mm	24 (17.9)	14 (19.7)			
≥ 14 mm	110 (82.1)	57 (80.3)			
WHO classification (n, %)			0.296		
APD	60 (44.4)	38 (53.5)			
SRC	75 (55.6)	33 (46.5)			

APD, poorly differentiated adenocarcinoma; *SRC*, signet ring cell carcinoma.

Table 8. Associated factors according to T staging in undifferentiated-type adenocarcinoma.

	T1a	T1b	*P*-value
Gross type (n, %)			0.003
Elevated	8 (5.8)	14 (19.2)	
Flat	57 (41.3)	19 (26.0)	
Depressed	73 (52.9)	40 (54.8)	
Sex (n, %)			0.012
Male	60 (43.5)	45 (61.6)	
Female	78 (56.5)	28 (38.4)	
Size (n, %)			0.003
1–13 mm	33 (24.1)	6 (8.2)	
≥ 14 mm	104 (75.9)	67 (91.8)	
Location of lesion (n, %)			0.276
Upper one-third	5 (3.6)	5 (6.8)	
Mid one-third	27 (19.6)	19 (26.0)	
Lower one-third	106 (76.8)	49 (67.1)	
WHO classification (n, %)			<0.001
APD	53 (38.4)	50 (68.5)	
SRC	85 (61.6)	23 (31.5)	

APD, poorly differentiated adenocarcinoma; *SRC*, signet ring cell carcinoma

IV. DISCUSSION

Although previous studies have reported the clinical efficacy of EUS in T-staging of EGC, the results are conflicting [7,24-26]. Some studies have reported that conventional endoscopy is comparable to EUS for the T-staging of EGC [6,27]. Various morphologic features, such as irregular surface and submucosal tumors, like marginal elevation, have been proposed as predictors of tumor invasion depth[28]. Identification and verification of additional morphological features of deep invasion in large datasets would allow a more complete depth prediction.

The sensitivity and overall AUC of EGC detection in the present study were 91.0% and 0.981, respectively, comparable to those in a previous report [10]. The overall AUC of T-staging by our lesion-based VGG-16 system was 0.851, which is higher than that previously reported for EUS prediction [6,26]. Unlike other studies, the present study also analyzed the factors affecting AI diagnosis [10,29]. The diagnostic accuracy of AI for T-staging was significantly affected by histopathologic differentiation. Undifferentiated-type histology was more frequently associated with an incorrect invasion depth diagnosis by AI. By reanalyzing only undifferentiated-type histology, T1b-EGC was significantly associated with an incorrect EGC invasion depth diagnosis by the AI. Interestingly, this finding was similar to that for the analysis in EUS. Previous studies have reported that the accuracy of EUS for depth prediction is poor in undifferentiated-type EGC or T1b-EGC [6,30]. Undifferentiated-type histology and T1b-EGC are two important factors for the decision to perform an extended ER. Therefore, these results can provide important directions for the development of an AI for EGC.

As it is critical that AI is properly trained, we performed extensive experimentation and discussion. There are some challenges in applying the loss function designed to train the classification model for a natural-image dataset to AI for EGC without modification. To overcome these difficulties, we proposed a novel loss function that computed a weighted sum of typical classification and

Grad-CAM losses. By applying the proposed loss function to EGC detection and EGC depth prediction models, the optimizer simultaneously minimized classification and localization losses in the activated Grad-CAM regions. Although there was no significant performance improvement in predicting the depth of EGCs between VGG-16 (AUC=0.844) and lesion-based VGG-16 (AUC=0.851), the trained lesion-based VGG-16 predicted the depth of EGCs by automatically activating EGC regions, whereas VGG-16 did not. The classification performance of VGG-16 trained by cross-entropy loss alone is still debatable regarding dataset bias, where the model considered non-EGC regions to optimize the objective. To the best of our knowledge, this is the first study to use a novel loss function that allows the optimizer to determine an optimum by simultaneously considering EGC depth prediction and localization losses of the activated regions. This model uses the proposed method to simultaneously provide prediction and localization.

To determine which proposed loss function made the CNN focus on the EGC region regardless of the network, we trained an 18-layer residual network (ResNet-18) as a CNN-based EGC depth prediction model [31]. We fine-tuned all weights for a ResNet-18 pre-trained on the ImageNet Dataset. The activation results of ResNet-18 are shown in Figure 7. As with VGG-16, ResNet-18 was also trained using two types of loss functions . As shown in the last two columns of Figure 7, the lesion-based ResNet-18 more accurately activated the EGCs as compared to ResNet-18.

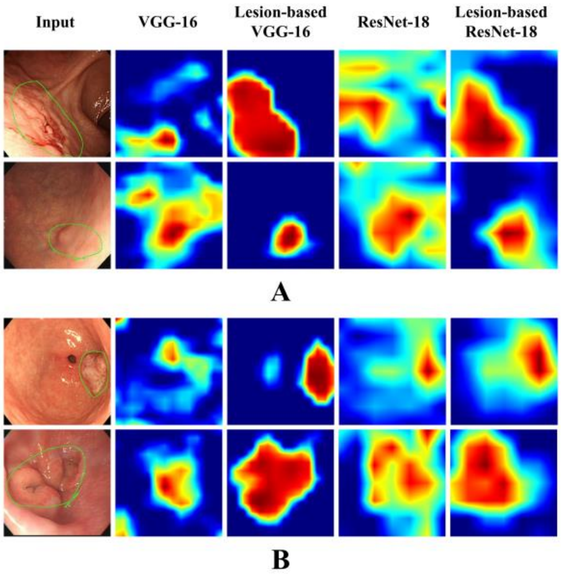

Figure 7. Comparisons of Grad-CAMs extracted from the last convolutional layer when a network was trained using two types of loss functions. The green lines on the input images indicate the actual EGC regions. A, T1a-EGC. B, T1b-EGC.

The present study has several limitations. First, we did not analyze the accuracy of EGC detection according to the background mucosa. That is, the background mucosa of the stomach is accompanied by chronic inflammatory changes such as chronic atrophic gastritis and intestinal metaplasia. These are important features that complicate EGC diagnosis. However, to overcome the differences in accuracy based on the features of the background mucosa, more than 9,800 non-EGC endoscopic images were learned. Second, the number of undifferentiated-type histology cases was relatively smaller than that of differentiated-type histology. The AI performance is related to the amount of data, and thus may have played an important role in the accurate prediction of EGC depth. It is possible that growth patterns or biological characteristics of undifferentiated histology areas are also affected. Similar findings were reported in previous EUS studies. Third, we did not compare the diagnostic accuracy of lesion-based VGG-16 to that of endoscopists for all images of the study, although endoscopists predicted the invasion depth for subsets of images in this study, with a sensitivity of 76% and overall accuracy of 73% (data not shown). Therefore, the proposed method may be a good tool for predicting the depth of EGC invasion. Finally, this is a retrospective study. Standardization of images is a very important part of the research involving image analysis. The images used were of good quality, and they appropriately characterized the lesions. However, they were not completely standardized with numerical analysis. To overcome the aforementioned limitations, we plan to perform research by using endoscopic video in the future.

V. CONCLUSION

In conclusion, AI may be a good tool not only for EGC diagnosis but also for the prediction of invasion depth, especially in differentiated-type EGC. To maximize the clinical usefulness, it is important to choose an appropriate method for AI application. The lesion-based model is the most appropriate training method for AI in EGC. EGC with undifferentiated-type histology and T1b-EGC is more frequently associated with an incorrect EGC invasion depth by AI. The development of a well-trained AI for undifferentiated-type histology and T1b-EGC is warranted. Further study is also necessary to understand the operating principles of AI and to validate these findings.

REFERENCES

1. Wang J, Yu JC, Kang WM, Ma ZQ. Treatment strategy for early gastric cancer. Surgical oncology 2012;21:119-123. Epub 2011/01/25.

2. Goto O, Fujishiro M, Kodashima S, Ono S, Omata M. Outcomes of endoscopic submucosal dissection for early gastric cancer with special reference to validation for curability criteria. Endoscopy 2009;41:118-122. Epub 2009/02/14.

3. Maruyama K. The Most Important Prognostic Factors for Gastric Cancer Patients: A Study Using Univariate and Multivariate Analyses. Scandinavian Journal of Gastroenterology 1987;22:63-68.

4. Mocellin S, Marchet A, Nitti D. EUS for the staging of gastric cancer: a meta-analysis. Gastrointestinal endoscopy 2011;73:1122-1134. Epub 2011/03/30.

5. Yanai H, Matsumoto Y, Harada T, et al. Endoscopic ultrasonography and endoscopy for staging depth of invasion in early gastric cancer: a pilot study. Gastrointestinal endoscopy 1997;46:212-216. Epub 1997/11/05.

6. Choi J, Kim SG, Im JP, Kim JS, Jung HC, Song IS. Comparison of endoscopic ultrasonography and conventional endoscopy for prediction of depth of tumor invasion in early gastric cancer. Endoscopy 2010;42:705-713. Epub 2010/07/24.

7. Pei Q, Wang L, Pan J, Ling T, Lv Y, Zou X. Endoscopic ultrasonography for staging depth of invasion in early gastric cancer: A meta-analysis. Journal of gastroenterology and hepatology 2015;30:1566-1573. Epub 2015/06/23.

8. Schmidt-Erfurth U, Sadeghipour A, Gerendas BS, Waldstein SM, Bogunovic H. Artificial intelligence in retina. Progress in retinal and eye research 2018;67:1-29. Epub 2018/08/05.

9. Alagappan M, Brown JRG, Mori Y, Berzin TM. Artificial intelligence in gastrointestinal endoscopy: The future is almost here. World journal of gastrointestinal endoscopy 2018;10:239-249. Epub 2018/10/27.

10. Hirasawa T, Aoyama K, Tanimoto T, et al. Application of artificial intelligence using a convolutional neural network for detecting gastric cancer in endoscopic images. Gastric cancer : official journal of the International Gastric Cancer Association and the Japanese Gastric Cancer Association 2018;21:653-660. Epub 2018/01/18.

11. Horie Y, Yoshio T, Aoyama K, et al. Diagnostic outcomes of esophageal cancer by artificial intelligence using convolutional neural networks. Gastrointestinal endoscopy 2018. Epub 2018/08/19.

12. Urban G, Tripathi P, Alkayali T, et al. Deep Learning Localizes and Identifies Polyps in Real Time With 96% Accuracy in Screening Colonoscopy. Gastroenterology 2018;155:1069-1078.e1068. Epub 2018/06/22.

13. Stone M. Cross-validatory choice and assessment of statistical predictions. Journal of the royal statistical society Series B (Methodological) 1974:111-147.

14. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:14091556 2014.

15. Russakovsky O, Deng J, Su H, et al. Imagenet large scale visual recognition challenge. International Journal of Computer Vision 2015;115:211-252.

16. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. ICCV2017:618-626.

17. De Boer P-T, Kroese DP, Mannor S, Rubinstein RY. A tutorial on the cross-entropy method. Annals of operations research 2005;134:19-67.

18. Zeiler MD, Fergus R. Visualizing and understanding convolutional networks. European conference on computer vision: Springer; 2014:818-833.

19. Simonyan K, Vedaldi A, Zisserman A. Deep inside convolutional networks: Visualising image classification models and saliency maps. arXiv preprint arXiv:13126034 2013.

20. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems2012:1097-1105.

21. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-cam: Visual explanations from deep networks via gradient-based localization. Proceedings of the IEEE International Conference on Computer Vision2017:618-626.

22. Milletari F, Navab N, Ahmadi S-A. V-net: Fully convolutional neural networks for volumetric medical image segmentation. 3D Vision (3DV), 2016 Fourth International Conference on: IEEE; 2016:565-571.

23. Kingma DP, Ba J. Adam: A method for stochastic optimization. arXiv preprint arXiv:14126980 2014.

24. Kim TY, Yi NH, Hwang JW, Kim JH, Kim GH, Kang MS. Morphologic pattern analysis of submucosal deformities identified by endoscopic ultrasonography for predicting the depth of invasion in early gastric cancer. Surgical endoscopy 2018. Epub 2018/10/20.

25. Han Y, Sun S, Guo J, et al. Is endoscopic ultrasonography useful for endoscopic submucosal dissection? Endoscopic ultrasound 2016;5:284-290. Epub 2016/11/03.

26. Kim J, Kim SG, Chung H, et al. Clinical efficacy of endoscopic ultrasonography for decision of treatment strategy of gastric cancer. Surgical endoscopy 2018;32:3789-3797. Epub 2018/02/13.

27. Choi J, Kim SG, Im JP, Kim JS, Jung HC, Song IS. Endoscopic prediction of tumor invasion depth in early gastric cancer. Gastrointestinal endoscopy 2011;73:917-927. Epub 2011/02/15.

28. Tsujii Y, Kato M, Inoue T, et al. Integrated diagnostic strategy for the invasion depth of early gastric cancer by conventional endoscopy and EUS. Gastrointestinal endoscopy 2015;82:452-459. Epub 2015/04/07.

29. Zhu Y, Wang QC, Xu MD, et al. Application of convolutional neural network in the diagnosis of the invasion depth of gastric cancer based on conventional endoscopy. Gastrointestinal endoscopy 2018. Epub 2018/11/20.

30. Kim JH, Song KS, Youn YH, et al. Clinicopathologic factors influence accurate endosonographic assessment for early gastric cancer. Gastrointestinal endoscopy 2007;66:901-908. Epub 2007/10/30.

31. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition2016:770-778.

ABSTRACT(IN KOREAN)

병변 기반의 컨볼루션 신경망을 이용한 조기위암의 내시경 탐지 및 깊이 예측의 향상

<지도교수 김지현>

연세대학교 대학원 의학과

윤 홍 진

조기 위암의 침범 깊이는 치료 방법을 결정하는 중요한 요소이다. 그러나 실제 임상에서 치료 전에 정확한 침범 깊이를 측정하는데 한계가 있다. 본 연구에서는 조기위암 발견 및 침범 깊이 예측에 최적화된 인공지능 모델을 개발하고 진단에 영향을 미치는 요인을 조사 하였다. 1705장의 조기위암 사진이 포함된 총 11,539의 내시경 이미지를 인공지능 모델을 이용하여 학습 및 테스트 하였다. 조기위암 발견 및 침범 깊이 예측에 대한 인공지능 모델의 정확도는 각각 0.981, 0.851 이었다. 여러 요인들 중에서 조직학적 미분화 암이 조기위암 침범 깊이의 인공지능 예측에서 유의하게 낮은 정확도를 보였다. 이 연구를 통해, 조기 위암 진단에 대한 병변 기반 인공지능 모델의 유용성을 확인하였고, 조직학적 미분화 암에 대해서는 추가적인 개선 및 검증이 필요함을 확인하였다.

핵심되는 말 : 조기위암, 인공지능, 컨볼루션 신경망, 내시경

PUBLICATION LIST

Yoon, H. J.; Kim, S. H.; Kim, J-H.; Keum J-S., A lesion-based convolutional neural network improves endoscopic detection and depth prediction of early gastric cancer. Journal of clinical medicine 2019, 8(9), 1310