

Received 20 February 2023, accepted 20 March 2023, date of publication 23 March 2023, date of current version 28 March 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3260983

RESEARCH ARTICLE

Enhanced Classification of Gastric Lesions and Early Gastric Cancer Diagnosis in Gastroscopy Using Multi-Filter AutoAugment

JUNG-WOO CHAE¹ AND HYUN-CHONG CHO^{1,2}, (Member, IEEE)

¹Department of Interdisciplinary Graduate Program for BIT Medical Convergence, Kangwon National University, Chuncheon 24341, South Korea

²Department of Electronics Engineering, Kangwon National University, Chuncheon 24341, South Korea

Corresponding author: Hyun-Chong Cho (hyuncho@kangwon.ac.kr)

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No. 2022R111A3053872) and was supported by “Regional Innovation Strategy (RIS)” through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (MOE) (2022RIS-005).

This work involved human subjects. Approval for all ethical and experimental procedures and protocols was granted by the Department of Internal Medicine of Gyeongsang National University Hospital, South Korea, under application no. GNUH 2022-05-033.

ABSTRACT Gastric cancer is high-risk cancer in terms of both incidence and mortality. However, if it is diagnosed early, there is a high chance of survival. Therefore, an early diagnosis of gastric cancer and precancerous lesions is very important. Gastroscopy is one of the best methods for diagnosing gastric cancer and precancerous lesions, but it relies on visual observation by medical specialists. Accordingly, factors such as the experience or fatigue of specialists can influence diagnosis results. To alleviate these problems, we propose a computer-aided diagnosis system that can improve the efficiency of diagnosis and reduce misdiagnoses by providing a second opinion. We aimed to classify healthy tissue, gastric lesions, and early gastric cancer using a Transformer-based deep-learning classification model called Vision Transformer, which has achieved the best performance in transfer learning. We also proposed a Multi-Filter AutoAugment (MFAA) method, which increases the classification performance of the model given small amounts of medical data. The medical data augmented using MFAA are better for training deep-learning models than conventionally augmented data; we effectively enhanced the classification performance of the model using MFAA. In experiments, the model achieved an F1-score of 0.87 and area under the curve of 0.94 in the classification of abnormalities (gastric lesions including early gastric cancer) and healthy tissue. In addition, it obtained an F1-score of 0.92 and area under the curve of 0.97 in the classification of early gastric cancer and non-cancerous gastric lesions.


INDEX TERMS Computer-aided diagnosis (CADx), deep-learning, early gastric cancer, gastric lesions, gastroscopy, medical data, multi-filter AutoAugment (MFAA).

I. INTRODUCTION

Gastric cancer, also known as stomach cancer, has the fifth highest incidence worldwide in both sexes according to the Global Cancer Observatory of the World Health Organization [1]. In addition, its mortality is ranked fourth. Considering both its incidence and mortality, it can be considered a high-risk cancer. There is a large difference in the survival

rate of patients with gastric cancer depending on when it is diagnosed. A survival rate of >90% is expected when gastric cancer is diagnosed early [2]. However, if diagnosis is delayed, the survival rate decreases sharply. Therefore, the early diagnosis of gastric cancer is an important factor for patient survival. However, early diagnosis of asymptomatic gastric cancer can be difficult [3].

Gastroscopy is one of the best methods for the early diagnosis of gastric cancer [4]. It is a common method in many countries, and a variety of high-performance equipment is

The associate editor coordinating the review of this manuscript and approving it for publication was Donato Impedovo .

used to develop this technique. Moreover, it has the advantage of being able to assist a medical specialist to diagnose early gastric cancer and precancerous lesions at a reasonable cost without any adverse effects on the patient. However, gastroscopy is limited in that it requires visual observation by a specialist. Therefore, factors such as fatigue and the skill of the specialist can influence the diagnosis results. Computer-aided diagnosis (CADx) systems can solve this problem by analyzing medical data and providing a second opinion to specialists [5]. Recently, research on deep-learning-based CADx systems has been actively conducted. Such systems quickly provide useful information to specialists based on accumulated medical data. Furthermore, the accuracy of the diagnosis can be improved by preventing subjective judgment or misdiagnosis by a specialist, thereby reducing the burden on the specialist. However, collecting medical data, which is the most important step in the development of CADx based on deep-learning, is difficult because of the process limitations. The premise that a lesion has occurred is first required, and then an approval process to protect patient information along with patient consent is essential [6]. The volume of medical data that can be obtained through this process is inevitably smaller than the volume of other types of data, and this is a disadvantage for deep-learning-based CADx systems, where the quantity and quality of the data are the most important factors affecting performance.

In this study, we propose an enhanced computer-aided gastroscopy diagnosis system that uses a Multi-Filter AutoAugment (MFAA) to address the problems mentioned above. The computer-aided gastroscopy diagnosis system uses a deep-learning-based classification model and aims to classify healthy tissue, gastric lesions, and early gastric cancer. MFAA is based on AutoAugment [7] and is a new data-augmentation method. This approach can reduce the quality degradation caused by conventionally augmenting medical data, which is a problem in CADx systems, where accuracy is the most important factor. In addition, it can effectively improve the performance of CADx systems with only a small amount of medical data.

II. RELATED WORK

Various studies have been conducted to diagnose gastric lesions or cancers using a CADx system. Muto et al. [8] recommended a diagnostic algorithm for early gastric cancer using magnifying endoscopy. The aim of this study was to distinguish between cancerous and non-cancerous but suspicious lesions. The vessel plus surface classification system developed by Yao et al. [9] was used. The system achieved an accuracy of 0.95, positive predictive value of 0.79, and negative predictive value of 0.99 in the diagnosis of gastric cancer. However, it was difficult for the system to diagnose lesions that exhibited contact bleeding or were covered with mucus.

Zhu et al. [10] developed a convolutional neural network (CNN) that automatically detected gastric cancer in endoscopic images. Their proposed diagnostic system was based

on a single-shot multi-box detector [11]. A total of 13,584 endoscopic images of gastric cancer were used for learning and 2,296 images were used for testing. It obtained an overall sensitivity of 0.92, and 161 non-cancerous lesions were detected as gastric cancer, resulting in a positive predictive value of 0.31.

Hirasawa et al. [12] constructed a CNN-based computer-aided detection system based on endoscopic images to determine invasion depth. Two classes, P0 and P1, were classified according to the invasion depth of the tumor. They used the ResNet-50 architecture [13], 790 images as the training dataset, and 203 images as the test dataset. The area under the curve (AUC) of the proposed method was 0.94, which was equivalent to that of experienced endoscopists and higher than that of junior endoscopists.

Horiuchi et al. [14] investigated the differences in diagnostic performance between expert endoscopists and a CADx system using magnifying endoscopy with narrow-band imaging. A total of 1,492 cancerous and 1,078 non-cancerous images were used to fine-tune a pre-trained GoogLeNet model [15]. A total of 174 videos (87 cancerous and 87 non-cancerous) were used to evaluate the diagnostic performance of the proposed system along with 11 experts. The proposed system achieved an AUC of 0.87, which was equivalent to or better than that of several experts.

Ali et al. [16] compared color features in different color spaces to detect abnormal areas in chromoendoscopy. A support vector machine [17] classifier was trained on color features and hybrid color–texture characteristics for gastric lesions. A total of 176 normal and gastric lesions images were used for the performance evaluation, and the proposed method achieved an accuracy of 0.87 and AUC of 0.91 in the classification of gastric lesions.

Iizuka et al. [18] trained a CNN based on an Inception-v3 network architecture [19] and a recurrent neural network consisting of two long short-term memory networks [20] on biopsy histopathology whole-slide images (WSIs) of the stomach and colon. The models were trained to classify WSIs as adenocarcinomas, adenomas, and non-neoplastic. The stomach dataset used for training and testing consisted of 4,128 WSIs and the colon dataset consisted of 4,036 WSIs. This approach yielded AUCs of 0.97 and 0.99 for gastric adenocarcinoma and adenoma, respectively, and 0.96 and 0.99 for colonic adenocarcinoma and adenoma, respectively.

Ueyama et al. [21] constructed an artificial intelligence (AI) assisted CNN CADx system to diagnose early gastric cancer using a ResNet-50 architecture based on magnifying narrow-band imaging (ME-NBI) images. It was trained and validated on a dataset of 5,574 ME-NBI images (3,797 images of early gastric cancers (EGCs) and 1,777 images of non-cancerous lesions). To evaluate the diagnostic accuracy, a separate test dataset of 2,300 ME-NBI images (1,430 images of EGCs and 870 images of non-cancerous lesions) was used. The proposed system required 60 s to analyze 2,300 test images and exhibited a sensitivity of 0.98.

Chen et al. [22] proposed a GasHis-Transformer model and its lightweight version, called the LW-GasHis-Transformer, to detect gastric cancer in histopathological images. A publicly available hematoxylin and eosin stained gastric histopathological image dataset was used for training and testing. They used image augmentation with a normalization process to speed up the model learning, and training was conducted using the global information module, following the concept of BoTNet-50 [23] and the local information module, which is based on the concept of Inception-v3. The results of the two models achieved accuracies of 97.97% and 96.43% on the gastric cancer histopathology dataset.

The related literature reveals that various CADx systems have been used with gastroscopy. Studies have been conducted to diagnose various factors such as early gastric cancer, gastric lesions, and the invasion depth of a tumor. All the studies have shown that CADx systems for gastroscopy can be helpful. They have also exhibited very high performance owing to the application of deep-learning techniques.

Our study proposes a model to classify healthy tissue, gastric lesions, and early gastric cancer using a very recent and best-performing deep-learning model, Vision Transformer (ViT). However, the amount of medical data used in the above studies is limited compared to the amount of data used in studies in other fields. Despite the long collection periods or the collaboration of several institutions, the amount of available data was not significant compared with that available in other fields. Accordingly, we also propose the MFAA, which can address the difficulty in medical data collection.

III. MATERIALS AND METHODS

This study aimed to develop a deep-learning-based CADx system for gastroscopy. This system consists of two parts.

First, abnormalities (gastric lesions, including early gastric cancer) and healthy tissue are classified, and then early gastric cancer and non-cancerous gastric lesions are classified. In addition, we use MFAA on all images to improve the classification performance even when only a small amount of medical data is used. The overall structure of the proposed method is illustrated in Fig. 1.

A. GASTROSCOPY DATASET

Gastroscopic images of healthy tissue, abnormalities, and early gastric cancer were collected from the Department of Internal Medicine of Gyeongsang National University Hospital, South Korea. All gastroscopic images were approved by the Institutional Review Board (IRB File No. GNUH 2022-05-033), and biopsy-verified. Gastroscopic images with a resolution of 1242×1080 pixels were obtained from 192 patients. For the images of healthy tissue, 600 gastroscopic images were obtained from 96 patients. For the images of abnormalities such as benign gastric ulcers, bleeding, blood clots, chronic gastritis, erosive gastritis, erythema, hemorrhages, metaplastic gastritis, polyps, submucosal tumors, and xanthoma, 300 gastroscopic images were

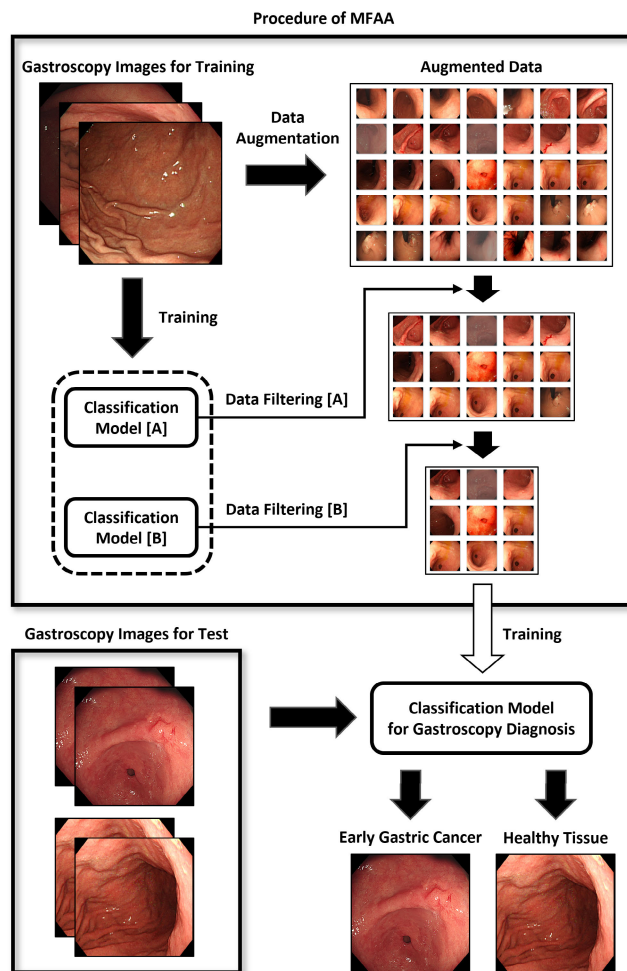


FIGURE 1. Structure of the proposed gastroscopy diagnosis system.

obtained from 48 patients. A further 300 gastroscopic images of early gastric cancer were obtained from 48 patients.

Two datasets were constructed using the obtained gastroscopic images. Dataset A was used for the classification of abnormalities and healthy tissue, while Dataset B was used for the classification of early gastric cancer and gastric lesions. The gastroscopic images included in each dataset were divided for training and performance evaluation. They were divided as evenly as possible, and care was taken not to include gastroscopic images of the same patient in both datasets. Moreover, to increase the reliability of the study, gastroscopic images of early gastric cancer obtained from the Catholic University of Korea St. Mary's Hospital provided by the AI Hub of the National Information Society Agency were used for performance evaluation [24]. A total of 3,285 gastroscopic images of early gastric cancer obtained from 225 patients were used; the resolution was 640×480 , and none were not used for training, only for performance evaluation. The configurations of Datasets A and B are shown in Table 1, and examples of the gastroscopic images are shown in Fig. 2.

TABLE 1. Specification of the gastroscopy dataset.

Dataset A	Train	Test	Total
Healthy Condition	300	300	600
Abnormalities	300	300	600
- Early gastric cancer	152	148	300
- Benign gastric ulcer	28	27	55
- Erosive gastritis	24	29	53
- Submucosal tumor	27	19	46
- Others	69	77	146
Dataset B	Train	Test	Total
Early Gastric Cancer	152	148	300
Early Gastric Cancer (AI Hub)	0	3,285	3,285
Gastric Lesions	148	152	300
- Benign gastric ulcer	28	27	55
- Erosive gastritis	24	29	53
- Submucosal tumor	27	19	46
- Others	69	77	146

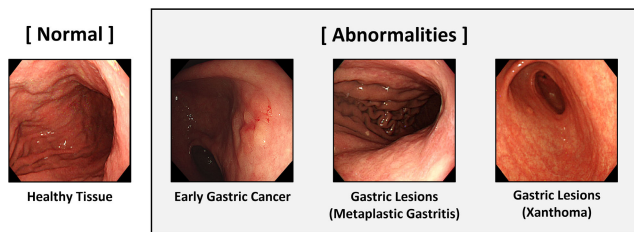


FIGURE 2. Examples of images obtained during gastroscopy.

B. CLASSIFICATION MODELS

ViT was used to classify the gastroscopic images [25]. Among the ViT models, ViT-H/14 pre-trained on the public ImageNet-21k dataset was used [26]. The ViT uses a Transformer architecture for classification using fixed-size image patches as the input. The detailed training procedure is as follows.

$$z_0 = [x_{class}; x_p^1 E; x_p^2 E; \dots; x_p^N E] + E_{pos} E \in R^{(P^2 \cdot C) \times D}, \quad E \in R^{(N+1) \times D} \quad (1)$$

Equation (1) corresponds to the input of the ViT, where x_{class} represents a classification token and $x_p^N E$ represents an image sequence divided into patches.

$$z_l = MSA(LN(z_{l-1})) + z_{l-1}, \quad l = 1 \dots L$$

$$z_l = MLP(LN(z_l)) + z_l, \quad l = 1 \dots L \quad (2)$$

Equation (2) is used in the transformer architecture. Layer normalization is applied to the previous input value and

multi-head attention is employed. Subsequently, a skip connection is added. In this procedure, z'_l is the output from the transformer, and z_l is the output of the multi-layer perceptron head of the ViT, which takes z'_l as its input. Finally, classification is performed as follows:

$$y = LN(z'_L) \quad (3)$$

ViT does not have inductive biases because of the characteristics of the transformer architecture. It also does not have the locality and translation equivariance characteristics of CNNs. Therefore, it is not necessary to use a large amount of data for training because it does not use a local receptive field but determines the position of an instance of a class by perceiving the global image. However, if pre-training is performed using a large amount of data, the model can identify both local and global features and perform better on highly complex images. This can be a strength in medical data, which are high in complexity and often used for transfer learning, mainly because of the small amount of data.

Moreover, the transformer architecture of the ViT models does not have parameter limitations. As a result, the possibility of overfitting is very small, and the probability of making a biased diagnosis using pre-training data is low. Therefore, it was determined that the use of a ViT is appropriate for medical tasks, where accurate and unbiased diagnoses are the most important factors.

An additional classification model for multi-filter data augmentation called big transfer (BiT) was used [27]. BiT is a model that aims to achieve the best performance in transfer learning. A ResNet architecture is used as the backbone, and during transfer learning, the upstream layers corresponding to pre-training and the downstream layers corresponding to fine-tuning, are separated. The BiT-HyperRule, which is a set of parameter values that obtains the highest transfer-learning and classification performance, was also reported. Accordingly, it was found that BiT achieves acceptable performance on medical data tasks. Among the BiT models, BiT-L, which is pre-trained on the public ImageNet-21k dataset, was used. In summary, ViT-H/14 was used as the main classification model and for filtering, and BiT-L was used for filtering the augmented medical data.

C. MFAA

In this study, we present MFAA, which can effectively improve classification performance even in situations where there are small amounts of data. MFAA is based on AutoAugment, which selects the optimal sub-policy for the training dataset. AutoAugment sets the augmentation policy S and validation accuracy R and then adjusts S to increase R through a recurrent neural network. Thus, the 25 sub-policies that obtained the best performance in the training dataset were selected. Each sub-policy consists of two augmentation operations, which are composed of operations such as changes to brightness, contrast, equalization, inversion, posterization, and shear. AutoAugment provides augmentation policies based on three datasets: CIFAR-10, ImageNet, and

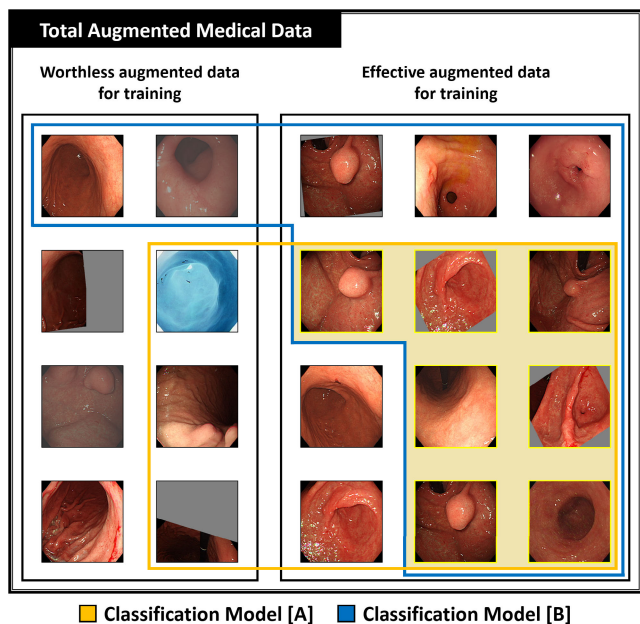


FIGURE 3. Operation of MFAA.

street view house numbers (SVHN). However, the use of an augmentation policy for the SVHN dataset, which consists of street view house numbers, was considered inappropriate for this study and was excluded [28]. Therefore, augmentation policies obtained using CIFAR-10 and ImageNet were applied to the gastroscopic images, and the number of gastroscopic images was increased 50-fold through the 25 sub-policies for CIFAR-10 and 25 sub-policies for ImageNet. Even if AutoAugment is used for data augmentation, the features in the augmented data necessary to classify the target object can be incorrectly added, transformed, or lost. Therefore, using all augmented data may interfere with the performance or may not result in significant improvement. MFAA can be a solution to this problem.

MFAA is configured by adding a data filtering procedure using the classification model for augmented data through AutoAugment. The weights of the classification model trained using the original data contain the necessary feature information of the target object. Using these points, we filter the augmented data. As a result, it is possible to leave only the appropriate augmented data generated by AutoAugment. Furthermore, classification models extract and use different features to classify objects depending on their architecture and purpose. Accordingly, when filtering using various classification models, as shown in Fig. 3, only augmented data with essential features for classifying objects are left, and this is the main advantage of MFAA. In the proposed study, two classification models, ViT and BiT, were trained and used for MFAA. Filtering through the two classification models for augmented gastroscopic images leaves only gastroscopic images with classification accuracies higher than 0.9. Descriptions of training methods for ViT and BiT are provided in the following sections. The detailed specifications

TABLE 2. Specifications of the augmented gastroscopy datasets.

Dataset A (Training)	Healthy Tissue	Abnormalities	Total
Original	300	300	600
AutoAugment	15,000	15,000	30,000
1-Filter	11,554	10,265	21,819
Ours (2-Filter, MFAA)	9,027	8,563	17,590
Dataset B (Training)	Early Gastric Cancer	Gastric Lesions	Total
Original	152	148	300
AutoAugment	7,600	7,400	15,000
1-Filter	6,224	5,083	11,307
Ours (2-Filter, MFAA)	5,708	3,747	9,455

of the dataset obtained using AutoAugment and MFAA are presented in Table 2.

D. ENVIRONMENTAL SETUP AND TRAINING

All training and performance evaluations were conducted on a system using Windows 10 x 64, CUDA 11.3 with cuDNN, Python 3.9.6, and PyTorch 1.12, with the following configuration: Intel® Core™ i9-12900KS Processor, NVIDIA RTX A6000, and 64GB RAM.

First, the model for the classification of abnormalities and healthy condition was trained. The ViT-H/14 model pre-trained on ImageNet-21k was used as the classification model. Gastroscopic images belonging to the training set of Dataset A and gastroscopic images augmented by MFAA were used for training. The training dataset was divided randomly using a ratio of 8:2 for training and validation. However, when the training dataset contained augmented data, the entire training dataset was not randomly divided into an 8:2 ratio. A division of 8:2 was applied to each set of the original and augmented data constituting the training dataset. This prevents biased configurations.

During the training of the classification model, the weights were selected using objective judgment and the selection was based on the validation loss. If the validation loss of the weights did not improve within five epochs or showed only a very small change, the first of the five epochs was considered the optimal point of training, and the weights for that epoch were selected. Using this approach, we generated and compared four methods (original, AutoAugment, 1-Filter, and our MFAA method) for the classification of abnormalities and healthy tissues.

Second, training for the classification of early gastric cancer and non-cancerous gastric lesions was also conducted. Gastroscopic images belonging to the training set of Dataset B and gastroscopic images augmented by MFAA were used for training. All other training procedures were

conducted using the same conditions as those used in the previous training method. Using this approach, we developed and compared the same four methods for the classification of early gastric cancer and gastric lesions.

For MFAA, two sets of weights trained using the ViT-H/14 classification model from the original gastroscopic images belonging to Dataset A and Dataset B were used. We used two sets of weights generated by changing the ViT-H/14 classification model to BiT-L classification model. The ViT-H/14 and BiT-L are model types based on the model sizes of ViT and BiT, respectively, and these two models yield stable classification performance because of their large model sizes compared with other model types.

IV. RESULTS AND DISCUSSION

The results of the proposed study were evaluated based on classification performance. Sensitivity, F1-score, and AUC were used as the key indicators for performance evaluation. Sensitivity was used to confirm the number of abnormalities correctly classified in the class of abnormalities and healthy tissues as well as the number of EGCs correctly classified in the class of early gastric cancer and gastric lesions. The F1-score was used to verify the overall performance instead of accuracy because the data used in the test differed in proportions. The AUC, which is mainly used in performance evaluation in medical research, was obtained by calculating the receiver operating characteristic (ROC) curve. The ROC curves obtained for the datasets by each method are shown in Fig. 4. The overall performance of the proposed method with detailed criteria such as the numbers of TP (true positives), TN (true negatives), FP (false positives), and FN (false negative) is summarized in Table 3. The highest performances in terms of precision, sensitivity, F1-score, and AUC of each part are indicated in bold.

In the first experiment, the classification performance of abnormalities and healthy tissues was evaluated. The abnormalities include gastric lesions and early gastric cancer. A performance evaluation was conducted on 300 gastroscopic images of abnormalities obtained from 48 patients and 300 gastroscopic images of healthy tissues obtained from the 48 patients included in Dataset A.

The first row shows the results of the trained weights obtained from the original data without augmentation. The second row lists the performance of the model trained on the original data and data augmented using AutoAugment. The third row lists the performance of the model trained on the original data and augmented data filtered by a single classification model. The last row lists the performance of the model trained on the original data and augmented data filtered through two classification models according to the proposed MFAA approach. The classification performance obtained using MFAA was the highest overall. Both the F1-score and AUC showed a performance improvement of approximately 2% compared to the original values. In the case of sensitivity, the original model yielded the highest performance but was very close to the value obtained using MFAA.

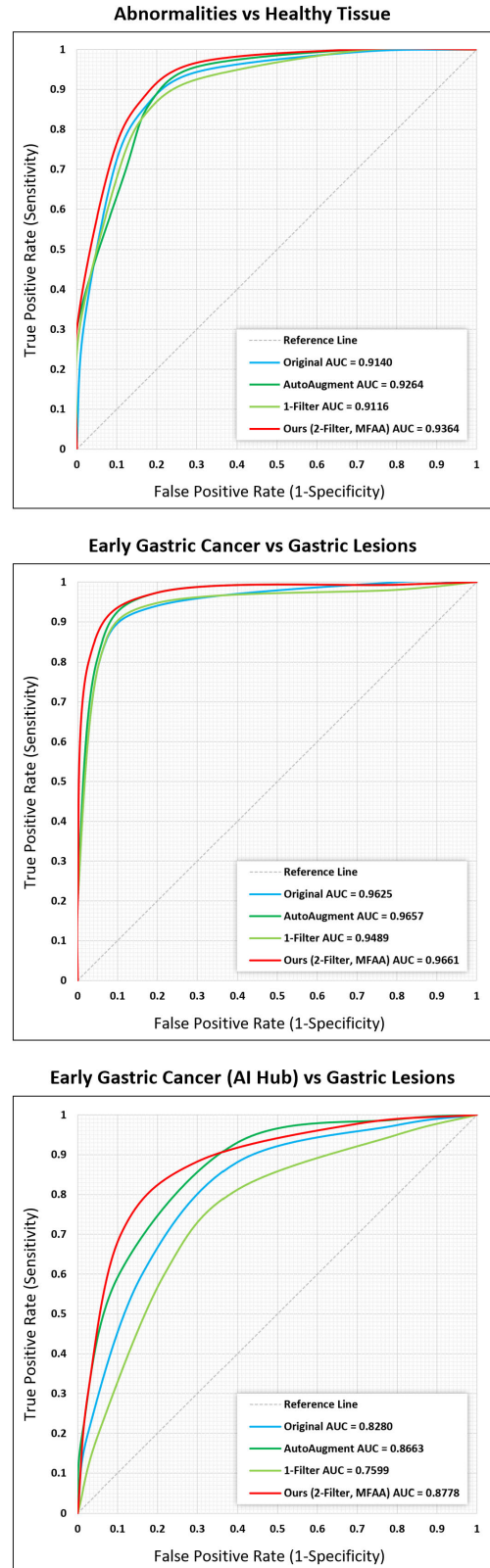


FIGURE 4. ROC curves according to the dataset and methods.

In this classification, healthy tissue has one category. However, abnormalities, although grouped into one category, are composed of various gastric lesions. Accordingly, unlike the

TABLE 3. Performance evaluation of gastroscopy classification.

Abnormalities vs Healthy Tissue	TP	FN	FP	TN	Precision	Sensitivity (=Recall)	F1-score	AUC
Original	271	29	67	233	0.8018	0.9033	0.8495	0.9140
AutoAugment	247	53	49	251	0.8345	0.8233	0.8289	0.9264
1-Filter	250	50	47	253	0.8418	0.8333	0.8375	0.9116
Ours (2-Filter, MFAA)	269	31	49	251	0.8459	0.8967	0.8706	0.9364
Early Gastric Cancer vs Gastric Lesions	TP	FN	FP	TN	Precision	Sensitivity (=Recall)	F1-score	AUC
Original	136	12	22	130	0.8608	0.9189	0.8889	0.9625
AutoAugment	137	11	14	138	0.9073	0.9257	0.9164	0.9657
1-Filter	140	8	30	122	0.8235	0.9459	0.8805	0.9489
Ours (2-Filter, MFAA)	139	9	16	136	0.8968	0.9392	0.9175	0.9661
Early Gastric Cancer (AI Hub) vs Gastric Lesions	TP	FN	FP	TN	Precision	Sensitivity (=Recall)	F1-score	AUC
Original	1,271	2,014	22	130	0.9830	0.3869	0.5553	0.8280
AutoAugment	1,240	2,045	13	139	0.9896	0.3775	0.5465	0.8663
1-Filter	1,299	1,986	30	122	0.9774	0.3954	0.5631	0.7599
Ours (2-Filter, MFAA)	2,269	1,016	16	136	0.9930	0.6907	0.8147	0.8778

augmentation for healthy tissue images, augmentation was achieved for each lesion and led to abnormalities including too many features. Therefore, worthless augmented data were added to the training dataset. This can be confirmed by the decrease in sensitivity caused by AutoAugment and 1-Filter. Therefore, it is concluded that some of the worthless augmented data were filtered out by MFAA, and the performance improved again in the proposed model (2-Filter, MFAA). Although the sensitivity was very slightly lower, a clear performance improvement in the F1-score and AUC was confirmed, which means that the proposed MFAA performed effectively in improving the classification of medical data.

In the second experiment, the early gastric cancer and gastric lesions classification performance was evaluated. This was performed to confirm whether early gastric cancer could be accurately diagnosed from data that includes various gastric lesions. The evaluation was conducted on 148 gastroscopic images of early gastric cancer obtained from 24 patients and 152 gastroscopic images of gastric lesions obtained from 24 patients. The evaluation process and arrangement of rows were the same as those described in the previous experiment. The results of the application of the proposed method also yielded the highest overall performance. Compared to the original values, the F1-score was

improved by approximately 3%, and the AUC was maximized with a slight increase. In the case of sensitivity, the 1-Filter augmented data obtained the best performance, which was approximately 1% higher than that of the proposed method. However, there was a difference of 3% in the F1-score and 2% in the AUC; it can be observed that the performance of the classification model based on the proposed method is much better than when the 1-Filter is used.

Finally, to increase the reliability of the proposed study, early gastric cancer and gastric lesions were classified using extra-gastroscopic images of early gastric cancer obtained from other institutions. We used 3,285 gastroscopic images of early gastric cancer from 225 patients provided by the AI Hub of the National Information Society Agency, South Korea. There was no separate training for this evaluation, and only tests were conducted. The evaluation process was the same as that described for the previous experiments. The only difference was that the gastroscopic images of early gastric cancer used in the performance evaluation were replaced with data from the AI Hub. Using the proposed method yielded the highest performance in terms of sensitivity, F1-score, and AUC. Sensitivity was very low for all methods except for the proposed method. MFAA led to a performance improvement of 30%. Even the overall performance of the F1-score was low, but MFAA substantially improved it by

approximately 25%. The AUC also improved by 5% compared to the original AUC.

The time complexity of the proposed study was confirmed through big O notation. The classification models of the proposed study have a process from training to testing. Training proceeds by repeatedly analyzing the dataset according to epoch and the testing proceeds to check the dataset once. Therefore, the dataset is set to n , and the original method can be expressed as (4):

$$f(n) + f(1) = f(n) = O(n) = n, \quad f(x) = O(g(x)) \quad (4)$$

In the case of the proposed MFAA, it consists of training twice to generate the weights to be used as filters, testing twice for filtering, training once through the MFAA dataset, and testing once. It also has an augmentation process that multiplies the dataset. The MFAA method can be expressed as (5):

$$3f(n) + 3f(1) + f(n) = f(n) = O(n) \quad (5)$$

To sum it up, execution time increases as the number of training and testing increase. However, the time complexity remains the same as $O(n)$. Further, the procedure of the MFAA method, which takes additional time, does not need to be repeated. Accordingly, there is no difference in execution time between the original method and the proposed MFAA method after training.

In conclusion, it was confirmed that the application of MFAA led to higher performance in most situations. This means that the overall performance of the model was improved in the classification of healthy tissues, gastric lesions, and early gastric cancer. Moreover, it was confirmed that the performance improvement was not as high when AutoAugment was applied or when only one classification model was used for data filtering. This supports the finding that the proposed MFAA, which leaves only data with essential features, works better. In the classification of early gastric cancer data from other institutions, MFAA resulted in high performance improvements in all metrics. Objectively, it can be concluded that the performance was not remarkably high compared to research in other fields. However, the training was conducted based on a very small number of gastroscopic images of early gastric cancer (148), and 3,285 gastroscopic images of early gastric cancer were classified. Therefore, it can be considered that the performance is relatively good, and the performance improvement obtained using the proposed method is encouraging.

V. CONCLUSION

In this study, we proposed a CADx system that can classify healthy tissues, gastric lesions, and early gastric cancer. The transformer-based deep-learning model ViT, which obtains the best classification performance and good characteristics for transfer learning using small amounts of medical data, was used as the classification model. We also proposed MFAA, which can effectively increase the classification performance of models trained on small amounts of medical

data. We experimentally verified the proposed gastroscopy classification using MFAA. The application of the proposed method to the data used to train the classification model led to high performance improvements in all classifications of healthy tissues, gastric lesions, and early gastric cancer, with an F1-score of 0.92 and AUC of 0.97. We also obtained very effective performance improvements in the classification of extra gastroscopy data obtained from an independent institution.

Although the proposed study obtained encouraging results, many challenges remain for future research. First, we intend to confirm the potential of the proposed MFAA with medical data from various fields beyond gastroscopy. Second, we plan to collect more gastroscopic images and apply MFAA to compare the performance improvements. In this study, two classification models were used for filtering, and we would like to determine whether a higher performance improvement is possible by increasing the number of classification models used for filtering. Finally, we would like to apply various classification models as well as ViT and BiT to gastroscopy diagnosis, select an optimized classification model, and conduct a study to improve the classification performance by changing the structure of the selected model.

REFERENCES

- [1] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray, "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA, Cancer J. Clinicians*, vol. 71, no. 3, pp. 209–249, May 2021.
- [2] J. Wang, J.-C. Yu, W.-M. Kang, and Z.-Q. Ma, "Treatment strategy for early gastric cancer," *Surgical Oncol.*, vol. 21, no. 2, pp. 119–123, Jun. 2012.
- [3] A. Axon, "Symptoms and diagnosis of gastric cancer at early curable stage," *Best Pract. Res. Clin. Gastroenterol.*, vol. 20, no. 4, pp. 697–708, Jan. 2006.
- [4] K. Yao, "The endoscopic diagnosis of early gastric cancer," *Ann. Gastroenterol.*, vol. 26, p. 11, Jan. 2013.
- [5] Y. Zhou, U. Teomete, O. Dandin, O. Osman, T. Dandinoglu, U. Bagci, and W. Zhao, "Computer-aided detection (CADx) for plastic deformation fractures in pediatric forearm," *Comput. Biol. Med.*, vol. 78, pp. 120–125, Nov. 2016.
- [6] K. J. Cios and G. William Moore, "Uniqueness of medical data mining," *Artif. Intell. Med.*, vol. 26, nos. 1–2, pp. 1–24, Sep. 2002.
- [7] E. D. Cubuk, B. Zoph, D. Mane, V. Vasudevan, and Q. V. Le, "AutoAugment: Learning augmentation policies from data," 2018, *arXiv:1805.09501*.
- [8] M. Muto, K. Yao, M. Kaise, M. Kato, N. Uedo, K. Yagi, and H. Tajiri, "Magnifying endoscopy simple diagnostic algorithm for early gastric cancer (MESDA-G)," *Digestive Endoscopy*, vol. 28, no. 4, pp. 379–393, May 2016.
- [9] K. Yao, G. Anagnostopoulos, and K. Ragnath, "Magnifying endoscopy for diagnosing and delineating early gastric cancer," *Endoscopy*, vol. 41, no. 5, pp. 462–467, May 2009.
- [10] Y. Zhu, Q.-C. Wang, M.-D. Xu, Z. Zhang, J. Cheng, Y.-S. Zhong, Y.-Q. Zhang, W.-F. Chen, L.-Q. Yao, P.-H. Zhou, and Q.-L. Li, "Application of convolutional neural network in the diagnosis of the invasion depth of gastric cancer based on conventional endoscopy," *Gastrointestinal Endoscopy*, vol. 89, no. 4, pp. 806–815, 2019.
- [11] W. Liu, D. Angelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot MultiBox detector," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 21–37.
- [12] T. Hirasawa, K. Aoyama, T. Tanimoto, S. Ishihara, S. Shichijo, T. Ozawa, T. Ohnishi, M. Fujishiro, K. Matsuo, J. Fujisaki, and T. Tada, "Application of artificial intelligence using a convolutional neural network for detecting gastric cancer in endoscopic images," *Gastric Cancer*, vol. 21, no. 4, pp. 653–660, 2018.

- [13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
- [14] Y. Horiuchi, T. Hirasawa, N. Ishizuka, Y. Tokai, K. Namikawa, S. Yoshimizu, A. Ishiyama, T. Yoshio, T. Tsuchida, J. Fujisaki, and T. Tada, "Performance of a computer-aided diagnosis system in diagnosing early gastric cancer using magnifying endoscopy videos with narrow-band imaging (with videos)," *Gastrointestinal Endoscopy*, vol. 92, no. 4, pp. 856–865, Oct. 2020.
- [15] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 1–9.
- [16] H. Ali, M. Sharif, M. Yasmin, and M. H. Rehmani, "Color-based template selection for detection of gastric abnormalities in video endoscopy," *Biomed. Signal Process. Control*, vol. 56, Feb. 2020, Art. no. 101668.
- [17] W. S. Noble, "What is a support vector machine?" *Nature Biotechnol.*, vol. 24, no. 12, pp. 1565–1567, 2006.
- [18] O. Iizuka, F. Kanavati, K. Kato, M. Rambeau, K. Arihiro, and M. Tsuneki, "Deep learning models for histopathological classification of gastric and colonic epithelial tumours," *Sci. Rep.*, vol. 10, no. 1, pp. 1–11, Jan. 2020.
- [19] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2818–2826.
- [20] R. C. Staudemeyer and E. Rothstein Morris, "Understanding LSTM—A tutorial into long short-term memory recurrent neural networks," 2019, *arXiv:1909.09586*.
- [21] H. Ueyama, Y. Kato, Y. Akazawa, N. Yatagai, H. Komori, T. Takeda, K. Matsumoto, K. Ueda, K. Matsumoto, M. Hojo, T. Yao, A. Nagahara, and T. Tada, "Application of artificial intelligence using a convolutional neural network for diagnosis of early gastric cancer based on magnifying endoscopy with narrow-band imaging," *J. Gastroenterol. Hepatol.*, vol. 36, no. 2, pp. 482–489, Feb. 2021.
- [22] H. Chen, C. Li, G. Wang, X. Li, M. Mamunur Rahaman, H. Sun, W. Hu, Y. Li, W. Liu, C. Sun, S. Ai, and M. Grzegorzec, "GasHis-transformer: A multi-scale visual transformer approach for gastric histopathological image detection," *Pattern Recognit.*, vol. 130, Oct. 2022, Art. no. 108827.
- [23] A. Srinivas, T.-Y. Lin, N. Parmar, J. Shlens, P. Abbeel, and A. Vaswani, "Bottleneck transformers for visual recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 16519–16529.
- [24] *AI Open Innovation Hub*. Accessed: Aug. 11/Sep. 19, 2022. [Online]. Available: <http://www.aihub.or.kr/>
- [25] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16×16 words: Transformers for image recognition at scale," 2020, *arXiv:2010.11929*.
- [26] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, and A. C. Berg, "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015.
- [27] A. Kolesnikov, L. Beyer, X. Zhai, J. Puigcerver, J. Yung, S. Gelly, and N. Houlsby, "Big transfer (BiT): General visual representation learning," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 491–507.
- [28] Y. Netzer, T. Wang, A. Coates, A. Bissacco, B. Wu, and A. Y. Ng, "Reading digits in natural images with unsupervised feature learning," in *Proc. NIPS Workshop Deep Learn. Unsupervised Feature Learn.*, 2011, pp. 1–9.



JUNG-WOO CHAE received the B.S. degree in electrical and electronic engineering and the M.S. degree in electrical and medical convergent engineering from Kangwon National University, South Korea, in 2019 and 2021, respectively, where he is currently pursuing the Ph.D. degree in the interdisciplinary graduate program in BIT medical convergence.



HYUN-CHONG CHO (Member, IEEE) received the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Florida, USA, in 2009. From 2010 to 2011, he was a Research Fellow with the University of Michigan, Ann Arbor, USA. From 2012 to 2013, he was a Chief Research Engineer with LG Electronics, South Korea. He is currently a Professor with the Department of Electronics Engineering and the Department of Interdisciplinary Graduate Program for BIT Medical, Kangwon National University, South Korea.

• • •