



Research Article

Development and Validation of Deep Learning Models for the Multiclassification of Reflux Esophagitis Based on the Los Angeles Classification

Hailong Ge,¹ Xin Zhou,² Yu Wang,¹ Jian Xu,¹ Feng Mo,¹ Chen Chao,¹ Jinzhou Zhu ³,
and Weixin Yu ¹

¹Department of General Surgery, Jintan Affiliated Hospital of Jiangsu University, Changzhou 213200, China

²Department of Gastroenterology, Jintan Affiliated Hospital of Jiangsu University, Changzhou 213200, China

³Department of Gastroenterology, The First Affiliated Hospital of Soochow University, Suzhou 215000, China

Correspondence should be addressed to Jinzhou Zhu; jzzhu@zju.edu.cn and Weixin Yu; ywx7097@163.com

Received 1 July 2022; Revised 16 August 2022; Accepted 6 February 2023; Published 18 February 2023

Academic Editor: Chin-Shiuh Shieh

Copyright © 2023 Hailong Ge et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This study is to evaluate the feasibility of deep learning (DL) models in the multiclassification of reflux esophagitis (RE) endoscopic images, according to the Los Angeles (LA) classification for the first time. The images were divided into three groups, namely, normal, LA classification A + B, and LA C + D. The images from the HyperKvasir dataset and Suzhou hospital were divided into the training and validation datasets as a ratio of 4 : 1, while the images from Jintan hospital were the independent test set. The CNNs- or Transformer-architectures models (MobileNet, ResNet, Xception, EfficientNet, ViT, and ConvMixer) were transfer learning via Keras. The visualization of the models was proposed using Gradient-weighted Class Activation Mapping (Grad-CAM). Both in the validation set and the test set, the EfficientNet model showed the best performance as follows: accuracy (0.962 and 0.957), recall for LA A + B (0.970 and 0.925) and LA C + D (0.922 and 0.930), Marco-recall (0.946 and 0.928), Matthew's correlation coefficient (0.936 and 0.884), and Cohen's kappa (0.910 and 0.850), which was better than the other models and the endoscopists. According to the EfficientNet model, the Grad-CAM was plotted and highlighted the target lesions on the original images. This study developed a series of DL-based computer vision models with the interpretable Grad-CAM to evaluate the feasibility in the multiclassification of RE endoscopic images. It firstly suggests that DL-based classifiers show promise in the endoscopic diagnosis of esophagitis.

1. Introduction

Gastroesophageal reflux disease (GERD) is a condition in which gastroesophageal reflux leads to esophageal mucosal lesions and troublesome symptoms [1, 2]. It is classified into reflux esophagitis (RE) with mucosal injuries and nonerosive reflux disease (NERD) only with symptoms [3]. Recently, the prevalence of RE has increased in Eastern Asia, due to the westernized lifestyle and diet [4, 5]. The severe complications of RE include ulcer bleeding and strictures. Even though RE-induced death is rare, these severe complications are related with significant morbidity and mortality rates [6].

According the Japan's 3rd guideline [3] and the Lyon Consensus [1], the paradigm of GERD diagnosis hinges on the identification of esophageal mucosal lesions and then the grading of the severity of esophagitis. In general, GERD is generally evaluated by clinical symptoms or responses to antisecretory therapy; however, the diagnosis required endoscopy and reflux monitoring [7]. Endoscopy is necessary to the grading of esophagitis, which plays a key role in the algorithms for the diagnosis and treatment of GERD [3]. The Los Angeles (LA) classification is the most widely used and validated scoring system to describe the endoscopic appearance of esophageal mucosa and stratify its severity [8].

Deep learning (DL) is a statistical learning method that empowers computers to extract features of raw data, including structured data, images, text, and audio, without human intervention. The remarkable progress of DL-based artificial intelligence (AI) has reshaped various aspects of clinical practice [9]. DL presents a significant advantage in the fields of computer vision to analyze medical images and videos containing gigantic quantities of information [10]. In gastroenterology, AI is increasingly being integrated into computer-aided diagnosis (CAD) systems to improve lesions detection and characterization in endoscopy [11]. To our best knowledge, there were no previous reports concerning the application of DL in the endoscopic classification of RE.

In this multicentral retrospective study, we aimed to evaluate the feasibility of DL models in the multi-classification of RE endoscopic images, according to the LA classification.

2. Methods

2.1. Datasets. Subjects who underwent the upper endoscopy were recruited from two hospitals as follows: (1) Suzhou: The First Affiliated Hospital of Soochow University and (2) Jintan: Affiliated Hospital of Jiangsu University, between 2015 and 2021. In the two centers, subjects were excluded if they have (1) esophagitis of other etiologies, e.g., pills-induced esophagitis, eosinophilic, radiation, and infectious esophagitis; (2) esophageal varices; (3) esophageal squamous cell cancer. This study was approved by the Ethics Committee of The First Affiliated Hospital of Soochow University and conducted in accordance with the Helsinki Declaration of 1975 as revised in 2000 (the IRB approval number 2022098). All participants signed statements of informed consent before inclusion. Besides, the Z-line endoscopic images were also obtained from an open dataset, HyperKvasir, which now is the largest dataset of the gastrointestinal endoscopy (<https://datasets.simula.no/hyper-kvasir/>) [12]. The dataset offers labeled/unlabeled/segmented image data and annotated video data from Bærum Hospital in Norway. The characteristic of the datasets was shown in Figure 1. Each endoscopic image of Z-line was determined and labeled as normal, LA classification A + B (LA A + B), or LA classification C + D (LA C + D) by three rich-experienced endoscopists, based on the LA classification. The endoscopic devices in our hospital include Olympus GIF-Q260, GIF-H290, and Fuji EG-601WR, while in the HyperKvasir dataset, they include Olympus and Pentax at the Department of Gastroenterology, Bærum Hospital.

2.2. Models

2.2.1. CNNs-Based Architectures. Pretrained convolutional neural networks (CNNs) include convolutional layers, average pooling layers, and fully connected layers, with ReLU activation. Besides, two dense layers (ReLU activation) and one dense layer (Softmax activation) were added on the top of the pretrained CNNs layers for feature extraction, as shown in Figure 2(a).

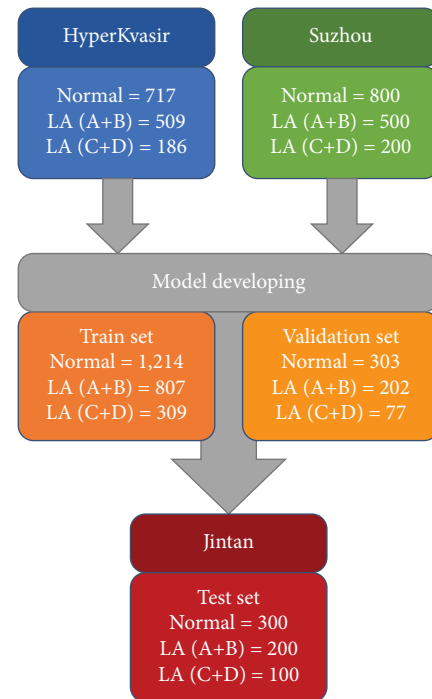


FIGURE 1: The characteristic of the datasets.

2.2.2. Transformer-Based Architectures. Transformer is characterized by synchronous input based on the self-attention mechanism. The Transformer encoder consists of three main components, namely, input embedding, multihead attention, and feed-forward neural networks. Similar as the CNNs, following them, three dense layers (ReLU or Softmax activation) were added on the top of the pretrained Transformer-based architectures.

2.2.3. Pretrained Models. The six CNNs-or Transformer-architectures models, i.e., MobileNet (MobileNet V1), ResNet (ResNet50 V2), Xception (Xception V1), EfficientNet (EfficientNet V2 small), ViT (ViT B/16), and ConvMixer (ConvMixer-768/32) were selected. These computer vision models were previously trained on the ImageNet database (<https://www.image-net.org>). The pretrained models and parameters were obtained from Keras or TensorFlow Hub (<https://hub.tensorflow.google.cn/>).

2.3. Training and Validation

2.3.1. Implementation. The CNNs-or Transformer-architectures models were transfer learning via Keras (TensorFlow framework as backbone). The Adam optimizer and the categorical cross-entropy cost function, with a fixed learning rate of 0.0001 and a batch size of 32, were compiled in the training of models. A link to the codes concerning the training procedure could be obtained here on https://osf.io/4tdhu/?view_only=b279429b6a284ad885da7cad79126df7.

2.3.2. Target Training. Endoscopic images of Z-line were saved as JPEG format. All images were rescaled to 331×331

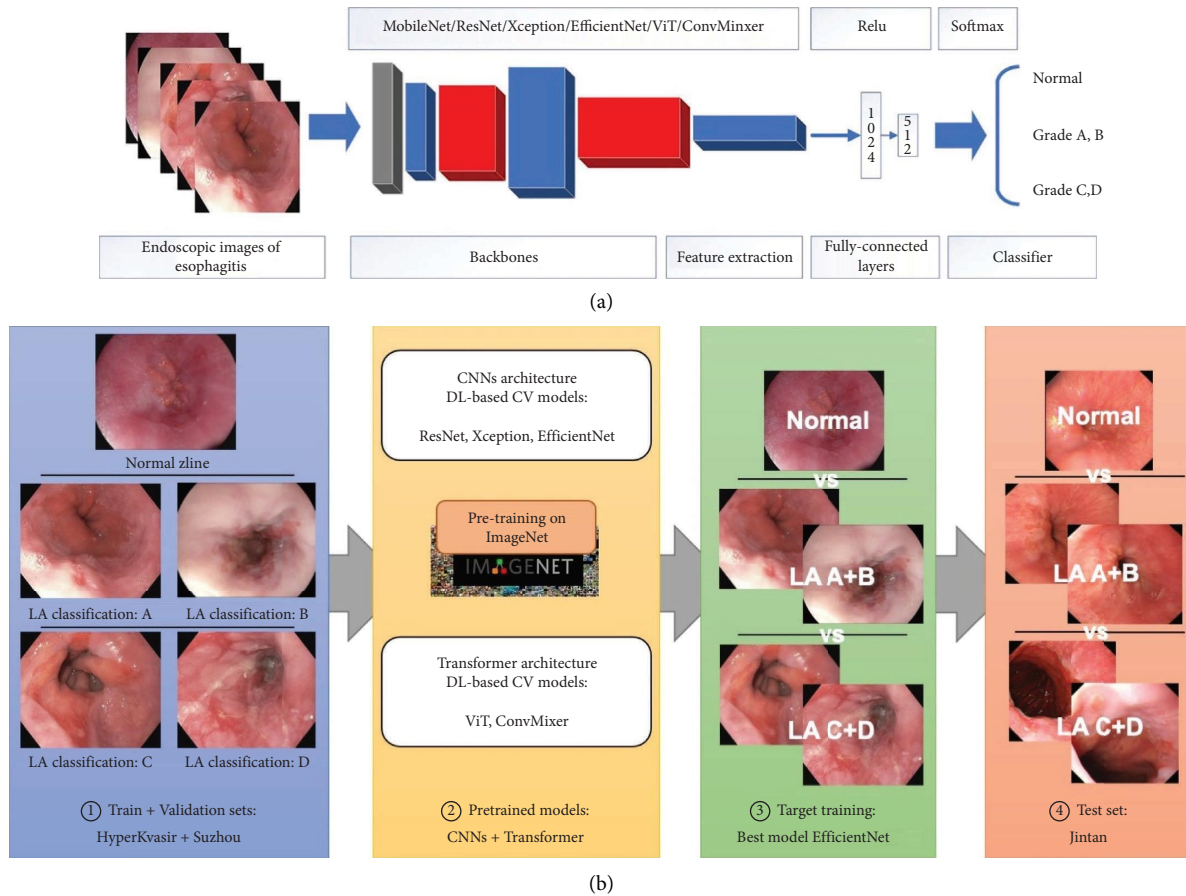


FIGURE 2: The flowchart of the study.

pixels and then the pixel values were normalized from 0 to 255 to 0 to 1. Based on the LA classification, the images were divided into three groups, namely, normal, LA A + B, and LA C + D. Images from the HyperKvasir dataset and Suzhou hospital were divided into the training and validation datasets as a ratio of 4:1. The flowchart of the study was plotted in Figure 2(b).

2.3.3. External Test. A total of 600 endoscopic images (as JPEG format) from Jintan hospital were the external test set, including 300 normal, 200 LA A + B, and 100 LA C + D (Figure 2(b)). The endoscopic devices in Jintan hospital include Olympus GIF-Q260 and GIF-H290.

2.3.4. Comparison with Endoscopists. To further evaluate the performance of the models, the images from the test dataset were determined by two endoscopists (junior, five-year endoscopic experience, and senior, more than ten-year experience).

2.3.5. Visualization of the Model. The visualization of the models was proposed using Gradient-weighted Class Activation Mapping (Grad-CAM) [13]. Grad-CAM uses the class-specific gradient information into the final convolutional layers of CNNs-based architectures to create

projecting maps of the key areas in the images without retraining. Based on the best multiclassification model, the Grad-CAM technology was to offer inferential explanation on the original images.

2.4. Statistical Analysis. The training of the models was performed on Python (version: 3.9) and TensorFlow (2.8.0). The performance was mainly evaluated by accuracy and recall. True positives (TP), true negatives (TF), false positives (FP), and false negatives (FN) were enumerated to assess the classifiers. Formulas were as follows: Accuracy = $(TP + TN) / (TP + FP + FN + TN)$, recall = $TP / (TP + FN)$, Marcor-recall = mean recalls, Matthew's correlation coefficient (MCC) = $(TP * TN - FP * FN) / \sqrt{(TP + FP) * (TP + FN) * (TN + FP) * (TN + FN)}$, and Cohen's kappa $k = (p_0 - p_e) / (1 - p_e)$ (p_0 : relative observed agreement among raters; p_e : hypothetical probability of chance agreement).

3. Results

3.1. Performance in the Validation Set. The confusion matrix of the six models in the validation set was plotted in Figure 3(a). The EfficientNet model showed the highest accuracy of 0.962, followed by the ConvMixer model (0.950) and Xception (0.938) (Table 1). The recalls for LA A + B and LA C + D of the EfficientNet model were 0.970 and 0.922,

TABLE 1: Performance metrics of models and endoscopists.

Models	Accuracy	Matthew's correlation coefficient	Cohen's kappa
<i>Validation dataset</i>			
MobileNet	0.916	0.859	0.820
ResNet	0.931	0.884	0.850
Xception	0.938	0.896	0.860
EfficientNet	0.962	0.936	0.910
ViT	0.933	0.888	0.850
ConvMixer	0.950	0.916	0.890
<i>Test dataset</i>			
MobileNet	0.916	0.821	0.780
ResNet	0.933	0.846	0.810
Xception	0.936	0.852	0.810
EfficientNet	0.957	0.884	0.850
ViT	0.938	0.854	0.820
ConvMixer	0.943	0.861	0.820
Junior endoscopist	0.916	0.820	0.780
Senior endoscopist	0.945	0.864	0.830

The bold figures indicate the highest numeric values.

while its Marco-recall was 0.946. In term of multiclass metrics, its MCC and Cohen's kappa were also highest (0.936 and 0.910).

3.2. Performance in the Test Set. The confusion matrix in the test set was plotted in Figure 3(b)). The EfficientNet model still presented the best performance. Its accuracy was 0.957, followed by ConvMixer (0.943) and Xception (0.936) (Table 1). Moreover, the recalls for LA A + B and LA C + D of the EfficientNet model were 0.925 and 0.930, while its Marco-recall reached 0.928, better than the other models. In term of multiclass metrics, its MCC and Cohen's kappa were still highest (0.884 and 0.850).

3.3. Comparison with the Endoscopists. In the test set, the junior endoscopist presented an accuracy of 0.916, recalls for LA A + B 0.885 and LA C + D 0.840, Marco-recall 0.863, MCC 0.820, and Cohen's kappa 0.780 (Table 1). In the meantime, the senior endoscopist showed an accuracy of 0.945, recalls for LA A + B 0.905 and LA C + D 0.890, Marco-recall 0.898, MCC 0.864, and Cohen's kappa 0.830.

3.4. The Grad-CAM Heatmap. According to the gradient information of the last convolution layer of the EfficientNet model, the Grad-CAM was plotted and highlighted the lesions of the original images (Figure 4). The left column displays the original endoscopic images. The middle column illustrates the Grad-CAM heatmap on the output of the last convolution layer. The right column shows the Grad-CAM heatmap added to the original endoscopic images, in which the highlighted regions reflect the lesions determined by the EfficientNet model.

The left column displays the original endoscopic images. The middle column illustrates the Grad-CAM heatmap on the output of the last convolution layer. The right column shows the Grad-CAM heatmap added to the original

endoscopic images, in which the highlighted regions reflect the lesions determined by the EfficientNet model.

4. Discussion

This study proposed a series of multiclassification computer vision models with the interpretable Grad-CAM to evaluate the feasibility of DL in the endoscopic images of RE, according to the LA classification. Six CNNs-or Transformer-architectures models were developed and the EfficientNet model showed practicable performance, better than the endoscopists.

In 1999, Lundell et al. [8] developed the LA classification to describe the mucosal appearance in endoscopy and to assess its correlation with the clinical changes in patients with RE. It was developed for the purpose of stratifying clinically relevant severity of RE. According to the LA classification, type A is defined as one (or more) mucosal break, no longer than 5 mm-long, that does not extend between the tops of two mucosal folds; type B is defined as one (or more) mucosal break, more than 5 mm, that still does not extend between the tops of two folds; type C is defined as one (or more) mucosal break that is continuous between the tops of two or more mucosal folds but which is no longer than 3/4 of the esophageal circumference; type D is defined as one (or more) mucosal break that is more than 3/4 of the circumference [8]. According to the Japan 2021 guideline [3] and the ACG 2021 guideline [7], RE is classified into mild RE (grade A or B of LA classification) and severe RE (grade C or D), in which the latter was defined as the high grade of RE, based on the Lyon Consensus [1]. The stratification is essential to the detailed diagnosis and the decision-making of therapy [3]. Thus, in this study, we labeled the images and trained the multiclassification models based on the forementioned guidelines. AI is being widely applied in a variety of clinical settings aiming to improve the management of the gastrointestinal diseases [14]. DL is a subset of machine learning that can automatically extract

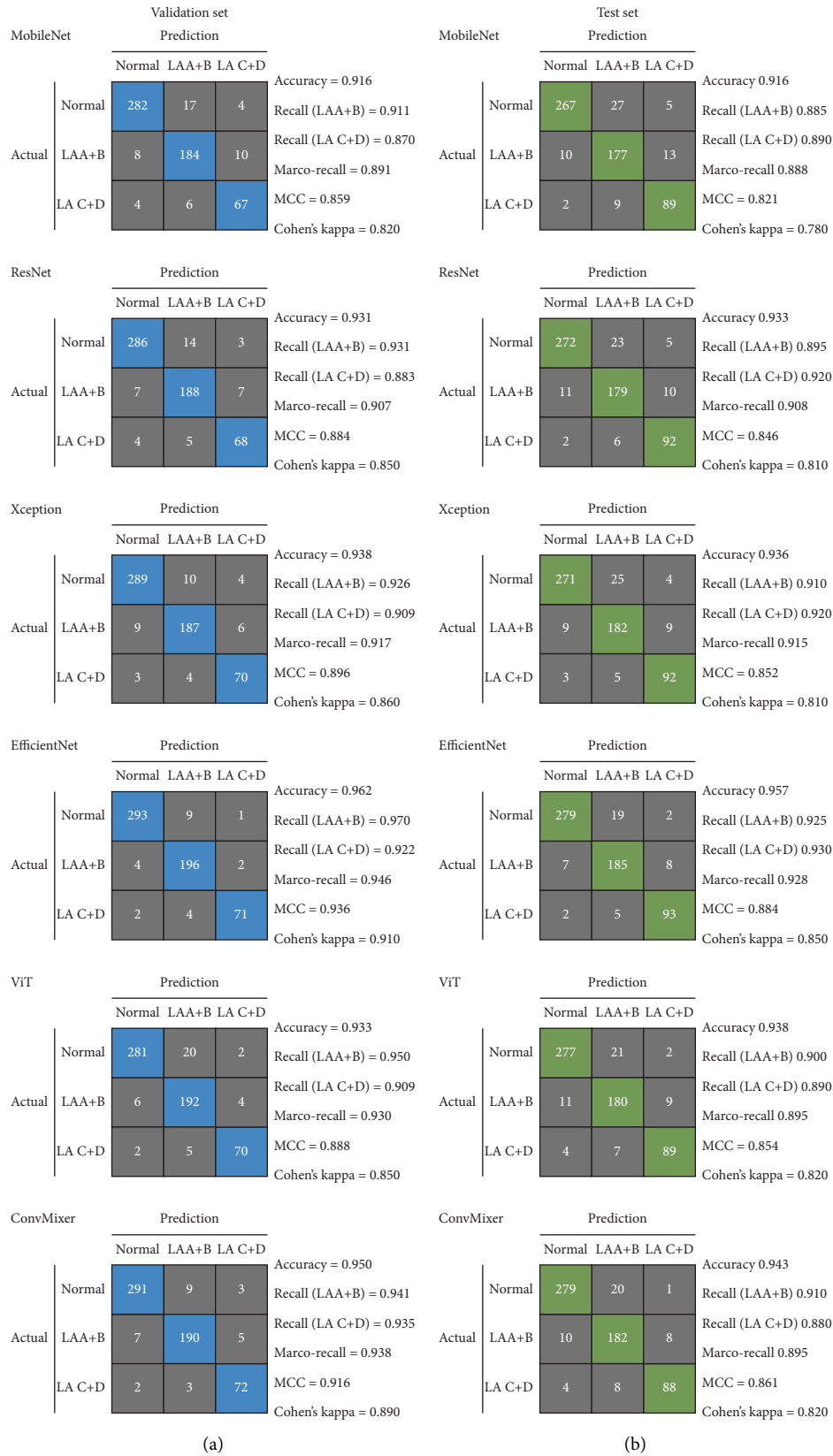


FIGURE 3: Confusion matrix of the models.

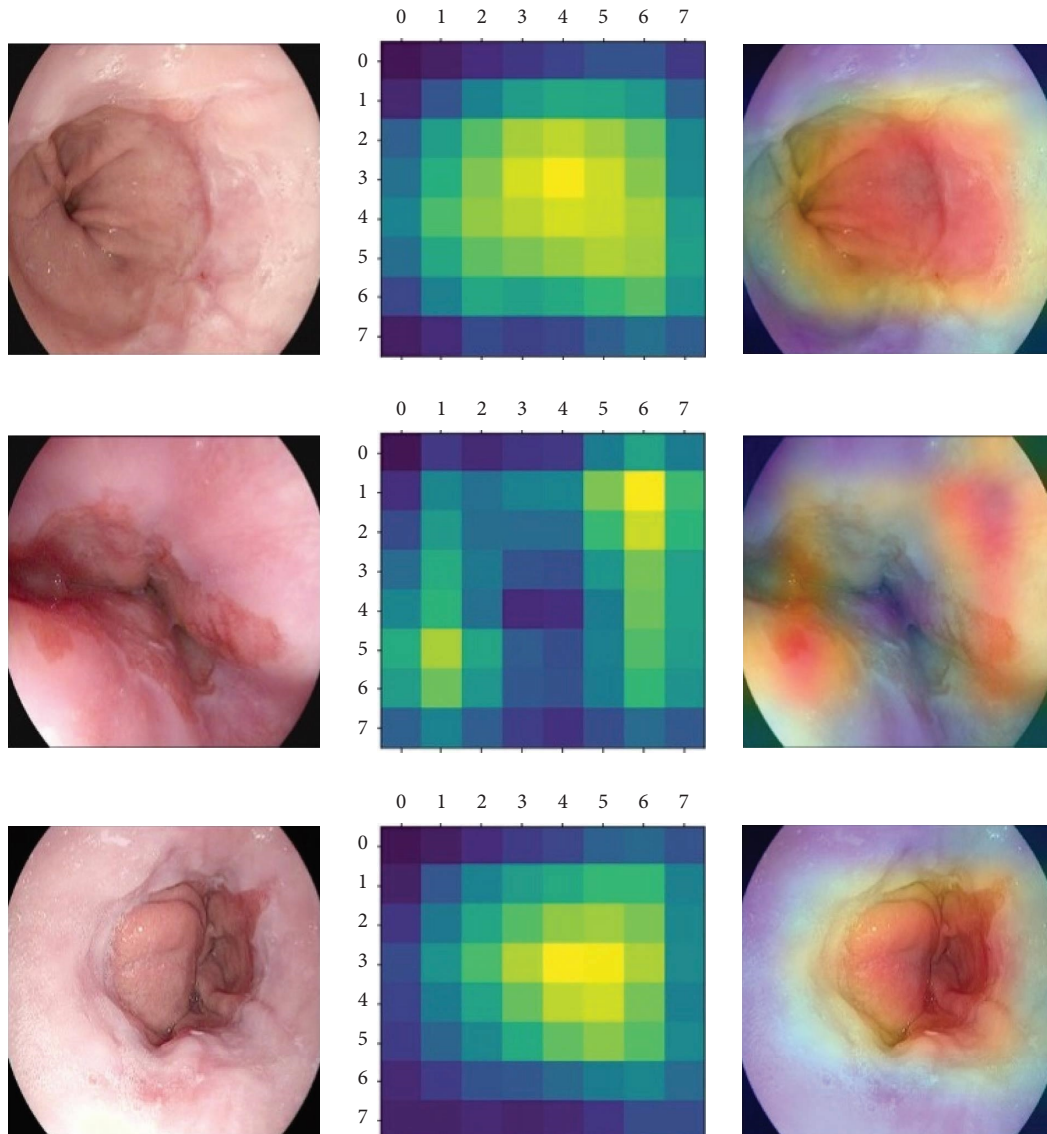


FIGURE 4: Grad-CAM heatmap.

features of input data via artificial neural networks, organized as CNNs and Transformer [15]. The past five years witness a series of studies assessing the performance of DL in the diagnosis of esophageal diseases [16–22]. The main application is the computer vision task, consisting of the detection and segmentation lesions in esophageal endoscopic images or video [23, 24]. The CAD system is designed to detect and differentiate lesions based on the mucosal/vascular pattern, to stratify the progression of the diseases or to assist the decision-making of therapy [20, 25, 26]. The remarkable advantage is reducing the workload of endoscopists and improving diagnostic accuracy [27, 28].

Recently, Visaggi et al. [29] performed a meta-analysis concerned machine learning in the diagnosis of esophageal diseases. According to their review, there were a total of 42 studies. Among them, nine were focused on Barrett's esophagus and three were about GERD [30–32]. In terms of DL, Ebigbo et al. [33] developed a real-time endoscopic

system to classify normal Barrett's esophagus and early esophageal adenocarcinoma, which showed an accuracy of 89.9%. Similarly, a CAD system by de Groof et al. [19] was used to improve the detection of dysplastic Barrett's esophagus. The ResNet/UNet-based system showed the performance of high accuracy detection and near-perfect segmentation, better than general endoscopists. One month ago, Tang et al. [17] trained a multitask DL model to diagnose esophageal lesions (normal vs. cancer vs. esophagitis). According to their report, the model achieved a high accuracy (93.43%) in complex classification, as well as a satisfied coefficient (77.84%) in semantic segmentation. Guimaraes et al. [18] proposed a CNNs-based multi-classification model (normal vs. eosinophilic esophagitis vs. candidiasis). In the test set, the model presented a fine global accuracy (0.915), higher than endoscopists.

In this multicenter study, six CNNs-or Transformer-architectures computer vision models were transfer

learning to the multiclassification of RE endoscopic images, according to the LA classification. The EfficientNet model displayed the highest accuracy and Marco-recall. EfficientNet is a CNNs architecture and scaling method that uniformly scales all dimensions with a set of fixed scaling coefficients [34]. There are various models designed to improve training efficiency, e.g., Transformer blocks in Transformers-architectures models. But expensive overhead depending on parameter size comes as an issue. EfficientNetV2 is the successor of EfficientNet, which is a family of models optimized for floating point operations and parameter efficiency. In 2021, Google used a combination of training-aware neural architecture search, scaling to further optimize the training speed and parameter efficiency to develop this new family [35]. EfficientNetV2 overcomes some of the training bottlenecks and outperforms the V1 models. Moreover, compared with Transformer-architectures models, EfficientNet shows advantage in this small dataset with limited computing power. In the comparison with the endoscopists, the EfficientNet model also showed advantages both in accuracy and recall. Interpretability for a DL model has been one of the essential respects. Computer scientists and medical practitioners are showing more concerns about the inference of AI during the development of models, especially in the field of computer vision. Therefore, lastly, we proposed the Grad-CAM technology to visualize the inferential explanation on the original images.

Our study has some limitations. To begin with, we only focus on esophagitis caused by reflex, rather than various etiologies, e.g., radiation, eosinophilic, and pill-induced esophagitis. Further studies, based on medical history and biopsy, are required to develop more complex classifiers for esophagitis. Besides, the images dataset was limited, while video files were not involved in the analyzation. This study still required more data for validation. Lastly, we did not deploy the models in endoscopic devices. We believe that this study may contribute to the future deployment in the actual practice.

5. Conclusions

In this study, we developed a series of DL-based computer vision models with the interpretable Grad-CAM to evaluate the feasibility of AI in the multiclassification of RE endoscopic images for the first time. It suggests that DL-based classifiers show promise in the endoscopic diagnosis of esophagitis. In the future, it is necessary to investigate the multimodal fusion in the classification of RE, integrating endoscopic images, clinical symptoms, esophageal pH monitoring, etc.

Data Availability

The dataset used to support the findings of the study is from HyperKvasir, which is now the largest open dataset of the gastrointestinal endoscopy (<https://datasets.simula.no/hyper-kvasir/>).

Ethical Approval

This study was approved by the Ethics Committee of The First Affiliated Hospital of Soochow University (the IRB approval number 2022098). All procedures performed in studies involving human participants were in accordance with the Helsinki Declaration of 1975 as revised in 2000. All subjects gave written informed consent before participation.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

Hailong Ge, Xin Zhou, and Jinzhou Zhu provided acquisition, analysis and interpretation of data, and statistical analysis. Yu Wang and Jian Xu performed development of methodology and writing. Feng Mo and Chen Chao were responsible for description and visualization of data. Weixin Yu contributed to study design. Jinzhou Zhu and Hailong Ge provided technical and material support. All authors have read and agreed to the published version of the manuscript. Hailong Ge, Xin Zhou, and Yu Wang contributed equally to this work.

Acknowledgments

This study was supported by the National Natural Science Foundation of China (82000540), Science and Technology Plan (Apply Basic Research) of Changzhou City (CJ20210006), Clinical Technology Development Fund of Jiangsu University (JLY20180134), Medical Education Collaborative Innovation Fund of Jiangsu University (JDY2022018), and Youth Program of Suzhou Health Committee (KJXW2019001).

References

- [1] C. P. Gyawali, P. J. Kahrilas, E. Savarino et al., "Modern diagnosis of GERD: the Lyon Consensus," *Gut*, vol. 67, no. 7, pp. 1351–1362, 2018.
- [2] L. Kia and I. Hirano, "Distinguishing GERD from eosinophilic oesophagitis: concepts and controversies," *Nature Reviews Gastroenterology & Hepatology*, vol. 12, no. 7, pp. 379–386, 2015.
- [3] K. Iwakiri, Y. Fujiwara, N. Manabe et al., "Evidence-based clinical practice guidelines for gastroesophageal reflux disease 2021," *Journal of Gastroenterology*, vol. 57, no. 4, pp. 267–285, 2022.
- [4] J. S. Nirwan, S. S. Hasan, Z. U. D. Babar, B. R. Conway, and M. U. Ghori, "Global prevalence and risk factors of gastro-oesophageal reflux disease (GORD): systematic review with meta-analysis," *Scientific Reports*, vol. 10, no. 1, p. 5814, 2020.
- [5] M. Sakaguchi, N. Manabe, N. Ueki et al., "Factors associated with complicated erosive esophagitis: a Japanese multicenter, prospective, cross-sectional study," *World Journal of Gastroenterology*, vol. 23, no. 2, pp. 318–327, 2017.
- [6] S. Ohashi, T. Maruno, K. Fukuyama et al., "Visceral fat obesity is the key risk factor for the development of reflux erosive esophagitis in 40–69-years subjects," *Esophagus*, vol. 18, no. 4, pp. 889–899, 2021.

- [7] P. O. Katz, K. B. Dunbar, F. H. Schnoll-Sussman, K. B. Greer, R. Yadlapati, and S. J. Spechler, "ACG clinical guideline for the diagnosis and management of gastroesophageal reflux disease," *American Journal of Gastroenterology*, vol. 117, no. 1, pp. 27–56, 2022.
- [8] L. R. Lundell, J. Dent, J. R. Bennett et al., "Endoscopic assessment of oesophagitis: clinical and functional correlates and further validation of the Los Angeles classification," *Gut*, vol. 45, no. 2, pp. 172–180, 1999.
- [9] K. Sumiyama, T. Futakuchi, S. Kamba, H. Matsui, and N. Tamai, "Artificial intelligence in endoscopy: present and future perspectives," *Digestive Endoscopy*, vol. 33, no. 2, pp. 218–230, 2021.
- [10] G. E. Tontini and H. Neumann, "Artificial intelligence: thinking outside the box," *Best Practice & Research Clinical Gastroenterology*, vol. 52-53, Article ID 101720, 2021.
- [11] P. Visaggi, N. de Bortoli, B. Barberio et al., "Artificial intelligence in the diagnosis of upper gastrointestinal diseases," *Journal of Clinical Gastroenterology*, vol. 56, no. 1, pp. 23–35, 2022.
- [12] H. Borgli, V. Thambawita, P. H. Smedsrud et al., "HyperKvasir, a comprehensive multi-class image and video dataset for gastrointestinal endoscopy," *Scientific Data*, vol. 7, no. 1, p. 283, 2020.
- [13] H. Jiang, J. Xu, R. Shi et al., "A multi-label deep learning model with interpretable grad-CAM for diabetic retinopathy classification," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference*, pp. 1560–1563, Montreal, Canada, July 2020.
- [14] C. Le Berre, W. J. Sandborn, S. Aridhi et al., "Application of artificial intelligence to gastroenterology and hepatology," *Gastroenterology*, vol. 158, no. 1, pp. 76–94 e2, 2020.
- [15] M. K. Sana, Z. M. Hussain, P. A. Shah, and M. H. Maqsood, "Artificial intelligence in celiac disease," *Computers in Biology and Medicine*, vol. 125, Article ID 103996, 2020.
- [16] M. R. Struyvenberg, A. J. de Groof, J. van der Putten et al., "A computer-assisted algorithm for narrow-band imaging-based tissue characterization in Barrett's esophagus," *Gastrointestinal Endoscopy*, vol. 93, no. 1, pp. 89–98, 2021.
- [17] S. Tang, X. Yu, C. F. Cheang et al., "Diagnosis of esophageal lesions by multi-classification and segmentation using an improved multi-task deep learning model," *Sensors*, vol. 22, no. 4, p. 1492, 2022.
- [18] P. Guimaraes, A. Keller, T. Fehlmann, F. Lammert, and M. Casper, "Deep learning-based detection of eosinophilic esophagitis," *Endoscopy*, vol. 54, no. 3, pp. 299–304, 2022.
- [19] A. J. de Groof, M. R. Struyvenberg, J. van der Putten et al., "Deep-learning system detects neoplasia in patients with barrett's esophagus with higher accuracy than endoscopists in a multistep training and validation study with benchmarking," *Gastroenterology*, vol. 158, no. 4, pp. 915–929, 2020.
- [20] A. Ebigbo, R. Mendel, A. Probst et al., "Real-time use of artificial intelligence in the evaluation of cancer in Barrett's oesophagus," *Gut*, vol. 69, no. 4, pp. 615–616, 2020.
- [21] X. L. Yuan, L. J. Guo, W. Liu et al., "Artificial intelligence for detecting superficial esophageal squamous cell carcinoma under multiple endoscopic imaging modalities: a multicenter study," *Journal of Gastroenterology and Hepatology*, vol. 37, no. 1, pp. 169–178, 2022.
- [22] X. X. Yang, Z. Li, X. J. Shao et al., "Real-time artificial intelligence for endoscopic diagnosis of early esophageal squamous cell cancer (with video)," *Digestive Endoscopy*, vol. 33, no. 7, pp. 1075–1084, 2021.
- [23] M. Tokat, L. van Tilburg, A. D. Koch, and M. C. Spaander, "Artificial intelligence in upper gastrointestinal endoscopy," *Digestive Diseases*, vol. 40, no. 4, pp. 395–408, 2021.
- [24] W. Du, N. Rao, J. Yong et al., "Improving the classification performance of esophageal disease on small dataset by semi-supervised efficient contrastive learning," *Journal of Medical Systems*, vol. 46, no. 1, p. 4, 2021.
- [25] I. Dickson, "Deep-learning AI for neoplasia detection in Barrett oesophagus," *Nature Reviews Gastroenterology & Hepatology*, vol. 17, no. 2, pp. 66–67, 2020.
- [26] J. Chen, Y. Jiang, T. S. Chang et al., "Detection of Barrett's neoplasia with a near-infrared fluorescent heterodimeric peptide," *Endoscopy*, vol. 54, no. 12, pp. 1198–1204, 2022.
- [27] P. Sharma and C. Hassan, "Artificial intelligence and deep learning for upper gastrointestinal neoplasia," *Gastroenterology*, vol. 162, no. 4, pp. 1056–1066, 2022.
- [28] G. Liu, J. Hua, Z. Wu et al., "Automatic classification of esophageal lesions in endoscopic images using a convolutional neural network," *Annals of Translational Medicine*, vol. 8, no. 7, p. 486, 2020.
- [29] P. Visaggi, B. Barberio, D. Gregori et al., "Systematic review with meta-analysis: artificial intelligence in the diagnosis of oesophageal diseases," *Alimentary Pharmacology & Therapeutics*, vol. 55, no. 5, pp. 528–540, 2022.
- [30] F. Pace, G. Riegler, A. de Leone et al., "Is it possible to clinically differentiate erosive from nonerosive reflux disease patients? A study using an artificial neural networks-assisted algorithm," *European Journal of Gastroenterology and Hepatology*, vol. 22, no. 10, pp. 1163–1168, 2010.
- [31] F. Pace, M. Buscema, P. Dominici et al., "Artificial neural networks are able to recognize gastro-oesophageal reflux disease patients solely on the basis of clinical data," *European Journal of Gastroenterology and Hepatology*, vol. 17, no. 6, pp. 605–610, 2005.
- [32] N. Horowitz, M. Moshkowitz, Z. Halpern, and M. Leshno, "Applying data mining techniques in the development of a diagnostics questionnaire for GERD," *Digestive Diseases and Sciences*, vol. 52, no. 8, pp. 1871–1878, 2007.
- [33] A. Ebigbo, R. Mendel, T. Rückert et al., "Endoscopic prediction of submucosal invasion in Barrett's cancer with the use of artificial intelligence: a pilot study," *Endoscopy*, vol. 53, no. 9, pp. 878–883, 2021.
- [34] A. Abedalla, M. Abdullah, M. Al-Ayyoub, and E. Benkhelifa, "Chest X-ray pneumothorax segmentation using U-Net with EfficientNet and ResNet architectures," *PeerJ Computer Science*, vol. 7, Article ID e607, 2021.
- [35] M. Tan and Q. V. Le, "EfficientNetV2: smaller models and faster Training2021," 2021, <https://ui.adsabs.harvard.edu/abs/2021arXiv210400298T>.