

Deep learning to predict esophageal variceal bleeding based on endoscopic images

Journal of International Medical Research

2023, Vol. 51(10) 1–13

© The Author(s) 2023

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/03000605231200371

journals.sagepub.com/home/imr



Yu Hong^{1,*}, Qianqian Yu^{2,*}, Feng Mo^{3,*},
Minyue Yin¹, Chang Xu¹, Shiqi Zhu¹, Jiaxi Lin¹,
Guoting Xu¹, Jingwen Gao¹, Lu Liu¹ and
Yu Wang³ 

Abstract

Objective: Esophageal varix (EV) bleeding is a particularly serious complications of cirrhosis. Prediction of EV bleeding requires extensive endoscopy experience; it remains unreliable and inefficient. This retrospective cohort study evaluated the feasibility of using deep learning (DL) to predict the 12-month risk of EV bleeding based on endoscopic images.

Methods: Six DL models were trained to perform binary classification of endoscopic images of EV bleeding. The models were subsequently validated using an external test dataset, then compared with classifications performed by two endoscopists.

Results: In the validation dataset, EfficientNet had the highest accuracy (0.910), followed by ConvMixer (0.898) and Xception (0.875). In the test dataset, EfficientNet maintained the highest accuracy (0.893), which was better than the endoscopists (0.800 and 0.763). Notably, one endoscopist displayed higher recall (0.905), compared with EfficientNet (0.870). When their predictions were assisted by artificial intelligence, the accuracies of the two endoscopists increased by 17.3% and 19.0%. Moreover, statistical agreement among the models was dependent on model architecture.

Conclusions: This study demonstrated the feasibility of using DL to predict the 12-month risk of EV bleeding based on endoscopic images. The findings suggest that artificial intelligence-aided diagnosis will be a useful addition to cirrhosis management.

³Department of General Surgery, Jintan Affiliated Hospital of Jiangsu University, Jintan, China

*These authors contributed equally to this work.

Corresponding author:

Yu Wang, Department of General Surgery, Jintan Affiliated Hospital of Jiangsu University, Jintan 320413, China.
Email: ywang030608@stu.suda.edu.cn

¹Department of Gastroenterology, The First Affiliated Hospital of Soochow University, Suzhou, China

²Department of Oncology, Jintan Affiliated Hospital of Jiangsu University, Jintan, China



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative

Commons Attribution-NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

Keywords

Computer vision, deep learning, endoscopy, esophageal variceal bleeding, gradient-weighted class activation mapping, cirrhosis.

Date received: 23 May 2023; accepted: 24 August 2023

Introduction

Variceal bleeding is a particularly serious complication of cirrhosis.¹ Variceal bleeding-related mortality remains high, with a prevalence approaching 20%.^{2–4} Esophageal varices (EVs) are the most common gastrointestinal varices.⁵ According to the Baveno consensus, high-risk varices in patients without prior EV bleeding are defined as: medium or large varices, small varices with red wale marks, or any size varices in patients with Child–Pugh class C disease.⁶ The consensus recommends endoscopic surveillance for patients with cirrhosis who have portal hypertension. During upper endoscopy, endoscopists are advised to perform a series of systematic assessments to record the locations, sizes, and appearances of EVs. Risk stratification based on these endoscopic assessments is essential for the clinical management of EVs; it can optimize medical resources and improve patient outcomes.⁷

In the past 15 years, substantial advances in deep learning (DL) algorithms have revolutionized image analysis patterns in various datasets.⁸ In clinical settings, computer vision (CV) tasks include helping radiologists and endoscopists to make decisions regarding large numbers of medical images.^{9–11} Convolutional neural networks (CNNs), a class of artificial neural networks, are characterized by the shared-weight architecture of sliding convolution kernels.¹² Recently, transformer-based neural networks have emerged as a competitive alternative with potential applications

in diverse CV tasks.¹³ Various DL models have been increasingly used in analyses of endoscopy datasets,^{14,15} including polyp detection (colonoscopy), early gastric cancer diagnosis (upper endoscopy), and automatic lesion detection (capsule endoscopy).^{16–21}

Although high-risk varices have been well-characterized since the late 1980 s,^{22,23} the prediction of EV bleeding risk requires extensive endoscopy experience; it remains unreliable and inefficient.^{2,5} Early identification of patients with cirrhosis who have high-risk EVs is important for improving patient outcomes and optimizing medical resources.¹ In this multicenter study, we evaluated the feasibility of using DL to predict the 12-month risk of EV bleeding based on model training via classification of endoscopic images.

Methods

Study design

This multicenter retrospective cohort study included randomly selected patients with cirrhosis who attended two hospitals—The First Affiliated Hospital of Soochow University, Suzhou, China (Suzhou hospital, training and validation datasets) and Jintan Affiliated Hospital of Jiangsu University, Changzhou, China (Jintan hospital, test dataset)—between 1 January 2015 and 1 January 2021 (Supplementary Table 1). All variceal images were acquired during maximal insufflation and distention of the

esophageal lumen, in the absence of peristaltic waves. The images were fully anonymized before analysis by the authors; the investigators also de-identified all patient information. This study protocol was approved by the Ethics Committee of The First Affiliated Hospital of Soochow University (approval number 2022098); the study was conducted in accordance with the Helsinki Declaration of 1975, as revised in 2013. All participants provided written informed consent for inclusion in the study. This article adheres to the STROBE guidelines.²⁴

At both centers, EVs were diagnosed based on the 2014 guidelines of the American Society for Gastrointestinal Endoscopy (ASGE 2014).²⁵ Patients were excluded if they had a history of EV bleeding, were receiving treatment for EVs, or had a non-cirrhotic etiology of portal hypertension. Treatments for EVs included nonselective beta blockers, sclerotherapy, endoscopic ligation, transjugular intrahepatic portosystemic shunt, and laparoscopic splenectomy.²⁶ Patients were followed up for at least 12 months to record EV outcomes (i.e., first occurrence of EV bleeding). Variceal bleeding was defined as follows: 1) active variceal bleeding on

endoscopy or 2) melena/hematemesis after exclusion of other causes of upper gastrointestinal bleeding (e.g., digestive ulcers or Dieulafoy disease).

The study protocol is shown in Figure 1. First, six CNN- or transformer-based DL models were selected for transfer learning in the CV task. After models had been pre-trained on ImageNet (detailed in next section), they were trained to perform binary classification of endoscopic images of EVs from Suzhou hospital (training and validation datasets). Moreover, the DL models were externally validated using images from Jintan hospital (test dataset), then compared with classifications performed by two endoscopists. Additionally, DL model inferences were visualized by gradient-weighted class activation mapping (Grad-CAM). Finally, artificial intelligence-aided performance was evaluated.

Models and pretraining

Six CNN- or transformer-based models were chosen: MobileNet (MobileNet V1), ResNet (ResNet-50 V2), Xception, EfficientNet (EfficientNet V2 s), ViT (ViT-B16), and

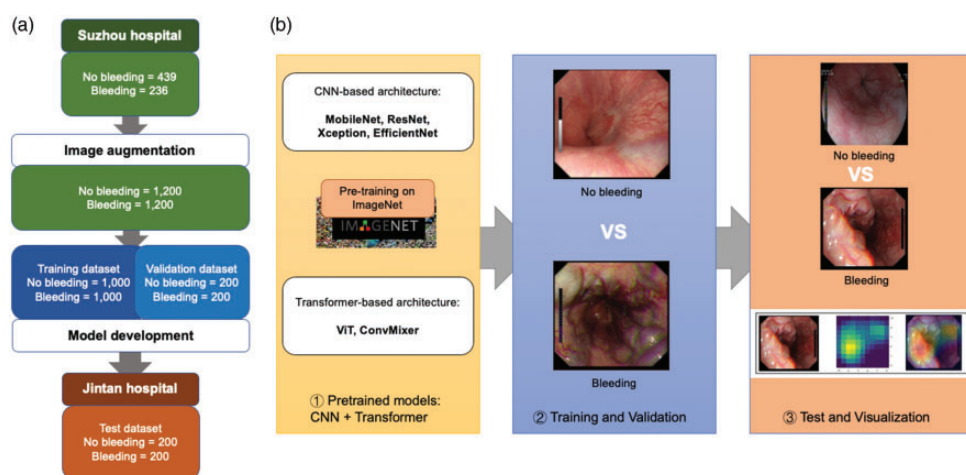


Figure 1. Study flowchart. (a) Datasets and (b) processes used in this study.

ConvMixer (ConvMixer-768/32). These models were pretrained on the ImageNet database (www.image-net.org), prior to transfer learning. Architectures and pre-trained parameters were downloaded from Keras (keras.io) or TensorFlow Hub (tfhub.dev). A flowchart for the model training process is provided in Supplementary Figure 1.

Target training and internal validation

As shown in Figure 1, 675 images of EVs were collected from Suzhou hospital and saved in JPEG format. All images were obtained from endoscopic systems established by Olympus Corporation (Tokyo, Japan) or Fujifilm Corporation (Tokyo, Japan). Images were divided into two classes according to whether bleeding occurred within 12 months after initial diagnosis: control (no bleeding, $n=439$) vs. bleeding ($n=236$). After image augmentation, the numbers of images were increased to 2000 in the training dataset (1000 vs. 1000) and 400 in the validation dataset (200 vs. 200). All images were rescaled and normalized to pixel dimensions of $331 \times 331 \times 3$. Considering the unique nature of endoscopy data, we used non-rigid transformation for endoscopic images. Rotational transformation and flip transformation were performed to augment the data. The augmentation procedure is shown in Supplementary Figure 2.

The parameter transfer learning procedure was implemented in Keras software (Google), using the TensorFlow framework as the backbone. Transfer learning was applied to the aforementioned CNN- or transformer-based architectures by combining the existing feature extraction layers (fixed layers) with additional activation layers (training layers) for binary classification. Fitted models were compiled using an Adam optimizer and the binary cross-entropy cost function, with a fixed learning

rate of 0.0001 and batch size of 32. Training was discontinued when accuracy in the validation dataset did not improve for three epochs. The maximum number of training epochs was limited to 100. This experiment was performed on a Mac mini (Apple M1, 8 cores, 16G RAM, macOS Ventura 13.0.1). The training code is available at <https://osf.io/ycxwr>.

External validation

Test dataset images were collected from Jintan hospital (control [no bleeding, $n=200$] vs. bleeding [$n=200$]). The criteria and procedure were consistent with the methodology regarding Suzhou hospital. Furthermore, the performances of two endoscopists (one from Jintan hospital [#1] and one from Suzhou hospital [#2]) were compared with the performances of the models. First, endoscopists with more than 10 years of experience independently classified the images based on the ASGE 2014 guidelines.²⁵ In accordance with the guidelines, large EVs (>5 mm) or EVs displaying red wales or spots were predicted to have a high risk of bleeding. Second, after independent judgement, the two endoscopists were given access to the best model's predictions; they subsequently re-classified the images, and this approached was regarded as artificial intelligence-aided prediction.

Model visualization

Grad-CAM has been proposed to visualize models by highlighting regions that are important for predictions from these models.²⁷ In this study, Grad-CAM heatmaps were created for inferential explanation of multiple output layers from the best model.

Evaluation criteria

DL models were fitted using Python software (version: 3.9; Python Software Foundation) and TensorFlow (2.8.0; Google). Model performances were evaluated in terms of accuracy, recall, precision, and F1-score. The algorithms are shown below (TP, true positive [correct assumption that an event has occurred]; TN, true negative [correct assumption that an event has not occurred]; FP, false positive [incorrect assumption that an event has not occurred]; FN, false negative [incorrect assumption that an event has not occurred]):

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN})$$

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP})$$

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN})$$

F1

$$\begin{aligned} & \text{— score} \\ & = (2 * \text{precision} * \text{recall}) / (\text{precision} + \text{recall}). \end{aligned}$$

Accuracy, precision, and recall are the main indicators of machine learning performance. The F1-score combines precision and recall into a single indicator. Statistical agreement among the DL models and the endoscopists was measured using Cohen's κ , which evaluates the level of agreement between two raters or judges when classifying items into mutually exclusive categories. Two-sided p -values <0.05 were considered statistically significant.

Results

Development and internal validation of DL models

The six CNN- or transformer-based models were trained by transfer learning, then used

for binary classification of the 12-month risk of EV bleeding based on endoscopic images. The performances of the six models in the validation dataset are shown in Figure 2a and Table 1. EfficientNet showed the highest accuracy (0.910), followed by ConvMixer (0.898) and Xception (0.875). The recall, precision, and F1-score of the EfficientNet model were 0.900, 0.918, and 0.905; these values were superior to the values of the other models. The confusion matrices of the models are shown in Figure 3a.

External validation of DL models

The performances of the six models in the test dataset are shown in Figure 2b and Table 1. EfficientNet maintained the highest accuracy (0.893), followed by ConvMixer (0.860) and Xception (0.850). The recall, precision, and F1-score of EfficientNet were 0.870, 0.911, and 0.890; these values were superior to the values of the other models. The confusion matrices of the models are shown in Figure 3b.

Comparison with endoscopists

The performances and confusion matrices of the two endoscopists in the test dataset are shown in Figure 2c, Figure 3c, and Table 1. The accuracies of the endoscopists were 0.800 and 0.763. Although their accuracies were lower than the DL model accuracies, endoscopist #1 achieved a recall of 0.905, which was superior to EfficientNet.

Artificial intelligence-aided performance

The performances of the two endoscopists with assistance from the best DL model (EfficientNet) are presented in Figure 2c, Figure 3d, and Table 1. For endoscopist #1, the accuracy increased by 17.3% to 0.938; recall was 0.920, precision was 0.953, and F1-score was 0.936.

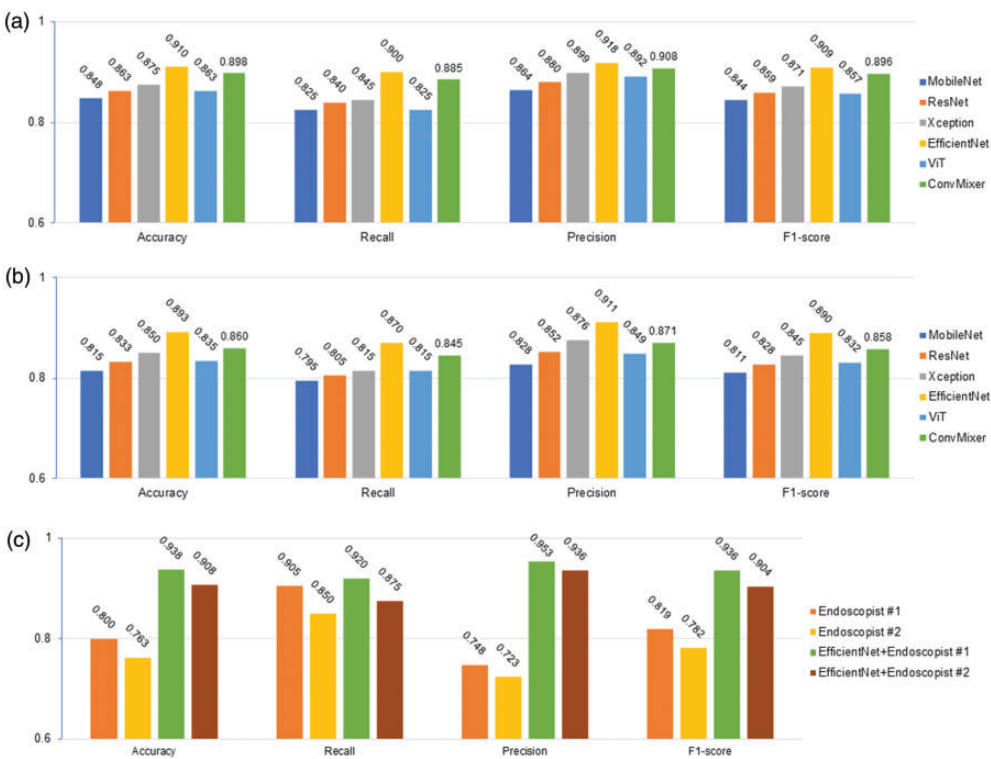


Figure 2. Performances of deep learning models and endoscopists in the validation and test datasets. a) Artificial intelligence models in validation dataset. b) Artificial intelligence models in test dataset. c) Endoscopists + best artificial intelligence model in test dataset.

For endoscopist #2, accuracy increased by 19.0% to 0.908, which was superior to EfficientNet.

Agreement between models and endoscopists

As shown in Figure 4, statistical agreement among the DL models and endoscopists was measured using Cohen’s κ . According to model architecture, the four CNN-based models displayed high levels of agreement (ranging from 0.745 to 0.945), whereas the κ value between the two Transformer-based models was 0.800. Finally, the two endoscopists demonstrated a κ value of 0.860.

Inferential explanation of the EfficientNet model

Based on the outputs of the EfficientNet model, Grad-CAM was used to highlight the artificial intelligence-inferred EVs on original images, as shown in Figure 5. Based on the best Grad-CAM model outputs, inferential explanations were established via heatmaps.

The left column shows the original endoscopic images. The middle column shows heatmaps established using the output of the last feature-extractor layer in the EfficientNet model. The right column shows the Grad-CAM heatmap overlaid on the original images, where highlighted regions reflect EVs identified by the model.

Table 1. Performances of deep learning models and endoscopists in each dataset.

Datasets	Models	Accuracy	Recall	Precision	F1-score
Training	MobileNet	0.969	0.966	0.972	0.969
	ResNet	0.979	0.977	0.981	0.979
	Xception	0.985	0.981	0.989	0.985
	EfficientNet	0.992	0.990	0.993	0.991
	ViT	0.981	0.982	0.979	0.981
	ConvMixer	0.986	0.983	0.988	0.985
Validation	MobileNet	0.848	0.825	0.864	0.844
	ResNet	0.863	0.840	0.880	0.859
	Xception	0.875	0.845	0.899	0.871
	EfficientNet	0.910	0.900	0.918	0.909
	ViT	0.863	0.825	0.892	0.857
	ConvMixer	0.898	0.885	0.908	0.896
Test	MobileNet	0.815	0.795	0.828	0.811
	ResNet	0.833	0.805	0.852	0.828
	Xception	0.850	0.815	0.876	0.845
	EfficientNet	0.893	0.870	0.911	0.890
	ViT	0.835	0.815	0.849	0.832
	ConvMixer	0.860	0.845	0.871	0.858
	Endoscopist #1	0.800	0.905	0.748	0.819
	Endoscopist #2	0.763	0.850	0.723	0.782
	EfficientNet + Endoscopist #1	0.938	0.920	0.953	0.936
	EfficientNet + Endoscopist #2	0.908	0.875	0.936	0.904

Discussion

In this study, we evaluated the feasibility of using DL models to classify images of EVs according to the 12-month risk of bleeding. We found that DL models demonstrated better classification accuracy and inference time, compared with endoscopists. Furthermore, artificial intelligence-aided prediction displayed considerable improvement.

In the late 1980s, the Japan Society for Portal Hypertension initially proposed a system to record the endoscopic appearance of EVs.²³ In 1995, the society revised the rules into the “General Rules for Recording Endoscopic Findings of Esophagogastric Varices.”²⁸ In 2010, information regarding portal hypertensive gastropathy and endoscopic ultrasonography was included in the second edition of the rules.²⁹ These rules are widely used in

Japan and other countries. Among Western countries, in 1988, the North Italian Endoscopic Club released a prediction index to determine 12-month risk of bleeding, based on Beppu’s Variceal Score and the Japanese rules.²²

Guidelines from the American Association for the Study of Liver Disease and the Baveno V consensus suggest the use of screening endoscopy for patients with cirrhosis to diagnose EVs and stratify bleeding risk.^{6,30} The optimal surveillance schedule for EVs is largely dependent on endoscopic appearance and the stage of portal hypertension. For patients with compensated cirrhosis and no varices on screening, endoscopy is recommended every 2 to 3 years; patients with small varices should undergo endoscopy every 2 years.²⁵ EVs may progress faster in patients with alcohol-induced or decompensated cirrhosis, or with high-risk signs on endoscopic

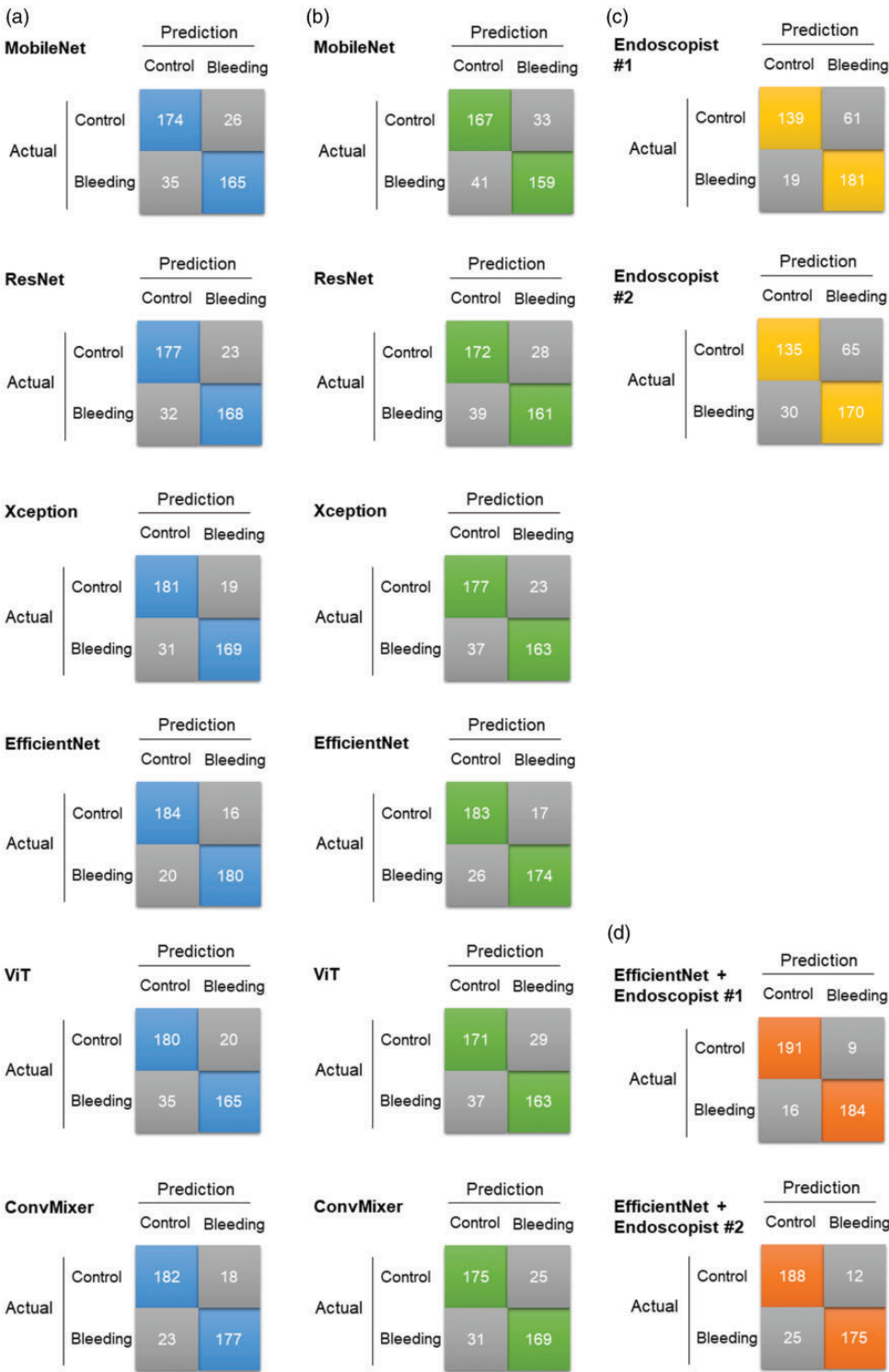


Figure 3. Confusion matrices of deep learning models and endoscopists in the validation and test datasets.
a) Artificial intelligence models in validation dataset. b) Artificial intelligence models in test dataset.
c) Endoscopists in test dataset. d) Endoscopists + artificial intelligence model in test dataset.

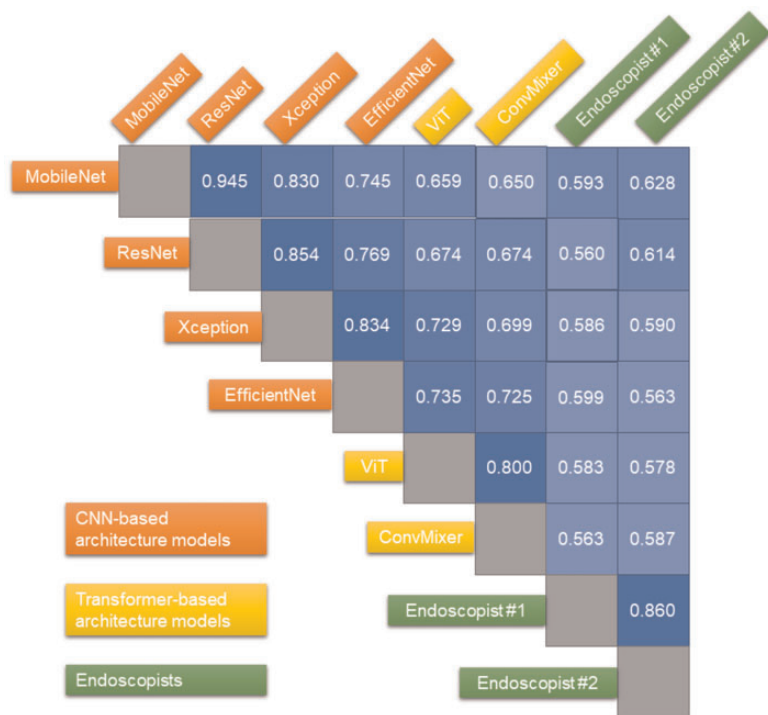


Figure 4. Statistical measures of agreement among the models and endoscopists. CNN, convolutional neural network.

examination; these patients should undergo endoscopy annually.²⁵ Current evidence indicates that the annual risk of EV bleeding increases by ~15% in patients with cirrhosis decompensation, large variceal size, or red signs in varices (e.g., red wale markings, cherry red spots, hematocystis, and telangiectasia).^{6,29} Early identification of high-risk EVs is needed to optimize medical resources and improve outcomes among patients with cirrhosis.¹

DL has been applied to various clinical datasets.³¹ In the past 5 years, DL-based computer-aided detection and diagnosis systems have demonstrated robust performance in digestive endoscopy.^{14,16,18,32} In the present study, six DL models were trained by transfer learning, then used for binary classification of the 12-month risk of EV bleeding based on endoscopic images.

EfficientNet showed the highest accuracy in the internal validation and external test datasets. EfficientNet is a CNN-based architecture that uses a compound scaling method, which uniformly scales all dimensions using a set of fixed scaling coefficients.³³ In various image classification competitions, EfficientNet has demonstrated superior transfer performance and state-of-the-art accuracy, while using considerably fewer parameters.³⁴

We also evaluated the performances of the DL models, compared with the performances of two endoscopists. Generally, the models showed superior accuracy and precision. However, the endoscopists displayed slightly better recall. In this study, two endoscopists classified images of EVs using the ASGE 2014 guidelines,²⁵ in which large EVs or varices with red wales

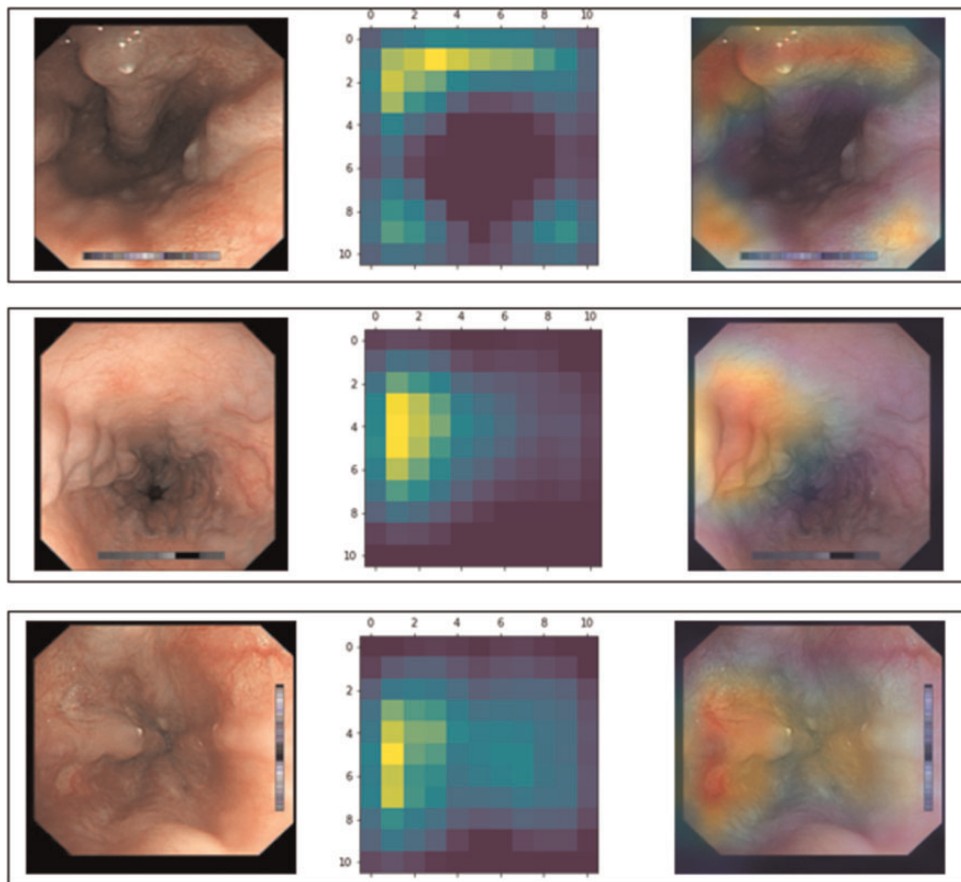


Figure 5. Visualization of EfficientNet model inferences by gradient-weighted class activation mapping (Grad-CAM). Left column) Original endoscopic images. Middle column) Heatmaps established using the output of the last feature-extractor layer in the EfficientNet model. Right column) Grad-CAM heatmap overlaid on the original images, where highlighted regions reflect esophageal varices identified by the model.

or spots are presumed to have a high risk of bleeding. This approach might increase the number of EVs with a predicted risk of bleeding, resulting in higher numbers of true positives and false positives, along with higher recall and lower precision. The F1-score is a combination indicator of the model's precision and recall. In terms of F1-score, EfficientNet and other DL models remained superior to the endoscopists. Furthermore, assessments of statistical agreement among the models and endoscopists revealed that the agreement was

dependent on model architecture. Variability between the two endoscopists (i.e., observer bias) is an important source of disagreement. Interpretability is an essential step of DL model development. Data scientists and medical practitioners often focus on the rationale for inferences by artificial intelligence, especially in CV tasks. Thus, Grad-CAM was used to provide an inferential explanation via heatmaps. Finally, the endoscopists re-classified the images based on the results of EfficientNet. The artificial intelligence-aided

predictions displayed substantial improvement compared with the predictions by endoscopists alone, which may be useful in clinical applications.

This study had some limitations. First, it only focused on endoscopic images, rather than other clinical data (e.g., liver function or radiological findings). Second, bleeding risk was predicted based on a single image, rather than multiple images from multiple EV sites. Further studies based on structured clinical data, radiological data, and multiple images are needed to establish more complex models for prediction of EV bleeding risk. Moreover, the rationale underlying agreements between various model architectures should be explored. Finally, this study did not include extensive independent datasets; such datasets should be included in future validation studies.

Conclusions

This study demonstrated the feasibility of using DL to predict the 12-month risk of EV bleeding based on endoscopic images. Artificial intelligence models demonstrated better classification accuracy and inference time, compared with endoscopists. Moreover, we found that agreement between models was architecture-dependent. The findings suggest that artificial intelligence-aided diagnosis will be useful in future cirrhosis management efforts.

Author contributions

YH, QY, and GX collected data. QY and FM wrote the manuscript. MY, JL, JG, LL, SZ, and CX analyzed data and trained the models. YH and YW contributed to the study design.

Data availability statement

The model training code is available at <https://osf.io/yxwvr>

Declaration of conflicting interest

The authors declare that there is no conflict of interest.

Funding

This study was supported by the Medical Education Collaborative Innovation Fund of Jiangsu University [JDY2022018] and the Science and Technology Plan (Apply Basic Research) of Changzhou City [CJ20210006].

ORCID iD

Yu Wang  <https://orcid.org/0009-0004-4063-8284>

References

1. Gines P, Krag A, Abraldes JG, et al. Liver cirrhosis. *Lancet* 2021; 398: 1359–1376. DOI: 10.1016/S0140-6736(21)01374-X.
2. De Franchis R and Baveno VIF. Expanding consensus in portal hypertension: report of the Baveno VI Consensus Workshop: stratifying risk and individualizing care for portal hypertension. *J Hepatol* 2015; 63: 743–752. DOI: 10.1016/j.jhep.2015.05.022.
3. GBD 2017 Cirrhosis Collaborators. The global, regional, and national burden of cirrhosis by cause in 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet Gastroenterol Hepatol* 2020; 5: 245–266. DOI: 10.1016/S2468-1253(19)30349-8.
4. Reverter E, Tandon P, Augustin S, et al. A MELD-based model to determine risk of mortality among patients with acute variceal bleeding. *Gastroenterology* 2014; 146: 412–419 e413. DOI: 10.1053/j.gastro.2013.10.018.
5. De Franchis R, Bosch J, Garcia-Tsao G, et al. Baveno VII - Renewing consensus in portal hypertension. *J Hepatol* 2021; 76: 959–974. DOI: 10.1016/j.jhep.2021.12.022.
6. De Franchis R. Baveno V Faculty. Revising consensus in portal hypertension: report of the Baveno V consensus workshop on methodology of diagnosis and therapy in portal hypertension. *J Hepatol* 2010; 53: 762–768. DOI: 10.1016/j.jhep.2010.06.004.

7. Garcia-Tsao G, Abraldes JG, Berzigotti A, et al. Portal hypertensive bleeding in cirrhosis: risk stratification, diagnosis, and management: 2016 practice guidance by the American Association for the study of liver diseases. *Hepatology* 2017; 65: 310–335. DOI: 10.1002/hep.28906.
8. Ali S, Dmitrieva M, Ghatwary N, et al. Deep learning for detection and segmentation of artefact and disease instances in gastrointestinal endoscopy. *Med Image Anal* 2021; 70: 102002. DOI: 10.1016/j.media.2021.102002.
9. Hosny A, Parmar C, Quackenbush J, et al. Artificial intelligence in radiology. *Nat Rev Cancer* 2018; 18: 500–510. DOI: 10.1038/s41568-018-0016-5.
10. Okagawa Y, Abe S, Yamada M, et al. Artificial intelligence in endoscopy. *Dig Dis Sci* 2021; 67: 1553–1572. DOI: 10.1007/s10620-021-07086-z.
11. Paderno A, Holsinger FC and Piazza C. Videomics: bringing deep learning to diagnostic endoscopy. *Curr Opin Otolaryngol Head Neck Surg* 2021; 29: 143–148. DOI: 10.1097/MOO.0000000000000697.
12. Choi J, Shin K, Jung J, et al. Convolutional neural network technology in endoscopic imaging: artificial intelligence for endoscopy. *Clin Endosc* 2020; 53: 117–126. DOI: 10.5946/ce.2020.054.
13. Mondal AK, Bhattacharjee A, Singla P, et al. xViTCOS: explainable vision transformer based COVID-19 screening using radiography. *IEEE J Transl Eng Health Med* 2022; 10: 1100110. DOI: 10.1109/JTEHM.2021.3134096.
14. Ang TL and Carneiro G. Artificial intelligence in gastrointestinal endoscopy. *J Gastroenterol Hepatol* 2021; 36: 5–6. DOI: 10.1111/jgh.15344.
15. Visaggi P, De Bortoli N, Barberio B, et al. Artificial intelligence in the diagnosis of upper gastrointestinal diseases. *J Clin Gastroenterol* 2022; 56: 23–35. DOI: 10.1097/mcg.0000000000001629.
16. Mori Y, Kudo SE, Misawa M, et al. Real-time use of artificial intelligence in identification of diminutive polyps during colonoscopy: a prospective study. *Ann Intern Med* 2018; 169: 357–366. DOI: 10.7326/M18-0249.
17. Misawa M, Kudo SE, Mori Y, et al. Accuracy of computer-aided diagnosis based on narrow-band imaging endocytoscopy for diagnosing colorectal lesions: comparison with experts. *Int J Comput Assist Radiol Surg* 2017; 12: 757–766. DOI: 10.1007/s11548-017-1542-4.
18. Saito H, Aoki T, Aoyama K, et al. Automatic detection and classification of protruding lesions in wireless capsule endoscopy images based on a deep convolutional neural network. *Gastrointest Endosc* 2020; 92: 144–151 e141. DOI: 10.1016/j.gie.2020.01.054.
19. Ding Z, Shi H, Zhang H, et al. Gastroenterologist-level identification of small-bowel diseases and normal variants by capsule endoscopy using a deep-learning model. *Gastroenterology* 2019; 157: 1044–1054 e1045. DOI: 10.1053/j.gastro.2019.06.025.
20. Wang P, Berzin TM, Glissen Brown JR, et al. Real-time automatic detection system increases colonoscopic polyp and adenoma detection rates: a prospective randomised controlled study. *Gut* 2019; 68: 1813–1819. DOI: 10.1136/gutjnl-2018-317500.
21. Min JK, Kwak MS and Cha JM. Overview of deep learning in gastrointestinal endoscopy. *Gut Liver* 2019; 13: 388–393. DOI: 10.5009/gnl18384.
22. North Italian Endoscopic Club for the S and Treatment of Esophageal Varices. Prediction of the first variceal hemorrhage in patients with cirrhosis of the liver and esophageal varices. A prospective multicenter study. *N Engl J Med* 1988; 319: 983–989. DOI: 10.1056/NEJM198810133191505.
23. The general rules for recording endoscopic findings on esophageal varices. *Jpn J Surg* 1980; 10: 84–87. DOI: 10.1007/BF02468653.
24. Von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Ann Intern Med* 2007; 147: 573–577. DOI: 10.7326/0003-4819-147-8-200710160-00010.

25. Hwang JH, Shergill AK, Acosta RD, et al. The role of endoscopy in the management of variceal hemorrhage. *Gastrointest Endosc* 2014; 80: 221–227. DOI: 10.1016/j.gie.2013.07.023.
26. Roberts D, Best LM, Freeman SC, et al. Treatment for bleeding oesophageal varices in people with decompensated liver cirrhosis: a network meta-analysis. *Cochrane Database Syst Rev* 2021; 4: CD013155. DOI: 10.1002/14651858.CD013155.pub2.
27. Jiang H, Xu J, Shi R, et al. A multi-label deep learning model with interpretable Grad-CAM for diabetic retinopathy classification. *Annu Int Conf IEEE Eng Med Biol Soc* 2020; 2020: 1560–1563. DOI: 10.1109/EMBC44109.2020.9175884.
28. Idezuki Y. General rules for recording endoscopic findings of esophagogastric varices (1991). Japanese Society for Portal Hypertension. *World J Surg* 1995; 19: 420–422; discussion 423. DOI: 10.1007/BF00299178.
29. Tajiri T, Yoshida H, Obara K, et al. General rules for recording endoscopic findings of esophagogastric varices (2nd edition). *Dig Endosc* 2010; 22: 1–9. DOI: 10.1111/j.1443-1661.2009.00929.x.
30. Riggio O, Angeloni S, Nicolini G, et al. Endoscopic screening for esophageal varices in cirrhotic patients. *Hepatology* 2002; 35: 501–502. DOI: 10.1053/jhep.2002.31308.
31. Dana J, Venkatasamy A, Saviano A, et al. Conventional and artificial intelligence-based imaging for biomarker discovery in chronic liver disease. *Hepatol Int* 2022; 16: 509–522. DOI: 10.1007/s12072-022-10303-0.
32. Faes L, Wagner SK, Fu DJ, et al. Automated deep learning design for medical image classification by health-care professionals with no coding experience: a feasibility study. *Lancet Digit Health* 2019; 1: e232–e242. DOI: 10.1016/S2589-7500(19)30108-6.
33. Abedalla A, Abdullah M, Al-Ayyoub M, et al. Chest X-ray pneumothorax segmentation using U-Net with EfficientNet and ResNet architectures. *PeerJ Comput Sci* 2021; 7: e607. DOI: 10.7717/peerj-cs.607.
34. Booyens A and Viriri S. Exploration of ear biometrics using EfficientNet. *Comput Intell Neurosci* 2022; 2022: 3514807. DOI: 10.1155/2022/3514807.