Performance Comparison of the Deep Learning and the Human Endoscopist for Bleeding Peptic Ulcer Disease

Hsu-Heng Yen^{1,2,3} · Ping-Yu Wu³ · Pei-Yuan Su¹ · Chia-Wei Yang¹ · Yang-Yuan Chen¹ · Mei-Fen Chen^{3,5} · Wen-Chen Lin⁵ · Cheng-Lun Tsai^{4,5} · Kang-Ping Lin^{3,5}

Received: 12 January 2021 / Accepted: 27 March 2021 / Published online: 30 April 2021
© The Author(s) 2021

Abstract

Purpose Management of peptic ulcer bleeding is clinically challenging. Accurate characterization of the bleeding during endoscopy is key for endoscopic therapy. This study aimed to assess whether a deep learning model can aid in the classification of bleeding peptic ulcer disease.

Methods Endoscopic still images of patients ($n = 1694$) with peptic ulcer bleeding for the last 5 years were retrieved and reviewed. Overall, 2289 images were collected for deep learning model training, and 449 images were validated for the performance test. Two expert endoscopists classified the images into different classes based on their appearance. Four deep learning models, including Mobile Net V2, VGG16, Inception V4, and ResNet50, were proposed and pre-trained by ImageNet with the established convolutional neural network algorithm. A comparison of the endoscopists and trained deep learning model was performed to evaluate the model's performance on a dataset of 449 testing images.

Results The results first presented the performance comparisons of four deep learning models. The Mobile Net V2 presented the optimal performance of the proposal models. The Mobile Net V2 was chosen for further comparing the performance with the diagnostic results obtained by one senior and one novice endoscopists. The sensitivity and specificity were acceptable for the prediction of "normal" lesions in both 3-class and 4-class classifications. For the 3-class category, the sensitivity and specificity were 94.83% and 92.36%, respectively. For the 4-class category, the sensitivity and specificity were 95.40% and 92.70%, respectively. The interobserver agreement of the testing dataset of the model was moderate to substantial with the senior endoscopist. The accuracy of the determination of endoscopic therapy required and high-risk endoscopic therapy of the deep learning model was higher than that of the novice endoscopist.

Conclusions In this study, the deep learning model performed better than inexperienced endoscopists. Further improvement of the model may aid in clinical decision-making during clinical practice, especially for trainee endoscopist.

Keywords Peptic ulcer · Bleeding · Deep learning · Artificial intelligence

1 Introduction

Peptic ulcer bleeding is a common gastrointestinal (GI) emergency with a 10% hospital mortality rate [1–3]. Important progress has been made in the treatment of this condition since the introduction of emergency endoscopy and the development of endoscopic therapy for hemostasis. The appearance of the ulcer base is probably the best available predictor of patient outcome. The classification of peptic GI bleeding was proposed by Forrest [4] in 1974. The classification differentiates among acute, recent (with risk of rebleeding), and almost-healed ulcerations. The goal of the Forrest classification is to make an immediate judgment of the risk of rebleeding and need for endoscopic intervention.

✉ Kang-Ping Lin
kplin@cycu.edu.tw

¹ Division of Gastroenterology, Department of Internal Medicine, Changhua Christian Hospital, Changhua, Taiwan

² General Education Center, Chienkuo Technology University, Changhua, Taiwan

³ Department of Electrical Engineering, Chung Yuan Christian University, Taoyuan, Taiwan

⁴ Department of Biomedical Engineering, Chung Yuan Christian University, Taoyuan, Taiwan

⁵ Technology Translation Center for Medical Device, Chung Yuan Christian University, 200 Chung Pei Road, Chung Li District, Taoyuan City 32023, Taiwan

This classification has been used since its introduction and has also been a standard for conducting various clinical trials [5–7]. Current guidelines [3, 8, 9] suggest that patients with high-risk ulcers such as active spurting (Forrest Ia), active oozing (Forrest Ib), or with a non-bleeding visible vessel (Forrest IIa) should receive an endoscopic therapy owing to the high risk of persistent bleeding or rebleeding. Peptic ulcers with an adherent clot (Forrest IIb) should be subjected to endoscopic clot removal to decide on further treatment plans. Ulcers with red spots (Forrest IIc) or a clean base can be observed without endoscopic therapy.

Accurate identification of such stigmata of hemorrhage is essential for endoscopists to deliver appropriate care; however, the ability of correct classification varies with endoscopists' experience. Laine et al. [10] reported that the rate of correctly identifying endoscopic stigmata of hemorrhage increased with endoscopic experience (performing five cases per month) from 59% to 73% before a training course. After the training course, the increase was related to the level of training: fellows, 15% increase; physicians with 0–20 years since training, 8% increase; and physicians with 20 years or more since training, 3% increase. Another study from Italy reported a high interobserver agreement for Forrest Ia/b lesions, but low agreement for Forrest II/III lesions [11]. The Canadian registry of patients with upper gastrointestinal bleeding revealed that only 47.8% of patients with high-risk stigmata received endoscopic therapy, whereas 9.8% of those with low-risk stigmata received endoscopic therapy [12], showing the wide variation of endoscopist practice in the real world.

Artificial intelligence (AI) is an emerging new technology that affects several aspects of healthcare. In endoscopy, AI is currently being used to detect lesions during endoscopic procedures such as colorectal lesions during colonoscopy [13], esophageal cancer during endoscopy [14], and small bowel ulcers during capsule endoscopy [15]. All these developments aim to provide a diagnostic efficacy that is similar or even superior to that of experienced endoscopists. The improved medical therapy such as the use of proton pump inhibitors and eradication of *Helicobacter pylori* infections have led to a decrease in the peptic ulcer disease rate. Achieving successful hemostasis and providing the best care depends on endoscopic skills and experience. A survey study from the UK demonstrates the decline over time in trainee experience for peptic ulcer bleeding from 76% in 1996 to 15% in 2011 [2]. The study also highlights a lack of trainee experience in more challenging cases, particularly in the out-of-hours period. Although the endoscopic skill such as injection, coagulation, or clipping to perform hemostasis can be improved by training with various *ex-vivo* model [16], the experience of determining the optimal management for a bleeding peptic ulcer can typically only be obtained by practice with real bleeding cases. Thus, there is a need to

develop a tool to assist trainees or young endoscopists during their management of bleeding peptic ulcers. Machine learning algorithms could predict the severity of a bleeding peptic ulcer with an acceptable accuracy from still endoscopic images.

Thus, in this study, we aimed to evaluate the performance of a deep learning model in classifying still endoscopic color images obtained from patients with bleeding peptic ulcers.

2 Materials and Methods

2.1 Patients and Data Preparation

The endoscopy records of patients who underwent endoscopic examination between January 2015 and January 2020 at the endoscopy center of Changhua Christian Hospital were retrospectively reviewed. The images were reviewed and retrieved for subsequent analysis by two expert endoscopists with 15 years of experience in therapeutic endoscopy. Inclusion criteria for analysis of the images were (a) images from patients with symptoms of gastrointestinal bleeding, i.e., hematemesis, anemia, or tarry stool; (b) bleeders were attributed to a peptic ulcer disease, i.e., gastric or duodenal ulcers; and (c) endoscopy performed with the Olympus 260 or 290 series system. Endoscopic images with a clear view of the pre-treatment peptic ulcers were included as the peptic ulcer group, whereas endoscopic images with a normal appearance of the gastric or duodenal mucosa were included as the control group. Images with bleeding from variceal hemorrhage, neoplasm, angiodysplasia, post polypectomy hemorrhage, or bleeding of unknown origin were excluded.

The endoscopic images of the bleeding peptic ulcers were first classified according to the Forrest classification after a consensus was reached between the two endoscopists as the ground truth for this study (expert 1). Next, we stratified the ulcers according to the clinical guideline as “no need of endoscopic therapy”, i.e., Forrest IIc and Forrest III lesions, and “need of endoscopic therapy”, i.e., Forrest Ia to Forrest IIb lesions. Those ulcer images requiring endoscopic therapy were further classified to determine the risk of endoscopic therapy after a review by two endoscopists. Ulcers with difficult locations such as the duodenum or the lesser curvature site or those with a big ulcer base or big visible vessels are considered high risk for endoscopic therapy as the primary endoscopic hemostasis may fail [17] and the remaining were considered low risk for endoscopic therapy.

The study complied with the World Medical Association Declaration of Helsinki for medical research involving human subjects, including research on identifiable human material and data and was approved by the institutional

review board of Changhua Christian Hospital (approval number: CCH IRB 200906).

2.2 Training of the Deep Learning Models

The endoscopic images were cropped to remove possible identification data for subsequent image training. The experiment was performed on the DeepQ AI Platform (<https://deepq.com/artificial-intelligence/>) for image classification task with different deep learning models. The DeepQ AI Platform is dedicated for medical imaging training with pre-trained established models using the ImageNet dataset such as ResNet-50, Inception-v4, VGG-16, and Mobile Net V2 models by the fine-tuning of these networks for data training. Considering the model performance and speed, the Mobile Net V2 model [18, 19] was chosen in this study for its potential use in the mobile circumference for emergent clinical consultation. A 5-fold cross-validation on the training set was performed such that the data was split randomly by the patient into five sets for training and validation. The batch size was 32, and training epoch was 100. Data augmentation was performed with a horizontal flip of 0.5.

2.3 Evaluation of the Model Performance

To evaluate the performance of the trained model system, we compared the accuracy of the proposed model with an additional testing dataset. For comparison with human endoscopists, two additional endoscopists (expert 2,3) with 10 years of endoscopic experience, one junior endoscopist with 2 years of experience (novice 1), and two novice endoscopists (novice 2 and 3) with one year of experience who were blinded to the study design reviewed the validation dataset for evaluating the accuracy.

3 Results

3.1 Characteristics of the Image Dataset for Analysis

A total of 2738 images from 1694 reviewed patients were selected for analysis as shown in Fig. 1. The images were split into the training (2289 images) and testing datasets (449 images). The model training task was evaluated for three-class tasks (normal vs. no therapy vs. therapy required) and four-class (normal vs. no therapy vs. low-risk therapy vs. high-risk therapy) tasks.

3.2 Results of Different Model Performance Metrics

The results of the performance comparison of the different models are presented in Table 1. The MobileNet V2 model had the shortest training time with a training accuracy of 90.59% for the four-class classification task and 94.09% for the three-class classification task.

The detailed result of the trained MobileNet V2 model for the prediction of the testing dataset is shown in Table 2. The prediction of normal, no therapy, and therapy in the 3-class

Table 1 Performance results of the different deep learning models

Model accuracy	4-class classification		3-class classification	
	Training accuracy (%)	Inference accuracy (%)	Training accuracy (%)	Inference accuracy (%)
Mobile net V2	90.59	79.73	94.09	83.07
VGG16	89.50	82.18	94.53	88.86
Inception V4	91.25	81.96	95.84	87.31
ResNet50	92.78	80.18	95.19	84.36

Fig. 1 Flow chart of images included in this study

Table 2 Model prediction results for the testing dataset

Label	TP	FP	FN	TN	F1 Score	Sensitivity	Specificity
3-Class Classification							
Normal	165	21	9	254	91.70%	94.83%	92.36%
No therapy	115	26	43	265	76.90%	72.78%	91.07%
Therapy	93	29	24	303	77.80%	79.49%	91.27%
4-Class Classification							
Normal	166	20	8	255	92.20%	95.40%	92.70%
No therapy	116	27	42	264	77.10%	73.42%	90.70%
High risk	9	4	11	425	54.50%	45.00%	99.10%
Low risk	67	40	30	312	65.70%	69.07%	88.60%

Fig. 2 The prediction in the 3-class classification

classification was high with an area under the receiver operating characteristic curve (AUROC) of 0.98, 0.92, and 0.91, respectively (Fig. 2). The prediction of normal, no therapy, high-risk therapy, and low-risk therapy in the 4-class classification was moderate to high with an AUROC of 0.99, 0.89, 0.92, and 0.88, respectively (Fig. 3). Examples of correct and incorrect labeled endoscopic images are illustrated in Figs. 4 and 5. The model in both classification tasks showed high sensitivity and specificity for determining the normal endoscopic images, but a lower sensitivity for determining the need for therapy. The sensitivity for determining peptic ulcers further decreased while stratifying the therapy groups into high- and low-risk groups.

3.3 Comparison of the Performance of the Deep Learning Model and Human Endoscopists

Tables 3 and 4 present the interobserver agreements on the testing image dataset with Cohen's kappa coefficient [20] for the 3-class classification and 4-class classification task. Interobserver agreement was high in the expert group for

Fig. 3 The prediction in the 4-class classification

the 3-class classification and substantial in the 4-class classification. The agreement of the deep learning model was substantial in the expert group and was higher than that of novice 2 and 3. The accuracy of the deep learning model in the 3-class classification for the determination of therapy was higher than that of novice 2 and 3 (Table 5).

4 Discussion

In the present study, we proposed a deep learning model for the classification of patients with bleeding peptic ulcer. A better prediction result from the deep learning model compared with novice human endoscopists was observed in our study. Our study is the first to show the potential use of the deep learning model in the management of peptic ulcer bleeding, particularly for young endoscopists, in the era of decreasing experience in managing such a gastrointestinal emergency [2, 8].

The development of AI has emerged to impact several aspects of human life in the twenty-first century. Since 2010,

(a)

(b)

Fig. 4 Examples of correct labeling results of the trained model. **a** A protruding non-bleeding vessel in a duodenal ulcer was correctly labelled as high risk for endoscopic therapy with the 4-class classifi-

cation model. **b** The heatmap showing an active bleeding in the stomach and was correctly labeled as endoscopic therapy required in the 3-class classification model

substantial progress has been made to extend its application to the health care field with the introduction of deep learning methods [21]. As the incidence of colorectal cancer is increasing, the need for colonoscopy screening has also increased [5, 6], but the manpower of endoscopists is lacking. Thus, the current AI technology in the field of endoscopy has mainly developed to aid in the detection and diagnosis of colon polyps to improve the quality results of

colonoscopy. The use of deep learning methods had been shown to be superior to that of the shape and context information detection method for classification, detection, segmentation, and tracking of polyps and an increase in the rate of accurate diagnosis with a reported accuracy of 80–96% [13, 22, 23].

The application of such technology in upper GI endoscopy had also been attempted [14, 24–26]. Yoshimasa et al.

Fig. 5 Examples of incorrect labeling results of the trained model. **a** Endoscopic view of the gastric antrum with mucus and bubbles was incorrectly labelled as an ulcer not needing endoscopic therapy. **b**

Endoscopic view of an ulcer in the gastric antrum with oozing was incorrectly labelled as no therapy

[26] from Japan utilized deep learning for the detection of esophageal cancer with a sensitivity of 98%. Rintaro et al. [25] reported a pilot study with 458 test images (225 dysplasia and 233 non-dysplasia) and correctly detected early neoplasia with a sensitivity of 96.4%, specificity of 94.2%, and accuracy of 95.4%. The human esophagus and colon are narrowed with less mucosal inflammation, and high-quality endoscopic images are easier to be obtained from

the esophagus and colon than from the stomach. Therefore, endoscopic detection of gastric lesions is usually more difficult than that of the esophagus and colon, especially in an emergency setting.

Zhang et al. [27] recently reported a diagnostic system based on a ResNet34 residual network structure of five gastric conditions, including peptic ulcer (PU), early gastric cancer and high-grade intraepithelial neoplasia, advanced

Table 3 The interobserver agreement of the testing dataset based on the 3-class classification evaluated with Cohen's kappa coefficient

	Expert 2	Expert 3	Novice1	Novice 2	Novice 3	Deep learning model
Expert 1	0.86535	0.85380	0.85380	0.56516	0.36524	0.74977
Expert 2		0.84838	0.83171	0.62997	0.43529	0.75338
Expert 3			0.77711	0.63133	0.33722	0.69696
Novice 1				0.53042	0.33722	0.75271
Novice 2					0.65071	0.50668
Novice 3						0.30499

Cohen's kappa coefficient: <0 as no agreement, 0–0.20 as slight, 0.21–0.40 as fair, 0.41–0.60 as moderate, 0.61–0.80 as substantial, and 0.81–1 as almost perfect agreement

Table 4 The interobserver agreement of the testing dataset based on the 4-class classification evaluated with Cohen's kappa coefficient

	Expert 2	Expert 3	Novice 1	Novice 2	Novice 3	Deep learning model
Expert 1	0.76234	0.70270	0.79002	0.45236	0.26978	0.69836
Expert 2		0.73999	0.70848	0.53714	0.37641	0.62796
Expert 3			0.72968	0.53594	0.39412	0.61998
Novice 1				0.43142	0.24761	0.69140
Novice 2					0.55031	0.39034
Novice 3						0.21629

Cohen's kappa coefficient: <0 as no agreement, 0–0.20 as slight, 0.21–0.40 as fair, 0.41–0.60 as moderate, 0.61–0.80 as substantial, and 0.81–1 as almost perfect agreement [20]

Table 5 Accuracy comparison of human endoscopists and the deep learning model

Operator	TP	FP	FN	TN	Sen (%)	Spe (%)	Accuracy (%)
<i>Accuracy for determining endoscopic therapy in the 3-class classification</i>							
Expert 2	109	27	8	305	93.16	91.87	92.20
Expert 3	106	52	11	280	90.6	84.34	85.97
Novice 1	90	12	27	320	76.92	96.39	91.31
Novice 2	112	70	5	262	95.73	78.92	83.30
Novice 3	114	117	3	215	97.44	64.76	73.27
Deep learning model	93	29	24	303	79.49	91.27	88.20
<i>Accuracy for determining high risk for endoscopic therapy in the 4-class classification</i>							
Expert 2	17	31	3	398	85.00	92.77	92.43
Expert 3	6	13	14	416	30.00	96.97	93.99
Novice 1	4	4	16	425	20.00	99.07	95.55
Novice 2	11	35	9	394	55.00	91.84	90.20
Novice 3	12	36	8	393	60.00	91.61	90.20
Deep learning model	9	4	11	425	45.00	99.07	96.66

gastric cancer, gastric submucosal tumors (SMTs), and normal gastric mucosa without lesions, with a diagnostic accuracy of 74.2–88.9%. Compared with these gastric conditions, a high-quality image from bleeding PU is usually more difficult to obtain compared with those obtained from non-emergency settings [28]. To the best of our knowledge, there is no previous study that has employed such deep learning for such endoscopic emergency, as we are the first to explore its potential use in clinical practice.

Patients presenting with PU bleeding is a clinical emergency requiring a prompt clinical decision for appropriate management. Endoscopic Forrest classification of the ulcer morphology is the corner stone of clinical trials for making a decision to perform endoscopic or medical therapy since its development [4–6, 11, 29]. However, the number of patients with PUs has decreased with the eradication of *H. pylori* infection [3]. Young endoscopists do not have sufficient experience in managing such diseases [1, 30];

thus, developing a computer-aided system may be helpful for such critical clinical decision-making. Current guidelines [3, 8, 9] suggest that endoscopic therapy should be provided to Forrest I, IIa, and IIb lesions; therefore, in the current study, we simplified the classification to “need” or “no need” of endoscopic therapy to fit the real-world clinical practice pattern. In addition, we also attempted to further stratify the risk of endoscopic therapy by experienced endoscopists based on the still endoscopic images. Our model shows high sensitivity and specificity for determining the normal endoscopic images, but low for determining the need of therapy or its difficulty. Compared with the higher sensitivity/specificity reported in other endoscopic situations, the lower sensitivity observed in our model was mainly explained by the complicated circumstances during PU bleeding, particularly for those with high-risk ulcers requiring endoscopic therapy, i.e., the image view of the lesion is difficult to be standardized in cases of emergency, the gastric contents are not clear, and the adjacent gastric/duodenal mucosa is inflamed or deformed such that it may increase the difficulty of such discrimination tasks. We performed our first experiment with different currently available deep learning models, and MobileNet V2 was chosen for subsequent study because of its acceptable accuracy and its shorter computing time that may be utilized in an emergency consultation setting.

A strength of this study was the comparison of the performance of the deep learning model with that of human endoscopists. The high interobserver agreement rate among experienced endoscopists and the low interobserver agreement rate among young endoscopists revealed the need for training to improve their experience to the expert level. Our trained model had a potential to be used as an aid for young endoscopists during their training process. In addition, we attempted to stratify the risk of endoscopic therapy based on the endoscopists’ experience in the current study. We found that discrepancies may exist in the endoscopists’ opinion, i.e., a good skilled endoscopist may consider a high-risk ulcer as a low-risk ulcer. In contrast, inexperienced endoscopists may also judge a high-risk ulcer as a low-risk ulcer owing to the lack of experience. Further studies are required to ensure better consensus among experts for this risk stratification classification for subsequent clinical use.

There are several limitations of the current study. First, the dataset came from one hospital in the past five years and only the Olympus endoscope system was utilized; thus, external validation is required. Second, the classification of the obtained images came mainly from two experts in our institution and may not reflect the opinion of other experts from other institutions. In addition, the images obtained for model training were captured from stored still images. PU bleeding is a dynamic process, and further studies are still required with video clips to better help the clinical practice.

5 Conclusion

In conclusion, we report the first use of a convolutional neural network for classifying endoscopic images of bleeding peptic ulcers. The performance of the model was better than that of endoscopist trainees. Further improvement of the model may aid in clinical decision-making during the training of young endoscopists.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s40846-021-00608-0>.

Acknowledgements The authors are grateful to the DeepQ company (<https://deepq.com>) for providing the DeepQ AI Platform for training and validation of the model utilized in this study. The authors received funding from the Changhua Christian Hospital (109-CCH-CYCU-001 and 109-CCH-IRP-008) for this study.

Declarations

Conflict of interest The authors have no conflict of interest to declare.

Ethical Approval The study was performed following the principles outlined in the Helsinki Declaration and it was approved by the Ethics Committee (IRB Number: 200906).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Waddell, K. M., Stanley, A. J., & Morris, A. J. (2017). Endoscopy for upper gastrointestinal bleeding: Where are we in 2017? *Frontline Gastroenterology*, 8(2), 94–97. <https://doi.org/10.1136/flgastro-2016-100791>.
2. Penny, H. A., Kurien, M., Wong, E., Ahmed, R., Ejenavi, E., Lau, M., Romaya, C., Gohar, F., Dear, K. L., Kapur, K., Hoeroldt, B., Lobo, A. J., & Sanders, D. S. (2016). Changing trends in the UK management of upper GI bleeding: Is there evidence of reduced UK training experience? *Frontline Gastroenterology*, 7(1), 67–72. <https://doi.org/10.1136/flgastro-2014-100537>.
3. Gralnek, I. M., Dumonceau, J. M., Kuipers, E. J., Lanas, A., Sanders, D. S., Kurien, M., Rotondano, G., Hucl, T., Dinis-Ribeiro, M., Marmo, R., Racz, I., Arezzo, A., Hoffmann, R.-T., Lesur, G., de Franchis, R., Aabakken, L., Veitch, A., Radaelli, F., Salgueiro, P., ... Hassan, C. (2015). Diagnosis and management of nonvariceal upper gastrointestinal hemorrhage: European Society of Gastrointestinal Endoscopy (ESGE) Guideline. *Endoscopy*, 47(10), a1–46. <https://doi.org/10.1055/s-0034-1393172>.

4. Forrest, J. A., Finlayson, N. D., & Shearman, D. J. (1974). Endoscopy in gastrointestinal bleeding. *Lancet*, 2(7877), 394–397. [https://doi.org/10.1016/s0140-6736\(74\)91770-x](https://doi.org/10.1016/s0140-6736(74)91770-x).
5. Yen, H. H., Yang, C. W., Su, W. W., Soon, M. S., Wu, S. S., & Lin, H. J. (2012). Oral versus intravenous proton pump inhibitors in preventing re-bleeding for patients with peptic ulcer bleeding after successful endoscopic therapy. *BMC Gastroenterology*, 12, 66. <https://doi.org/10.1186/1471-230X-12-66>.
6. Yen, H. H., Yang, C. W., Su, P. Y., Su, W. W., & Soon, M. S. (2011). Use of hemostatic forceps as a preoperative rescue therapy for bleeding peptic ulcers. *Surgical Laparoscopy, Endoscopy and Percutaneous Techniques*, 21(5), 380–382. <https://doi.org/10.1097/SLE.0b013e3182303007>.
7. Kim, D. S., Jung, Y., Rhee, H. S., Lee, S. J., Jo, Y. G., Kim, J. H., et al. (2016). Usefulness of the Forrest Classification to predict artificial ulcer rebleeding during second-look endoscopy after endoscopic submucosal dissection. *Clinical Endoscopy*, 49(3), 273–281. <https://doi.org/10.5946/ce.2015.086>.
8. Barkun, A. N., Almadi, M., Kuipers, E. J., Laine, L., Sung, J., Tse, F., Leontiadis, G. I., Abraham, N. S., Calvet, X., Chan, F. K. L., Douketis, J., Enns, R., Gralnek, I. M., Jairath, V., Jensen, D., Lau, J., Lip, G. Y. H., Loffroy, R., Maluf-Filho, F., ... Bardou, M. (2019). Management of nonvariceal upper gastrointestinal bleeding: Guideline recommendations from the international consensus group. *Annals of Internal Medicine*, 171(11), 805–822. <https://doi.org/10.7326/M19-1795>.
9. Sung, J. J., Chiu, P. W., Chan, F. K. L., Lau, J. Y., Goh, K. L., Ho, L. H., Jung, H.-Y., Sollano, J. D., Gotoda, T., Reddy, N., Singh, R., Sugano, K., Wu, K.-C., Wu, C.-Y., Bjorkman, D. J., Jensen, D. M., Kuipers, E. J., & Lanas, A. (2018). Asia-Pacific working group consensus on non-variceal upper gastrointestinal bleeding: An update 2018. *Gut*, 67(10), 1757–1768. <https://doi.org/10.1136/gutjnl-2018-316276>.
10. Laine, L., Freeman, M., & Cohen, H. (1994). Lack of uniformity in evaluation of endoscopic prognostic features of bleeding ulcers. *Gastrointestinal Endoscopy*, 40(4), 411–417. [https://doi.org/10.1016/s0016-5107\(94\)70202-0](https://doi.org/10.1016/s0016-5107(94)70202-0).
11. Mondardini, A., Barletti, C., Rocca, G., Garripoli, A., Sambataro, A., Perotto, C., Repici, A., & Ferrari, A. (1998). Non-variceal upper gastrointestinal bleeding and Forrest's classification: Diagnostic agreement between endoscopists from the same area. *Endoscopy*, 30(6), 508–512. <https://doi.org/10.1055/s-2007-1001335>.
12. Lu, Y., Barkun, A. N., & Martel, M. (2014). Adherence to guidelines: A national audit of the management of acute upper gastrointestinal bleeding The REASON registry. *Canadian Journal of Gastroenterology and Hepatology*, 28(9), 495–501. <https://doi.org/10.1155/2014/252307>.
13. Attardo, S., Chandrasekar, V. T., Spadaccini, M., Maselli, R., Patel, H. K., Desai, M., Capogreco, A., Badalamenti, M., Galtieri, P. A., Pellegatta, G., Fugazza, A., Carrara, S., Anderloni, A., Occhipinti, P., Hassan, C., Sharma, P., & Repici, A. (2020). Artificial intelligence technologies for the detection of colorectal lesions: The future is now. *World Journal of Gastroenterology*, 26(37), 5606–5616. <https://doi.org/10.3748/wjg.v26.i37.5606>.
14. Luo, H., Xu, G., Li, C., He, L., Luo, L., Wang, Z., Jing, B., Deng, Y., Jin, Y., Li, Y., Li, B., Tan, W., He, C., Seeruttun, S. R., Wu, Q., Huang, J., Huang, D.-W., Chen, B., Lin, S.-B., ... Xu, R.-H. (2019). Real-time artificial intelligence for detection of upper gastrointestinal cancer by endoscopy: A multicentre, case-control, diagnostic study. *Lancet Oncology*, 20(12), 1645–1654. [https://doi.org/10.1016/S1470-2045\(19\)30637-0](https://doi.org/10.1016/S1470-2045(19)30637-0).
15. Aoki, T., Yamada, A., Aoyama, K., Saito, H., Tsuboi, A., Nakada, A., Niikura, R., Fujishiro, M., Oka, S., Ishihara, S., Matsuda, T., Tanaka, S., Koike, K., & Tada, T. (2019). Automatic detection of erosions and ulcerations in wireless capsule endoscopy images based on a deep convolutional neural network. *Gastrointestinal Endoscopy*, 89(2), 357 e352–363 e352. <https://doi.org/10.1016/j.gie.2018.10.027>.
16. Lee, D. S., Ahn, J. Y., & Lee, G. H. (2019). A Newly designed 3-dimensional printer-based gastric hemostasis simulator with two modules for endoscopic trainees (with Video). *Gut Liver*, 13(4), 415–420. <https://doi.org/10.5009/gnl18389>.
17. Mullady, D. K., Wang, A. Y., & Waschke, K. A. (2020). AGA clinical practice update on endoscopic therapies for non-variceal upper gastrointestinal bleeding: Expert review. *Gastroenterology*, 159(3), 1120–1128. <https://doi.org/10.1053/j.gastro.2020.05.095>.
18. Singh, B., Toshniwal, D., & Allur, S. K. (2019). Shunt connection: An intelligent skipping of contiguous blocks for optimizing MobileNet-V2. *Neural Networks*, 118, 192–203. <https://doi.org/10.1016/j.neunet.2019.06.006>.
19. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. (2018, June 18–23). MobileNetV2: Inverted residuals and linear bottlenecks. In *2018 IEEE/CVF conference on computer vision and pattern recognition* (pp. 4510–4520). <https://doi.org/10.1109/CVPR.2018.00474>.
20. Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174.
21. Yang, H. C., Islam, M. M., & Jack Li, Y. C. (2018). Potentiality of deep learning application in healthcare. *Computer Methods and Programs in Biomedicine*, 161, A1. <https://doi.org/10.1016/j.cmpb.2018.05.014>.
22. Urban, G., Tripathi, P., Alkayali, T., Mittal, M., Jalali, F., Karnes, W., & Baldi, P. (2018). Deep learning localizes and identifies polyps in real time with 96% accuracy in screening colonoscopy. *Gastroenterology*, 155(4), 1069 e1068–1078 e1068. <https://doi.org/10.1053/j.gastro.2018.06.037>.
23. Tajbakhsh, N., Gurudu, S. R., & Liang, J. (2016). Automated polyp detection in colonoscopy videos using shape and context information. *IEEE Transactions on Medical Imaging*, 35(2), 630–644. <https://doi.org/10.1109/TMI.2015.2487997>.
24. Zhang, Y. H., Guo, L. J., Yuan, X. L., & Hu, B. (2020). Artificial intelligence-assisted esophageal cancer management: Now and future. *World Journal of Gastroenterology*, 26(35), 5256–5271. <https://doi.org/10.3748/wjg.v26.i35.5256>.
25. Hashimoto, R., Requa, J., Dao, T., Ninh, A., Tran, E., Mai, D., Lugo, M., El-Hage Chehade, N., Chang, K. J., Karnes, W. E., & Samarasena, J. B. (2020). Artificial intelligence using convolutional neural networks for real-time detection of early esophageal neoplasia in Barrett's esophagus (with video). *Gastrointestinal Endoscopy*, 91(6), 1264–1271. <https://doi.org/10.1016/j.gie.2019.12.049>.
26. Horie, Y., Yoshio, T., Aoyama, K., Yoshimizu, S., Horiuchi, Y., Ishiyama, A., Hirasawa, T., Tsuchida, T., Ozawa, T., Ishihara, S., Kumagai, Y., Fujishiro, M., Maetani, I., Fujisaki, J., & Tada, T. (2019). Diagnostic outcomes of esophageal cancer by artificial intelligence using convolutional neural networks. *Gastrointestinal Endoscopy*, 89(1), 25–32. <https://doi.org/10.1016/j.gie.2018.07.037>.
27. Zhang, L., Zhang, Y., Wang, L., Wang, J., & Liu, Y. (2020). Diagnosis of gastric lesions through a deep convolutional neural network. *Digestive Endoscopy*. <https://doi.org/10.1111/den.13844>.
28. Namikawa, K., Hirasawa, T., Nakano, K., Ikenoyama, Y., Ishioka, M., Shiroma, S., et al. (2020). Artificial intelligence-based diagnostic system classifying gastric cancers and ulcers: Comparison between the original and newly developed systems. *Endoscopy*, 52(12), 1077–1083. <https://doi.org/10.1055/a-1194-8771>.
29. de Groot, N. L., van Oijen, M. G., Kessels, K., Hemmink, M., Weusten, B. L., Timmer, R., Hazen, W. L., van Lelyveld, N., Vermeijden, R. R., Curvers, W. L., & Baak, B. C. (2014). Reassessment of the predictive value of the Forrest classification for peptic ulcer rebleeding and mortality: Can classification be simplified?

- Endoscopy*, 46(1), 46–52. <https://doi.org/10.1055/s-0033-1344884>.
30. Siau, K., Hawkes, N. D., & Dunckley, P. (2018). Training in endoscopy. *Current Treatment Options in Gastroenterology*, 16(3), 345–361. <https://doi.org/10.1007/s11938-018-0191-1>.