Peptic Ulcer Bleeding and AI

Peptic ulcer bleeding (PUB) is a common gastrointestinal (GI) emergency requiring prompt assessment, with a mortality rate of 2–10%.

1. Introduction

Peptic ulcer bleeding (PUB) is a common gastrointestinal (GI) emergency requiring prompt assessment, with a mortality rate of 2–10% [1][2][3]. Recently, with the reduced incidence of peptic ulcer disease and the advancement of endoscopic therapy, the bleeding-related hospitalization and mortality rates of PUB have decreased [4][5]. International guidelines have been updating the optimal management approach for patients suffering from PUB [6][7]. The cascade of management can be divided into three stages: pre-endoscopy, endoscopic, and post-endoscopy management. Pre-endoscopy management includes assessing patient's risk for hospitalization, providing adequate fluid and blood component resuscitation, prescribing medication such as a proton pump inhibitor (PPI), and identifying the timing of endoscopy (**Figure 1**). Endoscopic management includes assessing the nature of bleeder (e.g., peptic ulcer disease, malignancy, or variceal hemorrhage) and providing endoscopic therapy as appropriate. For post-endoscopy management, intravenous PPI infusion therapy is prescribed to reduce PUB recurrence. Furthermore, eradication of *Helicobacter pylori* infection decreases the recurrence of peptic ulcer disease, and long-term secondary PPIs are required for patients who are at risk for recurrent bleeding.
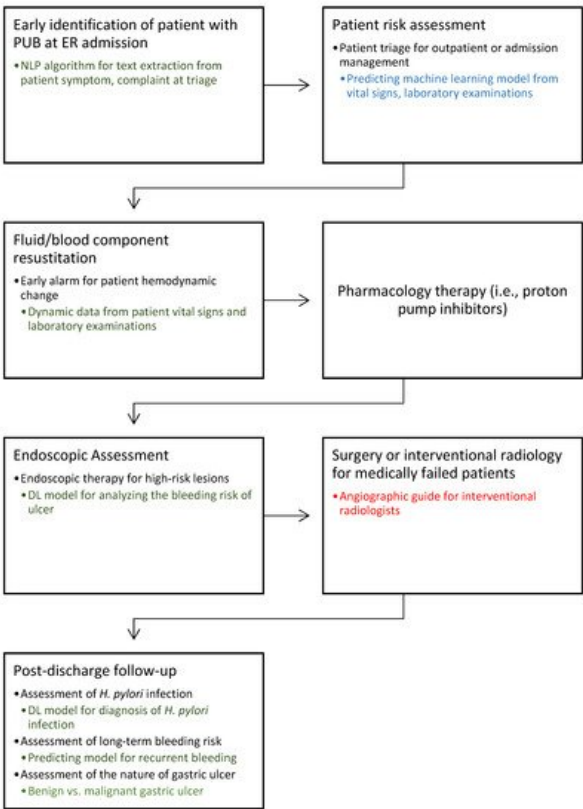
Figure 1. Cascade management of peptic ulcer bleeding. This diagram illustrates the potential role of artificial intelligence (AI) in the future management of peptic ulcer bleeding (PUB) based on text, data, and imaging. Blue: studies with multicenter clinical data validation; Green: studies with single-center clinical data validation; Red: no relevant research found in this field. AI, artificial intelligence; DL, deep learning; ER, emergency room; NLP, natural language processing; PUB, peptic ulcer bleeding.

Developed since the 1950s, artificial intelligence (AI) refers to computer programs that can simulate the human cognitive process in problem-solving and learning [8]. Through the machine learning (ML) approach, the computer can process large data to build various predicting models. Meanwhile, deep learning (DL) has further simulated human neuronal networks with improved performance, especially image processing, since 2010. A UK survey study demonstrated that the gastroenterology trainee experience for PUB management declined from 76% in 1996 to 15% in 2011 [2], owing to the decreased incidence of peptic ulcer disease. The use of AI technology for PUB could enhance the accuracy of patient triage, help achieve accurate therapeutic decisions, and prevent human errors caused by inexperience, especially in an emergency. In this review, we highlight the published literature in the last 5 years with keywords of "artificial intelligence", "peptic ulcer bleeding", "nonvariceal bleeding", "deep learning", or "machine learning" from a PubMed search to determine the current status and gain insight into the role of AI in PUB management.

2. Application of AI in the Pre-Endoscopy Period for Patient Risk Assessment

Upon presentation at the hospital, stratification of patients in terms of gastrointestinal bleeding (GIB) risk is recommended [6][7][9]. Accurately identifying ("phenotype") patients with GIB during initial assessment is the first step toward patient management, especially during these times of the COVID-19 pandemic. Shung et al. [10] used multiple natural language processing (NLP)-based approaches for automated phenotyping of patients in the emergency department. They found that the syntax-based NLP algorithm from patient triage information performed better than the systematized nomenclature of medicine code information for the patient's condition, which allows early use of patient triage to subsequent patient management.

In the past two decades, three widely validated scoring systems, namely, Glasgow–Blatchford score (GBS) for outpatient management [11], Rockall score for mortality [12], and the AIMS65 score [13][14][15], have been utilized for predicting low-risk patients. However, compared with these conventional scores [16], ML can potentially improve risk assessment for the need for transfusion, endoscopic evaluation, or hospital admission for observation. Clinical ML use is also more feasible than such conventional scores for busy clinicians through the automatic deployment of ML models with existing available electronic health records in many healthcare systems. In 2003–2008, nine small studies were conducted to investigate ML's potential for PUB risk assessment in comparison with the conventional scores [16]. The median areas under the curve (AUCs) were higher in artificial neural networks (0.93; range, 0.78–0.98) than in other ML models (0.81, range: 0.40–0.92) when predicting patient mortality, intervention requirement, or rebleeding. Moreover, ML generally provided a better prognostic performance in patients with GIB than conventional scores, and artificial neural networks tended to outperform other ML models.

In 2020, Seo et al. [17] prospectively analyzed 1439 PUB cases to compare the accuracy of ML and conventional scores for PUB patient instability including hypotension, rebleeding, and mortality. Four ML algorithms, namely, logistic regression with regularization, random forest classifier (RF), gradient boosting classifier (GB), and voting classifier (VC), were compared using the GBS and Rockall scores. The RF model was the most accurate in predicting mortality (AUROC: RF 0.917 vs. GBS 0.710), while the VC model was the most accurate for hypotension (VC 0.757 vs. GBS 0.668) and rebleeding within 7 days (VC 0.733 vs. GBS 0.694). The global feature importance analysis identified clinically significant variables, including blood urea nitrogen, albumin, hemoglobin, platelet, prothrombin time, age, and lactate. Thus, the ML models may be helpful in early predicting high-risk patients with initially stable upper GIB upon admission to the emergency department. However, ML performance relies on the quality of data, and these studies usually had a small sample size (<1000 cases) with no external validation data for their performance.

Shung et al. [18] were the first to conduct a large prospective international study for building an ML model for patients with PUB by testing and comparing the performance of the ML model and the conventional scoring system in 2020. They collected patient data from medical centers in four countries (US, Scotland, England, and Denmark; n = 1958) to build a model that can predict the need for hospital-based intervention (transfusion or hemostatic intervention) or 1 month mortality. Data from two Asia-Pacific sites (Singapore and New Zealand; n = 399) were externally validated. Only nonendoscopic features such as age, sex, clinical symptoms, and laboratory variables (hemoglobin, albumin, international normalized ratio, urea, and creatinine) were selected to build the model. The ML model showed a higher AUC (0.91) than GBS (0.88, p = 0.001), Rockall score (0.73, p < 0.001), and AIMS65 score (0.78, p < 0.001). In the external validation cohort, the ML still achieved a higher AUC (0.90) than GBS (0.87, p = 0.004), Rockall score (0.66, p < 0.001), and AIMS65 score (0.64 (p < 0.001). The proposed ML model improved the identification of low-risk patients who can be safely discharged early from the emergency department. Importantly, this ML model identified more than two times the number of patients with very low risk than the available best-performing clinical risk tool.

After presentation in the hospital, initially stable patients who are at risk for hemodynamic instability requiring blood transfusion must be identified during the dynamic monitoring of the patient status. Levi et al. [19] developed an ML model using publicly available intensive care unit (ICU) databases of 14,620 records with input variables, including several laboratory analyses and demographic information. Their model, which was based on the patient's vital signs and laboratory test changes in the first 5 h of ICU admission, showed a high level of accuracy (overall AUC, 0.80) in predicting the need for transfusion in the next 24 h of admission.

Therefore, such an algorithm is essential to provide improved risk assessment through the automatic retrieval of information from electronic health records, thereby allowing timely decision support in an already crowded clinical scenario.

3. Application of AI during Endoscopy

Forrest [20] described the endoscopic classification of PUB in 1974 (**Figure 2**). The classification requires endoscopist judgment of the risk for rebleeding and the need for endoscopic intervention. Current guidelines [3][6][7] suggest that patients who are highly at risk for ulcers, such as those with active spurting, active oozing, or a nonbleeding visible vessel, should receive endoscopic therapy because of the high risk for persistent bleeding or rebleeding, especially when only relying on drug therapy. However, the ability to make a correct classification varies with the endoscopist's experience, whereby an experienced endoscopist [21][22] can reportedly make better clinical judgment than clinical risk scores [23]. In the study of Laine et al. [24], the rate of correct identification of the endoscopic characteristic of hemorrhage increased as the endoscopic experience increased (performing five cases per month), from 59% to 73% before a training course. After the training course, the increase was related to the training level: fellows, 15% increase; physicians with 0–20 years of experience since training, 8% increase; physicians with an experience of 20 years or more since training, 3% increase. In an Italian study, Forrest Ia/b lesions showed a high interobserver agreement, whereas Forrest II/III lesions exhibited a low agreement [25].
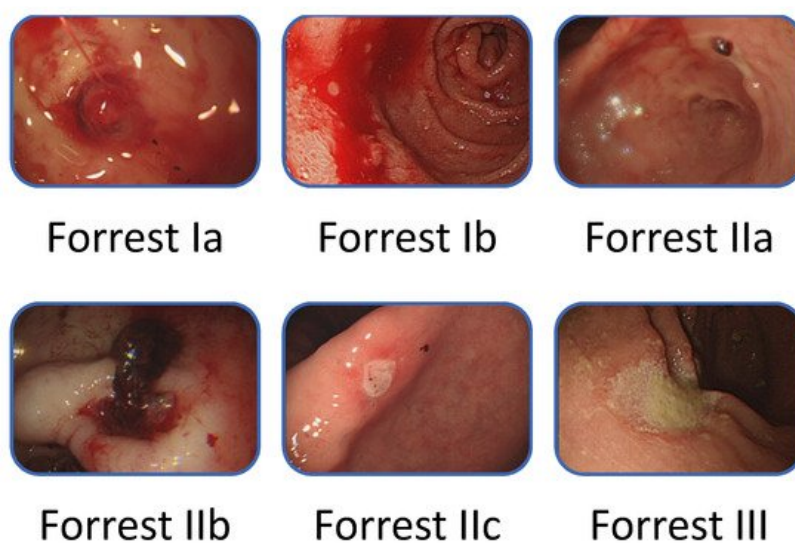
Figure 2. Forrest classification of bleeding peptic ulcers.

To explore whether AI is useful for identifying the endoscopic characteristics of hemorrhage during endoscopy, our study [26] initiated the proposal of a DL model that can classify endoscopic images with different bleeding risks according to the Forrest classification and using 2378 still endoscopic images from 1694 patients with PUB (**Figure 3**). The agreement of the model was moderate to substantial with the senior endoscopist on the testing dataset. The accuracy of the DL model was higher than that of a novice endoscopist. Therefore, the DL model has potential use, particularly in aiding young endoscopists in decision making during emergent endoscopy.
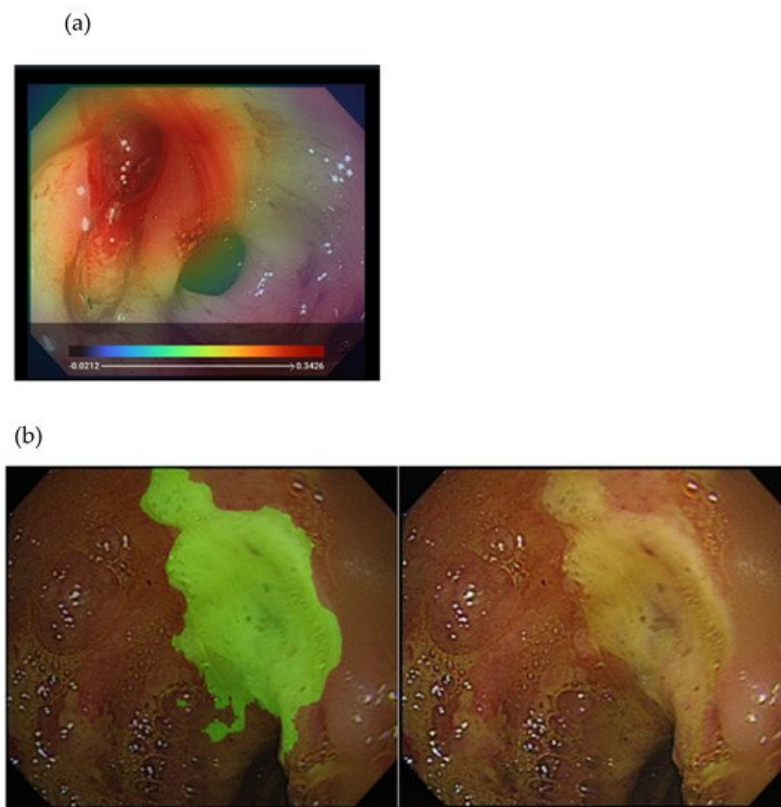
Figure 3. Illustration of the DL approach for analyzing endoscopy images in peptic ulcer disease: (**a**) heatmap image showing an active bleeder in the endoscopy image (**upper**); (**b**) segmentation of the ulcer area (**left**) from the original endoscopy image (**right**).

References

1. Waddell, K.M.; Stanley, A.J.; Morris, A.J. Endoscopy for upper gastrointestinal bleeding: Where are we in 2017? Frontline Gastroenterol. 2017, 8, 94–97.

2. Penny, H.A.; Kurien, M.; Wong, E.; Ahmed, R.; Ejenavi, E.; Lau, M.; Romaya, C.; Gohar, F.; Dear, K.L.; Kapur, K.; et al. Changing trends in the UK management of upper GI bleeding: Is there evidence of reduced UK training experience? Frontline Gastroenterol. 2016, 7, 67–72.

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.; et al. Diagnosis and management of nonvariceal upper gastrointestinal hemorrhage: European Society of Gastrointestinal Endoscopy (ESGE) Guideline. Endoscopy 2015, 47, a1–a46.

4. Lau, L.H.S.; Sung, J.J.Y. Treatment of upper gastrointestinal bleeding in 2020: New techniques and outcomes. Dig. Endosc. 2021, 33, 83–94.

5. Shivaraju, A.; Patel, V.; Fonarow, G.C.; Xie, H.; Shroff, A.R.; Vidovich, M.I. Temporal trends in gastrointestinal bleeding associated with percutaneous coronary intervention: Analysis of the 1998–2006 Nationwide Inpatient Sample (NIS) database. Am. Heart J. 2011, 162, 1062–1068.e5.

6. 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.; et al. Management of Nonvariceal Upper Gastrointestinal Bleeding: Guideline Recommendations From the International Consensus Group. Ann. Intern. Med. 2019, 171, 805–822.

7. 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.; et al. Asia-Pacific working group consensus on non-variceal upper gastrointestinal bleeding: An update 2018. Gut 2018, 67, 1757–1768.

8. El Hajjar, A.; Rey, J.F. Artificial intelligence in gastrointestinal endoscopy: General overview. Chin. Med. J. 2020, 133, 326–334.

9. Laine, L.; Barkun, A.N.; Saltzman, J.R.; Martel, M.; Leontiadis, G.I. ACG Clinical Guideline: Upper Gastrointestinal and Ulcer Bleeding. Am. J. Gastroenterol. 2021, 116, 899–917.

10. Shung, D.; Tsay, C.; Laine, L.; Chang, D.; Li, F.; Thomas, P.; Partridge, C.; Simonov, M.; Hsiao, A.; Tay, J.K.; et al. Early identification of patients with acute gastrointestinal bleeding using natural language processing and decision rules. J. Gastroenterol. Hepatol. 2021, 36, 1590–1597.

11. Stanley, A.J.; Ashley, D.; Dalton, H.R.; Mowat, C.; Gaya, D.R.; Thompson, E.; Warshow, U.; Groome, M.; Cahill, A.; Benson, G.; et al. Outpatient management of patients with low-risk upper-gastrointestinal haemorrhage: Multicentre validation and prospective evaluation. Lancet 2009, 373, 42–47.

12. Rockall, T.A.; Logan, R.F.; Devlin, H.B.; Northfield, T.C. Risk assessment after acute upper gastrointestinal haemorrhage. Gut 1996, 38, 316–321.

13. Kim, M.S.; Choi, J.; Shin, W.C. AIMS65 scoring system is comparable to Glasgow-Blatchford score or Rockall score for prediction of clinical outcomes for non-variceal upper gastrointestinal bleeding. BMC Gastroenterol. 2019, 19, 136.

14. Robertson, M.; Majumdar, A.; Boyapati, R.; Chung, W.; Worland, T.; Terbah, R.; Wei, J.; Lontos, S.; Angus, P.; Vaughan, R. Risk stratification in acute upper GI bleeding: Comparison of the AIMS65 score with the Glasgow-Blatchford and Rockall scoring systems. Gastrointest Endosc. 2016, 83, 1151–1160.

15. Hyett, B.H.; Abougergi, M.S.; Charpentier, J.P.; Kumar, N.L.; Brozovic, S.; Claggett, B.L.; Travis, A.C.; Saltzman, J.R. The AIMS65 score compared with the Glasgow-Blatchford score in predicting outcomes in upper GI bleeding. Gastrointest Endosc. 2013, 77, 551–557.

16. Shung, D.; Simonov, M.; Gentry, M.; Au, B.; Laine, L. Machine Learning to Predict Outcomes in Patients with Acute Gastrointestinal Bleeding: A Systematic Review. Dig. Dis. Sci. 2019, 64, 2078–2087.

17. Seo, D.W.; Yi, H.; Park, B.; Kim, Y.J.; Jung, D.H.; Woo, I.; Sohn, C.H.; Ko, B.S.; Kim, N.; Kim, W.Y. Prediction of Adverse Events in Stable Non-Variceal Gastrointestinal Bleeding Using Machine Learning. J. Clin. Med. 2020, 9, 2603.

18. Shung, D.L.; Au, B.; Taylor, R.A.; Tay, J.K.; Laursen, S.B.; Stanley, A.J.; Dalton, H.R.; Ngu, J.; Schultz, M.; Laine, L. Validation of a Machine Learning Model That Outperforms Clinical Risk Scoring Systems for Upper Gastrointestinal Bleeding. Gastroenterology 2020, 158, 160–167.

19. Levi, R.; Carli, F.; Arevalo, A.R.; Altinel, Y.; Stein, D.J.; Naldini, M.M.; Grassi, F.; Zanoni, A.; Finkelstein, S.; Vieira, S.M.; et al. Artificial intelligence-based prediction of transfusion in the intensive care unit in patients with gastrointestinal bleeding. BMJ Health Care Inform. 2021, 28, e100245.

20. Forrest, J.A.; Finlayson, N.D.; Shearman, D.J. Endoscopy in gastrointestinal bleeding. Lancet 1974, 2, 394–397.

21. Yen, H.H.; Yang, C.W.; Su, W.W.; Soon, M.S.; Wu, S.S.; Lin, H.J. Oral versus intravenous proton pump inhibitors in preventing re-bleeding for patients with peptic ulcer bleeding after successful endoscopic therapy. BMC Gastroenterol. 2012, 12, 66.

22. Yen, H.H.; Yang, C.W.; Su, P.Y.; Su, W.W.; Soon, M.S. Use of hemostatic forceps as a preoperative rescue therapy for bleeding peptic ulcers. Surg. Laparosc. Endosc. Percutan. Tech. 2011, 21, 380–382.

23. Brullet, E.; Garcia-Iglesias, P.; Calvet, X.; Papo, M.; Planella, M.; Pardo, A.; Junquera, F.; Montoliu, S.; Ballester, R.; Martinez-Bauer, E.; et al. Endoscopist's Judgment Is as Useful as Risk Scores for Predicting Outcome in Peptic Ulcer Bleeding: A Multicenter Study. J. Clin. Med. 2020, 9, 408.

24. Laine, L.; Freeman, M.; Cohen, H. Lack of uniformity in evaluation of endoscopic prognostic features of bleeding ulcers. Gastrointest. Endosc. 1994, 40, 411–417.

25. Mondardini, A.; Barletti, C.; Rocca, G.; Garripoli, A.; Sambataro, A.; Perotto, C.; Repici, A.; Ferrari, A. Non-variceal upper gastrointestinal bleeding and Forrest's classification: Diagnostic agreement between endoscopists from the same area. Endoscopy 1998, 30, 508–512.

26. Yen, H.-H.; Wu, P.-Y.; Su, P.-Y.; Yang, C.-W.; Chen, Y.-Y.; Chen, M.-F.; Lin, W.-C.; Tsai, C.-L.; Lin, K.-P. Performance Comparison of the Deep Learning and the Human Endoscopist for Bleeding Peptic Ulcer Disease. J. Med. Biol. Eng. 2021, 41, 504–513.