

Review

Open Access



Role of artificial intelligence in the detection, assessment and outcome of gastroesophageal varices

Gianluca Rompianesi, Francesca Pegoraro, Bianca Pacilio, Giusy Petti, Gianluca Benassai, Micaela Cappuccio, Roberto Montalti, Roberto Troisi

HPB, Minimally invasive, Robotic and Kidney Transplant Surgery Unit, Department of Clinical Medicine and Surgery - Federico II University Hospital of Naples, Naples 80131, Italy.

Correspondence to: Dr. Gianluca Rompianesi, HPB, Minimally invasive, Robotic and Kidney Transplant Surgery Unit, Department of Clinical Medicine and Surgery - Federico II University Hospital of Naples, Via S. Pansini 5, Naples 80131, Italy. E-mail: gianlucarompianesi@gmail.com

How to cite this article: Rompianesi G, Pegoraro F, Pacilio B, Petti G, Benassai G, Cappuccio M, Montalti R, Troisi R. Role of artificial intelligence in the detection, assessment and outcome of gastroesophageal varices. *Art Int Surg*. 2025;5:434-47. <https://dx.doi.org/10.20517/ais.2025.09>

Received: 9 Feb 2025 **First Decision:** 7 Jul 2025 **Revised:** 26 Jul 2025 **Accepted:** 11 Aug 2025 **Published:** 22 Aug 2025

Academic Editor: Gaya Spolverato **Copy Editor:** Pei-Yun Wang **Production Editor:** Pei-Yun Wang

Abstract

Gastroesophageal varices (GEVs) are one of the first clinically relevant consequences of PH, developing in 60%-80% of patients with liver cirrhosis. They are directly associated with a higher risk of decompensation and death. Screening endoscopy is the most common screening strategy in patients with cirrhosis. However, there is a tendency to find non-invasive predictors of GEVs to avoid costly and potentially harmful procedures safely. Artificial intelligence (AI)-driven predictive models effectively integrate diverse clinical, imaging, and laboratory data to provide non-invasive and precise risk stratification, reducing the reliance on endoscopic evaluations. Deep learning applications, particularly convolutional neural networks (CNNs), have proved highly effective in analyzing endoscopic images, thereby enhancing diagnostic accuracy beyond traditional visual inspection. Additionally, radiomics-based AI models utilizing computed tomography (CT) and elastography have enabled non-invasive risk assessment, improving predictions of bleeding risk and estimations of the hepatic venous pressure gradient (HVPG). Ethical considerations, such as data privacy and algorithmic bias, also require careful management. Future research should focus on prospective validation, real-world application studies, and the development of standardized AI frameworks to ensure the clinical applicability of these methods. AI-driven precision medicine has



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.



the potential to revolutionize the management of GEVs, offering more efficient, accurate, and individualized patient care while optimizing healthcare resource utilization.

Keywords: Artificial intelligence, esophageal varices, gastric varices, gastroesophageal varices, machine learning, deep learning, radiomics

INTRODUCTION

Portal hypertension (PH) is a severe complication of liver cirrhosis and a primary contributor to morbidity and mortality. PH arises due to increased portal inflow combined with high intrahepatic vascular resistances, leading to the development of upstream venous collaterals at anastomotic sites between the portal and systemic circulations^[1-3]. Hepatic venous pressure gradient (HVPG) is one of the most reliable prognostic factors and surrogate markers of PH, typically defined as a gradient > 5 mmHg^[2].

Gastroesophageal varices (GEVs) are one of the first clinically relevant consequences of PH, developing in 60%-80% of patients with liver cirrhosis over their lifetime when the HVPG exceeds 10 mmHg^[4,5].

GEVs are a key determinant of outcomes in cirrhotic patients, as they are directly associated with a higher risk of decompensation and death^[6]. Additionally, the presence and total surface area of portosystemic shunts have been identified as independent risk factors for mortality in cirrhotic patients and for post-hepatectomy liver failure in individuals undergoing liver resection for hepatocellular carcinoma (HCC)^[7,8].

Beyond their prognostic significance, GEVs carry substantial morbidity and mortality risks, with acute bleeding being the most feared complication. This can result in mortality rates of 15%-55%, with a recurrence likelihood of 33%-76% in the subsequent year^[9]. To enable early diagnosis and timely intervention, screening endoscopy is generally recommended for patients with cirrhosis, as GEVs can be present in up to 25% of those with compensated cirrhosis. Even patients with negative initial endoscopic screenings can develop GEVs at a rate of 5%-10% per year, with an annual GEV bleeding risk of 12%^[10-12]. The five-year mortality rate following the first bleeding episode is approximately 20% in patients with GEVs alone but exceeds 80% when accompanied by other signs of decompensation, such as bacterial infections or kidney failure^[13,14].

This risk must be balanced against the potential for unnecessary invasive procedures in low-risk patients, such as those with Child-Pugh A cirrhosis, who develop GEVs in fewer than half of cases. Recent guidelines aim to address this issue by optimizing screening strategies to avoid costly and potentially harmful procedures in patients with a low likelihood of GEVs^[10,11,15].

Thanks to groundbreaking advances in computational and information processing technologies, interest in and development of artificial intelligence (AI) have grown exponentially^[16].

AI encompasses a wide range of techniques, including artificial neural networks (ANN), machine learning (ML), deep learning (DL), and natural language processing algorithms. These algorithms can perform tasks that typically require human intelligence and are particularly suited to rapidly analyzing, predicting, summarizing, and identifying patterns in massive datasets^[17]. AI is increasingly being integrated into various medical fields, transforming healthcare by enhancing safety, accuracy, efficiency, and patient outcomes. Once optimized, AI algorithms can assist in numerous diagnostic and therapeutic applications, including early diagnosis, laboratory workflow optimization, precise drug dosing, prediction of treatment response, and personalized medicine^[18]. It is, therefore, unsurprising that over 755 AI-based medical devices have

been approved in the United States and Europe and are being gradually incorporated into routine patient care^[19].

Currently, GEV diagnosis and management rely heavily on resource-intensive and invasive procedures such as endoscopy^[20]. Endoscopic image evaluation is based on visual inspection, which is inherently subjective and not fully quantifiable, potentially leading to misclassification and inappropriate treatment decisions^[21]. AI applications could enhance diagnostic accuracy while reducing time, costs, and invasiveness, though their clinical value has yet to be fully established and validated.

In this review, we examine and summarize the current evidence on AI applications in all aspects of GEV management, exploring its potential benefits, limitations, and challenges. This paper aims to contribute to a better understanding of AI's role in GEV prediction, diagnosis, risk stratification, and outcome assessment.

METHODS

Search strategy and inclusion criteria

MEDLINE, Embase, and Web of Science were searched using the following terms: “artificial intelligence*” OR “AI” AND “varices*”. The last search was conducted on January 30, 2025, with no restrictions on language or publication status. Additional potentially relevant studies were identified from the reference list of selected studies. All studies reporting original data regarding all the possible roles of AI in GEV detection, characterization, and outcome prognostication were included.

VARICES DETECTION

ML models

Abd El-Salam *et al.* investigated the performance of six different ML algorithms as non-invasive predictors of GEVs in a population of 4,962 patients (80% training and 20% testing set) affected by hepatitis C virus (HCV)-related cirrhosis who underwent upper gastrointestinal endoscopy (UGIE)^[22]. The best results were achieved using the Bayesian network algorithm, which incorporated gender, platelet count, albumin, bilirubin, C-reactive protein (CRP), liver and spleen stiffness, and prothrombin concentration as independent predictors. This model achieved an accuracy of 68.9% and a 74.8% area under the curve (AUC).

Bayani *et al.* tested four different ML models - random forest (RF), ANN, support vector machine (SVM), and logistic regression - in a population of 490 cirrhotic patients^[23]. They evaluated the models based on accuracy, recall, precision, F1 score, and receiver operating characteristic (ROC) in predicting the presence and grade of GEVs. All models showed excellent results, but the RF model outperformed the others with a 99% average accuracy, recall values of 0.95-1, precision of 0.86-1, and F1 scores of 0.92-1. The same group published another analysis on the same population^[24] utilizing ensemble algorithms (CatBoost and XGBoost), which combine the predictions of several learning methods to improve performance by selecting the most impactful factors, clinical and laboratory values in this case, to predict GEV grades. CatBoost and XGBoost are high-performance, open-source gradient boosting ML algorithms used to make accurate predictions. They work by combining many simple decision trees, where each new tree helps fix the mistakes of the previous ones. CatBoost is generally preferred when dealing with categorical data, since it has a better performance thanks to its ability to automatically understand and work with these categories. The CatBoost model produced the best results, predicting all grades with 100% accuracy.

Şimşek *et al.* constructed a gradient-boosted ML algorithm to predict the presence of GEVs by analyzing patient demographics as well as clinical, endoscopic, laboratory, and radiological data in patients with

documented cirrhosis^[25]. The variables selected by the model included gender, ascites, encephalopathy, Child-Pugh score, and platelet count, resulting in a mean AUC of 0.68 for predicting GEVs.

Dong *et al.* developed a RF model to identify patients with GEVs and those needing treatment (VNT)^[26]. They created a score (EEndo score) based on the international normalized ratio, aspartate aminotransferase levels, platelet count, blood urea nitrogen, hemoglobin, and presence of ascites in a multicenter cohort of cirrhotic patients undergoing endoscopic screening. The EEndo score demonstrated an AUROC of 0.81-0.84, with the potential to safely reduce approximately one-third of all screening endoscopies while missing only 1.1% of GEVs that required treatment.

In 2020, Huang *et al.* conducted a multicenter trial on patients with compensated advanced chronic liver disease who underwent both endoscopy and non-contrast-enhanced computed tomography (CT) scans within the previous two weeks^[27]. From the CT images, they identified two radiomics signatures (rGEV and rHRV) capable of detecting the presence of GEVs (rGEV) with an AUROC of 0.871-0.941 and identifying high-risk varices (rHRV) with an AUROC of 0.831-0.836.

Three years later, the same group^[28] conducted another international multicenter cohort trial involving 2,794 compensated cirrhosis patients from 17 centers, including 1,283 patients from a single university hospital, constituting a “real-world cohort”. They developed an ML model that incorporated clinical and laboratory data into a light gradient-boosting machine algorithm, which utilized decision trees to calculate the value of each variable and identify rHRV, thereby reducing unnecessary endoscopic screenings. The ML model’s performance was compared with the Baveno VI criteria^[10] and demonstrated the ability to further reduce endoscopies across all cohorts, sparing an additional 13.9%-25.9% of patients from screening (all $P < 0.001$).

Noureddin *et al.* presented their ML algorithm at the 2021 AASLD Liver Meeting, which predicted HVP and GEVs using 457 parameters derived from liver histology, including quantitative assessments of septa, nodules, and fibrosis, in patients with compensated cirrhosis and HVP ≥ 6 mmHg^[29]. Their scores predicted clinically significant PH (SNOF score) with an AUC of 0.85 and the presence of GEVs (SNOF-V score) with an AUC of 0.86.

DL models

Chen *et al.* developed a deep convolutional neural network (DCNN) trained and tested on 14,718 images from a cohort of 3,021 patients with GEVs and 3,168 patients with a normal upper gastrointestinal tract to predict GEV presence and risk of rupture^[30]. Compared to endoscopy, the AI model showed significantly higher detection accuracy (92%-97% vs. 84%-94%), improved identification of red color signs and red spots, and comparable performance in assessing size, shape, color, and bleeding signs. The average time to analyze one image was 0.13 s for the DCNN model, compared to 18.75 s for endoscopists.

Procopet *et al.* analyzed data from a cohort of 202 patients with long-standing compensated chronic liver disease and a clinical suspicion of cirrhosis who underwent transjugular liver biopsy, HVP measurement, and liver stiffness measurement^[31]. They built an ANN to diagnose cirrhosis, clinically significant PH, and GEVs. Liver stiffness measurement proved to be the best non-invasive test for diagnosing cirrhosis, clinically significant PH, and GEVs, with C-statistics of 0.93, 0.94, and 0.90, respectively. The ANN demonstrated high diagnostic performance (accuracy > 80%) but was not statistically superior to liver stiffness measurement alone.

Yu *et al.* conducted a multicenter trial enrolling cirrhotic patients undergoing transjugular HVPG measurement and contrast-enhanced abdominal CT^[32]. They developed a DL network for 3D liver and spleen segmentation, as well as an automated ML-based CT radiomics HVPG quantitative model for HVPG estimation and multistage assessment. Although the study was limited by selection bias and a relatively small sample size (224 patients in the training dataset and 148 in the internal testing dataset), the model demonstrated excellent performance in liver and spleen segmentation and outperformed previously published tools^[33,34] in assessing HVPG stages (Spearman's rho = 0.616).

The results of AI techniques for the non-invasive detection of GEVs are promising, although some inconsistencies across studies have been highlighted, with variable AUC results. These discrepancies likely stem from heterogeneity in study design, patient populations, input variables, and algorithm selection. Some studies used relatively small cohorts (< 500 patients) or relied heavily on retrospective single-center data, which limits generalizability. Ensemble algorithms, such as CatBoost and XGBoost, reported perfect accuracy in predicting varices grades; however, they still lack external replication and are at risk of overfitting due to limited transparency in data preprocessing and feature selection.

PREDICTION OF VARICEAL BLEEDING

Variceal bleeding is a serious complication of PH and is associated with a significant risk of mortality. Reliable predictors that can assess and quantify the risk of bleeding could substantially improve patient prognosis. The currently available and most widely used scores are generally based on a combination of clinical and endoscopic variables^[35,36]. Hou *et al.* developed an ANN early-stage warning model to predict GEV bleeding in patients with liver cirrhosis^[37]. A total of 1,100 patients and 12 clinical and laboratory variables were included in the analysis to build an ANN model capable of better predicting one-year GEV bleeding, with an AUROC of 0.945 and a C-index of 0.936.

To identify a GEV bleeding risk stratification tool superior to traditional methods, Agarwal *et al.* developed and validated a supervised ML algorithm (XGBoost) using a cohort of 828 patients with GEVs and compensated advanced chronic liver disease^[38]. The group of patients classified as endoscopic low-risk + ML low-risk had a one-year bleeding rate ranging from 0.6% (derivation cohort) to 1.6% (external validation cohort), while the group classified as endoscopic high-risk + ML high-risk had a one-year bleeding rate ranging from 31.6% (external validation cohort) to 36.9% (derivation cohort). The accuracy of the ML algorithm ranged from 0.857 (external validation cohort) to 0.987 (derivation cohort), significantly higher than that of endoscopy (0.589).

An intriguing approach was explored by Wang *et al.*, in which endoscopic images from cirrhotic patients across different hospitals were analyzed in real time by a DCNN system (ENDOANGEL-GEV)^[39]. This system comprised six different models to segment and grade GEVs and classify red-colored signs. The model achieved an accuracy of 93%, with a sensitivity for detecting GEVs comparable to that of endoscopists (100% vs. 99.2%, $P = 1.000$). However, it showed significantly better performance in detecting and classifying grades of red color signs.

Wang *et al.* developed, trained, tested, and validated a series of automated multimodal ML models for predicting one-year GEV bleeding risk by integrating endoscopic images and clinical data from cirrhotic patients^[40]. The stacking model (consisting of six base models) provided the best AUC value (0.975), as well as sensitivity (0.952), accuracy (0.932), and F1 score (0.879).

Zhong *et al.* focused on predicting the one-year GEV rebleeding probability in patients with cirrhosis and GEV bleeding who underwent early transjugular intrahepatic portosystemic shunt procedures^[41]. The developed ANN showed that ALBI, PALBI, Child-Pugh, and MELD-based nomograms had similar prognostic performance (C-index ranging from 0.798 to 0.879).

Radiomics

Endoscopy is the most widely used method for detecting, characterizing, and grading GEVs. However, quantitative data can also be derived from radiological imaging and analyzed using AI algorithms (radiomics). In patients with endoscopically confirmed GEVs and cirrhosis, data from the portal phase of contrast-enhanced CT scans can be extracted (over 1,200 radiomic features from each region of interest per patient) and combined with clinical variables to predict the risk of GEV bleeding, achieving an AUC of 0.78 (95%CI: 0.68-0.87), as described by Liu *et al.*^[42]. More recently, Luo *et al.* developed a similar model to establish a radiomics signature (RadScore) based on five liver and three spleen CT features combined with clinical variables (albumin, fibrinogen, portal vein thrombosis, aspartate aminotransferase, and spleen thickness), achieving an AUC of 0.912^[43]. Slightly lower AUC values were obtained by Yang *et al.* with their integrated radiomics and clinical model in hepatitis B virus (HBV)-related cirrhotic patients (accuracy of 0.73)^[44].

Yan *et al.* achieved excellent results with their ML-based radiomics model for diagnosing high-bleeding-risk GEVs in cirrhotic patients, with an AUC of 0.736 in the external validation cohort^[45]. This model outperformed both the Baveno VI and expanded Baveno VI criteria, showing 49.0% and 32.8% higher efficiencies, respectively^[46].

Zhang *et al.* built a model to predict the risk of GEV bleeding by focusing on the spleen^[47]. They extracted 1,647 radiomic features, analyzed 12 of them, and obtained an AUC of 0.924 in the test set when combined with clinical high-risk bleeding factors.

A non-invasive method to predict GEVs requiring treatment in cirrhotic patients who underwent both abdominal enhanced CT and endoscopy was described by Lin *et al.*^[48]. The identified regions of interest for radiomics analysis included the liver, spleen, and lower esophagus to the gastric fundus. The resulting nomogram achieved an AUC of 0.947-0.973 in the validation sets.

ANN models have been shown to achieve high predictive accuracy in stratifying the risk of variceal bleeding, surpassing traditional clinical assessments. However, the reliability of these findings varies depending on model complexity, dataset size, and the rigor of validation. Hybrid strategies with multimodal approaches integrating imaging and clinical data proved to be valuable tools, providing high accuracy. Despite these advances, many studies still rely on limited patient numbers and lack validation of long-term outcomes.

OUTCOME

GEV bleeding is a significant determinant of poor outcomes in cirrhotic patients. Tseng *et al.* constructed a radiomics model based on liver, spleen, and combined features to perform a non-invasive calculation of portal vein pressure and predict patient outcome^[49]. The combined model achieved an AUROC of 0.855, an accuracy of 0.823, a specificity of 0.929, and a sensitivity of 0.735. Simsek *et al.* built an ML model capable of outperforming the MELD-Na and Child-Turcotte-Pugh scores in predicting overall survival^[50]. The model's AUC was 0.87, 0.85, and 0.76 for one-, three-, and twelve-month survival, respectively, and 0.91, 0.88, and 0.91 in patients with GEV bleeding at the same time points.

The results of AI for outcome prediction in patients with GEVs achieved high accuracy, suggesting a strong discriminative ability. These findings should still be interpreted cautiously until validated in diverse clinical settings with prospective cohorts and ideally compared against standard prognostic tools. The current results are often based on highly specialized imaging and radiomic analysis, which may not be readily available or standardized across institutions.

DISCUSSION

The application of AI in the management of GEVs has demonstrated significant promise in various aspects, including prediction, diagnosis, risk stratification, and outcome assessment. The studies reviewed illustrate the growing role of AI-based algorithms, particularly ML and DL models, in improving accuracy, efficiency, and non-invasiveness in diagnosing and managing GEVs.

One of the most striking advantages of AI in GEV prediction is its ability to integrate diverse clinical, laboratory, and imaging data to generate accurate, non-invasive models. ML models, such as RF and ensemble-based approaches, have shown outstanding predictive performance, in some cases achieving accuracy rates close to 100%. These models can identify patients at risk of developing GEVs and stratify them according to severity, thereby optimizing screening strategies and reducing the need for invasive endoscopic evaluations [Table 1]. Similarly, DL models, particularly convolutional neural networks (CNNs), have demonstrated remarkable accuracy in detecting GEVs and their features from endoscopic images, surpassing traditional visual inspection techniques.

The integration of AI with imaging techniques has also shown substantial potential. Radiomics-based approaches, which extract quantitative features from imaging modalities such as CT and elastography, have provided robust tools for non-invasive risk assessment. These models can effectively predict the presence of GEVs and their bleeding risk with high accuracy, outperforming conventional clinical criteria such as the Baveno VI guidelines [Table 2]. Moreover, AI-enhanced CT radiomics and liver stiffness-based assessments have facilitated more precise estimations of the HVPg, enabling better clinical decision making.

The risk stratification of GEV bleeding is another area where AI has demonstrated significant advancements. Traditional scoring systems, such as the Child-Pugh and MELD scores, have long been used for prognostic assessments; however, they often lack individualized precision. AI-driven models, particularly those employing ANN and supervised ML algorithms (like XGBoost), have exhibited superior predictive power, refining risk stratification and potentially guiding personalized treatment strategies. These approaches may enable timely prophylactic interventions, thereby reducing morbidity and mortality associated with GEV hemorrhage [Table 3].

Limitations and current challenges

Despite these promising findings, several challenges remain in the clinical integration of AI for GEV management. One major limitation is the variability in data quality and availability across different studies and healthcare settings. AI models require large, well-annotated datasets to ensure robustness and generalizability, which may not always be feasible in real-world clinical practice. Data bias is one of the most pervasive issues in AI healthcare applications. Models trained on datasets that lack demographic, geographic, or clinical diversity may yield biased predictions that disadvantage underrepresented populations.

Additionally, issues related to model transparency need to be addressed. The extent to which AI models and their development pipelines are accessible and interpretable to external reviewers is often limited and can

Table 1. Summary of AI models focusing on the prediction and detection of GEVs

Author	Country	AI type	Population	Number of patients	Patients in the training cohort	Patients in the validation cohort(s)	AUC training cohort	AUC validation cohort	Accuracy
Abd El-Salam <i>et al.</i> 2019 ^[22]	Egypt Saudi Arabia	SVM; RF; DT (C4.5); ANN (MLP); Naïve Bayes; Bayesian Net	HCV-related cirrhosis	4,962	80% (n = 3,970)	20% (n = 992)	NA	SVM: 67.1%; RF: 71.1%; C4.5: 69.0%; MLP: 71.6%; Naïve Bayes: 73.2%; Bayesian Net: 74.8%	SVM: 67.8%; RF: 66.3%; C4.5: 67.2%; MLP: 65.6%; Naïve Bayes: 66.7%; Bayesian Net: 68.9%
Bayani <i>et al.</i> 2022 ^[23]	Iran	RF; ANN; SVM; LR	Cirrhotic patients	490	NA	NA	NA	RF: 99.0%; ANN: 98.0%; SVM: 98.0%	RF: 99.0%
Bayani <i>et al.</i> 2022 ^[24]	Iran	ML (CatBoost; XGBoost)	Cirrhotic patients	490	NA	NA	NA	NA	CatBoost: 100%; XGBoost: 92.0%
Dong <i>et al.</i> 2019 ^[26]	USA	ML, RF model (Evendo score)	Cirrhotic patients	347	238 (retrospective cohort)	109 (prospective cohort), then 347 (internal validation by bootstrapping)	84.0% (presence of GEV); 75.0% (GEV needing treatment)	83.0% (presence of GEV in bootstrapped sample); 77.0% (GEV needing treatment)	NA
Huang <i>et al.</i> 2020 ^[27]	China	ML, radiomic signatures (rGEV for presence of GEV; rHRV for high-risk GEV)	Compensated advanced chronic liver disease	161	129 (80%)	32 (20%)	rGEV: 94.1%; rHRV: 83.6%	rGEV: 87.1%; rHRV: 83.1%	rGEV: 84.4%; rHRV: 84.4%
Huang <i>et al.</i> 2023 ^[28]	China Singapore India USA	ML (ML-EGD)	Compensated cirrhosis	2,794 (of which 1,283 “real world”, single Center cohort)	1,154 (90%)	129 (10%) + 966 (multicenter test cohort n.1) & 545 (international test cohort n.2)	73.7%	85.8%; 77.8% (test cohort n.1); 80.1% (test cohort n.2)	NA
Noureddin <i>et al.</i> 2021 ^[29]	USA	ML (SNOF for CS-PH; SNOF-V for presence of GEV)	NASH patients with compensated cirrhosis and HVP ≥ 6 mmHg	143	NA	NA	SNOF: 85.0%; SNOF-V: 86.0%	SNOF: 64.0%; SNOF-V: 72.0%	NA
Chen <i>et al.</i> 2021 ^[30]	China	DCNN (ENDOANGEL)	Patients diagnosed with GEV and normal esophagus and stomach	3,021	NA, 80% (different number of patients per significant variable)	NA (10% validation, 10% testing)	NA	97.4% (EV identification); 94.7% (size); 93.1% (red spots); 92.5% (red signs)	99.4% (identification of EV); 92.5% (size); 92.2% (bleeding signs); 92.0% (GV identification)
Procopet <i>et al.</i> 2015 ^[31]	Romania France Switzerland	ANN (including and excluding liver stiffness, LS)	Compensated long-lasting chronic liver diseases and a clinical suspect of cirrhosis	202	158 (78%)	44 (22%)	90.0% (LS and HVPg)	NA	ANN + LS: 81.8%; ANN-LS: 77.3%
Wang <i>et al.</i> 2023 ^[40]	China	Multi-modality DLRP (DLRP-SM)	Compensated advanced chronic liver disease	265 (1,136 liver stiffness + 1,042 spleen stiffness)	191 (808 liver images) + 187 (742 spleen images)	74 (328 liver images) + 72 (300 spleen images)	97.0% (GEV identification); 97.0% (high-risk GEV)	91.0% (GEV identification); 88.0% (high-risk GEV)	NA
Yu <i>et al.</i> 2022 ^[32]	China UK Turkey	DL network, radiomics (aHVPg)	Cirrhotic patients with HVPg measurements	372	224 (60%)	148 (40%)	90.0% (HPVG > 12 mmHg)	77.0% (HPVG > 12 mmHg)	84%

AI: Artificial intelligence; GEVs: gastroesophageal varices; AUC: area under the curve; SVM: support vector machine; RF: random forest; DT: decision tree; ANN: artificial neural network; MLP: multi-layer perceptron; HCV: hepatitis C virus; NA: not available; LR: logistic regression; ML: machine learning; CS-PH: clinically significant portal hypertension; HVP: hepatic venous pressure gradient; DCNN: deep convolutional neural network; GV: gastric varice; LS: liver stiffness; DLRP: deep learning risk prediction model; DL: deep learning.

represent a critical issue. Many algorithms operate as “black boxes”, with limited disclosure of their training data sources, feature selection criteria, or validation procedures^[51]. This opacity poses a challenge for clinicians and regulators seeking to evaluate the reliability, reproducibility, and clinical applicability of AI tools. Most high-performance models, especially deep neural networks, offer limited interpretability, making it challenging for clinicians to understand how conclusions are derived from input features^[52]. This lack of explainability raises concerns related to informed consent, accountability, and shared decision making.

Another important consideration is the need for prospective validation and regulatory approval of AI-driven tools before their routine clinical implementation. While many of the reviewed models have demonstrated excellent performance in retrospective and validation cohorts, prospective studies are required to confirm their effectiveness in diverse patient populations. Furthermore, ethical concerns, including data privacy and algorithmic bias, must be carefully managed to ensure equitable AI applications in hepatology.

Addressing these challenges requires a concerted effort across research, policy, and clinical implementation. Strategies include the use of diverse and representative datasets, transparent reporting standards such as TRIPOD-AI and CONSORT-AI^[53], and the development of inherently interpretable models or post-hoc explainability techniques.

CONCLUSION

AI has emerged as a transformative tool in the prediction, diagnosis, risk stratification, and outcome assessment of GEVs, offering innovative solutions to long-standing clinical challenges. The integration of ML and DL algorithms with non-invasive diagnostic modalities holds significant potential for reducing reliance on resource-intensive endoscopic procedures, improving early detection, and refining treatment strategies for patients with liver cirrhosis.

While the reviewed studies highlight the impressive diagnostic accuracy and predictive power of AI models, further research is required to overcome current limitations and facilitate clinical translation. Future efforts should focus on prospective multicentric validation, real-world implementation studies, and the development of standardized AI frameworks. Addressing these challenges will pave the way for AI-driven precision medicine, ultimately enhancing patient outcomes and optimizing healthcare resource utilization in the management of PH and its complications.

Table 2. Summary of AI models focusing on gastroesophageal variceal bleeding prediction

Author	Country	AI type	Population	Number of patients	Patients in the training cohort	Patients in the validation cohort(s)	AUC training cohort	AUC validation cohort	Accuracy
Agarwal <i>et al.</i> 2021 ^[38]	India	ML (XGBoost)	Compensated advanced chronic liver disease	828	497	149 (internal validation); 182 (external validation)	NA	NA	98.7% (derivation cohort); 93.7% (internal validation cohort); 85.7% (external validation cohort)
Wang <i>et al.</i> 2023 ^[40]	China	CNN; multimodal MLs (GBM, GLM, XGBoost, RF, DL, Stacking)	Cirrhotic patients	341	275	66 (validation); 161 (test)	GBM: 99.8%; GLM: 87.6%; XGBoost: 97.0%; RF: 100%; DL: 100%; Stacking: 100%	GBM: 99.0% (validation), 96.8% (test); GLM: 90.3% (validation), 84.9% (test); XGBoost: 99.4% (validation), 93.5% (test); RF: 98.3% (validation), 97.6% (test); DL: 98.4% (validation, 93.7% (test); Stacking: 99.8% (validation), 97.5% (test)	GBM: 95.2% (validation), 88.8% (test); GLM: 87.9% (validation), 83.2% (test); XGBoost: 97.0% (validation), 87.6% (test); RF: 97.0% (validation), 91.9% (test); DL: 89.4% (validation, 88.8% (test); Stacking: 97.0% (validation), 93.2% (test)
Liu <i>et al.</i> 2022 ^[42]	China	Radiomics (CT + clinical, RadScore, Nomogram)	Endoscopic-proven GEVs and cirrhosis	317	222 (70%)	95 (30%)	CT + clinical: 87.0%; RadScore: 70.0%; Nomogram: 89.0%	CT + clinical: 76.0%; RadScore: 66.0%; Nomogram: 78.0%	CT + clinical: 78.0%; RadScore: 61.0%; Nomogram: 68.0%
Luo <i>et al.</i> 2023 ^[43]	China	Radiomics (RadScore; Clinical + radiomics)	Cirrhotic patients	211	149 (70%)	62 (30%)	RadScore: 81.7%; Clinical + radiomics: 92.5%	RadScore: 74.1%; Clinical + radiomics: 91.2%	NA
Yang <i>et al.</i> 2019 ^[44]	China	Radiomics (clinical, radiomics and radiomics + clinical model)	HBV-related cirrhosis	295	236	59	Clinical: 64.0%; Radiomics: 82.0%; Radiomics + clinical: 83.0%	Clinical: 61.0%; Radiomics: 61.0%; Radiomics + clinical: 64.0%	Clinical: 54.0%; Radiomics: 66.0%; Radiomics + clinical: 73.0%
Yan <i>et al.</i> 2022 ^[45]	China USA	ML-based radiomics model	Cirrhotic patients	796	391	405 (external validation)	Mild EV: 94.3%; HREV: 98.3%	Mild EV: 73.2% (internal validation), 65.4% (external validation); HREV: 83.4% (internal validation), 73.6% (external validation)	Mild EV: 70.5% (internal validation), 64.1% (external validation); HREV: 94.7% (internal validation), 74.3% (external validation)
Zhang <i>et al.</i> 2022 ^[47]	China	Radiomics (clinical, radiomics and radiomics + clinical model)	Cirrhotic patients with PH	100	77	33	Radiomics + clinical: 94.7%; Radiomics: 94.1%; Clinical: 50.6%	Radiomics + clinical: 92.4%; Radiomics: 80.2%; Clinical: 73.0%	NA
Hou <i>et al.</i> 2023 ^[37]	China	ANN	Cirrhotic patients	1,100	999	101	95.9%	94.5%	93.6%
Wang <i>et al.</i> 2022 ^[39]	China	DCNN (ENDOANGEL-GEV)	Cirrhotic patients	NA	6,034 endoscopic images	11,009 endoscopic images (Dataset 2) + 161 prospective patients (Dataset 3)	NA	NA	Model 3: 93.4% (red signs) Model 4: 94.84% (EV grade I), 93.67% (EV grade II); 93.88% (EV grade III); 94.62% (red signs in prospective cohort); 94.92% (red signs in GV in prospective cohort)

Zhong <i>et al.</i> 2021 ^[41]	China	ANN (4 different nomograms)	Patients with cirrhosis and variceal bleeding undergoing early TIPS procedures	259	NA	NA	NA	NA	Internal validation Nomogram 1: 87.9%; Nomogram 2: 82.9%; Nomogram 3: 87.4%; Nomogram 4: 79.8% External validation Nomogram 1: 72.0%; Nomogram 2: 71.9%; Nomogram 3: 71.8%; Nomogram 4: 70.3%
--	-------	-----------------------------	--	-----	----	----	----	----	--

AI: Artificial intelligence; AUC: area under the curve; ML: machine learning; NA: not available; CNN: convolutional neural network; GBM: gradient boosting machine; GLM: generalized linear model; RF: random forest; DL: deep learning; CT: computed tomography; GEVs: gastroesophageal varices; HBV: hepatitis B virus; EV: esophageal varice; HREV: high-risk esophageal varice; PH: portal hypertension; ANN: artificial neural network; DCNN: deep convolutional neural network; GV: gastric varice; TIPS: transjugular intrahepatic portosystemic shunt.

Table 3. Summary of AI models focusing on GEVs' outcomes

Author	Country	AI type	Population	Number of patients	Patients in the training cohort	Patients in the validation cohort(s)	AUC training cohort	AUC validation cohort	Accuracy
Tseng <i>et al.</i> 2020 ^[49]	China	Radiomics (rPVP Liver, rPVP Spleen, rPVP Liver and Spleen)	Patients with PH admitted for a TIPS procedure	169	107	62	NA	rPVP Liver: 78.9%; rPVP Spleen: 83.2%; rPVP Liver and Spleen: 85.5%	rPVP Liver: 75.8%; rPVP Spleen: 79.0%; rPVP Liver and Spleen: 82.3%
Simsek <i>et al.</i> 2021 ^[50]	Turkey	ML	Cirrhotic patients	124	80% (50 random iterations)	20% (50 random iterations)	NA	87.0% (1-month survival); 85.0% (3-month survival); 76.0% (1-year survival); With VB: 91.0% (1-month survival); 88.0% (3-month survival); 91.0% (1-year survival)	NA

AI: Artificial intelligence; GEVs: gastroesophageal varices; AUC: area under the curve; PH: portal hypertension; TIPS: transjugular intrahepatic portosystemic shunt; NA: not available; ML: machine learning; VB: variceal bleeding.

DECLARATIONS

Authors' contributions

Made substantial contributions to the conception and design of the study and performed data analysis and interpretation: Rompianesi G, Pegoraro F, Montalti R, Troisi R

Performed data acquisition, as well as providing administrative, technical, and material support: Pacilio B, Petti G, Benassai G, Cappuccio M

Availability of data and materials

Not applicable.

Financial support and sponsorship

None.

Conflicts of interest

Rompianesi G is a Junior Editorial Board member of the journal *Artificial Intelligence Surgery*. Rompianesi G was not involved in any steps of editorial processing, notably including reviewer selection, manuscript handling, or decision making. The other authors declared that there are no conflicts of interest.

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Copyright

© The Author(s) 2025.

REFERENCES

1. Garcia-Tsao G, Bosch J. Management of varices and variceal hemorrhage in cirrhosis. *N Engl J Med.* 2010;362:823-32. DOI PubMed
2. Bosch J, Abraldes JG, Fernández M, García-Pagán JC. Hepatic endothelial dysfunction and abnormal angiogenesis: new targets in the treatment of portal hypertension. *J Hepatol.* 2010;53:558-67. DOI PubMed
3. Iwakiri Y. Pathophysiology of portal hypertension. *Clin Liver Dis.* 2014;18:281-91. DOI PubMed PMC
4. Garcia-Tsao G, Abraldes JG, Berzigotti A, Bosch J. 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;65:310-35. DOI PubMed
5. Groszmann RJ, Garcia-Tsao G, Bosch J, et al; Portal Hypertension Collaborative Group. Beta-blockers to prevent gastroesophageal varices in patients with cirrhosis. *N Engl J Med.* 2005;353:2254-61. DOI PubMed
6. D'Amico G, Garcia-Tsao G, Pagliaro L. Natural history and prognostic indicators of survival in cirrhosis: a systematic review of 118 studies. *J Hepatol.* 2006;44:217-31. DOI PubMed
7. Praktijnjo M, Simón-Talero M, Römer J, et al; Baveno VI-SPSS group of the Baveno Cooperation. Total area of spontaneous portosystemic shunts independently predicts hepatic encephalopathy and mortality in liver cirrhosis. *J Hepatol.* 2020;72:1140-50. DOI PubMed
8. Rompianesi G, Han HS, Fusai G, et al. Pre-operative evaluation of spontaneous portosystemic shunts as a predictor of post-hepatectomy liver failure in patients undergoing liver resection for hepatocellular carcinoma. *Eur J Surg Oncol.* 2025;51:108778. DOI PubMed
9. Baiges A, Hernández-Gea V, Bosch J. Pharmacologic prevention of variceal bleeding and rebleeding. *Hepatol Int.* 2018;12:68-80. DOI PubMed
10. de Franchis R, Bosch J, Garcia-Tsao G, Reiberger T, Ripoll C; Baveno VII Faculty. Baveno VII - Renewing consensus in portal hypertension. *J Hepatol.* 2022;76:959-74. DOI PubMed PMC
11. Kaplan DE, Ripoll C, Thiele M, et al. AASLD Practice Guidance on risk stratification and management of portal hypertension and varices in cirrhosis. *Hepatology.* 2024;79:1180-211. DOI PubMed
12. Seo YS. Prevention and management of gastroesophageal varices. *Clin Mol Hepatol.* 2018;24:20-42. DOI PubMed PMC
13. Augustin S, Muntaner L, Altamirano JT, et al. Predicting early mortality after acute variceal hemorrhage based on classification and regression tree analysis. *Clin Gastroenterol Hepatol.* 2009;7:1347-54. DOI PubMed
14. D'Amico G, Pasta L, Morabito A, et al. Competing risks and prognostic stages of cirrhosis: a 25-year inception cohort study of 494 patients. *Aliment Pharmacol Ther.* 2014;39:1180-93. DOI PubMed
15. Mattos ÂZ, Schacher FC, John Neto G, Mattos AA. Screening for esophageal varices in cirrhotic patients - Non-invasive methods. *Ann Hepatol.* 2019;18:673-8. DOI PubMed
16. Alowais SA, Alghamdi SS, Alsuhebany N, et al. Revolutionizing healthcare: the role of artificial intelligence in clinical practice. *BMC Med Educ.* 2023;23:689. DOI PubMed PMC
17. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. *Future Healthc J.* 2019;6:94-8. DOI PubMed PMC
18. Rompianesi G, Pegoraro F, Ceresa CD, Montalti R, Troisi RI. Artificial intelligence in the diagnosis and management of colorectal cancer liver metastases. *World J Gastroenterol.* 2022;28:108-22. DOI PubMed PMC
19. Muehlematter UJ, Bluethgen C, Vokinger KN. FDA-cleared artificial intelligence and machine learning-based medical devices and their 510(k) predicate networks. *Lancet Digit Health.* 2023;5:e618-26. DOI PubMed

20. European Association for the Study of the Liver. EASL Clinical Practice Guidelines for the management of patients with decompensated cirrhosis. *J Hepatol.* 2018;69:406-60. DOI PubMed
21. 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 PubMed
22. Abd El-Salam SM, Ezz MM, Hashem S, et al. Performance of machine learning approaches on prediction of esophageal varices for Egyptian chronic hepatitis C patients. *Inf Med Unlocked.* 2019;17:100267. DOI
23. Bayani A, Asadi F, Hosseini A, et al. Performance of machine learning techniques on prediction of esophageal varices grades among patients with cirrhosis. *Clin Chem Lab Med.* 2022;60:1955-62. DOI PubMed
24. Bayani A, Hosseini A, Asadi F, et al. Identifying predictors of varices grading in patients with cirrhosis using ensemble learning. *Clin Chem Lab Med.* 2022;60:1938-45. DOI PubMed
25. Şimşek C, Tekin E, Sahin H, Sahin TK, Balaban YH. Artificial intelligence to predict esophageal varices in patients with cirrhosis. *Acibadem Univ Saglik Bilim Derg.* 2021;12:625-9. Available from: https://www.researchgate.net/publication/352791299_Artificial_Intelligence_to_Predict_Esophageal_Varices_in_Patients_with_Cirrhosis. [Last accessed on 18 Aug 2025]
26. Dong TS, Kalani A, Aby ES, et al. Machine learning-based development and validation of a scoring system for screening high-risk esophageal varices. *Clin Gastroenterol Hepatol.* 2019;17:1894-901.e1. DOI PubMed
27. Huang Y, Huang F, Yang L, et al. Development and validation of a radiomics signature as a non-invasive complementary predictor of gastroesophageal varices and high-risk varices in compensated advanced chronic liver disease: a multicenter study. *J Gastroenterol Hepatol.* 2021;36:1562-70. DOI PubMed
28. Huang Y, Li J, Zheng T, et al. Development and validation of a machine learning-based model for varices screening in compensated cirrhosis (CHESS2001): an international multicenter study. *Gastrointest Endosc.* 2023;97:435-44.e2. DOI PubMed
29. Nouredin M, Goodman Z, Tai D, et al. Machine learning liver histology scores correlate with portal hypertension assessments in nonalcoholic steatohepatitis cirrhosis. *Aliment Pharmacol Ther.* 2023;57:409-17. DOI PubMed PMC
30. Chen M, Wang J, Xiao Y, et al. Automated and real-time validation of gastroesophageal varices under esophagogastroduodenoscopy using a deep convolutional neural network: a multicenter retrospective study (with video). *Gastrointest Endosc.* 2021;93:422-32.e3. DOI PubMed
31. Procopet B, Cristea VM, Robic MA, et al. Serum tests, liver stiffness and artificial neural networks for diagnosing cirrhosis and portal hypertension. *Dig Liver Dis.* 2015;47:411-6. DOI PubMed
32. Yu Q, Huang Y, Li X, et al. An imaging-based artificial intelligence model for non-invasive grading of hepatic venous pressure gradient in cirrhotic portal hypertension. *Cell Rep Med.* 2022;3:100563. DOI PubMed PMC
33. Qi X, An W, Liu F, et al. Virtual hepatic venous pressure gradient with CT angiography (CHESS 1601): a prospective multicenter study for the noninvasive diagnosis of portal hypertension. *Radiology.* 2019;290:370-7. DOI PubMed
34. Simbrunner B, Marculescu R, Scheiner B, et al. Non-invasive detection of portal hypertension by enhanced liver fibrosis score in patients with different aetiologies of advanced chronic liver disease. *Liver Int.* 2020;40:1713-24. DOI PubMed PMC
35. North Italian Endoscopic Club for the Study 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;319:983-9. DOI PubMed
36. Merkel C, Zoli M, Siringo S, et al. Prognostic indicators of risk for first variceal bleeding in cirrhosis: a multicenter study in 711 patients to validate and improve the North Italian Endoscopic Club (NIEC) index. *Am J Gastroenterol.* 2000;95:2915-20. DOI PubMed
37. Hou Y, Yu H, Zhang Q, et al. Machine learning-based model for predicting the esophagogastric variceal bleeding risk in liver cirrhosis patients. *Diagn Pathol.* 2023;18:29. DOI PubMed PMC
38. Agarwal S, Sharma S, Kumar M, et al. Development of a machine learning model to predict bleed in esophageal varices in compensated advanced chronic liver disease: a proof of concept. *J Gastroenterol Hepatol.* 2021;36:2935-42. DOI PubMed
39. Wang J, Wang Z, Chen M, et al. An interpretable artificial intelligence system for detecting risk factors of gastroesophageal variceal bleeding. *NPJ Digit Med.* 2022;5:183. DOI PubMed PMC
40. Wang Y, Hong Y, Wang Y, et al. Automated multimodal machine learning for esophageal variceal bleeding prediction based on endoscopy and structured data. *J Digit Imaging.* 2023;36:326-38. DOI PubMed PMC
41. Zhong BY, Tang HH, Wang WS, et al. Performance of artificial intelligence for prognostic prediction with the albumin-bilirubin and platelet-albumin-bilirubin for cirrhotic patients with acute variceal bleeding undergoing early transjugular intrahepatic portosystemic shunt. *Eur J Gastroenterol Hepatol.* 2021;33:e153-60. DOI PubMed
42. Liu H, Sun J, Liu G, Liu X, Zhou Q, Zhou J. Establishment of a non-invasive prediction model for the risk of oesophageal variceal bleeding using radiomics based on CT. *Clin Radiol.* 2022;77:368-76. DOI PubMed
43. Luo R, Gao J, Gan W, Xie WB. Clinical-radiomics nomogram for predicting esophagogastric variceal bleeding risk noninvasively in patients with cirrhosis. *World J Gastroenterol.* 2023;29:1076-89. DOI PubMed PMC
44. Yang JQ, Zeng R, Cao JM, et al. Predicting gastro-oesophageal variceal bleeding in hepatitis B-related cirrhosis by CT radiomics signature. *Clin Radiol.* 2019;74:976.e1-9. DOI PubMed
45. Yan Y, Li Y, Fan C, et al. A novel machine learning-based radiomic model for diagnosing high bleeding risk esophageal varices in cirrhotic patients. *Hepatol Int.* 2022;16:423-32. DOI PubMed
46. Augustin S, Pons M, Maurice JB, et al. Expanding the Baveno VI criteria for the screening of varices in patients with compensated

- advanced chronic liver disease. *Hepatology*. 2017;66:1980-8. DOI PubMed
47. Zhang Y, Zhao Q, Wen J. Splenic CT radiomics nomogram predicting the risk of upper gastrointestinal hemorrhage in cirrhosis. *J Radiat Res Appl Sci*. 2023;16:100486. DOI
 48. Lin Y, Li L, Yu D, et al. A novel radiomics-platelet nomogram for the prediction of gastroesophageal varices needing treatment in cirrhotic patients. *Hepatol Int*. 2021;15:995-1005. DOI PubMed
 49. Tseng Y, Ma L, Li S, et al. Application of CT-based radiomics in predicting portal pressure and patient outcome in portal hypertension. *Eur J Radiol*. 2020;126:108927. DOI PubMed
 50. Simsek C, Sahin H, Emir Tekin I, Koray Sahin T, Yasemin Balaban H, Sivri B. Artificial intelligence to predict overall survivals of patients with cirrhosis and outcomes of variceal bleeding. *Hepatol Forum*. 2021;2:55-9. DOI PubMed PMC
 51. Wiens J, Saria S, Sendak M, et al. Do no harm: a roadmap for responsible machine learning for health care. *Nat Med*. 2019;25:1337-40. DOI PubMed
 52. Tjoa E, Guan C. A survey on explainable artificial intelligence (XAI): toward medical XAI. *IEEE Trans Neural Netw Learn Syst*. 2021;32:4793-813. DOI PubMed
 53. Collins GS, Moons KGM. Reporting of artificial intelligence prediction models. *Lancet*. 2019;393:1577-9. DOI PubMed