ScieRxiv

NOT PEER-REVIEWED

Review

Artificial intelligence in gangrenous cholecystitis: Advances in risk factor identification and preoperative prediction—A narrative review

Wang Ping, Chen Xingyu, Hao Tianjiao, Chen Jisong

Department of Hepatopancreatobiliary Surgery, Taizhou Fourth People's Hospital, Taizhou 225300, Jiangsu, China.

Correspondence to: Dr. Wang Ping, Department of Hepatopancreatobiliary Surgery, Taizhou Fourth People's Hospital, Taizhou 225300, Jiangsu, China. Email: 348781503@qq.com

Received: 28 June 2026 | Approved: 14 July 2026 | Online: 14 July 2026

Abstract

Gangrenous cholecystitis (GC) represents the most severe pathological subtype of acute cholecystitis, yet its preoperative diagnostic rate remains below 10%. Conventional scoring systems based on logistic regression demonstrate limited predictive performance. A comprehensive literature search was conducted across PubMed, Web of Science, Embase, and CNKI databases for publications from January 2020 to December 2025. The search strategy combined Medical Subject Headings (MeSH) terms and free-text keywords, including "gangrenous cholecystitis", "gallbladder gangrene", "artificial intelligence", "machine learning", "deep learning", "prediction model", and "risk factor", linked by Boolean operators (AND/OR). Inclusion criteria comprised: (1) original research articles on AI/ML/DL models for GC prediction or diagnosis; (2) studies reporting model performance metrics (AUC, sensitivity, specificity); and (3) publications in English or Chinese.

Exclusion criteria encompassed case reports, editorials, conference abstracts, and studies lacking quantitative model evaluation. Two independent reviewers screened titles and abstracts, followed by full-text review of potentially eligible articles; disagreements were resolved through consensus with a third senior reviewer. Data extraction captured study design, sample size, AI algorithm type, input features, and performance metrics. Given the substantial heterogeneity across study designs, a narrative synthesis approach was adopted. Machine learning models—particularly XGBoost and Random Forest—achieved AUC values up to 0.944 in preoperative GC prediction, substantially outperforming traditional scoring systems. Deep learning models integrating non-contrast and contrast-enhanced CT attained an AUC of 0.965 in multicenter validation. Explainable AI (XAI) employing SHAP (SHapley Additive exPlanations) identified novel risk factors, including hypokalemia and hyponatremia. AI demonstrates considerable potential in GC risk stratification; however, prospective validation, multicenter.

Keywords: Gallbladder gangrene, gangrenous cholecystitis, artificial intelligence, machine learning, deep learning, explainable AI, prediction model, review

INTRODUCTION

Acute cholecystitis is among the most common acute abdominal conditions encountered in clinical practice. Epidemiological data from China indicate an annual incidence of approximately 0.15%-0.23%[1]. Among its subtypes, gangrenous cholecystitis (GC) is the most fulminant pathological variant, characterized pathologically by full-thickness ischemic necrosis of the gallbladder wall, with a reported prevalence of 2%-30% among acute cholecystitis cases[2,3]. The perioperative complication rate of GC is approximately two- to three-fold higher than that of non-gangrenous cholecystitis, and GC is associated with markedly prolonged hospital stays and increased healthcare expenditures[2,4].

Preoperative diagnosis of GC has long posed a formidable challenge in hepatobiliary surgery. A retrospective cohort study by El Asmar *et al.* demonstrated that among 122 patients with GC, the preoperative diagnostic accuracy was merely 7.4%, with over 90% of GC cases being missed prior to surgery[2]. This diagnostic shortfall results in delayed

intervention—once GC is confirmed, emergent surgery is mandated, whereas patients without GC may be managed conservatively or scheduled for elective cholecystectomy[5,6]. Consequently, accurate preoperative identification of patients at high risk for GC is of critical importance for optimizing clinical decision-making and improving patient outcomes. The present review employs a narrative review methodology. A systematic search was conducted across PubMed, Web of Science, Embase, and the China National Knowledge Infrastructure (CNKI) databases for publications spanning January 2020 to December 2025. The search strategy combined Medical Subject Headings (MeSH) terms with free-text keywords, including “gangrenous cholecystitis”, “gallbladder gangrene”, “artificial intelligence”, “machine learning”, “deep learning”, and “prediction model”, connected by Boolean operators (AND/OR). Inclusion criteria were: (1) original research on AI/ML/DL models for GC prediction or diagnosis; (2) studies reporting model performance metrics (AUC, sensitivity, specificity); and (3) publications in English or Chinese. Exclusion criteria comprised case reports, editorials, conference abstracts, and studies lacking quantitative model evaluation. Two independent reviewers screened the literature, with disagreements resolved through consensus discussion. Given the considerable heterogeneity in study design, algorithm type, and outcome measures among the included studies, a narrative synthesis—rather than a quantitative meta-analysis—was adopted.

Traditional identification of GC risk factors has relied predominantly on univariate analysis and multivariable logistic regression. Currently established risk factors include advanced age, male sex, diabetes mellitus, cardiovascular disease, elevated white blood cell (WBC) count, elevated C-reactive protein (CRP), gallbladder distension, and pericholecystic fluid[2,7,8]. Clinical scoring systems constructed from these risk factors, such as the Yacoub score, have been applied but demonstrate only modest predictive performance (AUC 0.75-0.78)[9]. These scoring tools harbor several inherent limitations—they assume linear relationships among variables, incorporate a limited number of predictors, and largely fail to integrate imaging information[10].

In recent years, artificial intelligence (AI) technologies have achieved breakthrough advances in medicine[11,12]. Machine learning (ML) algorithms can discern non-linear

relationships within high-dimensional data[13,14], deep learning (DL) models can automatically extract features from imaging data[15], and explainable AI (XAI) provides methodological support for bridging the gap between "black-box" models and clinical trust[16]. In the GC domain, two landmark studies in 2025—Ma *et al.*[17] achieving an AUC of 0.944 using XGBoost on a single-center cohort of 1,006 patients, and Hu *et al.*[18] validating a multicenter model using Random Forest—signal that AI-based GC prediction has entered the clinical validation phase. Contemporaneously, Guo *et al.*[19] integrated self-supervised learning with multicenter CT imaging, attaining a model AUC of 0.965, thereby introducing imaging-based deep learning into preoperative GC diagnosis and pioneering an entirely novel technical pathway.

The present review aims to systematically examine the application of AI technologies in GC risk factor identification and prediction over the past five years, spanning three dimensions: machine learning models for structured clinical data, deep learning models for imaging data, and emerging AI technologies. We further discuss current challenges and future directions, with the goal of providing a reference framework for precision diagnosis and treatment of GC.

TRADITIONAL RISK FACTORS FOR GANGRENOUS CHOLECYSTITIS

Demographic and comorbidity factors

Traditional risk factors for GC have been extensively investigated. With respect to demographics, advanced age is the most consistently identified independent risk factor. El Asmar *et al.*[2] reported in a retrospective cohort study that, after adjusting for sex, diabetes, coronary heart disease, and hypertension, each one-year increase in age was associated with an approximately 3% increase in the risk of acute gangrenous cholecystitis (OR = 1.03, 95% CI: 1.01-1.05, $p = 0.009$). Liu Nanbin *et al.*[20] similarly confirmed that age ≥ 60 years constitutes an independent risk factor. Male sex represents another independent predictor. Multivariable logistic regression analysis by El Asmar *et al.*[2] showed that the risk of developing acute gangrenous cholecystitis in males was approximately 1.9 times that in females (OR = 1.897, 95% CI: 1.024-3.513, $p = 0.042$). Notably, some heterogeneity persists across studies regarding whether male sex qualifies as an independent predictor;

certain investigations failed to retain male sex as an independent risk factor in multivariable analysis[21], a finding potentially attributable to differences in population characteristics, sample size, and collinearity—warranting further exploration.

With respect to comorbidities, diabetes mellitus (DM) is the most extensively studied comorbidity-related risk factor, with multiple studies confirming that DM patients face a significantly elevated risk of progression from acute cholecystitis to gangrene[2,22]. El Asmar *et al.*[2] identified DM as an independent predictor of GC complicated by sepsis (OR = 2.67, 95% CI: 1.08-6.60, $p = 0.033$). Cardiovascular disease constitutes an important risk factor for GC, with its mechanism primarily mediated through gallbladder wall hypoperfusion secondary to atherosclerotic changes in the cystic artery[23]. Underlying cardiovascular pathology—such as coronary atherosclerosis and heart failure—can intrinsically cause thickening of the gallbladder microvasculature and luminal narrowing. In the setting of superimposed systemic hypoperfusion events (e.g., acute myocardial infarction, open-heart surgery, aortic surgery), cystic artery perfusion is further compromised, leading to gallbladder wall (particularly mucosal layer) ischemia and necrosis, with rapid progression to gangrenous cholecystitis[23]. Furthermore, vagal reflexes triggered by cholecystitis may exacerbate myocardial ischemia, establishing a vicious "ischemia-inflammation" cycle that further elevates the risk of GC[23]. Portal hypertension-induced gallbladder wall edema, congestion, and coagulopathy also jointly contribute to increased GC risk and surgical complexity[24].

Laboratory parameters

With respect to laboratory parameters, elevated WBC count is the most classical biomarker for GC; the Tokyo Guidelines 2018 (TG18) incorporate WBC > 18 × 10^9/L as a criterion for severe acute cholecystitis[6]. Procalcitonin (PCT) exhibits substantially higher specificity for cholecystitis complicated by severe systemic infection (Tokyo Guidelines Grade III) compared with CRP (92.1% *vs.* 69.7%), achieving a negative predictive value of 96.5% at a cutoff of ≥ 5.095 ng/mL[25]. Elevated D-dimer (>1.18 μg/mL) is an independent risk factor for gallbladder gangrene (OR = 4.03), implicating local circulatory disturbance and gallbladder wall microthrombosis in the pathogenesis of GC[21]. The neutrophil-to-

lymphocyte ratio (NLR) and the Systemic Immune-Inflammation Index (SII) are composite inflammatory indices that have attracted considerable attention in recent years. Zheng *et al.*[21] demonstrated that SII significantly outperformed CRP, WBC, and PCT in predicting GC (AUC: 0.880 *vs.* 0.797/0.726/0.773, $p < 0.05$), a finding subsequently corroborated by Ardelean *et al.*[26] (SII AUC = 0.889). Non-neoplastic elevation of CA19-9 in GC is a phenomenon increasingly recognized in recent years. Zheng *et al.*[21] identified CA19-9 > 18.8 U/mL as an independent risk factor for gallbladder gangrene (OR = 7.21), with the putative mechanism involving severe inflammation-induced damage to biliary epithelial cells and increased CA19-9 release[21,27]. Notably, thrombocytosis during the active phase of GC (platelet count > 351 × 10^9/L) reflects a systemic inflammation-driven hypercoagulable state and serves as an independent risk factor for gallbladder gangrene[21]; conversely, when GC is complicated by sepsis or disseminated intravascular coagulation (DIC), platelet counts may decline markedly due to consumptive coagulopathy[28].

Imaging features

With respect to imaging, gallbladder distension is a strong predictor of GC. El Asmar *et al.*[2] determined, through ROC curve analysis, that a transverse gallbladder diameter > 3.85 cm represents the optimal cutoff for differentiating GC from non-gangrenous cholecystitis. Multivariable logistic regression demonstrated that patients with a transverse diameter > 3.85 cm faced an approximately 3.42-fold increased risk of GC compared with those with a diameter ≤ 3.85 cm (OR = 3.42, 95% CI: 1.786-6.562, $p < 0.001$). Zheng *et al.*[21] likewise confirmed that gallbladder distension is an independent risk factor for gallbladder gangrene (multivariable OR = 6.47, 95% CI: 2.19-19.10, $p < 0.001$), with a detection rate of 63.7% in the gangrene group, substantially higher than the 18.5% observed in the non-gangrene group.

Gallbladder wall thickening (≥ 4 mm) is a classical diagnostic criterion for acute cholecystitis[6]; however, it should be noted that this sign per se is not specific for GC—it can occur in both simple and gangrenous cholecystitis, reflecting inflammatory edema of the gallbladder wall[29]. The key imaging findings truly indicative of GC are structural loss within necrotic regions of the gallbladder wall, manifested on ultrasound as loss of the

mucosal line/wall interface echogenicity, wall de-lamination, asymmetric wall thickening, or intraluminal floating necrotic membranes, with an incidence exceeding 90%[29,30]. Color Doppler ultrasound (CDUS) and contrast-enhanced ultrasound (CEUS) further enhance the specificity of GC diagnosis by demonstrating focal perfusion defects in the gallbladder wall—CEUS-demonstrated wall perfusion defects can achieve a specificity of 100% (PPV 100%)[31]. On contrast-enhanced CT, GC manifests as heterogeneous wall enhancement, focal non-enhancing areas, or even wall discontinuity (irregular or absent wall)[32]; in a study of 31 GC cases, focal reduced enhancement was observed in 80% of patients undergoing contrast-enhanced scanning[30]. In essence, wall thickening reflects the degree of inflammation, whereas wall thinning/structural loss/perfusion defects represent the specific signs indicative of gangrenous transformation[6,31,32].

Limitations of traditional scoring systems

Several GC prediction scoring systems have been developed based on the aforementioned factors. Wu *et al.*[33], based on a retrospective analysis of 5,243 cholecystectomy patients, proposed a simplified scoring system (0-5 points) comprising four factors: age (stratified scoring: ≤ 45 years = 0, 46-65 years = 1, > 65 years = 2), heart rate > 90 bpm, WBC > 13,000/mm^2, and gallbladder wall thickening ≥ 4 mm, achieving a modified AUC of 0.77. Earlier, Yacoub *et al.*[34] proposed a five-factor score (0-6.5 points) including male sex (2 points), age > 45 years (1 point), heart rate > 90 bpm (1 point), WBC > 13,000/mm^3 (1.5 points), and gallbladder wall thickening > 4.5 mm (1 point); however, male sex lost statistical significance in Wu *et al.*'s multivariable analysis. Bouassida *et al.*[9] further proposed a new scoring system in 2021 incorporating six factors—ASA score 3-4, body temperature, symptom duration, WBC, male sex, and pericholecystic fluid—achieving an AUC of 0.83, outperforming both prior scoring systems.

Nevertheless, these scoring systems harbor inherent limitations: (1) reliance on the linearity assumption of logistic regression, rendering them incapable of capturing interaction effects and non-linear relationships among variables; (2) incorporation of a limited number of variables, resulting in suboptimal information utilization; (3) lack of deep integration of imaging features; (4) absence of external and prospective validation in the majority of cases;

and (5) inability to provide individualized contribution analysis, yielding insufficient clinical interpretability. These limitations create the methodological space for AI technologies. Moreover, from a deeper methodological perspective, these traditional scoring systems suffer from additional deficiencies: (6) absence of a temporal dimension—all scores are derived from single cross-sectional data at admission and cannot capture the dynamic trajectory of inflammation from onset to necrosis, even though temporal change patterns per se may harbor important predictive information; (7) failure to integrate the increasingly rich repertoire of novel composite inflammatory biomarkers, such as NLR, SII, and D-dimer, which have been validated as independent predictors of GC in multiple studies[21,26]; and (8) incomplete reporting of model calibration and decision curve analysis (DCA) for published scoring systems, hindering assessment of their net clinical benefit in real-world settings. These limitations collectively provide ample methodological rationale and clinical demand-driven impetus for the introduction of AI technologies.

MACHINE LEARNING PREDICTION MODELS BASED ON STRUCTURED DATA

Application of traditional machine learning algorithms

In 2025, Ma *et al.*[17] published the largest GC machine learning prediction study to date. This investigation enrolled 1,006 patients with cholecystitis (897 in the training/validation set, 109 in the external test set), including 208 in the GC group and 689 in the non-GC group. The authors collected 26 clinical features and constructed and compared five models: decision tree, support vector machine (SVM), random forest (RF), XGBoost, and AdaBoost. Through five-fold cross-validation, XGBoost was identified as the optimal model, achieving an AUC of 94.4% (95% CI: 88.80%-98.39%) on the external test set (n = 109), significantly outperforming the preoperative clinical diagnostic accuracy of 86.95%. The study also developed a streamlined model comprising only seven variables (WBC, NLR, D-dimer, fibrinogen, gallbladder width, gallbladder wall thickness, and hypokalemia/hyponatremia), which maintained high predictive performance while reducing data acquisition costs, and released a Streamlit-based online prediction application (https://gangrenous-cholecystitis-prediction-model.streamlit.app/), enabling real-time probability prediction and visualization of variable contributions.

In the same year, Hu *et al.*[18] published the first multicenter GC machine learning prediction model. Employing multicenter retrospective data, the study enrolled 1,044 patients (521 in the training set, 223 in the test set, 300 in the external validation set), including 210 GC patients. Through dual variable selection using LASSO regression and the Boruta algorithm, 10 predictors were ultimately identified, and a random forest-based prediction model was developed and externally validated, achieving an AUC of 0.818 (95% CI: 0.763-0.873) on the external validation set. SHAP (SHapley Additive exPlanations) analysis identified gallbladder wall thickening, CRP, pericholecystic fluid, WBC, and impacted stones as the five most important predictive features. Multicenter external validation confirmed the model's generalizability, providing a reliable tool for preoperative GC risk stratification.

These two studies represent the current state of the art in structured-data ML prediction for GC. Notably, Ma *et al.*[17] found that commonly used oversampling methods such as SMOTE paradoxically exacerbated overfitting in GC prediction—a finding of methodological significance for future research. Table 1 systematically summarizes the model architectures and performance metrics of representative studies in the GC-AI prediction domain. It is particularly noteworthy that although both Ma *et al.*[17] and Hu *et al.*[18] employed tree ensemble models (XGBoost *vs.* Random Forest), a marked performance gap exists between them (AUC 0.944 *vs.* 0.818). This discrepancy may be attributable to the following factors: (1) differences in data source—Ma *et al.* used single-center data, which may have led to overfitting center-specific patterns, whereas Hu *et al.* provided more conservative and reliable performance estimates through multicenter external validation; (2) differences in the predictor variable set—Ma *et al.*'s model incorporated hypokalemia/hyponatremia, a highly discriminative variable (GC group 87.7% *vs.* non-GC group 48.1%), which may have substantially enhanced discriminative capacity; and (3) differences in sample size—Ma *et al.*'s training set (n = 897) was approximately 1.7 times larger than that of Hu *et al.* (n = 521). We recommend that future studies conduct head-to-head model benchmarking under a unified variable set and standardized

preprocessing pipeline to more reliably assess the relative merits of different algorithms in GC prediction.

Table 1. Summary of representative AI models for gangrenous cholecystitis prediction.

Study	Year	Algorithm	N (Total / GC)	AUC	Sensitivity	Specificity	Key Features
Ma *et al.* [17]	2025	XGBoost	1,006 / 208	0.944	—	—	Single-center; 7-variable simplified model; online App; SHAP
Hu *et al.* [18]	2025	Random Forest	1,044 / 210	0.818	—	—	Multicenter (3 cohorts); LASSO + Boruta; SHAP
Guo *et al.* [19]	2025	seResNet-50 (SSL)	1,228 / —	0.965*	0.88*	0.95*	Multicenter CT; non-contrast + contrast fusion; SSL pre-training
Guo *et al.* [19]	2025	seResNet-50 (SSL)	val: 143 / 164	0.879 /0.887[d]	0.77/0.76[d]	0.83/0.84[d]	Independent validation; outperformed radiologists (AUC

							0.721/0.703)
Bouassida *et al.* [9]	2021	Logistic score	—	0.83	—	—	Traditional scoring (6 factors); no imaging integration
Chen *et al.* [41]	2026	Stacking ensemble + radiomics	—	0.82[s]	0.714[s]	0.831[s]	Clinical-radiomics fusion; suppurative cholecystitis
Jayanthi *et al.* [39]	2026	SE-CapsNet + BiLSTM	10,692 images	—	0.976[g]	0.993[g]	Ultrasound DL; 9-class; GC as independent class (C4)

*Training set; [d] = Validation set I / Validation set II performance; [s] = External test set; [g] = Test set (C4: membranous & gangrenous cholecystitis subclass). AUC = area under the ROC curve; GC = gangrenous cholecystitis; SSL = self-supervised learning; SHAP = SHapley Additive exPlanations; Se = sensitivity; Sp = specificity.

Introduction of explainable artificial intelligence

Although machine learning models have demonstrated excellent performance in GC prediction, their "black-box" decision-making mechanisms impede clinical adoption. Surgeons need to understand why a model makes a given prediction before incorporating it into clinical decision-making. Explainable AI (XAI) serves as a critical bridge across this chasm by providing transparent explanations of model decisions[35]. Mainstream XAI methods such as SHAP quantify the contribution of each clinical feature to the prediction

outcome, thereby visualizing the model's decision-making process and facilitating clinicians' understanding and trust in model predictions.

In the study by Ma *et al.*[17], SHAP analysis yielded an unexpected discovery—hypokalemia/hyponatremia emerged as an important predictor of GC (present in 87.7% of the GC group *vs.* 48.1% of the non-GC group), a factor that had received scant attention in traditional research and whose identification was enabled through the application of SHAP. In Hu *et al.*'s[18] multicenter model, SHAP ranking identified gallbladder wall thickening as the most important predictive feature, surpassing traditional markers such as WBC and CRP—a finding highly consistent with the pathophysiological mechanism of ischemic necrosis of the gallbladder wall.

The value of SHAP extends beyond feature importance ranking to its ability to explain each individual prediction. Ma *et al.*'s[17] online application not only provides a predicted probability but also displays the positive or negative contribution of each variable to the prediction, enabling surgeons to assess whether the model's prediction aligns with clinical intuition, thereby fostering trust. This individualized explanation is a capability that traditional scoring systems cannot offer.

Beyond SHAP, LIME (Local Interpretable Model-agnostic Explanations) represents another important model-agnostic XAI method, which explains individual predictions by fitting a locally linear approximation model in the neighborhood of the sample to be explained[36]. However, as of 2025, no published report on the application of LIME in GC prediction exists. Future research may explore the combined use of SHAP and LIME—employing SHAP for global feature importance ranking and LIME to assist with fine-grained explanations for specific high-risk patients—to further enhance the clinical interpretability and individualized decision-support capabilities of GC prediction models.

Development of online prediction tools

The clinical value of AI prediction models ultimately hinges on their accessibility and usability in real-world clinical settings. In recent years, online prediction tools for GC have begun to proliferate:

Ma *et al.*'s[17] Streamlit-based online application is currently the most comprehensive GC prediction tool, generating a GC probability and visualizing the contribution of each variable upon entry of seven clinical parameters.

Hu *et al.*[18] designed a clinical decision-support tool based on their multicenter model, aimed at assisting surgeons in rapidly assessing GC risk.

In the broader domain of biliary surgery, Cicerone *et al.*[37] developed two machine learning-based risk prediction tools in 2025—CholeSurgRisk I (preoperative model, AUC = 0.8456) and CholeSurgRisk II (comprehensive model, AUC = 0.8903)—for predicting the risk of major complications following laparoscopic cholecystectomy for acute calculous cholecystitis. The dissemination of the preoperative online tool "CholeSurgRisk I" also provides a reference paradigm for the development of GC prediction models.

The emergence of these online tools signals that AI-based GC prediction is transitioning from academic research to clinical practice; however, their real-world effectiveness and clinical impact remain unknown and require evaluation through prospective studies.

DEEP LEARNING MODELS BASED ON IMAGING DATA

CT imaging deep learning

CT is the most commonly employed imaging modality for preoperative GC evaluation. In 2025, Guo *et al.*[19] published the first multicenter GC imaging deep learning study to date. This investigation enrolled 1,228 patients from three centers, yielding 7,368 CT images (921 in the training set, 143 in independent validation set I, and 164 in validation set II). A self-supervised learning (SSL) strategy was implemented using the seResNet-50 framework, integrating both non-contrast and contrast-enhanced CT modalities. The SSL strategy involved pre-training on large volumes of unlabeled CT images, followed by fine-tuning on

limited labeled data, thereby effectively mitigating the bottleneck of scarce medical image annotations. The results demonstrated that the fusion model achieved an AUC of 0.965 on the training set and 0.879 and 0.887 on independent validation sets I and II, respectively—significantly outperforming the contrast-enhanced CT unimodal model (0.791/0.810) and the non-contrast CT unimodal model (0.756/0.730) (DeLong test, $P < 0.001$). The training set sensitivity and specificity were 88% and 95%, respectively; validation set I sensitivity and specificity were 77% and 83%; and validation set II sensitivity and specificity were 76% and 84%. Notably, the diagnostic performance of the fusion model in validation set I (AUC 0.879) significantly surpassed that of two senior radiologists (AUC 0.721 and 0.703, respectively). This study confirms the substantial potential of multimodal CT fusion SSL strategies for preoperative GC imaging diagnosis.

This study demonstrates that the performance of CT imaging deep learning in GC diagnosis has surpassed that of conventional radiological interpretation; nevertheless, the two approaches remain complementary—AI can detect subtle imaging features imperceptible to the human eye, whereas the macroscopic judgment and clinical experience of radiologists remain indispensable.

Ultrasound image AI analysis

Ultrasound is the first-line imaging modality for acute cholecystitis; however, AI analysis of ultrasound images poses considerably greater challenges than CT, owing to the operator-dependent nature of ultrasound image quality and low standardization. Despite these challenges, preliminary progress has been achieved in this domain.

Das *et al.*[38] utilized clinical tabular data from 319 patients with gallstone disease (38 features) sourced from the UCI database to develop a predictive model for gallstone disease onset based on the CatBoost classifier and the Sea Lion Optimization Algorithm (SLOA). The SLOA simultaneously performed dual tasks—selecting 19 features from the original 38 and optimizing CatBoost hyperparameters. Through five-fold cross-validation, the SLOA_CB model achieved an accuracy of 80.42% (AUC 0.82-0.92). The study also employed three XAI methods—SHAP, LIME, and DiCE—to construct a multi-layered

interpretability framework: SHAP revealed, at the global level, that CRP and vitamin D were the most important predictive features; LIME provided case-specific local explanations; and DiCE performed counterfactual analysis, indicating which feature modifications would reverse the prediction. Although this study targeted gallstone disease rather than GC, its combined "SLOA optimization + multi-method XAI" framework provides methodological reference for clinical data modeling in GC.

A study that genuinely addressed the ultrasound imaging level was that of Jayanthi *et al.*[39], which constructed a hybrid deep learning model—HDLMFE-ADGDT—covering nine gallbladder diseases based on 10,692 ultrasound images. The model employed SE-CapsNet for feature extraction combined with CNN-BiLSTM for sequence classification, achieving an overall accuracy of 99.09% (sensitivity 95.87%, specificity 99.49%). Of particular relevance to GC, this study designated category 4 (C4) as an exclusive classification for membranous and gangrenous cholecystitis, attaining a test-set sensitivity of 97.64% and specificity of 99.31%—making this the first study to include GC as an independent category within a multi-class ultrasound DL model. In comparison with classical architectures such as VGG16 (97.78%), ResNet152 (85.30%), and InceptionV3 (86.50%), the SE-CapsNet plus BiLSTM hybrid approach demonstrated clear superiority in ultrasound image classification of gallbladder diseases.

Given that ultrasound is the most commonly used initial imaging examination for acute cholecystitis, the development of GC-specific ultrasound AI models holds significant clinical value and broad application prospects.

Radiomics

The radiomics paradigm involves extracting high-throughput quantitative features from medical images and converting this imaging information into mineable data. Gupta *et al.*[40] conducted an early exploration of CT texture analysis in gallbladder cancer, employing TexRAD software to process portal venous phase CT images from 38 patients with suspected gallbladder cancer. Filter-histogram-based texture analysis identified consistent texture parameters—mean intensity and kurtosis—as predictors of adenocarcinoma

histological grade, with kurtosis significantly associated with overall survival (p = 0.020). Although this study focused on gallbladder cancer rather than inflammatory gallbladder disease, its methodological framework—comprising ROI delineation, multi-scale texture feature extraction, and association analysis between texture parameters and clinical outcomes—provides a transferable technical pathway and exploratory experience for subsequent radiomics-based investigation of GC.

Chen *et al.*[41] integrated clinical features with CT radiomic features in 2026, employing a stacking ensemble strategy to construct an explainable machine learning model for preoperative prediction of acute suppurative cholecystitis (ASC). In this multicenter retrospective study, a clinical model (AUC = 0.75), a radiomics model (AUC = 0.76), and a clinical-radiomics fusion model (AUC = 0.82) were separately constructed, with the fusion model achieving a sensitivity of 71.4% and specificity of 83.1% on the external test set—significantly outperforming the unimodal models. The model also provided feature-level explanations through SHAP analysis. This study offers methodological reference for predicting cholecystitis severity from a clinical-radiomics fusion perspective. The advantage of radiomics lies in its ability to quantify textural features imperceptible to the naked eye—such as necrosis heterogeneity and indirect characterization of gallbladder wall microvascular density—thereby providing a new dimension for the imaging evaluation of GC.

Multimodal fusion strategies

Multimodal fusion represents a frontier direction in current medical AI. Guo *et al.*'s[19] study has confirmed that the fusion of non-contrast and contrast-enhanced CT significantly outperforms either modality alone. The underlying mechanism resides in the complementary imaging information provided by the two modalities—non-contrast CT reflects baseline tissue density characteristics, whereas contrast-enhanced CT reveals perfusion status. Given that the core pathology of GC is ischemic necrosis, non-enhancing regions of the gallbladder wall on contrast-enhanced CT carry direct diagnostic significance for GC.

Looking forward, more advanced multimodal fusion strategies warrant exploration: combined clinical data plus imaging feature models—fusing ML-extracted clinical risk factors with DL-extracted imaging features at the feature or decision level holds promise for constructing a more comprehensive GC prediction framework; ultrasound plus CT plus clinical parameter trimodal fusion—with ultrasound providing real-time bedside assessment, CT delivering in-depth imaging information, and clinical parameters offering systemic status evaluation, the synergy of the three modalities may enable whole-process GC risk monitoring spanning from the emergency department to the operating room.

LARGE LANGUAGE MODELS AND EMERGING AI TECHNOLOGIES: POTENTIAL APPLICATIONS

Potential roles of large language models in GC diagnosis and treatment

Large language models (LLMs), as the representative of next-generation AI technologies, are profoundly transforming the landscape of medical information processing. In the GC domain, potential LLM applications include:

Electronic health record text mining: Natural language processing (NLP) techniques can autonomously parse unstructured electronic health records to extract latent risk factor information. For example, textual descriptions such as “right upper quadrant tenderness with rebound” and “fever for 3 days” documented in admission notes can, following NLP parsing, serve as supplementary information for GC prediction, enriching the feature space of the model.

Clinical decision support: LLMs can integrate clinical practice guidelines such as the Tokyo Guidelines with individualized risk assessments to provide surgical decision-making recommendations regarding the timing of GC surgery. However, LLMs carry the risk of generating false or misleading medical information (i.e., “hallucination”), warranting a high degree of vigilance in clinical settings.

Patient education: LLMs can automatically generate GC-related educational materials and preoperative discussion talking points, enhancing patients' understanding of the disease and treatment options.

Federated learning and multicenter collaboration

The generalizability of GC-AI models necessitates diverse training data; however, medical data privacy protection imposes stringent constraints on cross-institutional data sharing. Federated learning offers a promising solution by enabling each healthcare institution to train models locally, sharing only model parameters rather than raw data. Hu *et al.*'s[18] multicenter model still employed a conventional centralized data integration strategy. The future incorporation of a federated learning framework holds the potential to include heterogeneous data from more centers while preserving data privacy, further enhancing model generalizability and fairness.

Surgical video AI analysis

Intraoperative real-time AI assistance is a frontier domain in surgical AI. In 2022, Ward, Hashimoto *et al.*[42] developed an AI model based on laparoscopic cholecystectomy videos capable of automatically identifying the severity of gallbladder inflammation (Parkland Grading Scale, PGS) and predicting surgical progress. This study employed deep learning to analyze surgical video frames and successfully predicted three key surgical metrics: (1) operative duration ($R^2 = 0.61$); (2) attainment of the Critical View of Safety (CVS); and (3) risk of gallbladder injury. This study represents the first application of AI video analysis to automated assessment of gallbladder inflammation severity, providing surgeons with a novel tool for preoperative risk assessment and surgical difficulty prediction. It carries potential methodological value for complex GC-related gallbladder surgery, such as cases involving Calot's triangle fibrosis and gangrenous transformation of the gallbladder wall.

Although no published studies have reported intraoperative real-time AI identification of GC, the feasibility of this technical pathway has been validated—intraoperative recognition of GC wall necrosis features could assist surgeons in adjusting surgical strategy early in the

procedure (e.g., conversion from laparoscopic cholecystectomy to open surgery or subtotal cholecystectomy).

Temporal data analysis

All current GC prediction models are static, cross-sectional predictors—rendering judgments based on single-time-point data at admission. However, GC development is a dynamic process, with a temporal window extending from the onset of inflammation to full-thickness wall necrosis. Temporal models such as recurrent neural networks (RNNs) and long short-term memory networks (LSTMs) can perform a task beyond the capabilities of current static models: dynamically tracking the temporal evolution of inflammatory markers. For instance, if WBC and CRP exhibit a sustained upward trajectory over a 24-h period, temporal models may issue an early warning even before absolute values reach conventional thresholds—a task that static models cannot accomplish but dynamic models may achieve.

CHALLENGES AND LIMITATIONS

Data-level challenges

GC labels depend on postoperative pathology: Definitive diagnosis of GC requires postoperative pathological examination, meaning that the training labels for ML models are derived exclusively from patients who underwent surgery; unrecognized GC may exist among patients who improved with conservative management, introducing selection bias.

Class imbalance: The incidence of GC among acute cholecystitis cases ranges from approximately 2% to 30%—a relatively low prevalence. Although oversampling methods such as SMOTE are widely employed to address imbalanced data, Ma *et al.*[17] found that these methods paradoxically exacerbated overfitting in GC prediction. To date, no generally accepted optimal solution exists for effectively handling class imbalance without introducing synthetic data.

Multicenter data heterogeneity: Variations in laboratory testing methodologies, imaging equipment parameters, and clinical practices across centers may lead to performance

degradation when models are deployed in new environments. Establishing unified data acquisition standards and preprocessing pipelines is critical to addressing this challenge.

Model-level challenges

Overfitting risk: Training sample sizes for GC prediction models are relatively limited, whereas deep learning models may involve millions of parameters, rendering overfitting risk non-trivial. Ma *et al.*[17] mitigated this issue through external test set validation and variable reduction strategies; however, more rigorous cross-validation and independent validation remain necessary.

The "black-box" problem and limitations of XAI: Although methods such as SHAP provide model explanations, these explanations may themselves be misleading—correlation does not imply causation, and the "important features" revealed by SHAP may not necessarily reflect true pathophysiological mechanisms. Furthermore, current XAI methods are primarily suited for tree-based models on structured data and relatively simple imaging models; explanations for complex deep learning architectures remain inadequate.

Lack of prospective randomized controlled validation: All published GC-AI models are retrospective studies, lacking prospective validation and—more importantly—randomized controlled trials (RCTs) to evaluate the actual improvement in patient outcomes attributable to AI-assisted decision-making. This constitutes the greatest barrier to the translation of AI from research into clinical practice.

Clinical translation challenges

Clinician trust and adoption rates: The clinical adoption of AI prediction models depends on clinician trust. Multiple survey studies have demonstrated that surgeons' acceptance of AI-assisted decision-making correlates positively with model interpretability[43,44]. XAI methods such as SHAP help build trust by providing individualized explanations; however, truly large-scale clinical adoption still requires the accumulation of more extensive validation data.

Legal and ethical boundaries: When AI predictions conflict with clinical judgment, which should prevail? Who bears legal liability for decisions made with AI assistance? These questions are particularly thorny in the context of emergency surgical decision-making—delays in GC management can be fatal, yet AI-driven over-intervention may also expose patients to unnecessary surgical risk.

Accessibility across different tiers of hospitals: Current GC-AI research originates almost exclusively from large tertiary medical centers, and model performance in primary care settings remains uncertain. Given that early identification of GC is most urgently needed precisely at the primary care level, adapting AI tools to resource-limited environments constitutes a critical challenge for clinical translation.

FUTURE DIRECTIONS

Multimodal large models

Future GC-AI prediction should evolve toward multimodal integration—combining clinical text (NLP-extracted), laboratory parameters (temporal analysis), imaging features (DL-extracted), and surgical video (real-time analysis) within a unified predictive framework. Such multimodal large models could fully leverage complementary information from diverse data sources, thereby enabling an upgrade from “single-time-point prediction” to “whole-disease-course dynamic risk assessment”.

Prospective randomized controlled trials

The clinical value of AI-assisted GC decision-making ultimately requires validation through prospective RCTs. A possible study design would randomize patients with acute cholecystitis into an AI-assisted decision-making arm and a conventional decision-making arm, with the primary endpoints being the preoperative GC identification rate and the time from admission to surgery, and secondary endpoints including complication rates, length of hospital stay, and healthcare costs. Only when RCTs demonstrate superior outcomes in the AI-assisted arm compared with the conventional arm can GC-AI tools truly earn recommendation status in clinical practice guidelines.

Edge computing deployment

Deploying GC prediction models in primary care hospitals and emergency departments also requires addressing the practical constraint of limited computational resources. Edge computing technology enables models to run on local devices without reliance on cloud servers, meeting the real-time prediction demands of emergency settings. Ma *et al.*'s[17] seven-variable streamlined model is particularly well suited for such deployment—lightweight, requiring few input variables and low computational demand.

Digital twins

Constructing a gallbladder digital twin model based on patient-specific data and comparing the trajectory of GC progression under different intervention strategies could provide an intuitive prediction tool for individualized treatment decisions. This concept currently remains a forward-looking vision in the GC domain, with no published studies to date; however, with advances in organ-level multi-physics modeling, it may become a reality in the future.

AI-driven GC risk stratification clinical pathway

Establishing a GC risk stratification pathway based on AI predictions: low risk (AI-predicted probability < 10%) → close observation, conservative management, or elective surgery; intermediate risk (10%-50%) → contrast-enhanced CT evaluation, multidisciplinary team discussion, early surgery; high risk (> 50%) → emergent surgical intervention, avoidance of delay. This pathway transforms AI prediction from an "academic tool" into a "clinical standard", holding the potential to systematically improve the efficiency of GC diagnosis and treatment.

Guideline integration

Incorporating AI prediction tools into international standards such as the Tokyo Guidelines represents an important step toward clinical standardization of GC-AI. We recommend that future guideline revisions include recommendations for AI-assisted assessment, specifying the applicable scenarios, limitations, and human-AI collaborative decision-making workflows for AI tools.

CONCLUSIONS

AI technologies are profoundly reshaping the paradigm of GC risk factor identification and prediction. In the domain of structured clinical data, interpretable ML models based on algorithms such as XGBoost and Random Forest have demonstrated predictive performance surpassing both traditional scoring systems and preoperative clinical judgment, with AUC values reaching as high as 0.944. In the imaging domain, DL models based on self-supervised learning and multimodal CT fusion (AUC 0.965) have pioneered entirely novel approaches to non-invasive preoperative GC diagnosis. XAI methods such as SHAP have revealed counterintuitive risk factors—including hypokalemia and hyponatremia—and enabled individualized prediction explanations, serving as a critical bridge between "black-box" models and clinical trust. Nevertheless, the translation from research to clinical practice continues to face numerous challenges, including label bias, model overfitting, lack of prospective validation, and barriers to clinical adoption. In the future, the development of multimodal large models, prospective RCTs, edge computing deployment, and AI-driven clinical pathways will propel GC-AI from a "prediction tool" toward a "decision-making partner", ultimately realizing precision prevention and individualized treatment of GC.

DECLARATIONS

Authors' contributions

Conceptualization, Methodology, Investigation, Writing - original draft: P Wang;
Data curation, Formal analysis, Investigation: XY Chen;
Data curation, Software, Validation, Visualization: TT Hao;
Supervision, Project administration, Writing - review & editing: JS Chen.

Availability of data and materials

All data analyzed in this review are derived from publicly available published literature. The search strategy and list of included studies are available from the corresponding author upon reasonable request.

Financial support and sponsorship

None.

Conflict of interest

Not applicable.

Ethical approval and consent to participate

This study is a narrative review that exclusively analyzed published literature; no human participants, animal subjects, or primary patient data were involved. Therefore, ethical approval was not required. Not applicable. This study is a review of published literature and does not involve the recruitment of human participants.

Consent for publication

Not applicable. This manuscript does not contain any individual person's data in any form.

Copyright

REFERENCES

[1] Branch of Biliary Surgery, Chinese Society of Surgery, Chinese Medical Association. Guidelines for the diagnosis and treatment of acute biliary tract infections (2021 edition). *Chinese Journal of Surgery,* 2021. 6(59): pp. 422-9.[DOI: 10.1097/CM9.0000000000003258]

[2] El Asmar, N., et al., Predictors of acute gangrenous cholecystitis and its complications: a retrospective cohort study. *BMC Surgery,* 2025. 26(1): p. 39.[PMID: 41382140 PMCID: PMC12817750 DOI: 10.1186/s12893-025-03356-z]

[3] Fagan, S.P., et al., Prognostic factors for the development of gangrenous cholecystitis. *American Journal of Surgery,* 2003. 186(5): pp. 481-5.[PMID: 14599611 DOI: 10.1016/j.amjsurg.2003.08.001]

[4] V, K., et al., Risk Factors Associated With Gangrenous Cholecystitis: A Cohort Study From Eastern India. *Cureus,* 2024. 16(11): p. e74126.[PMID: 39712675 PMCID: PMC11662193 DOI: 10.7759/cureus.74126]

[5] Pisano, M., et al., 2020 World Society of Emergency Surgery updated guidelines for the diagnosis and treatment of acute calculus cholecystitis. *World Journal of Emergency Surgery,* 2020. 15(1): p. 61.[PMID: 33153472 PMCID: PMC7643471 DOI: 10.1186/s13017-020-00336-x]

[6] Okamoto, K., et al., Tokyo Guidelines 2018: flowchart for the management of acute cholecystitis. *Journal of Hepato-Biliary-Pancreatic Sciences,* 2018. 25(1): pp. 55-72.[PMID: 29045062 DOI: 10.1002/jhbp.516]

[7] Fabbri, N., et al., Enhancing the management of acute and gangrenous cholecystitis: a systematic review supported by the TriNetX database. *Translational Gastroenterology and Hepatology,* 2025. 10(0): p. 16.[PMID: 39944582 PMCID: PMC11811563 DOI: 10.21037/tgh-24-27]

[8] Córdova-Velázquez, C.A., et al., Risk Factors Associated with Acute Gangrenous Cholecystitis. *Acta Gastroenterologica Latinoamericana,* 2024. 54(2): pp. 147-52.[PMID: 41341977 PMCID: PMC12199649 DOI: 10.52787/agl.v54i2.402]

[9] Bouassida, M., et al., Acute Gangrenous Cholecystitis: Proposal of a Score and Comparison with Previous Published Scores. *Journal of Gastrointestinal Surgery,* 2021. 25(6): pp. 1479-86.[PMID: 32607855 DOI: 10.1007/s11605-020-04707-2]

[10] Rajkomar A, Dean J, Kohane I. Machine learning in medicine. *The New England Journal of Medicine,* 2019. 380(14): pp. 1347-58.[PMID: 26572668 PMCID: PMC5831252 DOI: 10.1161/CIRCULATIONAHA.115.001593]

[11] Fahim, Y.A., et al., Artificial intelligence in healthcare and medicine: clinical applications, therapeutic advances, and future perspectives. *European Journal of Medical Research,* 2025. 30(1): p. 848.[PMID: 40988064 PMCID: PMC12455834 DOI: 10.1186/s40001-025-03196-w]

[12] Mohajer-Bastami, A., et al., Artificial intelligence in healthcare: applications, challenges, and future directions. A narrative review informed by international, multidisciplinary expertise. *Frontiers in Digital Health,* 2025. 7(0): p. 1644041.[PMID: 41306935 PMCID: PMC12645148 DOI: 10.3389/fdgth.2025.1644041]

[13] Sharma A, Liu J, Shao S. Advances in AI and machine learning for predictive medicine. *Journal of Human Genetics*, 2024. 69(10): pp. 487-97.[PMID: 38424184 PMCID: PMC11422165 DOI: 10.1038/s10038-024-01231-y]

[14] Badawy, M., N. Ramadan and H.A. Hefny, Healthcare predictive analytics using machine learning and deep learning techniques: a survey. *Journal of Electrical Systems and Information Technology,* 2023. 10(1): p. 40.[DOI: 10.1186/s43067-023-00108-y]

[15] Li, M., et al., Medical image analysis using deep learning algorithms. *Frontiers in Public Health,* 2023. 11(0): p. 1273253.[PMID: 38026291 PMCID: PMC10662291 DOI: 10.3389/fpubh.2023.1273253]

[16] Rosenbacke, R., et al., How Explainable Artificial Intelligence Can Increase or Decrease Clinicians' Trust in AI Applications in Health Care: Systematic Review. *JMIR AI,* 2024. 3(0): p. e53207.[PMID: 39476365 PMCID: PMC11561425 DOI: 10.2196/53207]

[17] Ma Y, Luo Y, Wu Y, et al. An explainable predictive machine learning model of gangrenous cholecystitis based on clinical data: a retrospective single center study. *World Journal of Emergency Surgery,* 2025. 20(1): p. 1.[PMID: 39757162 PMCID: PMC11702024 DOI: 10.1186/s13017-024-00571-6]

[18] Hu, Y., Y. Chen and H. Zhao, Development and Validation of an Explainable Machine Learning Model for Gangrenous Cholecystitis Prediction: A Multicenter Retrospective Study. *Journal of Inflammation Research,* 2025. 18(0): pp. 17747-58.[PMID: 41439126 PMCID: PMC12719635 DOI: 10.2147/JIR.S552926]

[19] Guo Q, Liu Y, Huang Y, et al. Self-supervised learning model integrates plain and contrast-enhanced CT for preoperatively identifying gangrenous cholecystitis: a multicenter retrospective cohort study. *International Journal of Surgery,* 2025. 111(11): pp. 7622-32.[PMID: 40844296 PMCID: PMC12626600 DOI: 10.1097/JS9.0000000000003063]

[20] Liu NB, Li K, Wei YH, Xu Y, Chen QN. Risk factor analysis of acute rapidly progressive gangrenous cholecystitis. *Abdominal Surgery,* 2021. 2(34): pp. 149-53.

[21] Zheng, J., et al., The Risk Factor Analysis of Gallbladder Gangrene in Acute Acalculous Cholecystitis: A Single-Center Retrospective Study. *Gastroenterology Research and Practice,* 2025. 2025(0): p. 8696872.[PMID: 41163918 PMCID: PMC12566959 DOI: 10.1155/grp/8696872]

[22] Yang CX, Wang LL, Pan GZ, You FP, Yuan QZ. Risk factors for gangrenous cholecystitis: a meta-analysis. *Modern Medicine,* 2024. 1(52): pp. 9-17.

[23] Saragò, M., et al., Acute acalculous cholecystitis and cardiovascular disease, which came first? After two hundred years still the classic chicken and eggs debate: A review of

literature. *Annals of Medicine & Surgery,* 2022. 78(0): p. 103668.[PMID: 35734727 PMCID: PMC9206910 DOI: 10.1016/j.amsu.2022.103668]

[24] Botezatu, C., et al., Cholecystectomy in the Context of Cirrhosis, Sclero-Atrophic Cholecystitis, and Gangrenous Cholecystitis: A Literature Review. *Medicina (Kaunas),* 2025. 61(8): p. 1314.[PMID: 40870359 PMCID: PMC12387634 DOI: 10.3390/medicina61081314]

[25] Hou Q, Liu H, Chen L, et al. Procalcitonin and CRP as biomarkers for diagnosis and severity assessment of acute cholecystitis. *Open Medicine (Warsaw),* 2025. 20(1): p. 20251258.[PMID: 40989157 PMCID: PMC12452066 DOI: 10.1515/med-2025-1258]

[26] Feier, C.V.I., et al., Diagnostic and Prognostic Value of AISI, SII, and SIRI in Predicting Gangrenous Evolution of Acute Lithiasic Cholecystitis. *Diagnostics (Basel),* 2026. 16(3): p. 441.[PMID: 41681759 PMCID: PMC12896793 DOI: 10.3390/diagnostics16030441]

[27] Wang S, Zhang D, Zhou W, et al. High serum CA19-9 predicts severe cholecystitis in calculous cholecystitis patients. *BMC Gastroenterology,* 2025. 25(1): p. 33.[PMID: 39849359 PMCID: PMC11756138 DOI: 10.1186/s12876-025-03616-5]

[28] Satoh, K., et al., Practical approach to thrombocytopenia in patients with sepsis: a narrative review. *Thrombosis Journal,* 2024. 22(1): p. 67.[PMID: 39039520 PMCID: PMC11265094 DOI: 10.1186/s12959-024-00637-0]

[29] Corr, P., Sonography of gangrenous cholecystitis. *Journal of Emergencies, Trauma, and Shock,* 2012. 5(1): pp. 82-3.[DOI: 10.4103/0974-2700.93112]

[30] Sureka, B., et al., Gangrenous cholecystitis: Analysis of imaging findings in histopathologically confirmed cases. *The Indian Journal of Radiology & Imaging,* 2018. 28(1): pp. 49-54.[PMID: 29692527 PMCID: PMC5894319 DOI: 10.4103/ijri.IJRI_421_16]

[31] Kawai, R., et al., Contrast-enhanced ultrasonography with Sonazoid for diagnosis of gangrenous cholecystitis. *Journal of Medical Ultrasonics,* 2016. 43(2): pp. 193-9.[PMID: 27033865 DOI: 10.1007/s10396-015-0677-6]

[32] Bennett, G.L., et al., CT findings in acute gangrenous cholecystitis. *AJR. American Journal of Roentgenology,* 2002. 178(2): pp. 275-81.[PMID: 11804880 DOI: 10.2214/ajr.178.2.1780275]

[33] Wu B, Tian J, Bin W, et al. Predicting gangrenous cholecystitis. *HPB (Oxford),* 2014. 16(9): pp. 801-6.[PMID: 24635779 PMCID: PMC4159452 DOI: 10.1111/hpb.12226]

[34] Yacoub, W.N., et al., Prediction of patients with acute cholecystitis requiring emergent cholecystectomy: a simple score. *Gastroenterology Research and Practice,* 2010. 2010(0): p. 901739.[PMID: 20631896 PMCID: PMC2902011 DOI: 10.1155/2010/901739]

[35] Kim C, Um S, Gwon S, et al. Transparency of medical artificial intelligence systems. *Nature Reviews Bioengineering,* 2026. 4(1): pp. 11-29.[PMID: 42027791 PMCID: PMC13102313 DOI: 10.1038/s44222-025-00363-w]

[36] Ribeiro MT, Singh S, Guestrin C. Why Should I Trust You: Explaining the Predictions of Any Classifier. *Proceedings of the 22nd ACM SIGKDD (KDD 2016),* 2016: pp. 1115-24.[DOI: 10.48550/arXiv.1602.04938]

[37] Cicerone, O., et al., Development and validation of machine learning tools for predicting postoperative complications in acute calculous cholecystitis. *Updates in Surgery,* 2025.[DOI: 10.1007/s13304-025-02476-8]

[38] Das, P., et al., Gallstone disease classification using SLOA-optimized CatBoost classifier with explainable AI. *PLoS ONE,* 2026. 21(6): p. e0342945.[DOI: 10.1371/journal.pone.0342945]

[39] Jayanthi, S., et al., Gallbladder disease diagnosis from ultrasound using squeeze-and-excitation capsule network with convolutional bidirectional long short-term memory. *Scientific Reports,* 2026. 16(1): p. 3147.[DOI: 10.1038/s41598-025-32978-9]

[40] Gupta, P., et al., Computed tomography texture-based radiomics analysis in gallbladder cancer: initial experience. *Clinical and Experimental Hepatology,* 2021. 7(4): pp. 406-14.[PMID: 35402717 PMCID: PMC8977886 DOI: 10.5114/ceh.2021.111173]

[41] Chen, G., et al., Explainable machine learning model integrating clinical and radiomic features for predicting acute suppurative cholecystitis. *World Journal of Gastroenterology,* 2026. 32(11).

[42] Ward, T.M., et al., Artificial intelligence prediction of cholecystectomy operative course from automated identification of gallbladder inflammation. *Surgical Endoscopy,* 2022. 36(9): pp. 6832-40.[PMID: 35031869 DOI: 10.1007/s00464-022-09009-z]

[43] Hur, S., et al., Comparison of SHAP and clinician friendly explanations reveals effects on clinical decision behaviour. *NPJ Digital Medicine,* 2025. 8(1): p. 578.[PMID: 41006498 PMCID: PMC12475050 DOI: 10.1038/s41746-025-01958-8]

[44] Sophie Isabelle, L.M.M.S., An integrative review on the acceptance of artificial intelligence among healthcare professionals in hospitals. *NPJ Digital Medicine,* 2023. 1(6): p. 111.[PMID: 37301946 PMCID: PMC10257646 DOI: 10.1038/s41746-023-00852-5]