Journal of Inflammation Research

Dovepress
Taylor & Francis Group

Open Access Full Text Article

ORIGINAL RESEARCH

Development and Validation of an Explainable Machine Learning Model for Gangrenous Cholecystitis Prediction: A Multicenter Retrospective Study

Yilong Hu[1,*], Yunfeng Chen[2,*], Hailiang Zhao[3]

[1]Department of General Surgery of the International Medical Center, The Fourth Affiliated Hospital of Soochow University, Suzhou, Jiangsu, People's Republic of China; [2]Department of Ultrasound, Nanjing Gaochun People's Hospital, Nanjing, Jiangsu, People's Republic of China; [3]Department of Gastroenterology, Affiliated Hospital of Youjiang Medical University for Nationalities, Baise, Guangxi, People's Republic of China

*These authors contributed equally to this work

Correspondence: Hailiang Zhao, Department of Gastroenterology, Affiliated Hospital of Youjiang Medical University for Nationalities, No. 18, Zhongshan 2nd Road, Youjiang District, Baise City, Guangxi, 533000, People's Republic of China, Email 1120060639@163.com

Journal of Inflammation Research downloaded from https://www.dovepress.com/
For personal use only.

Purpose: To develop and externally validate an interpretable machine learning model for preoperative prediction of Gangrenous cholecystitis (GC) using multicenter clinical data.

Patients and Methods: This retrospective multicenter study included 744 patients with cholecystitis who underwent cholecystectomy at one institution, split into training (n=521) and testing (n=223) cohorts, and a temporal external validation cohort of 300 patients from a second center. Twenty preoperative variables were screened by LASSO regression and Boruta algorithm; predictors selected by both were used to construct six machine learning models. Model performance was assessed via AUC, calibration, and decision curve analysis. SHAP analysis provided model interpretability.

Results: The Random Forest (RF) model demonstrated superior predictive performance, achieving an AUC of 0.893 in the training set, 0.875 in the testing set, and 0.818 in external validation. Calibration and decision curve analyses indicated excellent agreement and clinical benefit. SHAP analysis identified gallbladder wall thickening, C-reactive protein, pericholecystic fluid, white blood cell count, and impacted stone as the most influential predictors, ensuring transparency of model decisions.

Conclusion: In our multicenter cohorts, this interpretable machine learning model showed good discrimination for preoperative risk stratification of gangrenous cholecystitis and acceptable generalizability between centers. By integrating clinical, laboratory, and imaging features and providing explainability, the approach may assist perioperative decision-making when used alongside clinical judgment. Prospective, multicenter evaluations and clinical impact studies are warranted before routine clinical adoption.

Keywords: gangrenous cholecystitis, machine learning, risk prediction, model interpretability

Introduction

Gangrenous cholecystitis (GC) is a severe and life-threatening complication of acute cholecystitis, characterized by transmural necrosis of the gallbladder wall and associated with a substantially increased risk of morbidity and mortality.[1] GC often presents with non-specific clinical and laboratory findings, making early and accurate diagnosis challenging, particularly in its early phases.[2,3] Delays in diagnosis or misidentification of high-risk cases may result in rapid clinical deterioration, higher rates of complications such as perforation or sepsis, and increased lengths of hospital stay.[4,5]

Despite advances in imaging and laboratory medicine, the preoperative identification of GC remains suboptimal.[6] Traditional risk stratification models and clinical scoring systems offer limited sensitivity and specificity, often relying on a small number of variables without accounting for the complex, multifactorial pathophysiology underlying progression

Received: 10 July 2025
Accepted: 9 December 2025
Published: 17 December 2025

© 2025 Hu et al. This work is published and licensed by Dove Medical Press Limited. The full terms of this license are available at https://www.dovepress.com/terms.php and incorporate the Creative Commons Attribution – Non Commercial (unported, v4.0) License (http://creativecommons.org/licenses/by-nc/4.0/). By accessing the work you hereby accept the Terms. Non-commercial uses of the work are permitted without any further permission from Dove Medical Press Limited, provided the work is properly attributed. For permission for commercial use of this work, please see paragraphs 4.2 and 5 of our Terms (https://www.dovepress.com/terms.php).

from simple to gangrenous cholecystitis.[7–9] In addition, there remains wide variation in the utilization and interpretability of these risk models across different institutions and patient cohorts.[10]

In recent years, machine learning (ML) algorithms have demonstrated superior performance across a range of medical prediction tasks, including disease diagnosis, prognosis, and risk assessment. These models are capable of incorporating high-dimensional data and uncovering non-linear relationships among clinical variables which may not be apparent through traditional statistical approaches. Furthermore, the development of explainable artificial intelligence techniques has made it possible to interpret and visualize the contributions of individual features, thereby enhancing clinical acceptability and transparency of ML-driven predictions.

However, few studies have systematically applied interpretable machine learning approaches to predict the occurrence of GC in patients with acute cholecystitis, especially using large, multicenter cohorts with robust external validation. The identification of high-risk patients prior to surgery could aid in tailored perioperative management and timely intervention, potentially improving outcomes and resource allocation.

Therefore, the objective of this study was to develop and externally validate an explainable machine learning model for the prediction of gangrenous cholecystitis using retrospective, multicenter data. By integrating routinely available demographic, clinical, and laboratory variables, we aim to provide a reliable, interpretable tool for risk stratification of GC patients suitable for real-world clinical practice.

Materials and Methods

Study Design and Participants

In this multicenter, retrospective study, we initially reviewed the medical records of 965 patients with cholecystitis who underwent cholecystectomy at Nanjing Gaochun People's Hospital between January 1, 2023, and May 31, 2025. After applying exclusion criteria, we identified 744 patients eligible for primary analysis. These patients were randomly allocated in a 7:3 ratio to the Training cohort (n=521) and Testing cohort (n=223) using stratified random sampling based on outcome status to ensure proportional representation of gangrenous cholecystitis cases in both groups. For a temporal external validation of our machine learning model, we included an additional cohort of 300 patients meeting similar selection criteria from Nanjing Yimin Hospital, spanning January 1, 2020, to May 31, 2025. All surgical procedures were performed by experienced hepatobiliary surgeons with over 10 years of clinical expertise, consisting of either open cholecystectomy or laparoscopic cholecystectomy. This study was approved by the Ethics Committee of Nanjing Gaochun People's Hospital and conducted in accordance with the Declaration of Helsinki. Due to the retrospective nature of the study and the use of de-identified data, the requirement for informed consent was waived by the institutional review board.

Identification of Research Variables and Participants

We identified 20 preoperative clinical factors that may influence the development of gangrenous cholecystitis. These factors encompass essential patient characteristics including age (stratified as ≥70 years), sex, body mass index (BMI) and ASA classification. Clinical presentation variables evaluated comprise fever, heart rate and vomiting. Comorbidities examined include diabetes mellitus and hypertension. Preoperative laboratory parameters assessed consist of white blood cell count (WBC), C-reactive protein (CRP), total bilirubin, Gamma-Glutamyl transferase (GGT),D-dimer, fibrinogen and platelet count. Imaging characteristics considered were gallbladder wall thickening, pericholecystic fluid presence, impacted stone, and acalculous cholecystitis. All continuous laboratory values were measured using standardized clinical protocols. Quantitative variables were analyzed as continuous without arbitrary thresholds. Data standardization and quality control: Nearest preoperative laboratory values were compiled, converted to uniform units, and screened for plausibility; unresolved discrepancies were excluded. We used prespecified, binary definitions to reduce subjectivity. Gallbladder wall thickness is measured on right-upper-quadrant ultrasound as the maximal anterior wall thickness perpendicular to the wall; wall thickening is defined as ≥3 mm. Impacted stone is defined as an immobile calculus at the gallbladder neck or cystic duct that does not change position with patient repositioning and shows posterior acoustic shadowing on ultrasound; CT or MRCP findings consistent with a calcific or filled defect causing obstruction are

considered confirmatory when available.Narrative reports were coded with a short manual, with disagreements adjudicated by a senior radiologist.

Inclusion criteria were as follows: (1) patients diagnosed with acute cholecystitis or acute exacerbation of chronic cholecystitis at our institution and who received comprehensive clinical management within the same facility; (2) patients who underwent cholecystectomy; and (3) individuals with complete and accessible clinical data, including age, surgical records, and length of hospital stay.

Exclusion criteria included: (1) patients with a prior diagnosis of chronic cholecystitis presenting for elective surgical intervention; (2) individuals previously diagnosed with acute cholecystitis who had undergone percutaneous transhepatic gallbladder drainage (PTGBD) and were now admitted for elective laparoscopic cholecystectomy; (3) patients with concomitant acute biliary or pancreatic disorders, such as obstructive jaundice secondary to choledocholithiasis, acute cholangitis, or acute pancreatitis; (4) those undergoing additional surgical procedures, including choledochotomy and lithotripsy, choledochoscopic exploration and lithotripsy, biliary–intestinal anastomosis, appendectomy, or similar interventions; and (5) patients with incomplete or missing clinical data.

The pathological diagnosis of acute cholecystitis, with or without gangrenous transformation, was established by experienced pathologists based on examination of cholecystectomy specimens. Gangrenous cholecystitis was defined by the presence of extensive transmural necrosis of the gallbladder wall, characterized by prominent infiltration of neutrophils and mononuclear cells. Additional histopathological findings included mural infarction with associated hemorrhage as well as mucosal necrosis subsequently replaced by acute inflammatory infiltrates and granulation tissue.

Feature Selection

To address multicollinearity and identify robust predictors from the initial 20 variables, we implemented a dual-method feature selection approach. First, least absolute shrinkage and selection operator (LASSO) regression was applied with 10-fold cross-validation to optimize the regularization parameter (λ), minimizing binomial deviance (Figure 1A). This L1-penalized method shrinks non-predictive coefficients to zero. Second, we employed the Boruta algorithm, a Random Forest-based wrapper method that iteratively compares feature importance against synthetic "shadow" attributes to statistically confirm relevance (Figure 1B). Final variable inclusion required consensus between methods: Only features selected by both LASSO and Boruta were retained. This stringent intersection strategy prioritized generalizability while mitigating overfitting, yielding a refined subset for subsequent modeling.

Model Construction and Comparison

After feature selection, the common variables identified by both the LASSO regression and Boruta algorithm were utilized as input predictors for model development. Six distinct machine learning models,including Random Forest (RF), Support Vector Machine (SVM), Naive Bayes (NB), Extreme Gradient Boosting (XGB), logistic regression (LR), and Decision Tree (DT), were constructed to predict the risk of gangrenous cholecystitis (GC). In this study, we employed

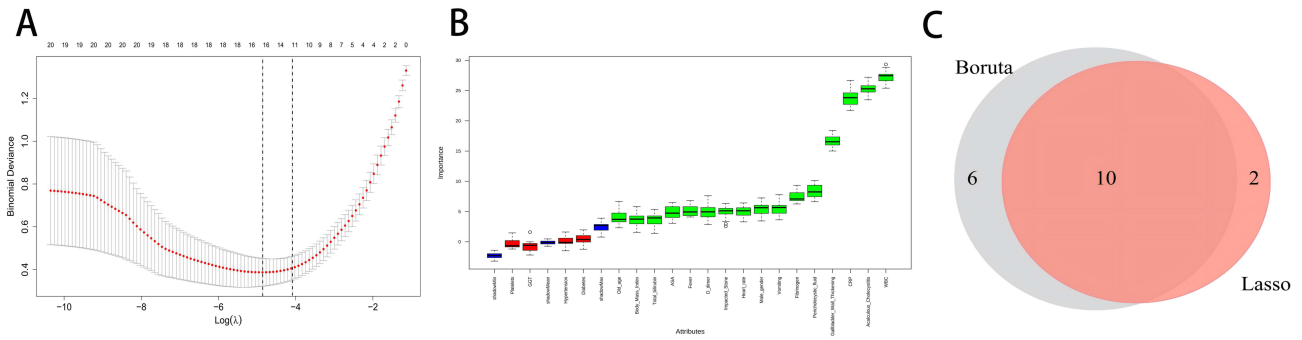

Figure 1 Features selection. (**A**) Boruta identified relevant features. (**B**) Selection of tuning parameters for the LASSO regression approach. (**C**) Venn diagram showing the overlap between predictors selected by LASSO and Boruta. Ten predictors were selected by both methods, six were unique to LASSO, and two were unique to Boruta. The final feature set was defined by the intersection.

https://doi.org/10.2147/JIR.S552926

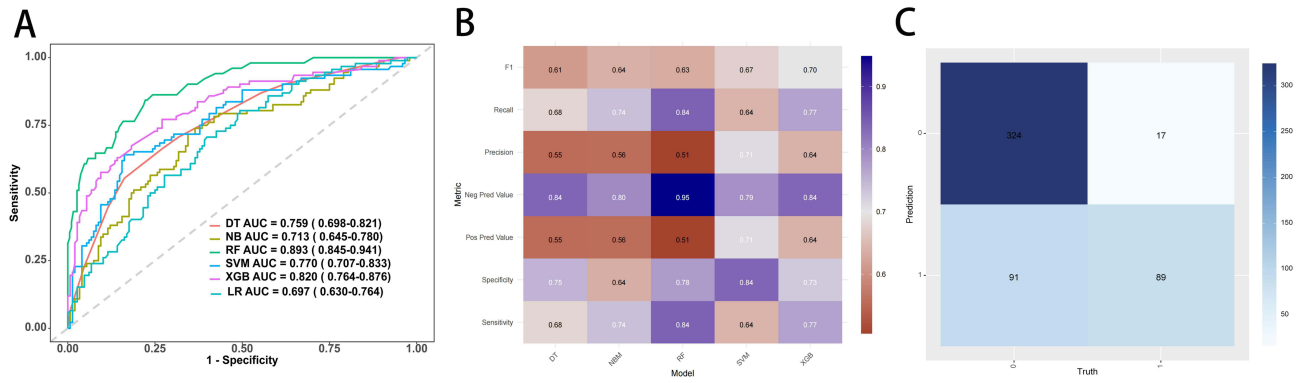

Figure 2 Performance metrics of models in the training cohort. (**A**) ROC curves and AUC values for the five model. (**B**) Comparative analysis metrics across all five models. (**C**) Confusion matrix of the RF model.

a 10-fold cross-validation methodology for model selection. Model performance in the training cohort was comprehensively evaluated using metrics such as the area under the receiver operating characteristic curve (AUC), accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) (Figure 2A and B).

Comparative analysis revealed that the RF model outperformed the other algorithms across all main metrics, with the highest AUC value (0.893), as well as superior specificity and sensitivity. The optimal cut-off point for the RF model was determined through Youden's index. The confusion matrix of the RF model illustrated robust discrimination in classifying patients within the training cohort (Figure 2C). Subsequently, the trained models were evaluated independently on the testing and external validation cohorts. To assess calibration fidelity and clinical utility, we performed calibration curves and decision curve analysis (DCA).

The SHAP to Model Interpretation

Interpreting machine learning models presents significant challenges due to their inherent complexity. The SHAP (Shapley Additive Explanations) method, grounded in game theory, addresses the "black box" nature of these models by quantitatively assessing and ranking the importance of input features in relation to model predictions. By calculating the contribution of each feature to individual and overall prediction outcomes, SHAP offers both local and global interpretability, thereby enhancing the transparency and comprehensibility of machine learning models.

Statistical Analyses

All statistical analyses and data visualizations were performed using R (version 4.4.2) and JD_DCPM (V6.11, Jingding Medical Technology Co., Ltd). Continuous variables were assessed for normality via the Shapiro–Wilk test. Normally distributed data are presented as mean ± standard deviation, with group comparisons conducted using Student's *t*-tests. Non-normally distributed variables are expressed as median and interquartile range [M (Q1, Q3)] and analyzed via the Mann–Whitney *U*-test. Categorical variables are reported as frequencies (percentages) and evaluated using Chi-square tests or Fisher's exact tests (for cell counts <5). Statistical significance was defined as a two-tailed p-value < 0.05.

Result

Patient Characteristics for Training, Testing, and External Validation Cohorts

A total of 1044 patients were enrolled, divided into the training (n=521), testing (n=223), and external validation (n=300) cohorts. Baseline demographic and clinical characteristics, including sex, age, BMI, hypertension, diabetes, heart rate, fever, vomiting, laboratory values (platelets, GGT, total bilirubin, WBC, CRP, D-dimer, fibrinogen), imaging features (impacted stone, gallbladder wall thickening, pericholecystic fluid, acalculous cholecystitis), and ASA classification, were well balanced across all groups (all p>0.05). The incidence of gangrenous cholecystitis was also similar (20.3% vs 19.7% vs 20.0%, p=0.98). No significant differences were observed among the cohorts, supporting the comparability and robustness of subsequent analyses (Table 1). Comparison of baseline characteristics between gangrenous and non-

Table 1 Baseline Characteristics of the Training, Testing, and External Validation Sets

Variables	Training Cohort N=521	Testing Cohort N=223	Validation Cohort N=300	P value
Male, n (%)	261 (50.1%)	109 (48.9)	152 (50.7)	0.842
Age (≥70), n (%)	105 (20.2)	49 (22.0)	58 (19.3)	0.682
BMI, mean (SD), kg/m2	24.5 ± 3.2	24.2 ± 3.3	24.7 ± 3.1	0.156
Hypertension, n (%)	102 (19.6)	48 (21.5)	63 (21.0)	0.784
Diabetes, n (%)	76 (14.6)	36 (16.1)	47 (15.7)	0.832
Heart rate(>90),n (%)	50 (9.6)	24 (10.8)	33 (11.0)	0.758
Platelets,median(IQR), 10^9/L	152 (125–183)	148 (121–178)	155 (127–185)	0.214
GGT,median(IQR),U/L	39 (30–51)	42 (32–55)	38 (29–49)	0.086
Total bilirubin,median(IQR),μmol/L	14 (10–21)	16 (12–23)	15 (11–20)	0.142
WBC,median(IQR),10^9/L	7.2 (5.2–9.3)	6.8 (4.9–8.7)	7.1 (5.3–9.1)	0.327
CRP,median(IQR),mg/L	5.3 (2.4–10.2)	4.7 (2.1–9.5)	5.5 (2.5–10.8)	0.184
D-dimer,median(IQR),ug/mL	1.05 (0.85–1.35)	1.02 (0.80–1.32)	1.08 (0.88–1.40)	0.314
Fibrinogen,median(IQR),ug/mL	3.05 (2.52–3.54)	2.98 (2.48–3.45)	3.12 (2.58–3.62)	0.198
Fever,n (%)	206 (39.5)	94 (42.2)	124 (41.3)	0.795
Vomiting,n (%)	51 (9.8)	25 (11.2)	29 (9.7)	0.766
Impacted stone,n (%)	28 (5.4)	12 (5.4)	14 (4.7)	0.903
Gallbladder wall thickening,n (%)	158 (30.3)	70 (31.4)	87 (29.0)	0.812
Pericholecystic fluid,n (%)	127 (24.4)	58 (26.0)	79 (26.3)	0.806
Acalculous cholecystitis,n (%)	12 (2.3)	3 (1.3)	5 (1.7)	0.654
ASA classification(>2),n (%)	132 (25.3)	54 (24.2)	76 (25.3)	0.965
GC,n (%)	106 (20.3)	44 (19.7)	60 (20.0)	0.98

Notes: Data are presented as n (%) for categorical variables and as mean ± SD or median (IQR) for continuous variables, as appropriate. The symbol (%) denotes the percentage of patients within the corresponding cohort.

Abbreviations: BMI, body mass index; GGT, Gamma-Glutamyl transferase; WBC, white blood cell count; CRP, C-reactive protein; ASA, American Society of Anesthesiologists; GC, Gangrenous cholecystitis.

gangrenous cholecystitis is summarized in Table 2. Patients with gangrenous cholecystitis had significantly higher inflammatory and coagulation indicators, including WBC, CRP, total bilirubin, fibrinogen, and D-dimer levels, as well as a higher incidence of gallbladder wall thickening, pericholecystic fluid, impacted stones, and acalculous cholecystitis.

Feature Selection

To identify the most informative predictors for model construction, we applied both the Boruta and LASSO feature selection methods. The Boruta algorithm identified 16 important variables, highlighting those most strongly associated with the outcome. In parallel, LASSO regression reduced the candidate set to 12 variables by penalizing less informative features. Notably, 10 variables were consistently selected by both approaches, namely: gallbladder wall thickening, C-reactive protein, pericholecystic fluid, white blood cell count, impacted stone, acalculous cholecystitis, fibrinogen, fever, D-dimer, and total bilirubin (Figure 1C). These overlapping features, representing clinical, laboratory, and imaging parameters, formed the foundation of subsequent model development to ensure both statistical rigor and clinical relevance.

Model Development and Performance

Following feature selection, Six machine learning models, including Random Forest (RF), Support Vector Machine (SVM), Naive Bayes (NB), Extreme Gradient Boosting (XGB), logistic regression (LR), and Decision Tree (DT), were constructed using the intersected predictors. As shown in Figure 2A, the RF model demonstrated the best overall performance in the training set, achieving the highest AUC (0.893). Notably, Figure 2B presents a comprehensive comparison of the sensitivity, specificity, and accuracy for each algorithm. The RF model consistently outperformed the others, with higher sensitivity and specificity values, indicating its ability to correctly identify both gangrenous and non-gangrenous cholecystitis cases with minimal misclassification. The balanced and outstanding results across these key metrics further highlight the robustness and practical potential of the RF model. Additionally, the confusion matrix in

Table 2 Baseline Comparison Between Gangrenous and Non-gangrenous Cholecystitis in the Overall Cohort

Variables	Gangrenous Cholecystitis (n = 210)	Non-Gangrenous Cholecystitis (n = 834)	P value
Male, n (%)	110 (52.4)	412 (49.4)	0.438
Age (≥ 70 years), n (%)	47 (22.4)	165 (19.8)	0.396
BMI, mean (SD), kg/m^2	24.7 ± 3.3	24.5 ± 3.2	0.481
Hypertension, n (%)	45 (21.4)	168 (20.1)	0.689
Diabetes, n (%)	36 (17.1)	123 (14.7)	0.379
Heart rate > 90, n (%)	25 (11.9)	82 (9.8)	0.384
Platelets, median (IQR), $10^9/L$	149 (122–178)	153 (126–183)	0.188
GGT, median (IQR), U/L	42 (32–58)	39 (30–50)	0.076
Total bilirubin, median (IQR), μmol/L	22 (15–30)	14 (10–20)	< 0.001
WBC, median (IQR), $10^9/L$	13.5 (11.0–16.2)	7.0 (5.2–9.1)	< 0.001
CRP, median (IQR), mg/L	18.5 (9.8–28.4)	4.8 (2.2–9.7)	< 0.001
D-dimer, median (IQR), μg/mL	1.65 (1.25–2.10)	1.02 (0.82–1.32)	< 0.001
Fibrinogen, median (IQR), μg/mL	3.85 (3.30–4.35)	2.98 (2.48–3.45)	< 0.001
Fever, n (%)	112 (53.3)	312 (37.4)	< 0.001
Vomiting, n (%)	25 (12.0)	80 (9.6)	0.281
Impacted stone, n (%)	18 (8.6)	36 (4.3)	0.011
Gallbladder wall thickening, n (%)	92 (43.8)	223 (26.7)	< 0.001
Pericholecystic fluid, n (%)	78 (37.1)	186 (22.3)	< 0.001
Acalculous cholecystitis, n (%)	10 (4.8)	10 (1.2)	0.002
ASA classification > 2, n (%)	56 (26.7)	206 (24.7)	0.535

Notes: Data are presented as n (%) for categorical variables and as mean ± SD or median (IQR) for continuous variables, as appropriate. The symbol (%) denotes the percentage of patients within the corresponding cohort.
Abbreviations: BMI, body mass index; GGT, Gamma-Glutamyl transferase; WBC, white blood cell count; CRP, C-reactive protein; ASA, American Society of Anesthesiologists.

Figure 2C further confirmed the high reliability of the RF model in patient classification. Collectively, these results indicate that the RF model is the most effective approach among those evaluated, providing a strong foundation for subsequent external validation and clinical application.

Model Performance on Both the Testing and External Validation Sets

The Random Forest (RF) model demonstrated robust performance in the testing cohort (n=223), achieving an AUC of 0.875 (95% CI: 0.827–0.922; Figure 3A). The calibration curve (Figure 3B) indicated near-perfect alignment between predicted probabilities and observed outcomes, while decision curve analysis (Figure 3C) confirmed significant clinical utility across relevant threshold probabilities. The confusion matrix (Figure 3D) further validated high classification accuracy. When externally validated on an independent cohort (n=300), the model maintained strong performance with an AUC of 0.818 (95% CI: 0.763–0.873; Figure 4A). Though minor calibration deviations appeared at extreme probabilities (Figure 4B), overall concordance remained clinically acceptable DCA (Figure 4C) demonstrated preserved net benefit across clinically meaningful threshold probabilities, and the confusion matrix (Figure 4D) showed consistent classification performance. The observed AUC decrease of 0.057 between cohorts likely reflects institutional heterogeneity but confirms the model's robustness for multicenter deployment. Both evaluations support the RF model's utility for gangrenous cholecystitis (GC) risk stratification, with testing cohort performance indicating optimal calibration and external validation confirming generalizability.

Model Interpretability

To ensure clinical transparency and robust interpretability, the Random Forest (RF) model's predictive behavior for gangrenous cholecystitis (GC) was explored using SHapley Additive exPlanations (SHAP) analysis (Figure 5). The mean absolute SHAP value ranking (Figure 5A) shows that the five most influential predictors are gallbladder wall thickening,

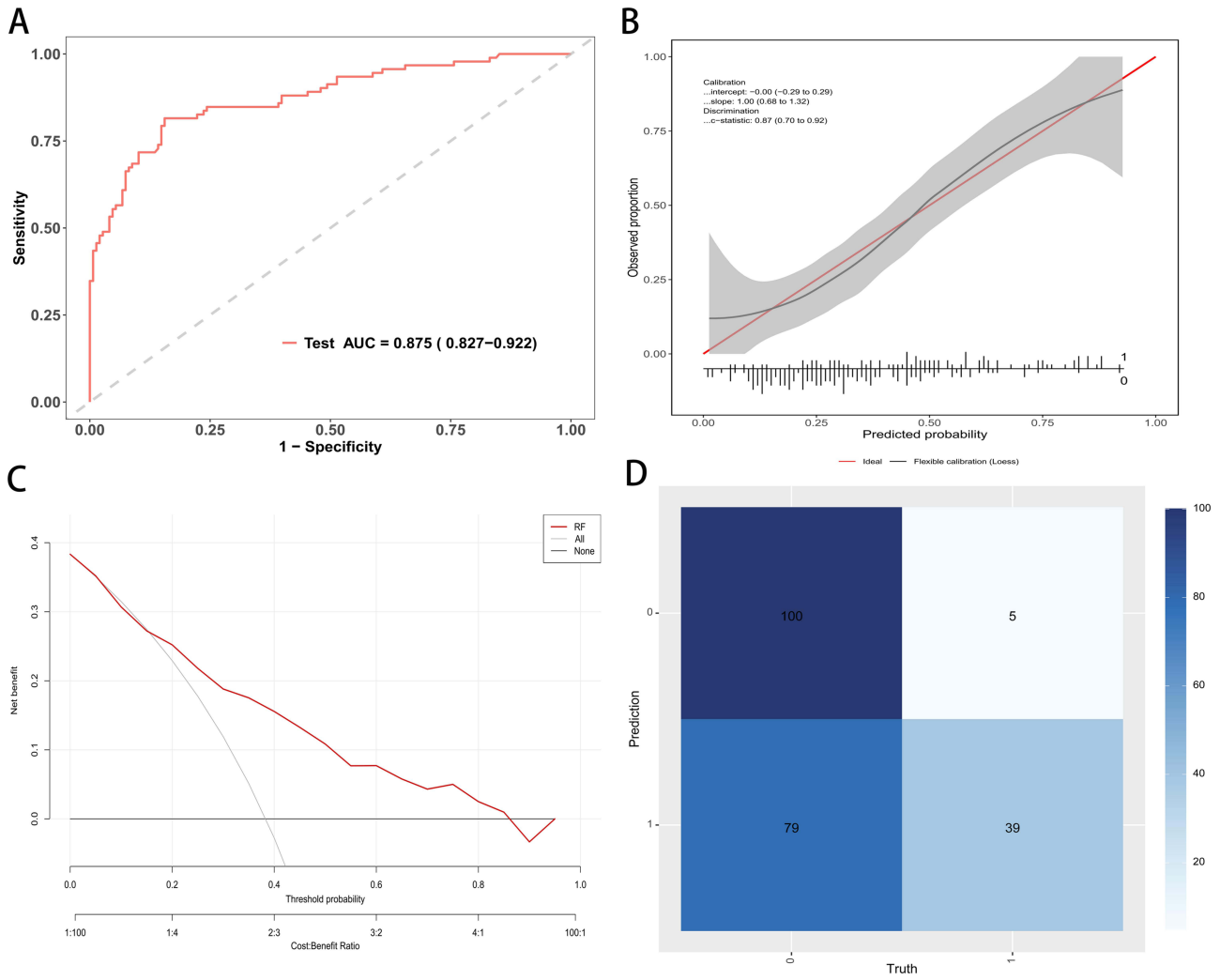

Figure 3 RF mode evaluation in testing cohort. (**A**) Calibration curve of the testing set. (**B**) Analysis of decision curves for the test cohort. (**C**) Confusion matrix of the RF model in testing set. (**D**) Testing set ROC curves and AUC values.

C-reactive protein (CRP), pericholecystic fluid, white blood cell (WBC) count, and the presence of an impacted stone. These features provided the greatest contributions to the model's overall risk stratification.

The SHAP summary plot (Figure 5B) illustrates how each key variable influences the predicted risk of gangrenous cholecystitis at the individual level. High values of gallbladder wall thickening and CRP markedly increase risk, while low values are protective. Pericholecystic fluid, WBC, and impacted stone also contribute to increased risk, though their effects can vary and show some non-linear patterns across patients.The plot also reveals that some variables have both high and low values associated with positive or negative SHAP values, reflecting complex and potentially interacting influences within the model.

Discussion

This study presents a rigorously developed and externally validated Random Forest (RF) model for preoperative prediction of gangrenous cholecystitis (GC), an acute and life-threatening complication of cholecystitis. Our model achieved high discriminative capability (AUC 0.875 in testing and 0.818 in external validation), together with sound calibration and broad clinical utility as demonstrated by decision curve analysis. Through the integration of routinely accessible clinical, laboratory, and imaging features, and a focus on model interpretability via SHAP values, we address

https://doi.org/10.2147/JIR.S552926

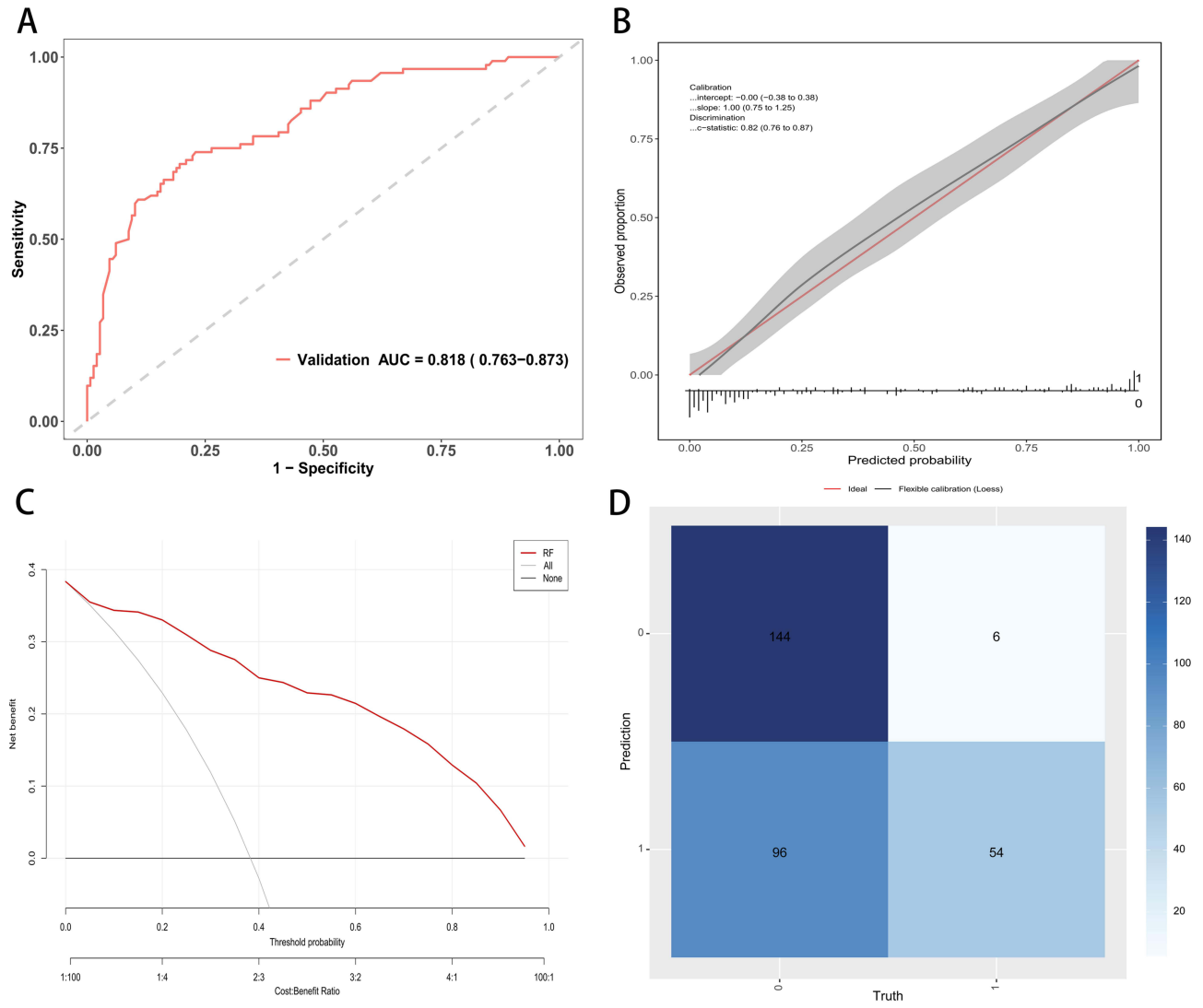

Figure 4 RF mode evaluation in the temporal external validation cohort. (**A**) Calibration curve of the external validation set. (**B**) Decision curve analysis of the external validation set. (**C**) Confusion matrix of the RF model in the external validation set (**D**) External validation set ROC curves and AUC values.

important limitations of previous GC prediction tools and provide a translational solution for contemporary acute care settings.

The urgent differentiation of GC from non-gangrenous acute cholecystitis is clinically challenging yet critically important, given the higher risks of necrosis, perforation, and sepsis, and the distinct management pathways.[11,12] Traditional clinical risk scores and single-parameter predictors, such as systemic inflammatory response syndrome (SIRS) criteria or isolated laboratory thresholds, have demonstrated variable accuracy and limited generalizability.[13–15] In contrast, ML-based methods can leverage complex, nonlinear interactions among features, which is particularly advantageous in the heterogeneous and multifactorial progression of GC. Our results underscore this added value: the superior AUC compared to historical data likely results from the RF model's capacity to combine subtle and otherwise sub-threshold clinical clues into a robust risk signature. Several prediction models and risk scores for GC have been proposed. These studies generally demonstrated good discrimination but were predominantly single-center, often based on smaller cohorts, and rarely included external validation or formal assessment of calibration and clinical net benefit. Our model builds on this literature by using only preoperative, routinely collected variables, validating performance in an external hospital cohort, and providing a comprehensive evaluation that includes discrimination, calibration plots, and decision-curve analysis.

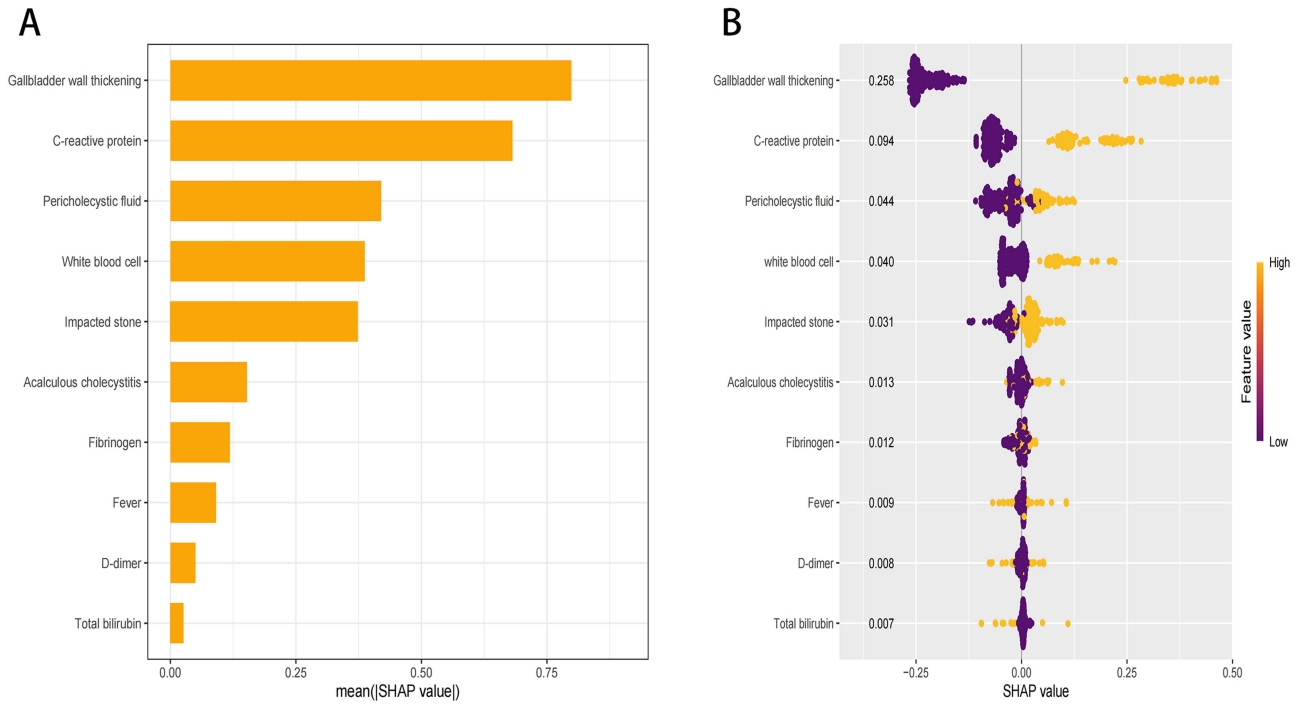

Figure 5 SHAP analysis of the RF model. (**A**) Mean absolute SHAP values corresponding to each clinical attribute. (**B**) SHAP values depicting the impact of various clinical characteristics on the model's output.

Importantly, our model highlights the clinical and pathophysiological relevance of its most predictive features: gallbladder wall thickening, CRP, pericholecystic fluid, WBC count, and the presence of an impacted stone. Each of these variables reflects established mechanisms of GC progression. For instance, gallbladder wall thickening and pericholecystic fluid are hallmarks of advanced local inflammation and impending transmural necrosis.[16–19] CRP and WBC elevations signal systemic inflammatory activation, reflecting the transition from localized to gangrenous disease.[20–22] The identification of impacted stone as a key contributor echoes the importance of obstructive pathophysiology in precipitating ischemic injury to the gallbladder wall.[23,24] Notably, the model's nuanced handling of these features, such as its ability to demonstrate nonlinear thresholds and variable risk increments depending on the context, makes it especially well suited to real-world patient presentations, which often do not fit classic patterns. From a methodological perspective, our modeling strategy was designed to limit overfitting while retaining clinically meaningful predictors. We combined two complementary feature-selection procedures: LASSO, which applies coefficient shrinkage and embedded variable selection within a penalized regression framework, and Boruta, a wrapper algorithm that uses repeated resampling and comparison with permuted "shadow" variables to identify robust predictors. Concordance between LASSO and Boruta helped us prioritize variables that were consistently informative across resamples and centers, thereby reducing the risk that the final model would rely on spurious or site-specific signals.

The final model also demonstrated good interpretability through SHAP analysis. The top contributors, such as gallbladder wall thickening, C-reactive protein (CRP), and pericholecystic fluid, are pathophysiologically plausible and readily accessible in routine practice. Gallbladder wall thickening and pericholecystic fluid reflect local edema, ischemia, and perivesicular inflammation, which are hallmarks of progression toward gangrene, while elevated CRP captures the systemic inflammatory burden. The directions of effect indicated by SHAP values align with clinical experience and existing literature, which supports the face validity of the model. By linking model output to familiar clinical and imaging features, these explanations may facilitate clinician acceptance and integration of the tool into decision-making.

Model interpretability, achieved through comprehensive SHAP analysis, is a unique strength of this study. Beyond ranking feature importance, SHAP summary plots characterize the direction, magnitude, and nonlinear nature of predictor effects at both the global and patient-specific levels. This is crucial for clinical translation: surgeons require

not only actionable risk stratification but also confidence in how a model arrives at each risk estimate.[25,26] The detailed interpretability framework allows physicians to audit, understand, and trust model outputs, supporting responsible clinical decision-making in the high-pressure environment of emergency surgery.[27] By demonstrating that key model drivers are consistent with current pathophysiological understanding, we mitigate one of the major obstacles to AI adoption in healthcare—the "black box" problem. To facilitate bedside use, the model will output a risk band together with a plain-language explanation of the top contributors, such as the presence of gallbladder wall thickening, an elevated C-reactive protein level, or pericholecystic fluid. High-risk predictions trigger predefined steps: expedited surgery, allocation to an experienced surgeon or team, enhanced preoperative preparation (advanced anesthesia monitoring, blood availability), and postoperative intensive care unit (ICU) planning; discordant cases prompt targeted imaging review. Thus, explainability functions as an auditable rationale for triage and resource allocation rather than an additional graph to interpret. We plan a prospective evaluation to determine final operating thresholds and quantify the impact on workflow and outcomes.

Moreover, the external validation on a geographically distinct cohort confirms that the RF model retains both accuracy and calibration in novel patient populations, a critical benchmark for robust prediction tools. The modest reduction in performance observed between internal and external sets is anticipated and reflects the inevitable heterogeneity among institutions, imaging protocols, and local practice patterns.[28,29] Nevertheless, the model's preservation of meaningful net clinical benefit emphasizes its potential for multicenter deployment and global impact.

Despite these strengths, several limitations of our study warrant discussion. In the external validation cohort there were approximately 60 GC events. For a 10 variable model, this event count limits statistical power and the precision of performance estimates, especially calibration and decision curve analyses. Although the EPV rule mainly applies to model development and was satisfied in our training cohort, the limited number of events at validation warrants cautious interpretation of the external results. Current guidance recommends external validations with at least 100 events and 100 non events to obtain precise estimates. We therefore plan larger, contemporaneous, multicenter validations and, if needed, model updating and recalibration. Although we employed multicenter external validation, all centers were located within the same national context, possibly restricting generalizability to international or community hospital populations. Future collaboration with additional regions and healthcare systems will be important. The retrospective design introduces risks of selection bias, incomplete data, and confounding issues inherent to all observational electronic health record studies. We attempted to mitigate these through rigorous data cleaning, standardized variable definitions, and inclusion of only preoperatively available features; however, residual confounding cannot be fully excluded. The model's performance in certain subgroups, such as elderly patients, those with atypical presentation, or severely delayed diagnosis, was not specifically analyzed and might merit separate evaluation. In addition, while SHAP analysis provides unprecedented transparency, the translation of model explanation into actionable bedside decisions remains an evolving field, and the risk of over-reliance on automated outputs must be balanced with clinical judgment. The external validation cohort was accrued in an earlier calendar period than the main cohort, producing a temporal external validation. This design better probes robustness to temporal dataset shift but may also limit generalizability to current practice, as secular changes in referral patterns, diagnostic protocols, and treatment thresholds can alter case mix and outcome prevalence. The modest reduction in AUC observed externally may partly reflect such shifts. Future work will include contemporaneous, prospective multicenter validation and periodic model updating/recalibration to mitigate temporal drift.

Future directions include broadening the scope of validation to international, community-based, and lower-resource hospitals; integrating additional biomarkers, advanced imaging features, or time-series trends to further refine risk estimation; and conducting prospective clinical impact studies to measure how model deployment affects triage, time to intervention, surgical outcomes, and resource utilization. Embedding the model within hospital electronic medical records with real-time data flows, and combining with natural language processing (NLP) for unstructured data capture, could further enhance utility. Ongoing recalibration and monitoring for model drift are essential if practice environments or patient demographics change.

Conclusion

Our multicenter retrospective study suggests that an interpretable machine-learning model can help sort patients by preoperative risk of gangrenous cholecystitis in our datasets. Used with clinical judgment, it may support timely, individualized decisions. Prospective, multicenter studies are still needed to confirm generalizability, define decision thresholds, and show any benefit for patient outcomes before routine adoption.

Acknowledgments

This study was generously supported by Jingding Medical Tech, to whom we extend our sincere gratitude. We especially thank them for providing authorization and technical support for the JD_DCPM software. The team at Jingding Medical Tech offered invaluable assistance in data processing.

Disclosure

The author(s) report no conflicts of interest in this work.

References

1. Ma Y, Luo M, Guan G. et al. An explainable predictive machine learning model of gangrenous cholecystitis based on clinical data: a retrospective single center study. *World J Emerg Surg*. 2025;20:1.
2. Roozbahani M, Shahmoradi M, Mehri J, Qolampoor A, Nasiri B, Pakmehr F. Risk factors of gangrenous cholecystitis in patients with acute cholecystitis: a cross-sectional study. *Int J Adv Biol Biomed Res*. 2019;7(3):267–274. doi:10.33945/sami/ijabbr.2019.3.7
3. Ares JÁ D, Martínez García R, Estellés Vidagany N, et al. Can inflammatory biomarkers help in the diagnosis and prognosis of gangrenous acute cholecystitis? A prospective study. *Rev Esp Enferm Dig*. 2021;113(1):41–44. doi:10.17235/reed.2020.7282/2020
4. Alghamdi K, Rizk H, Jamal W, et al. Risk factors of gangrenous cholecystitis in general surgery patient admitted for cholecystectomy in King Abdul-Aziz University Hospital (KAUH), Saudi Arabia. *Mater Sociomed*. 2019;31(4):286–289. doi:10.5455/msm.2019.31.286-289
5. Khan L, Zaidan F, Masmali A, et al. Gangrenous cholecystitis: a grim complication of acute cholecystitis. *SAS J Med*. 2018;4(8):116–118. doi:10.36347/sasjm.2018.v04i08.002
6. Patel R, Tse JR, Shen L, et al. Improving diagnosis of acute cholecystitis with US: new paradigms. *RadioGraphics*. 2024;44:e240032.
7. Bozada-Gutiérrez K, Trejo-ávila M, Chávez-Hernández F, et al. Surgical treatment of acute cholecystitis in patients with confirmed COVID-19: ten case reports and review of literature. *World J Clin Cases*. 2022;10(4):1296–1310. doi:10.12998/wjcc.v10.i4.1296
8. Shirah B, Shirah H, Saleem M, Chughtai M, Elraghi M, Shams M. Predictive factors for gangrene complication in acute calculous cholecystitis. *Ann Hepatobiliary Pancreat Surg*. 2019;23(3):228. doi:10.14701/ahbps.2019.23.3.228
9. Portinari M, Scagliarini M, Valpiani G, et al. Do I need to operate on that in the middle of the night? Development of a nomogram for the diagnosis of severe acute cholecystitis. *J Gastrointest Surg*. 2018;22(6):1016–1025. doi:10.1007/s11605-018-3708-y
10. Kim HY, Lee JH, Kim SG, et al. Ultrasonographic predictors of acute gangrenous cholecystitis in patients treated with laparoscopic cholecystectomy: a single center retrospective study. *Scand J Gastroenterol*. 2025;60:174–183.
11. Marinova P. Predictors for gangrene and perforation of gallbladder wall in patients with acute cholecystitis. *J Biomed Clin Res*. 2023;16 (2):146–152. doi:10.2478/jbcr-2023-0020
12. Milone M, Vertaldi S, Bracale U, et al. Robotic cholecystectomy for acute cholecystitis. *Medicine*. 2019;98(30):e16010. doi:10.1097/md.0000000000016010
13. Alam W, Karam K. Gangrenous cholecystitis as a potential complication of COVID-19: a case report. *Clin Med Insights Case Rep*. 2021;14:11795476211042459. doi:10.1177/11795476211042459
14. Sureka B, Rastogi A, Mukund A, Thapar S, Bhadoria A, Chattopadhyay T. Gangrenous cholecystitis: analysis of imaging findings in histopathologically confirmed cases. *Indian J Radiol Imaging*. 2018;28(1):49–54. doi:10.4103/ijri.ijri_421_16
15. Gojayev A, Karakaya E, Erkent M, et al. A novel approach to distinguish complicated and non-complicated acute cholecystitis: decision tree method. *Medicine*. 2023;102(19):e33749. doi:10.1097/md.0000000000033749
16. Gupta P, Dutta U, Rana P, et al. Gallbladder reporting and data system (GB-RADS) for risk stratification of gallbladder wall thickening on ultrasonography: an international expert consensus. *Abdom Radiol*. 2021;47(2):554–565. doi:10.1007/s00261-021-03360-w
17. Wang X, Zhang H, Bai Z, Xie X, Feng Y. Current status of artificial intelligence analysis for the diagnosis of gallbladder diseases using ultrasonography: a scoping review. *Transl Gastroenterol Hepatol*. 2025;10:12. doi:10.21037/tgh-24-61
18. Pruthi H, Chabbra M, Soundararajan R, et al. Role of dual energy computed tomography in evaluation of suspected wall thickening type of gallbladder cancer. *Clin Exp Hepatol*. 2022;8(1):92–95. doi:10.5114/ceh.2022.114188
19. Singh T, Gupta P. Role of dual-energy computed tomography in gallbladder disease: a review. *J Gastrointest Abdom Radiol*. 2022;5(2):107–113. doi:10.1055/s-0042-1743173
20. Chen J, Gao Q, Huang X, Wang Y. Prognostic clinical indexes for prediction of acute gangrenous cholecystitis and acute purulent cholecystitis. *BMC Gastroenterol*. 2022;22(1):491. doi:10.1186/s12876-022-02582-6
21. Chen L, Chen X. The role of different systemic inflammatory indexes derived from complete blood count in differentiating acute from chronic calculus cholecystitis and predicting its severity. *J Inflamm Res*. 2024;17:2051–2062. doi:10.2147/jir.s453146
22. Şakalar Ş, Özakın E, Çevik AA, et al. Plasma procalcitonin is useful for predicting the severity of acute cholecystitis. *Emerg Med Int*. 2020;2020:8329310. doi:10.1155/2020/8329310
23. Mehrzad M, Jehle C, Roussel L, Mehrzad R. Gangrenous cholecystitis: a silent but potential fatal disease in patients with diabetic neuropathy. A case report. *World J Clin Cases*. 2018;6(15):1007–1011. doi:10.12998/wjcc.v6.i15.1007

24. Liu Y, Xue D, Peng Y. The value of ultrasonography in predicting acute gangrenous cholecystitis. *Curr Med Imaging*. 2022;18(12):1257–1260. doi:10.2174/1573405618666220321124627
25. Gomes CA, Soares C, Di Saverio S, et al. Gangrenous cholecystitis in male patients: a study of prevalence and predictive risk factors. *Ann Hepatobiliary Pancreat Surg*. 2019;23(1):34–40. doi:10.14701/ahbps.2019.23.1.34
26. Gong Q, Chen X, Liu F, Cao Y. Machine learning-based integration develops a neutrophil-derived signature for improving outcomes in hepatocellular carcinoma. *Front Immunol*. 2023;14:1216585. doi:10.3389/fimmu.2023.1216585
27. K V, Pushpaketu N, Badhai S, et al. Risk factors associated with gangrenous cholecystitis: a cohort study from Eastern India. *Cureus*. 2024;16(11): e74126. doi:10.7759/cureus.74126
28. Wu B, Buddensick TJ, Ferdosi H, et al. Predicting gangrenous cholecystitis. *HPB*. 2014;16(9):801–806. doi:10.1111/hpb.12226
29. Sezikli İ, Tutan MB, Turhan VB, Özkan M, Topçu R. Medical management or surgery for acute cholecystitis: enhancing treatment selection with decision trees. *Turk J Trauma Emerg Surg*. 2024;30(7):883–891. doi:10.14744/tjtes.2024.64796

Journal of Inflammation Research

Dovepress
Taylor & Francis Group

Publish your work in this journal

The Journal of Inflammation Research is an international, peer-reviewed open-access journal that welcomes laboratory and clinical findings on the molecular basis, cell biology and pharmacology of inflammation including original research, reviews, symposium reports, hypothesis formation and commentaries on: acute/chronic inflammation; mediators of inflammation; cellular processes; molecular mechanisms; pharmacology and novel anti-inflammatory drugs; clinical conditions involving inflammation. The manuscript management system is completely online and includes a very quick and fair peer-review system. Visit http://www.dovepress.com/testimonials.php to read real quotes from published authors.

Submit your manuscript here: https://www.dovepress.com/journal-of-inflammation-research-journal