Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

# Development of a Deep Learning Model Integrating Computed Tomography Images and Blood Data for the Diagnosis of Acute Cholecystitis

**Takeshi Horaguchi**
Gifu University Graduate School of Medicine
**Ryosuke Nomura**
National Cancer Center
**Shunsuke A Sakai**
National Cancer Center
**Nao Saito**
National Cancer Center
**Kyoko Kurihara**
National Cancer Center
**Masaki Ohira**
National Cancer Center
**Ritsuki Takaha**
Gifu University Graduate School of Medicine
**Noriki Mitsui**
Gifu University Graduate School of Medicine
**Yuji Hatanaka**
Gifu University Graduate School of Medicine
**Ryoma Yokoi**
Gifu University Graduate School of Medicine
**Shoma Nagata**
Gifu University School of Medicine
**Masashi Kuno**
Central Japan International Medical Center
**Hirokatsu Hayashi**
Central Japan International Medical Center
**Masahiro Fukada**
Gifu University Graduate School of Medicine
**Yuta Sato**
Gifu University Graduate School of Medicine
**Itaru Yasufuku**

Gifu University Graduate School of Medicine
**Ryuichi Asai**
Gifu University Graduate School of Medicine
**Hideaki Bando**
National Cancer Center Hospital East
**Riu Yamashita**
National Cancer Center
**Nobuhisa Matsuhashi**
nobuhisa517@hotmail.com

Gifu University Graduate School of Medicine

---

**Research Article**

**Keywords:** acute cholecystitis, artificial intelligence, computed tomography, deep learning, laboratory data

**Posted Date:** November 1st, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-10986715/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

**Additional Declarations:** No competing interests reported.

---

## Abstract

# Background

In this study, we aimed to develop and evaluate an artificial intelligence-based diagnostic model for the diagnosis of acute cholecystitis (AC) using non-contrast CT images and clinical data.

# Methods

This retrospective study included 197 patients (100 AC, 97 non-AC) treated between January 2016 and December 2025 at a single center. Patients were randomly divided into training (n = 139) and test (n = 58) datasets. Three models were constructed: an imaging-based deep learning model, a clinical data-based machine learning model, and a hybrid machine learning model integrating deep learning-derived imaging features with clinical data. CT images were preprocessed, and gallbladder regions were segmented. Clinical variables included white blood cell counts and levels of C-reactive protein and liver function markers. Model performance was evaluated using accuracy, precision, recall, specificity, F1 score, and area under the receiver operating characteristic curve (AUC). Statistical comparisons were performed using Welch's t-test and Chi-square test.

# Results

The imaging-based model achieved accuracy 0.741, precision 0.703, recall 0.867, specificity 0.607, F1 score 0.776, and AUC 0.863. The blood-based model achieved accuracy 0.914, precision 0.931, recall 0.900, specificity 0.929, F1 score 0.915, and AUC 0.946. The hybrid model showed the highest performance, with accuracy 0.931, precision 0.882, recall 1.000, specificity 0.857, F1 score 0.938, and AUC 0.996.

# Conclusions

A hybrid model integrating CT imaging and clinical data improved diagnostic performance for AC compared with single-modality models.

## Background

Acute cholecystitis (AC) is a leading cause of acute abdomen in emergency departments, accounting for 3–10% of patients presenting with abdominal pain [1,2]. Rapid progression can lead to severe complications, such as gangrenous cholecystitis, perforation, abscess formation, and sepsis, necessitating early diagnosis and immediate therapeutic decision-making [3]. The Tokyo Guidelines 2018 provide a widely accepted framework for diagnosis and severity grading by integrating local signs,

systemic inflammation, and imaging features [1]. Although the Tokyo Guidelines demonstrate high diagnostic performance, their application still requires appropriate assessment and interpretation of these findings, particularly imaging findings. In emergency care settings, patients with AC are often initially evaluated by junior physicians, emergency physicians, or clinicians without specialized expertise in hepatobiliary disease. In such situations, particularly when immediate consultation with a radiologist or hepatobiliary specialist is not available, objective support for integrating clinical, laboratory, and imaging findings may facilitate initial assessment and triage.

Ultrasonography is an established first-line imaging modality for patients with suspected acute cholecystitis because of its non-invasive nature, accessibility, and real-time utility [4]. However, its diagnostic performance may be affected by examiner experience and technical factors, such as patient body habitus and overlying bowel gas [5]. CT provides complementary and comprehensive evaluation, particularly when ultrasonography is inconclusive or technically limited, when complications are suspected, or when alternative causes of acute abdominal pain need to be assessed [6]. Recent studies have indicated that CT has diagnostic performance comparable to that of ultrasonography, with a sensitivity and specificity of 83.9% and 94.0%, respectively, compared to 79.0% and 93.6% for ultrasonography [7].

Although non-contrast CT is not the preferred initial imaging modality for patients with a specific clinical suspicion of AC, it may already have been obtained during the evaluation of nonspecific abdominal pain or may be selected when contrast-enhanced CT is unsuitable. Non-contrast CT can provide valuable information regarding gallbladder morphology and pericholecystic changes. Because some imaging findings may be subtle or overlap with noninflammatory conditions, integration with routinely available laboratory data may enhance the diagnostic information obtained from non-contrast CT [6,8].

Recent advances in artificial intelligence, particularly with deep learning-based models, have shown significant potential in medical image analysis by extracting subtle features that are not easily detectable by humans [9–13]. In acute abdominal imaging, artificial intelligence, including deep learning approaches, can improve diagnostic accuracy, reduce interpretation time, and support triage [14]. Deep learning-based models using ultrasonographic images have also been reported for the diagnosis of acute cholecystitis, suggesting the potential of artificial intelligence-assisted imaging assessment in this clinical setting [15]. CT provides quantitative attenuation information and cross-sectional anatomical data, making it another promising modality for artificial intelligence-based image analysis [16,17]. In particular, recent approaches that decompose Hounsfield units (HU) into three RGB channels for model input have been shown to improve the predictive performance of CT-based analyses [18]. Furthermore, the clinical diagnosis of AC routinely incorporates laboratory data, such as elevated levels of inflammatory markers, suggesting that a multimodal approach integrating imaging with clinical metadata may further enhance diagnostic performance.

Because the imaging component of the Tokyo Guidelines may require expertise that is not always available at the point of initial assessment, we asked whether an AI-based approach could reproduce guideline-based diagnostic classifications using routinely available inputs without requiring expert image interpretation by non-specialists. We developed and evaluated three models for diagnosing AC: an image-based model using non-contrast CT, a blood test-based model using routinely available laboratory data, and a hybrid model integrating deep learning-derived imaging features with blood test data. Using diagnoses established by specialists in accordance with the Tokyo Guidelines 2018 and supported by surgical pathology and/or clinical follow-up as the reference standard, we evaluated the diagnostic performance and interpretability of these models and further examined a sensitivity-oriented operating point for supporting initial assessment and triage in emergency care settings where specialist consultation is not immediately available.

## Methods

# Study Population

Patients were identified through a retrospective review of consecutive clinical records. The exclusion criteria included incomplete clinical/laboratory data or suboptimal CT image quality (e.g., severe motion artifacts). A total of 197 patients were enrolled, comprising 100 patients with AC and 97 control participants without AC. The underlying diagnoses of the 97 non-AC control patients are summarized in Supplementary Table S2 (Additional file 2). The diagnosis of AC was established in accordance with the Tokyo Guidelines 2018 criteria, integrating clinical symptoms, laboratory markers, and imaging findings. Surgical pathology and/or clinical follow-up served as the reference standards. All CT images had corresponding radiology reports. For each patient assigned to the training dataset, three reviewers—a board-certified radiologist (SN, with 13 years of clinical experience) and two board-certified gastrointestinal surgeons (RT, with 8 years of clinical experience, and TH, with 13 years of clinical experience)—independently selected a single representative axial CT slice. When the selected slices differed among the reviewers, the final slice was determined by consensus. The representative image was defined as the slice considered most morphologically informative for gallbladder evaluation. In most cases, this corresponded to the slice showing the largest gallbladder diameter. In non-AC cases, the slice showing the largest visible gallbladder was similarly selected. Biochemical data were extracted from electronic medical records at the time of CT imaging.

## Data Preprocessing and Image Transformation

DICOM images were processed using the pydicom library (version 3.0.1). Raw pixel values were converted to HU using rescale slope and intercept parameters. To enhance soft-tissue contrast, images were windowed to a range of − 70 to 170 HU, followed by normalization to 8-bit grayscale.

To leverage the transfer learning capabilities of models pretrained on natural images, we generated three-channel pseudo-color RGB images by applying three distinct HU window settings (18): Red (R), −70 to 50 HU; Green (G), − 10 to 110 HU; and Blue (B), 50 to 170 HU. The resulting images were saved in PNG format for model input.

White saturation (blowout) artifacts were corrected using a rule-based image processing approach implemented with OpenCV (version 4.13.0.92) and NumPy (version 1.26.4). First, candidate pixels for white saturation were identified based on intensity and color information. Specifically, pixels with a mean RGB intensity ≥ 210 were selected, as well as pixels with a value (V) ≥ 210 and saturation (S) ≤ 70 in the HSV color space. Next, connected component analysis with 8-connectivity was applied to the binary mask, and only components with an area ≥ 10 pixels were retained. To refine the detected regions, morphological closing was performed using an elliptical kernel (7 × 7), followed by dilation using an elliptical kernel (5 × 5) for two iterations. For each detected region, a surrounding ring-shaped area was defined by subtracting the original mask from its dilated version. The pixel values within the white saturation regions were then replaced by the median RGB values computed from the corresponding ring regions, enabling local intensity interpolation while preserving surrounding structural information.

## Gallbladder Segmentation and Region of Interest Localization

Automated gallbladder segmentation was performed using TotalSegmentator [19] (version 2.12.0). A bounding box was generated from the segmentation mask, incorporating a 15-pixel margin in all directions to capture the pericholecystic environment. In four cases where automated segmentation was suboptimal, manual correction was performed by a specialist to ensure the gallbladder remained the central region of interest.

## Deep Learning Framework

Input images were resized to 128 × 128 pixels and converted into three-channel (RGB) format. The imaging model employed an EfficientNet-B0 backbone pretrained on ImageNet, implemented using TensorFlow (version 2.16.2). The architecture included a Global Average Pooling layer, a fully connected (dense) layer with 128 units (GELU activation), a dropout layer (rate = 0.4), and a final sigmoid output layer for binary classification.

Training was executed in two phases:

1. Warm-up: The backbone was frozen, and the classification head was trained using the AdamW optimizer (learning rate = $1 \times 10^{-4}$)
2. Fine-tuning: The final 10 layers of the backbone were unfrozen and optimized using Adam (learning rate = $1 \times 10^{-5}$), while Batch Normalization layers remained frozen to preserve internal covariate shift stability

Training was conducted with a batch size of 32 for up to 300 epochs, utilizing early stopping (patience = 10) based on validation loss. Following the training phase, this model served as a feature extractor: the 128-dimensional latent vectors from the penultimate dense layer were extracted and utilized as input features for the subsequent downstream machine learning models.

## Model Architecture and Training

We developed and compared three diagnostic frameworks: the Imaging Model (a deep learning model using CT data); the Blood Model (a machine learning model using blood data); and the Hybrid Model (a machine learning model integrating deep learning-derived imaging features and clinical data). The dataset was partitioned at the patient level into a training set (71%, n = 139) and a test set (29%, n = 58) to prevent data leakage. A single representative axial CT slice was used for each patient during model training. For evaluation in the independent test dataset, all eligible axial CT slices containing the gallbladder were analyzed. The maximum predicted probability across all slices was used as the patient-level imaging score, and a patient was classified as positive if at least one slice had a predicted probability of acute cholecystitis ≥ 0.50.

Automated machine learning was implemented via the PyCaret library (version 3.3.2) to identify the optimal classification architecture. The input feature space consisted of standardized biochemical laboratory data, the 128-dimensional latent vectors extracted from the deep learning backbone, or a concatenated vector of both for the hybrid approach.

To ensure a robust assessment of generalizability, model performance was rigorously evaluated on an independent test set. The primary metrics for discriminative capacity included accuracy, precision, recall (sensitivity), specificity, and F1 score. Additionally, the area under the receiver operating characteristic curve (AUC) was calculated to evaluate the overall diagnostic performance.

# Blood-Based and Hybrid Machine Learning Models

Clinical variables (white blood cell count, platelet count, levels of C-reactive protein, total bilirubin, aspartate aminotransferase, alanine aminotransferase, gamma-glutamyl transpeptidase, alkaline phosphatase, and creatinine, and the prothrombin time–international normalized ratio) were standardized using z-score normalization. For the Blood Model, an Extra Trees classifier was selected as the final estimator. For the Hybrid Model, 128-dimensional feature vectors extracted from the bottleneck layer of the deep learning model were reduced using principal component analysis, retaining components that explained more than 85% of the cumulative variance (n = 31; Supplementary Figure S1, Additional file 1). These features were then concatenated with the clinical variables, and a Random Forest classifier was used as the final estimator for the Hybrid Model. Model selection for both models was performed using PyCaret based on cross-validated performance. Additional optimization analyses are shown in Supplementary Figure S5 (Additional file 1).

# Statistical Analysis

Baseline characteristics were compared between the AC and non-AC groups using Welch's two-sample t-test for continuous variables and the chi-square test for categorical variables. All statistical analyses were implemented in R statistical software(version 4.5.3), with a $p$-value < 0.05 considered statistically significant.

During manuscript preparation, ChatGPT (OpenAI) was used to assist with language, formatting, and structural organization. All output was reviewed and edited by the authors, who take full responsibility for the final manuscript.

## Results

# Patient Characteristics and Dataset Partitioning

A total of 197 patients (100 with AC and 97 without AC) were enrolled in this study (Table 1). Baseline demographic characteristics were generally comparable between the two groups, with no significant differences in sex, age, height, or serum creatinine levels. However, patients with AC had significantly higher body weight and body mass index than those without AC. Laboratory findings also differed significantly between the groups: the AC group had higher white blood cell counts, C-reactive protein levels, total bilirubin levels, liver enzyme levels (aspartate aminotransferase, alanine aminotransferase, gamma-glutamyl transpeptidase, and alkaline phosphatase), and the prothrombin time–international normalized ratio, while platelet counts were significantly lower.

Table 1
Demographic and clinical characteristics of the study cohort

| **Sex** | **Non-AC (n = 97)** | **AC (n = 100)** | ***p*-value** |
|---|---|---|---|
| | | | 0.101 |
| Male | 76 (78.4%) | 68 (68.0%) | |
| Female | 21 (21.6%) | 32 (32.0%) | |
| Age | 67.4 ± 12.6 | 66.3 ± 14.3 | 0.624 |
| Height [cm] | 163.8 ± 9.0 | 161.5 ± 9.6 | 0.086 |
| Weight [kg] | 55.1 ± 9.1 | 60.1 ± 11.1 | < 0.001 |
| BMI [kg / $m^2$] | 20.5 ± 3.2 | 23.0 ± 3.6 | < 0.001 |
| WBC [ /μL] | 6750 ± 2335 | 12571 ± 5412 | < 0.001 |
| PLT [$\times 10^3$/μL] | 24.7 ± 8.3 | 21.6 ± 9.4 | 0.014 |
| CRP [mg/dL] | 0.673 ± 1.49 | 12.6 ± 10.5 | < 0.001 |
| T-Bil [mg/dL] | 0.80 ± 0.30 | 1.71 ± 1.2 | < 0.001 |
| AST [U/L] | 20.6 ± 10.0 | 75.9 ± 138.6 | < 0.001 |
| ALT [U/L] | 18.4 ± 14.5 | 63.3 ± 118.5 | < 0.001 |
| γGTP [U/L] | 44.1 ± 38.1 | 120.5 ± 236.4 | < 0.001 |

Data are presented as numbers for categorical variables and means ± standard deviations for continuous variables. AC, acute cholecystitis; ALP, alkaline phosphatase; ALT, alanine aminotransferase; AST, aspartate aminotransferase; BMI, body mass index; Cre, creatinine; CRP, C-reactive protein; PLT, platelet; PT–INR, prothrombin time-international normalized ratio; T-Bil, total bilirubin; WBC, white blood cell; γGTP, gamma-glutamyl transferase

| Sex | Non-AC (n = 97) | AC (n = 100) | *p*-value |
|---|---|---|---|
| | | | 0.101 |
| ALP [U/L] | $95.2 \pm 54.2$ | $245.5 \pm 236.4$ | < 0.001 |
| Cre [mg/dL] | $0.933 \pm 0.446$ | $0.9494 \pm 0.646$ | 0.878 |
| PT–INR [-] | $0.951 \pm 0.124$ | $1.08 \pm 0.232$ | < 0.001 |

Data are presented as numbers for categorical variables and means ± standard deviations for continuous variables. AC, acute cholecystitis; ALP, alkaline phosphatase; ALT, alanine aminotransferase; AST, aspartate aminotransferase; BMI, body mass index; Cre, creatinine; CRP, C-reactive protein; PLT, platelet; PT–INR, prothrombin time-international normalized ratio; T-Bil, total bilirubin; WBC, white blood cell; γGTP, gamma-glutamyl transferase

# Image Processing and Feature Extraction

Three diagnostic architectures were developed and evaluated: the Imaging Model, the Blood Model, and the Hybrid Model integrating both data streams (Fig. 1). To optimize deep learning performance, CT DICOM images underwent a standardized preprocessing pipeline. Raw data were converted to HU and normalized to 8-bit grayscale. To enhance feature representation, three-channel RGB inputs were generated using varying HU window settings. The gallbladder was segmented using TotalSegmentator, with regions of interest extracted including a pericholecystic margin; manual refinement was performed where necessary. These processed images served as input for EfficientNet, from which 128-dimensional latent feature vectors were extracted for downstream machine learning classification (Supplementary Figure S2A–D, Additional file 1).

# Performance of the Imaging Model

On the independent test dataset, the Imaging Model, which utilized 128-dimensional features extracted using EfficientNet, achieved an accuracy of 0.741, precision of 0.703, recall of 0.867, specificity of 0.607, and F1 score of 0.776 and an AUC of 0.863 (Fig. 2A, B). Gradient-weighted Class Activation Mapping (Grad-CAM) visualization confirmed that, in true-positive cases, the model consistently localized its attention to the gallbladder and its immediate anatomical environment (Supplementary Figure S3A, Additional file 1). In true-negative cases, minimal activation was observed in the gallbladder region, reflecting a lack of pathological morphological markers (Supplementary Figure S3B, Additional file 1). However, in misclassified cases (false positives and false negatives), attention maps were less localized, suggesting that relying solely on morphological features like gallbladder volume or wall thickness may be insufficient for definitive diagnosis in atypical presentations (Supplementary Figure S3C, D, Additional file 1).

# Performance of the Blood Model

The Blood Model integrated ten laboratory variables: white blood cell count, platelet count, levels of C-reactive protein, total bilirubin, aspartate aminotransferase, alanine aminotransferase, gamma-glutamyl transpeptidase, alkaline phosphatase, and creatinine, and the prothrombin time–international normalized ratio. Following an automated machine learning comparison using PyCaret, the Extra Trees Classifier was identified as the optimal architecture.

On the test dataset, this model achieved an accuracy of 0.914, precision of 0.931, recall of 0.900, specificity of 0.929, F1 score of 0.915 and an AUC of 0.946 (Fig. 3A, B). Feature importance analysis identified the C-reactive protein level as the primary contributor, followed by the white blood cell count and total bilirubin level (Fig. 3C), underscoring the centrality of systemic inflammatory biomarkers in the classification task.

Analysis of five misclassified cases (three false negatives and two false positives) revealed that these instances lacked a clear correlation between clinical presentation and gallbladder morphology (Supplementary Figure S4A, B, Additional file 1). Notably, false-positive cases exhibited markedly elevated C-reactive protein levels (> 9 mg/dL) without anatomical AC evidence, while false-negative cases showed low C-reactive protein levels (< 1 mg/dL) despite potential clinical indicators (Supplementary Table S1, Additional file 2). This suggests the Blood Model is highly sensitive to inflammatory signals but lacks specificity in the absence of structural confirmation.

# Performance of the Hybrid Model

Finally, we developed a hybrid machine learning model using a Random Forest classifier to fuse imaging-derived features with clinical laboratory data. This integrated approach achieved superior performance on the test dataset, with an accuracy of 0.931, precision of 0.882, recall of 1.000, specificity of 0.857, and an F1 score of 0.938 (Fig. 4A). The AUC reached 0.996 (Fig. 4B). Feature importance analysis revealed that both imaging-derived and laboratory features contributed meaningfully to the model's predictions (Fig. 4C). Among these, the most influential variable was the first principal component of image features, followed by key inflammatory markers such as the C-reactive protein level and white blood cell count. Additional contributions from total bilirubin and liver-related enzyme levels further highlighted the complementary role of biochemical indicators. The four false-positive cases are shown in Fig. 4D, and their laboratory parameters are summarized in Table 2.

Table 2
Laboratory parameters in four misclassified cases

| Case | WBC [/μL] | PLT [$\times 10^4$/μL] | CRP [mg/dL] | T-Bil [mg/dL] | AST [U/L] | ALT [U/L] | γGTP [U/L] | ALP [U/L] | Cre [mg/dL] | PT–INR |
|---|---|---|---|---|---|---|---|---|---|---|
| D1 | 9730 | 15.8 | 1.94 | 0.9 | 22 | 11 | 81 | 100 | 0.92 | 0.95 |
| D2 | 8670 | 25 | 0.19 | 0.7 | 22 | 18 | 26 | 104 | 1.29 | 0.95 |
| D3 | 5260 | 19.4 | 0.34 | 1.1 | 19 | 11 | 26 | 58 | 0.95 | 0.95 |
| D4 | 7290 | 29.4 | 0.41 | 0.6 | 14 | 12 | 72 | 187 | 1.00 | 1.14 |

Data are presented as values for each case. WBC, white blood cell count; PLT, platelet count; CRP, C-reactive protein; T-Bil, total bilirubin; AST, aspartate aminotransferase; ALT, alanine aminotransferase; γGTP, gamma-glutamyl transferase; ALP, alkaline phosphatase; Cre, creatinine; PT–INR, prothrombin time–international normalized ratio

# Discussion

In this study, we addressed the challenge of accurately diagnosing AC in emergency settings, where diagnostic decisions must balance accuracy with practical constraints such as invasiveness, cost, and time. These considerations highlight the clinical importance of diagnostic support based on non-contrast CT data. We developed three models: an Imaging Model, a Blood Model, and a Hybrid Model. The Hybrid Model achieved the best overall performance, underscoring the value of integrating imaging and clinical data.

The Imaging Model appeared to be influenced by morphologic features such as gallbladder distension, even without clear inflammatory findings. In contrast, the Blood Model was vulnerable in cases with attenuated inflammatory responses, such as chemotherapy-related cytopenia, potentially leading to false-negative predictions. Our results indicate complementary limitations of imaging and laboratory data, supporting the use of a hybrid approach to improve diagnostic robustness. Despite this advantage, the Hybrid Model still produced four false-positive cases. Review of these cases suggested that the mechanisms of false-positive classification were heterogeneous. In D2 and D3, gallbladder distension was more prominent than in most control subjects, whereas gallbladder wall thickening and systemic inflammatory findings were not evident. These findings suggest that physiologic gallbladder distension, particularly in the setting of poor oral intake or fasting, may have mimicked morphologic features of acute cholecystitis and contributed to false-positive classification [20,21]. In contrast, D1 did not show marked gallbladder distension, but the white blood cell count and C-reactive protein level were relatively higher than in most control subjects. This finding suggests that nonspecific systemic inflammation related to the underlying malignancy or general condition may have influenced the laboratory component of the Hybrid Model [22]. In D4, neither gallbladder distension nor systemic inflammatory findings were prominent. Although cholestatic enzymes are not included in the diagnostic criteria for acute cholecystitis, hepatobiliary abnormalities, including cholestatic liver enzyme elevations, can occur in

patients receiving parenteral nutritional support and may act as laboratory confounders in multimodal prediction [23]. These findings indicate that the false-positive results of the Hybrid Model were not attributable to a single factor but rather to different combinations of morphologic, inflammatory, and laboratory confounders. Further refinement incorporating inflammation-specific imaging features, nutritional status, and clinical context may improve specificity.

While point-of-care ultrasound remains widely used, its diagnostic performance is often influenced by operator expertise and availability, with a reported sensitivity and specificity of approximately 79.0% and 93.6%, respectively [7,15]. To address these limitations, CT-based frameworks have been developed, with reported AUCs ranging from 0.7933 to 0.8569 [10]. In the present study, the Hybrid Model, which integrated CT-derived imaging features with routinely available laboratory data, achieved an AUC of 0.996 and a recall of 1.00, indicating that no cases of acute cholecystitis were missed in this cohort. This high sensitivity may be clinically advantageous in settings where minimizing false-negative diagnoses is critical. By reducing reliance on operator-dependent modalities, this approach may improve consistency in clinical decision-making and support triage, particularly in emergency settings where initial assessments are performed by non-specialists.

This study was subject to several limitations. First, the retrospective design may have introduced selection and information biases [24]. Second, the single-center design limits the generalizability of the present findings. External validation using independent cohorts from other institutions is required to assess the robustness of the model across different patient populations, CT scanners, and acquisition protocols [25,26]. Finally, the reference standard, derived from the Tokyo Guidelines 2018 criteria and composite clinical data, carries an inherent risk of label uncertainty. Future research should focus on prospective validation and the implementation of domain adaptation techniques to account for inter-institutional variability. Furthermore, ensuring transparency in artificial intelligence model development is paramount; subsequent reporting should adhere to the TRIPOD + AI [27], CLAIM 2024 [28], and STARD-AI [29] guidelines to ensure that these tools are both reproducible and ethically sound.

# Conclusions

In conclusion, a deep learning-based hybrid model using non-contrast CT images showed potential as a diagnostic support tool for acute cholecystitis. In particular, it may serve as an objective and clinically implementable adjunct in emergency settings where junior residents and non-gastrointestinal specialists are involved in the initial assessment.

# Abbreviations

AC
acute cholecystitis
AI
artificial intelligence

ALP
alkaline phosphatase
ALT
alanine aminotransferase
AST
aspartate aminotransferase
AUC
area under the receiver operating characteristic curve
BMI
body mass index
Cre
creatinine
CRP
C-reactive protein
CT
computed tomography
DICOM
Digital Imaging and Communications in Medicine
GELU
Gaussian error linear unit
Grad-CAM
gradient-weighted class activation mapping
HSV
hue,saturation,and value
HU
Hounsfield unit
PCA
principal component analysis
PLT
platelet count
PT-INR
prothrombin time-international normalized ratio
RGB
red,green,and blue
ROC
receiver operating characteristic
WBC
white blood cell count
γGTP
gamma-glutamyl transpeptidase.

# Declarations

### Ethics approval and consent to participate

The study protocol was approved by the Institutional Review Board of Gifu University Graduate School of Medicine (Approval No. 2024-007) and was conducted in accordance with the Declaration of Helsinki. This retrospective cohort study included cases from January 2016 to December 2025. Because this was a retrospective study using anonymized data, the requirement for written informed consent was waived.

### Consent for publication

Not applicable.

# Competing interests

The authors declare that they have no competing interests.

# Additional files

Additional file 1 (.pdf). Supplementary Figures S1–S5: cumulative explained variance of image-derived principal components; CT preprocessing; representative Grad-CAM visualizations; discordant blood-model cases; and hybrid-model optimization.

Additional file 2 (.docx). Supplementary Tables S1–S2: laboratory parameters in five cases misclassified by the Blood Model and diagnoses of the 97 non-acute cholecystitis control cases.

# Funding

This study was supported by the 2024 “Clinical Research Grant for Young Physicians Engaged in Abdominal Emergency Care” from the Japanese Association for Abdominal Emergency Medicine. The funding body had no role in the study design; data collection, analysis, or interpretation; manuscript preparation; or the decision to submit for publication.

# Author Contribution

TH conceived and designed the study. TH, RN, and NS drafted the manuscript. RT, NMi, YH, RYo, MK, HH, MF, YS, IY, and RA contributed to data acquisition. RN, SAS, NS, KK, MO, HB, and RYa contributed to data analysis, interpretation, and model development. RYa critically revised the manuscript for important intellectual content. NMa supervised the study and finalized the manuscript. All authors read and

approved the final manuscript.For clarity, NMi denotes Noriki Mitsui, NMa denotes Nobuhisa Matsuhashi, RYo denotes Ryoma Yokoi, and RYa denotes Riu Yamashita.


# Acknowledgement

We would like to thank Editage (www.editage.jp) for the English language editing. Computational analyses were performed on the KASHIWARP server at the National Cancer Center.


# Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request. All scripts used in this study are available on GitHub (https://github.com/RyosukeNomural/model_for_acute_cholecystitis).

## References


1. Yokoe M, Hata J, Takada T, et al. Tokyo Guidelines 2018: diagnostic criteria and severity grading of acute cholecystitis (with videos). J Hepatobiliary Pancreat Sci. 2018;25:41–54. 10.1002/jhbp.515.
2. O'Connell RM, Hardy N, Ward L, et al. Management and patient outcomes following admission with acute cholecystitis in Ireland: a national registry-based study. Surgeon. 2024;22:364–8. 10.1016/j.surge.2024.08.004.
3. Mencarini L, Vestito A, Zagari RM, Montagnani M. The diagnosis and treatment of acute cholecystitis: a comprehensive narrative review for a practical approach. J Clin Med. 2024;13:2695. 10.3390/jcm13092695.
4. Markou-Pappas N, Frassini S, Di Gioia G, Tomasoni M, Dughetti S, Ansaloni L. Enhancing surgical decision-making: the critical need for structured ultrasound training in residency. A case report of acute necrotic cholecystitis. Case Rep Emerg Surg Trauma. 2025;3:47. 10.4081/crest.2025.47.
5. Sidhu PS, Clevert DA, Deganello A, et al. Controversies in contrast-enhanced ultrasound (CEUS): pregnancy, paediatric, abdominal trauma, complex renal cysts, and endovascular aortic repair follow-up. Insights Imaging. 2025;16:179. 10.1186/s13244-025-02055-w.
6. Shakespear JS, Shaaban AM, Rezvani M. CT findings of acute cholecystitis and its complications. AJR Am J Roentgenol. 2010;194:1523–9. 10.2214/AJR.09.3640.
7. de Oliveira GS, Torri GB, Gandolfi FE, et al. Computed tomography versus ultrasound for the diagnosis of acute cholecystitis: a systematic review and meta-analysis. Eur Radiol. 2024;34:6967–79. 10.1007/s00330-024-10783-8.
8. Abd El Bagi ME, Almutairi BM, Alsolamy SJ. Imaging of non-traumatic acute abdominal pain in adults presenting to the emergency department. Saudi Med J. 2016;37:324–7. 10.15537/smj.2016.3.13068.

9. Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. Med Image Anal. 2017;42:60–88. 10.1016/j.media.2017.07.005.
10. Chen BQ, Zang W, Liu JX, Yang Y, Zhang XL, Ju RH. Deep learning enables accurate diagnosis of acute cholecystitis and prediction of suppuration using noncontrast CT. *iScience* 2025;28(12):114180. 10.1016/j.isci.2025.114180
11. Olsen GH, Goodman ED, Aklilu JG, et al. Using artificial intelligence to model expert panel diagnosis of cholecystitis severity. Surg Endosc. 2025;39:6560–8. 10.1007/s00464-025-12015-6.
12. Park JJ, Kim KA, Nam Y, Choi MH, Choi SY, Rhie J. Convolutional neural network-based diagnosis of appendicitis via CT scans in patients with acute abdominal pain presenting in the emergency department. Sci Rep. 2020;10:9556. 10.1038/s41598-020-66674-7.
13. Rajpurkar P, Park A, Irvin J, et al. AppendiXNet: deep learning for diagnosis of appendicitis from a small dataset of CT exams using video pretraining. Sci Rep. 2020;10:3958. 10.1038/s41598-020-61055-6.
14. Yao J, Chu LC, Patlas M. Applications of artificial intelligence in acute abdominal imaging. Can Assoc Radiol J. 2024;75:761–70. 10.1177/08465371241250197.
15. Yu CJ, Yeh HJ, Chang CC, et al. Lightweight deep neural networks for cholelithiasis and cholecystitis detection by point-of-care ultrasound. Comput Methods Programs Biomed. 2021;211:106382. 10.1016/j.cmpb.2021.106382.
16. Lambin P, Rios-Velazquez E, Leijenaar R, et al. Radiomics: extracting more information from medical images using advanced feature analysis. Eur J Cancer. 2012;48:441–6. 10.1016/j.ejca.2011.11.036.
17. Rajpurkar P, Acosta JN, Dogra S et al. a2z-1 for multi-disease detection in abdomen-pelvis CT: external validation and performance analysis across 21 conditions. arXiv [Preprint]. 2024. 10.48550/arXiv.2412.12629
18. Seto T, Takeuchi M, Hashimoto M et al. Classification of esophageal cancer CT images using deep learning. Proc Annu Conf Jpn Soc Artif Intell 2019;JSAI2019:2N4J1302. 10.11517/pjsai.jsai2019.0_2n4j1302
19. Wasserthal J, Breit HC, Meyer MT, et al. TotalSegmentator: robust segmentation of 104 anatomic structures in CT images. Radiol Artif Intell. 2023;5:e230024. 10.1148/ryai.230024.
20. Alsaadi M. Sonographic evaluation of gallbladder distension based on interval fasting hours: a prospective study. Int J Biomed. 2025;15(3):535–9. 10.21103/Article15(3)_OA13.
21. Huffman JL, Schenker S. Acute acalculous cholecystitis: a review. Clin Gastroenterol Hepatol. 2010;8:15–22. 10.1016/j.cgh.2009.08.034.
22. Allin KH, Nordestgaard BG. Elevated C-reactive protein in the diagnosis, prognosis, and cause of cancer. Crit Rev Clin Lab Sci. 2011;48(4):155–70. 10.3109/10408363.2011.599831.
23. Żalikowska-Gardocka M, Przybyłkowski A. Review of parenteral nutrition-associated liver disease. Clin Exp Hepatol. 2020;6(2):65–73. 10.5114/ceh.2019.95528.

24. Sedgwick P. Retrospective cohort studies: advantages and disadvantages. BMJ. 2014;348:g1072. 10.1136/bmj.g1072.
25. Traverso A, Wee L, Dekker A, Gillies R. Repeatability and reproducibility of radiomic features: a systematic review. Int J Radiat Oncol Biol Phys. 2018;102:1143–58. 10.1016/j.ijrobp.2018.05.053.
26. Da-ano R, Masson I, Lucia F, et al. Performance comparison of modified ComBat for harmonization of radiomic features for multicenter studies. Sci Rep. 2020;10:10248. 10.1038/s41598-020-66110-w.
27. Collins GS, Moons KGM, Dhiman P, et al. TRIPOD + AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ. 2024;385:e078378. 10.1136/bmj-2023-078378.
28. Tejani AS, Klontzas ME, Gatti AA, et al. Checklist for artificial intelligence in medical imaging (CLAIM): 2024 update. Radiol Artif Intell. 2024;6:e240300. 10.1148/ryai.240300.
29. Sounderajah V, Guni A, Liu X, et al. The STARD-AI reporting guideline for diagnostic accuracy studies using artificial intelligence. Nat Med. 2025;31:3283–9. 10.1038/s41591-025-03953-8.

## Figures

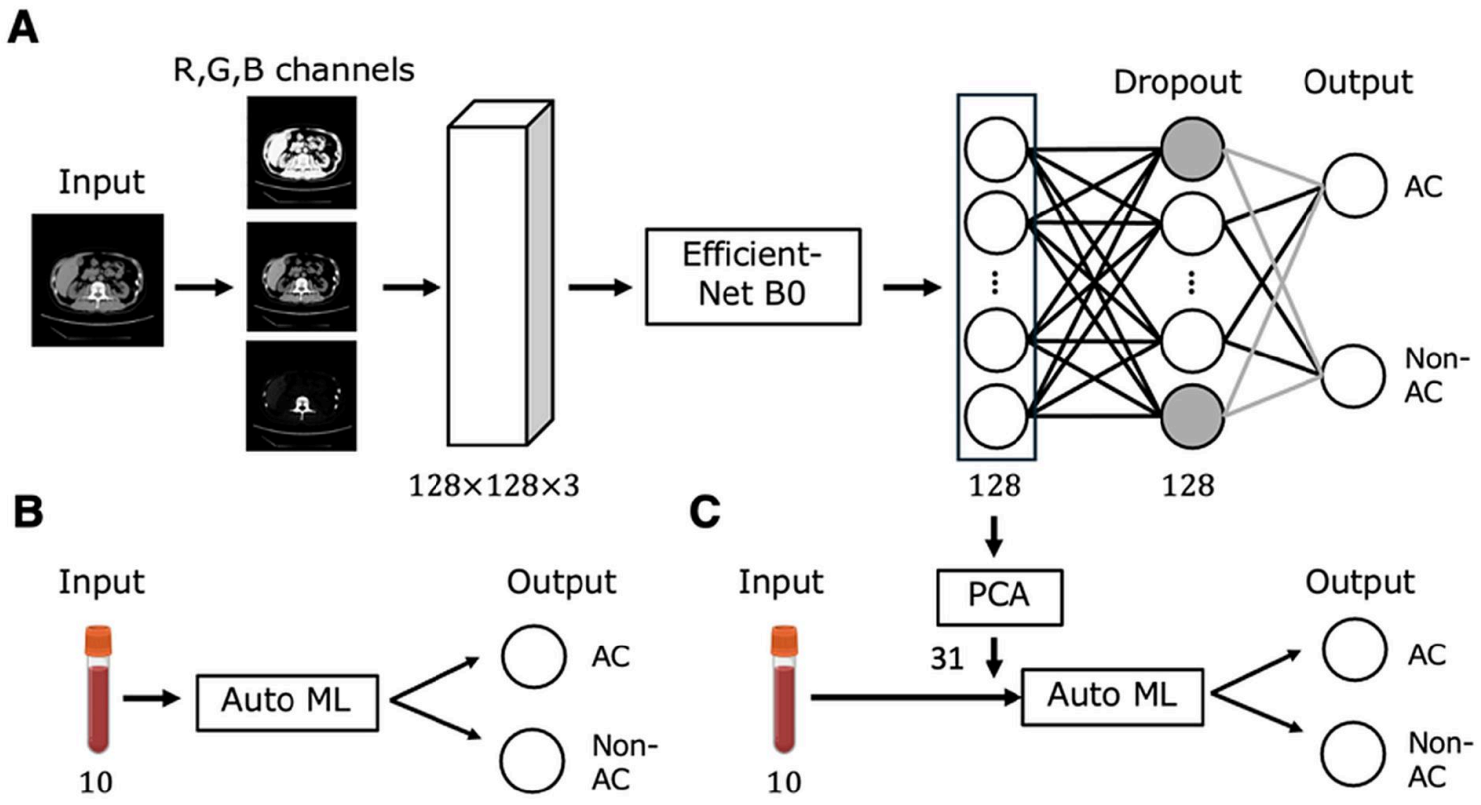



**Figure 1**

Overview of the Imaging, Blood, and Hybrid Models for Acute Cholecystitis Classification.

(A) Overview of Imaging Model. Input CT images were converted into three-channel (RGB) format and resized to 128 × 128 × 3 pixels before being processed by an EfficientNet-B0-based deep learning model. The model generated a binary output for acute cholecystitis (AC) versus non-acute cholecystitis (non-AC). (B) Overview of Blood Model. Ten clinical blood variables were used as input for a machine learning classifier to distinguish between AC and non-AC. (C) Overview of Hybrid Model. A 128-dimensional latent feature vector extracted from the imaging model was reduced to 31 dimensions by principal component analysis and combined with the 10 blood variables. The integrated features were then used as input for a machine learning classifier to distinguish between AC and non-AC.

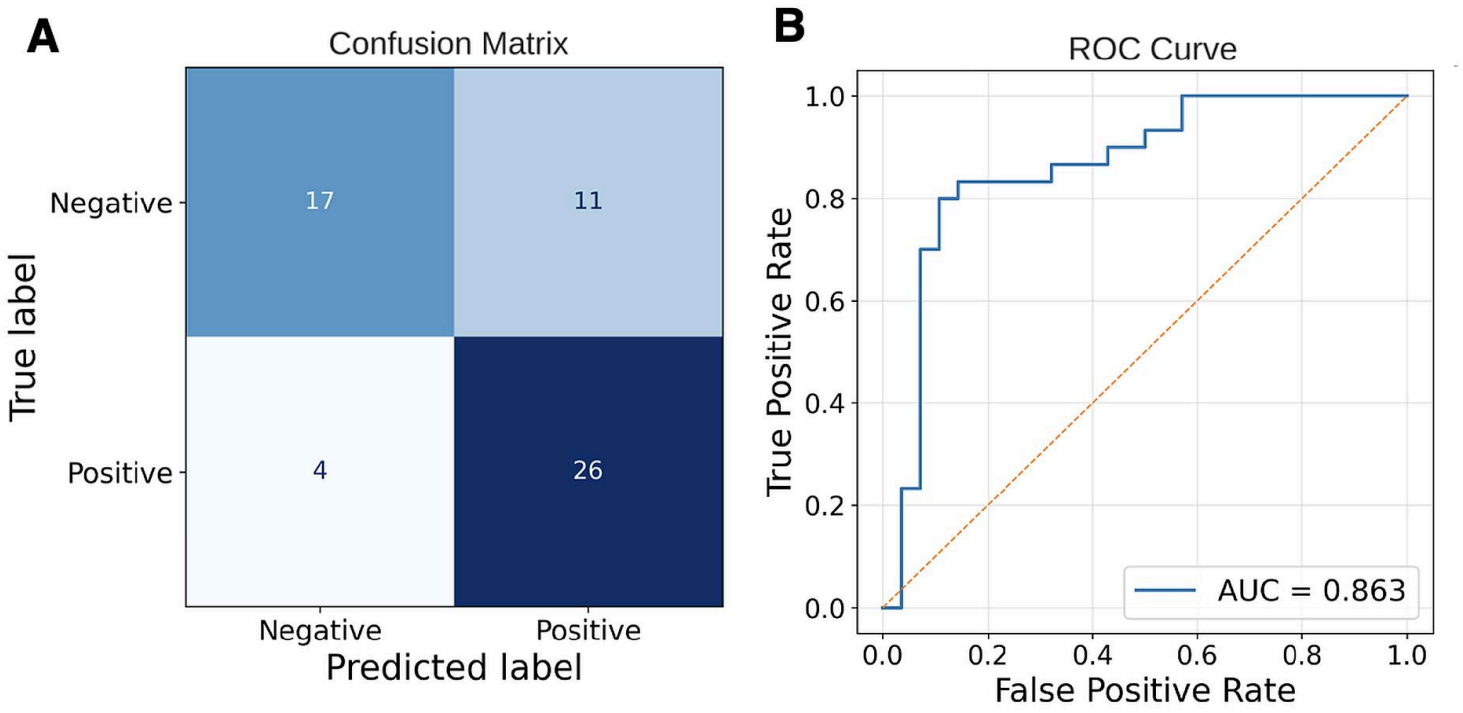



**Figure 2**

Diagnostic Performance of the Imaging Model for Acute Cholecystitis.

(A) Confusion matrix showing the true and predicted labels. (B) Receiver operating characteristic (ROC) curve evaluating the classification performance. The dashed line represents the performance of a random classifier. AUC, area under the curve.

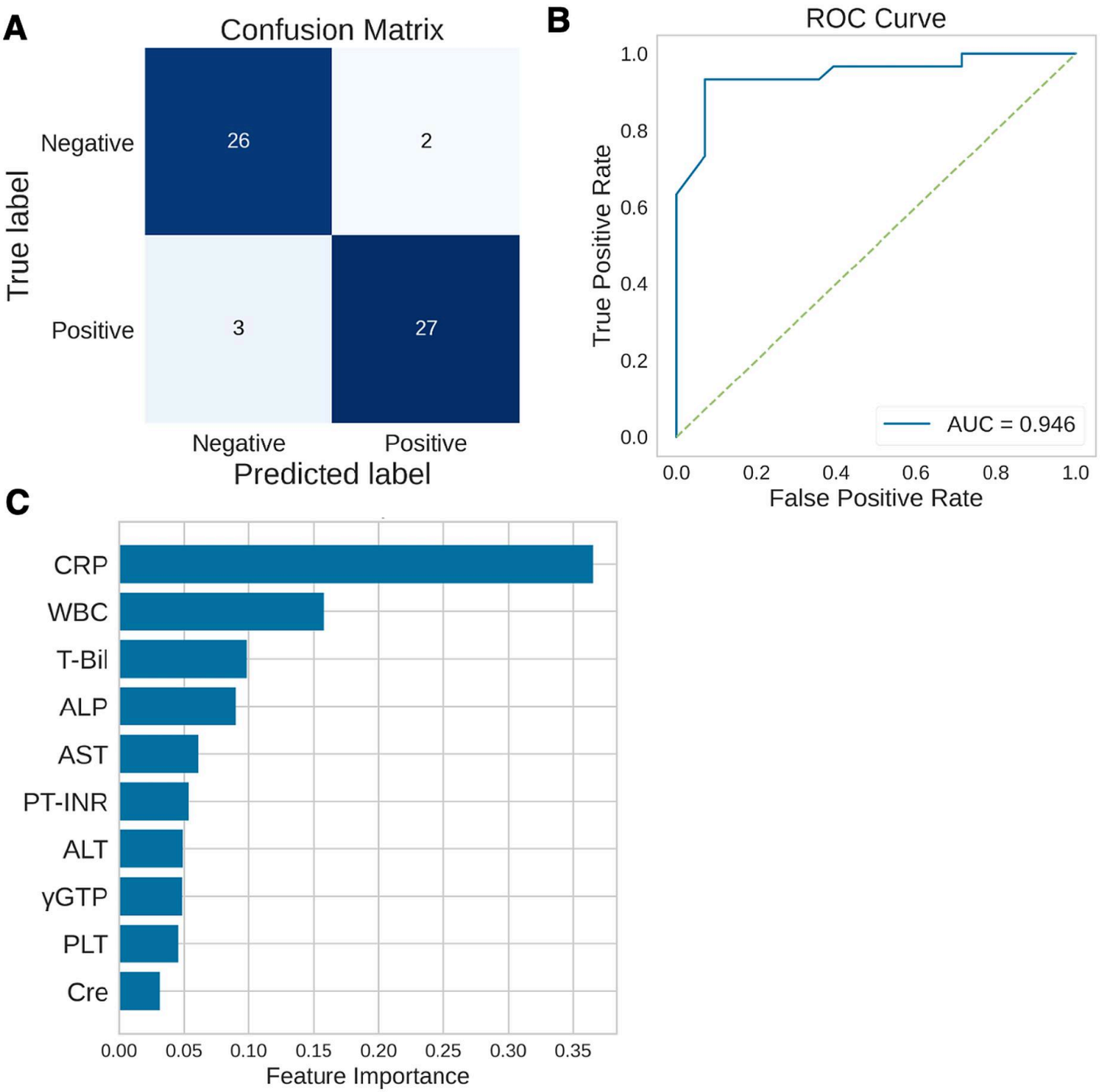



**Figure 3**

Diagnostic Performance and Feature Importance of the Blood Model for Acute Cholecystitis.

(A) Confusion matrix showing the true and predicted labels for the model. (B) Receiver operating characteristic (ROC) curve evaluating the classification performance. The dashed line represents the performance of a random classifier. (C) Bar chart displaying the relative feature importance of all laboratory variables contributing to the model's predictions.

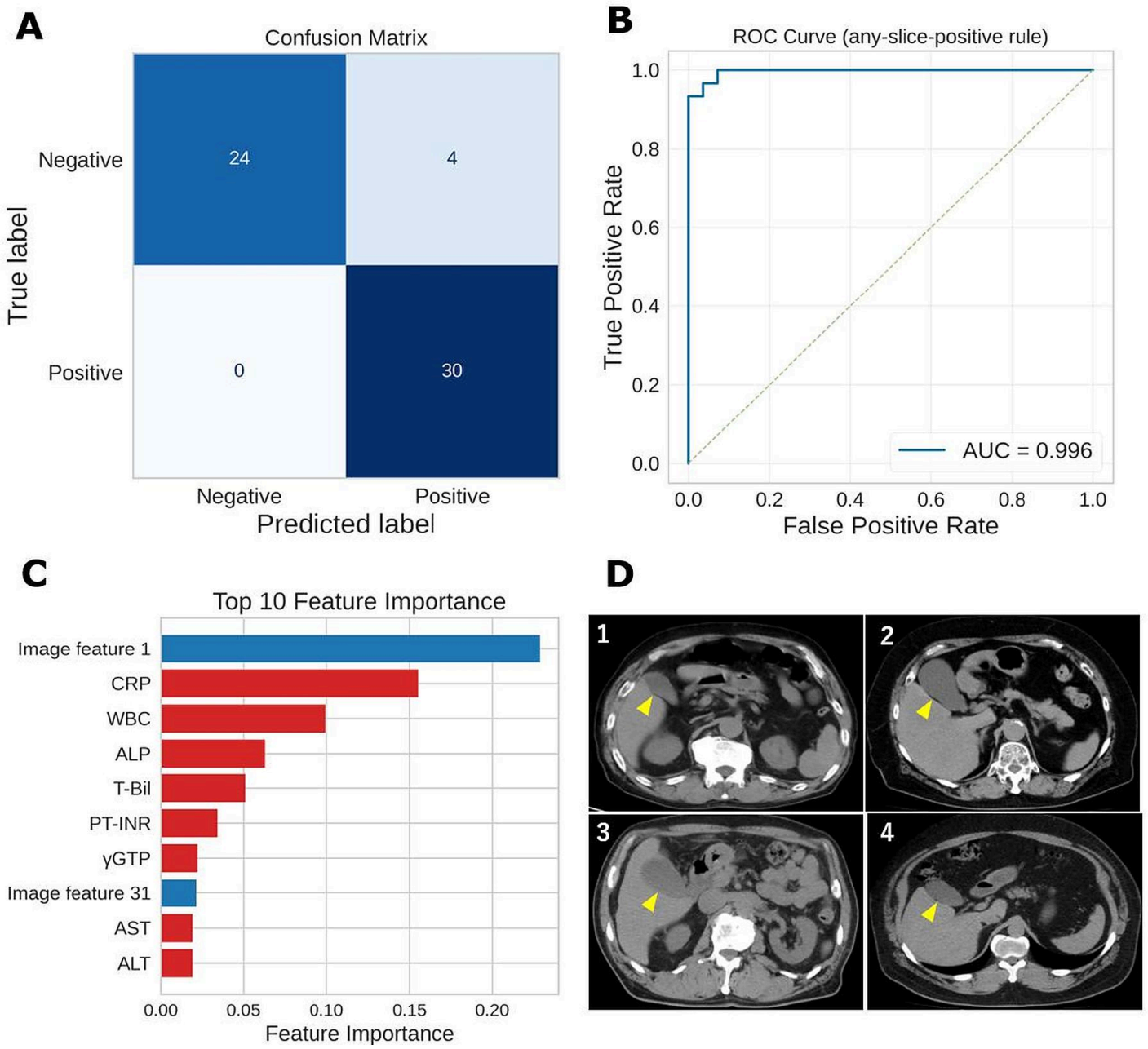



**Figure 4**

Diagnostic Performance, Feature Importance, and False-Positive Cases of the Hybrid Model for Acute Cholecystitis.

(A) Confusion matrix showing the true and predicted labels. (B) Receiver operating characteristic (ROC) curve evaluating the classification performance. The dashed line represents the performance of a random classifier. (C) Feature importance plot displaying the top 10 variables, including laboratory parameters and image-derived principal component analysis features, contributing to the model's predictions. (D) Four representative cases incorrectly classified as positive by the model. For each case, axial CT images are shown with the gallbladder indicated by yellow arrows.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile2SupplementaryTables.docx.docx
- Additionalfile1SupplementaryFigures.pdf