frontiers in Medicine

TYPE Original Research
PUBLISHED 09 October 2026
DOI 10.3389/fmed.2026.1848334

Check for updates

OPEN ACCESS

EDITED BY
Anna Corti,
Polytechnic University of Milan, Italy

REVIEWED BY
Nikolaos Pararas,
National and Kapodistrian University of Athens, Greece
Horia Minea,
"Grigore T. Popa" University of Medicine and Pharmacy, Romania
Oussama Metrouh,
Beth Israel Deaconess Medical Center and Harvard Medical School, United States

*CORRESPONDENCE
Xiaowei Han
hxw2002hxw@163.com
Danqiong Wang
sara_good@126.com

RECEIVED 06 April 2026
REVISED 18 August 2026
ACCEPTED 21 September 2026
PUBLISHED 09 October 2026

CITATION
Zhang G, Liu X, Huang W, Zhu X, Xia C, Xia H, Wei Y, Han X and Wang D (2026) Radiomics-based interpretable machine learning model for identifying high-risk patients with small bowel obstruction: a multicenter study.
Front. Med. 13:1848334.
doi: 10.3389/fmed.2026.1848334

COPYRIGHT
© 2026 Zhang, Liu, Huang, Zhu, Xia, Xia, Wei, Han and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Radiomics-based interpretable machine learning model for identifying high-risk patients with small bowel obstruction: a multicenter study

Guozheng Zhang[1], Xiaohui Liu[1], Weitao Huang[1], Xisong Zhu[1], Cong Xia[2], Hua Xia[3], Yuguo Wei[4], Xiaowei Han[1]* and Danqiong Wang[5]*

[1]Department of Radiology, The Quzhou Affiliated Hospital of Wenzhou Medical University Quzhou People's Hospital, Quzhou, China, [2]Department of Radiology, Quzhou Kecheng People's Hospital, Quzhou, China, [3]Department of Radiology, Quzhou TCM Hospital at the Junction of Four Provinces, Quzhou, China, [4]Department of Medical Affairs, Pharmaceutical Diagnosis, GE Healthcare, Hangzhou, China, [5]Department of Critical Care Medicine, The Quzhou Affiliated Hospital of Wenzhou Medical University Quzhou People's Hospital, Quzhou, China

Background: Early identification of high-risk small bowel obstruction remains clinically challenging, especially when prompt intervention is needed to prevent bowel ischemia and related adverse outcomes.
Methods: In this multicenter retrospective study, 274 patients with small bowel obstruction from three hospitals were included. Patients from Hospital A were randomly divided into training and test cohorts, and patients from the other two hospitals formed an external validation cohort. Radiomics features were extracted from venous-phase contrast-enhanced CT images, and logistic regression, naive Bayes, Extra Trees, Light Gradient Boosting Machine (LightGBM), and multilayer perceptron models were developed. Independent clinical predictors were identified and integrated with radiomics features to build a combined model.
Results: LightGBM showed the most stable internal performance, with AUC values of 0.934 and 0.870 in the training and test cohorts, respectively. Lower diastolic blood pressure and bowel wall thickening were independent predictors of high-risk status. The combined model achieved AUC values of 0.942, 0.867, and 0.803 in the training, test, and external validation cohorts, respectively, with an external validation accuracy of 0.833 and sensitivity of 0.946.
Conclusion: CT-based radiomics can identify patients with small bowel obstruction at high risk, and a combined model integrating radiomics and clinical predictors may support earlier risk stratification and intervention.

KEYWORDS
computed tomography, machine learning, radiomics, risk stratification, small bowel obstruction

Introduction

Small bowel obstruction (SBO) is a common acute abdominal emergency characterized by substantial clinical heterogeneity, and adhesive SBO accounts for a large proportion of cases (1–3). Although many patients improve with nonoperative management such as gastrointestinal decompression, fasting, and fluid resuscitation, some progress to bowel ischemia, necrosis, strangulation, or failure of conservative treatment and ultimately require emergency surgery (4–6). Delayed recognition of these high-risk patients increases complications, bowel loss, and mortality (7). In routine practice, this decision is often difficult because symptoms, laboratory results, and early imaging findings do not always evolve in parallel. Clinicians therefore need not only confirmation of obstruction, but also early tools that can separate patients likely to respond to conservative management from those who may deteriorate and require urgent escalation of care.

Computed tomography (CT) remains the cornerstone of imaging evaluation for SBO because it can identify the level, cause, and severity of obstruction while also depicting high-risk findings such as bowel ischemia, mesenteric involvement, and ascites (8–10). Previous studies have shown that bowel wall thickening, decreased enhancement, mesenteric edema, mesenteric fluid, and ascites are associated with strangulation or surgically relevant ischemia (9–14). However, qualitative CT interpretation still depends heavily on reader experience and may be challenged by subtle findings, overlapping inflammatory changes, and interobserver variability. In borderline cases, the visible CT signs may be insufficiently specific or may appear only after clinically important bowel compromise has already developed. As a result, conventional CT assessment alone may not fully capture the underlying biologic heterogeneity of obstructed bowel segments or provide sufficiently reproducible risk stratification across centers.

Radiomics extracts high-throughput quantitative features from medical images and provides a complementary analytic dimension for disease-risk assessment and treatment decision-making (15–17). Prior studies have demonstrated broad predictive potential for radiomics across multiple diseases, and increasing attention has been directed toward interpretability, generalizability, and clinical translatability (18–21). In SBO, previous investigations have attempted to predict failure of conservative management, strangulation, ischemia, and early surgical need using CT findings, clinical scores, inflammatory biomarkers, and early machine-learning approaches (22–25). Nevertheless, multicenter, interpretable, and externally validated radiomics models for identifying patients with SBO at high risk remain relatively scarce. This gap is clinically important because a model that performs only in a single-center development cohort, or that cannot explain why it flags a patient as high risk, is unlikely to influence real emergency decision-making.

Accordingly, we developed radiomics, clinical, and combined models based on CT radiomics features, conventional CT findings, and clinical variables and validated these models in a multicenter cohort. SHAP was further used to improve model interpretability and to provide quantitative evidence for the early identification of high-risk SBO. We specifically aimed to determine whether integrating radiomics with key clinical predictors could improve risk stratification beyond either approach alone and whether such performance could be retained in an external validation cohort drawn from other institutions. The distinguishing contribution of the present work is therefore not radiomics or machine learning alone, but their combination in a three-hospital design that directly compares five algorithms, evaluates clinical, radiomics, and combined models, uses SHAP for both global and patient-level explanations, and tests generalizability in an independent external cohort.

Materials and methods

Study design and patient selection

This multicenter retrospective study enrolled 274 patients diagnosed with small bowel obstruction (SBO) who received treatment at three hospitals between January 2015 and December 2023. In the total cohort, 162 patients presented with adhesive obstruction, 69 with internal hernia, and 43 with inguinal hernia. Of these patients, 148 who underwent surgical intervention were assigned to the high-risk group, whereas the remaining 126 patients who achieved clinical improvement following conservative management constituted the low-risk group. Hospital A contributed 198 patients, who were randomly split 7:3 into a training cohort ($n = 138$) and a test cohort ($n = 60$); the remaining 76 patients from Hospitals B and C formed the external validation cohort. Patients were classified as high risk if they underwent surgery or required surgery after failed conservative management and as low risk if they responded well to conservative treatment. Surgical indications were evaluated from the medical and operative records using prespecified criteria, including peritoneal signs, CT evidence of strangulation, ischemia or closed-loop obstruction, hemodynamic deterioration, and failure of nonoperative management. Among the high-risk patients, underwent immediate surgery and underwent surgery after failed nonoperative management (Figure 1).

The inclusion criteria were as follows: patients with preoperative small bowel obstruction (SBO), preoperative contrast-enhanced computed tomography (CT) examinations, and complete clinical and imaging datasets. Poor image quality was defined as severe motion artifacts, inadequate venous phase enhancement, or incomplete coverage of the lesional intestinal segments, which precluded reliable CT evaluation or radiomic segmentation. Image quality was independently assessed by two abdominal radiologists, and any disagreements were resolved via consensus. Patients were also excluded if the time interval between CT scanning and the definitive treatment decision exceeded 72 h.

The institutional review boards of the participating hospitals approved the study protocol and waived the requirement for informed consent because of the retrospective design. The study was prepared with reference to the STROBE and CLAIM recommendations for observational and machine-learning research.

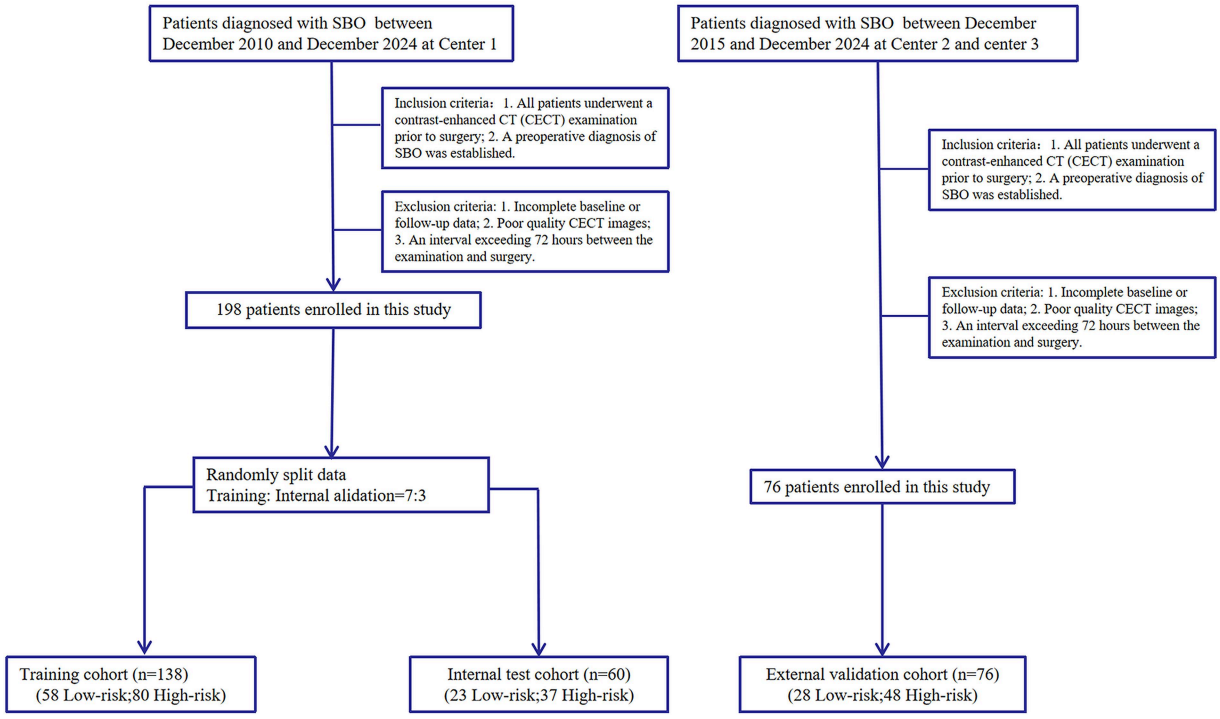


FIGURE 1
Flowchart illustrating study enrollment for patients with small bowel obstruction (SBO) at three centers, detailing inclusion and exclusion criteria, enrollment numbers, random data split, and division into training, internal test, and external validation cohorts with low-risk and high-risk groups.

Clinical and CT signs

Demographic, clinical, vital-sign, and laboratory variables were collected, including sex, age, white blood cell count, C-reactive protein, blood pressure, heart rate, respiratory rate, abdominal pain pattern, and cessation of flatus or defecation. CT variables included bowel wall thickening, decreased enhancement, maximum bowel diameter, the small-bowel feces sign, mesenteric edema, mesenteric fluid, and ascites. These candidate variables were selected to capture clinically relevant hemodynamic status, inflammatory burden, and CT signs of bowel compromise.

CT signs, ROI segmentation, and feature extraction

All images were reviewed by two abdominal radiologists with 15 and 10 years of experience, respectively, who were blinded to outcome. Disagreements were resolved by consensus, and uncertain obstruction sites were adjudicated by a senior abdominal radiologist with 30 years of experience.

All patients underwent contrast-enhanced CT, and venous-phase source images in Digital Imaging and Communications in Medicine format were used for radiomics analysis. Regions or volumes of interest were manually delineated around the bowel segment related to obstruction to encompass the affected bowel wall and key pathologic regions while excluding obvious luminal contents, gas, and ascites whenever possible.

After standardized image preprocessing, radiomics features were extracted, including first-order, texture, shape, wavelet, and other filter-derived features. Feature robustness was assessed before model development so that reproducible features could be retained.

Radiomics reproducibility details

Radiomics feature extraction was performed using PyRadiomics v3.0.1, following the IBSI-1 standard configuration. Six feature classes were computed: first-order statistics, shape-based features, gray-level co-occurrence matrix (GLCM), gray-level run-length matrix (GLRLM), gray-level size-zone matrix (GLSZM), gray-level dependence matrix (GLDM), and neighborhood gray-tone difference matrix (NGTDM). Images were resampled to $1 \times 1 \times 1$ mm^3 voxel size (B-spline for image, nearest-neighbour for segmentation mask), with intensity discretization using a bin width of 25 HU. No additional filtering was performed. Feature robustness was assessed by ICC using repeated manual segmentations; only features with ICC $\geq$0.75 were kept. For machine-learning preparation, Z-score normalization was performed using only training-set statistics to avoid leakage. Features with >5% missing voxels were excluded and minor missing entries were median-imputed. Class imbalance was addressed via balanced class weights within the classifier. Hyperparameters were optimized via stratified 5-fold cross-validation on the training cohort (Figure 2).

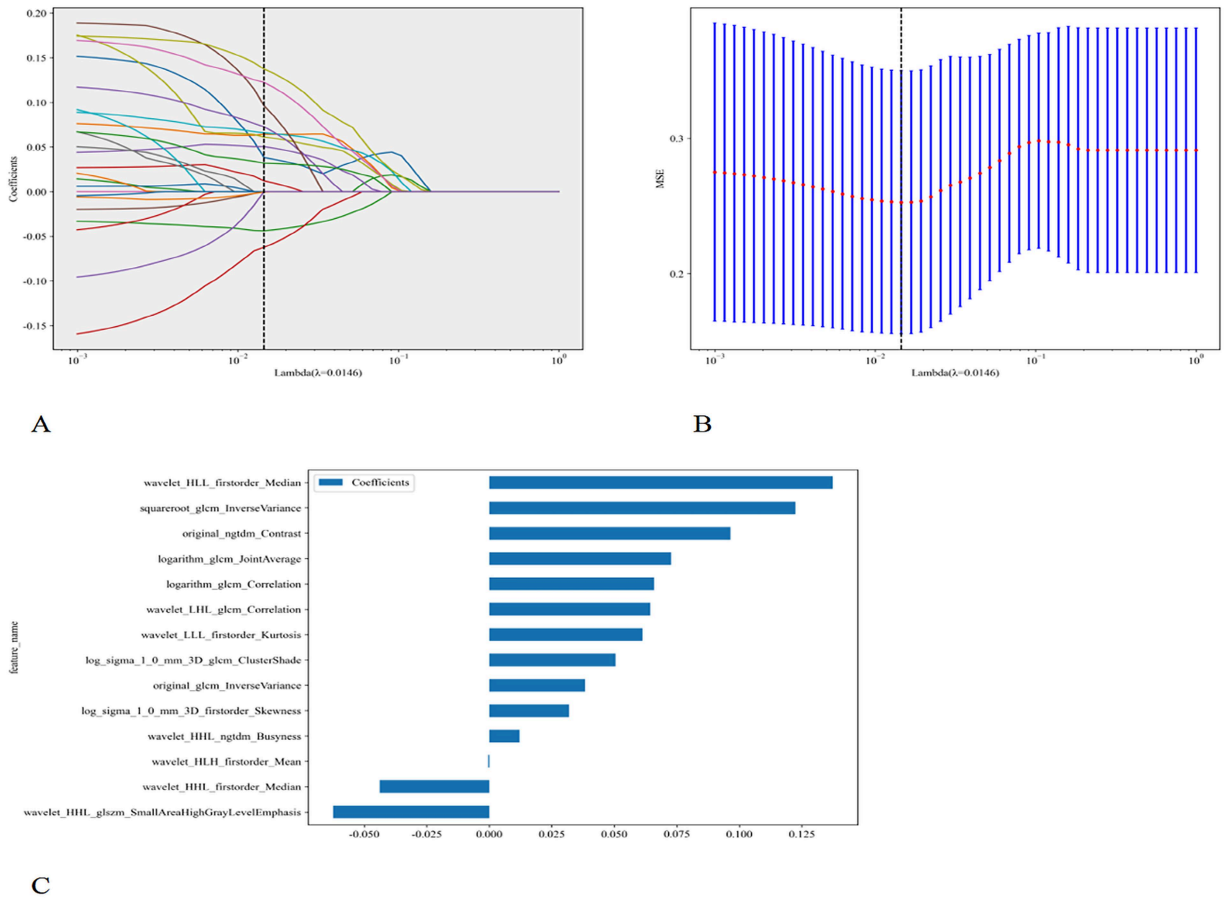


FIGURE 2
Panel A shows a line plot with multiple colored coefficient paths shrinking towards zero as the penalty parameter lambda increases, with a vertical dashed line marking lambda selection. Panel B contains a mean squared error plot versus lambda, with error bars and a dashed line at the selected lambda. Panel C displays a horizontal bar chart ranking feature names by coefficient values, with the highest absolute values at the top and bottom.

Model development and interpretability analysis

Radiomics features in the training cohort were standardized before further analysis. For each pair of redundant features with an absolute Pearson correlation coefficient greater than 0.9, one feature was retained. Least absolute shrinkage and selection operator regression with cross-validation was then used to identify features with nonzero coefficients; 14 features were retained. Based on these features, five machine-learning classifiers were developed: logistic regression, naive Bayes, Extra Trees, Light Gradient Boosting Machine (LightGBM), and multilayer perceptron.

Univariable and multivariable logistic regression analyses were used to identify independent clinical predictors and construct the clinical model. These predictors were then integrated with radiomics features to develop the combined model. The best-performing radiomics classifier was further interpreted using SHapley Additive exPlanations (SHAP) summary and feature-importance analyses together with representative case-level visualizations.

Statistical analysis

Continuous variables were expressed as mean and standard deviation, and categorical variables were expressed as counts and percentages. Group comparisons were performed using the chi-square test, Fisher exact test, or appropriate parametric or nonparametric tests. Model performance was evaluated using AUC, accuracy, sensitivity, specificity, positive predictive value, and negative predictive value. A two-sided p value < 0.05 was considered statistically significant.

Results

Patient characteristics

A total of 274 patients were included across the training, test, and external validation cohorts. Baseline characteristics were generally similar between the training and test cohorts, with significant differences only in cessation of defecation ($p = 0.035$) and the feces sign ($p = 0.038$). Compared with the training cohort, the external validation cohort differed in respiratory rate ($p < 0.001$), cause of obstruction ($p = 0.017$), colicky pain ($p = 0.023$), continuous pain ($p < 0.001$), mild tenderness ($p = 0.001$), rebound tenderness ($p < 0.001$), decreased enhancement ($p = 0.049$), and mesenteric edema ($p = 0.006$), indicating moderate intercenter heterogeneity.

Performance of radiomics models

In the training cohort, LightGBM showed the best overall performance among the radiomics models, with an accuracy of 0.877, an AUC of 0.934, a sensitivity of 0.937, a specificity of 0.793, a positive predictive value of 0.862, and a negative predictive value of 0.902. Extra Trees ranked second with an AUC of 0.917, whereas logistic regression, multilayer perceptron, and naive Bayes yielded AUCs of 0.890, 0.877, and 0.811, respectively.

In the test cohort, LightGBM again achieved the highest AUC (0.870), indicating good internal generalizability, whereas naive Bayes achieved the highest accuracy (0.817). The AUCs of Extra Trees, logistic regression, and multilayer perceptron were 0.814, 0.793, and 0.781, respectively.

In the external validation cohort, naive Bayes showed the best performance, with an AUC of 0.825 and an accuracy of 0.803, whereas LightGBM achieved an AUC of 0.816 and an accuracy of 0.789. Overall, LightGBM was the most stable model across the development cohorts, whereas naive Bayes performed slightly better in external validation.

Independent clinical predictors and model comparison

Univariable analysis identified age, systolic blood pressure, diastolic blood pressure, white blood cell count, cause of obstruction, vomiting, nausea, peritoneal fluid, colicky pain, continuous abdominal pain, bowel diameter, bowel wall thickening, mesenteric fluid, the feces sign, cessation of defecation, cessation of flatus, and decreased enhancement as variables associated with high-risk status.

In multivariable logistic regression, only diastolic blood pressure and bowel wall thickening remained significant. Lower diastolic blood pressure was associated with high-risk status (OR, 0.955; 95% CI, 0.924-0.988; $p = 0.024$), whereas bowel wall thickening was an independent positive predictor (OR, 1.486; 95% CI, 1.188-1.861; $p = 0.004$).

In the training cohort, the AUCs of the clinical, radiomics, and combined models were 0.848, 0.934, and 0.942, respectively. The combined model performed best, with an accuracy of 0.877, a sensitivity of 0.887, and a specificity of 0.862.

In the test cohort, the AUCs of the clinical, radiomics, and combined models were 0.739, 0.870, and 0.867, respectively. The combined model achieved the highest accuracy (0.817), whereas the radiomics model had the highest AUC.

In the external validation cohort, the AUCs of the clinical, radiomics, and combined models were 0.731, 0.803, and 0.803, respectively. The combined model yielded the highest accuracy (0.833) and sensitivity (0.946), suggesting greater safety for screening high-risk patients (Figure 3).

Calibration plots showed that the predicted probabilities of the clinical, radiomics, and combined models generally followed the observed event rates in the training, internal test, and external validation cohorts (Appendix Figure 6, panels D-F), although

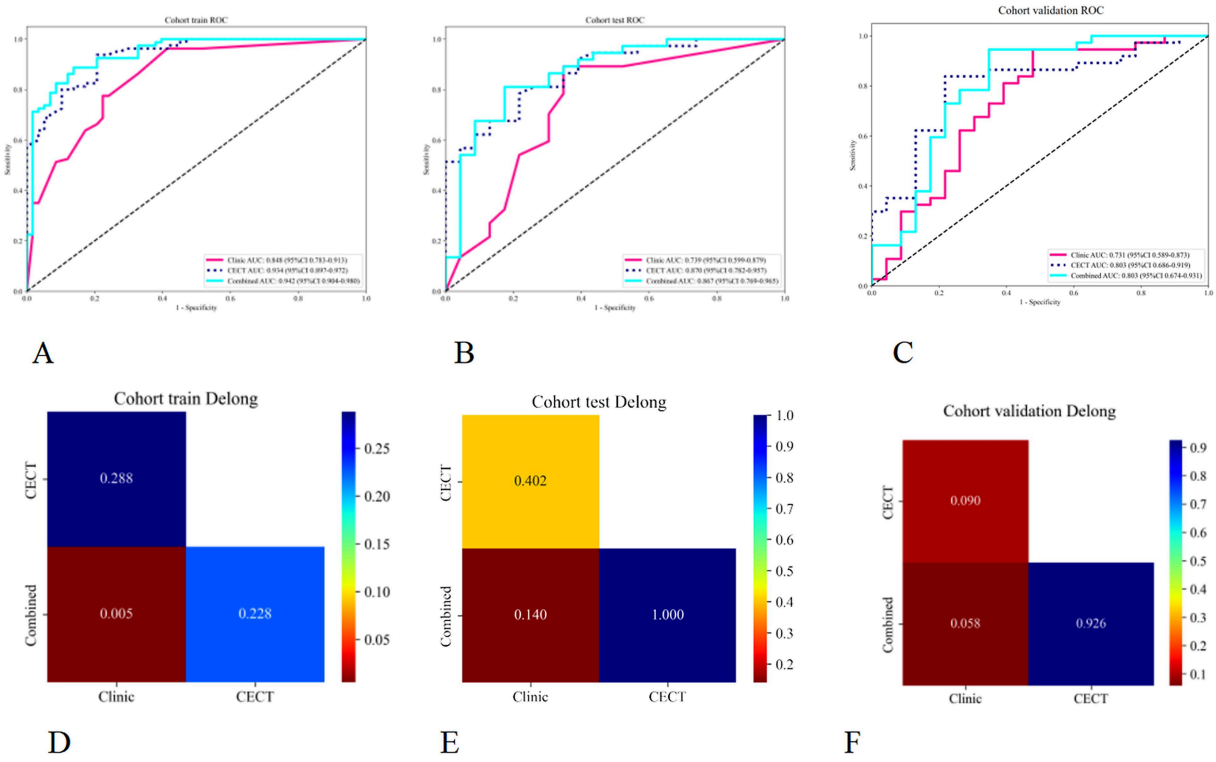


FIGURE 3
Panel A shows a line chart ROC curve for cohort train with AUC values for Clinic, Radiomics, and Combined models. Panel B shows a ROC curve for cohort test. Panel C shows a ROC curve for cohort validation. Panels D, E, and F display color-coded Delong test heatmaps for cohort train, test, and validation, with values comparing Clinic, Radiomics, and Combined models.

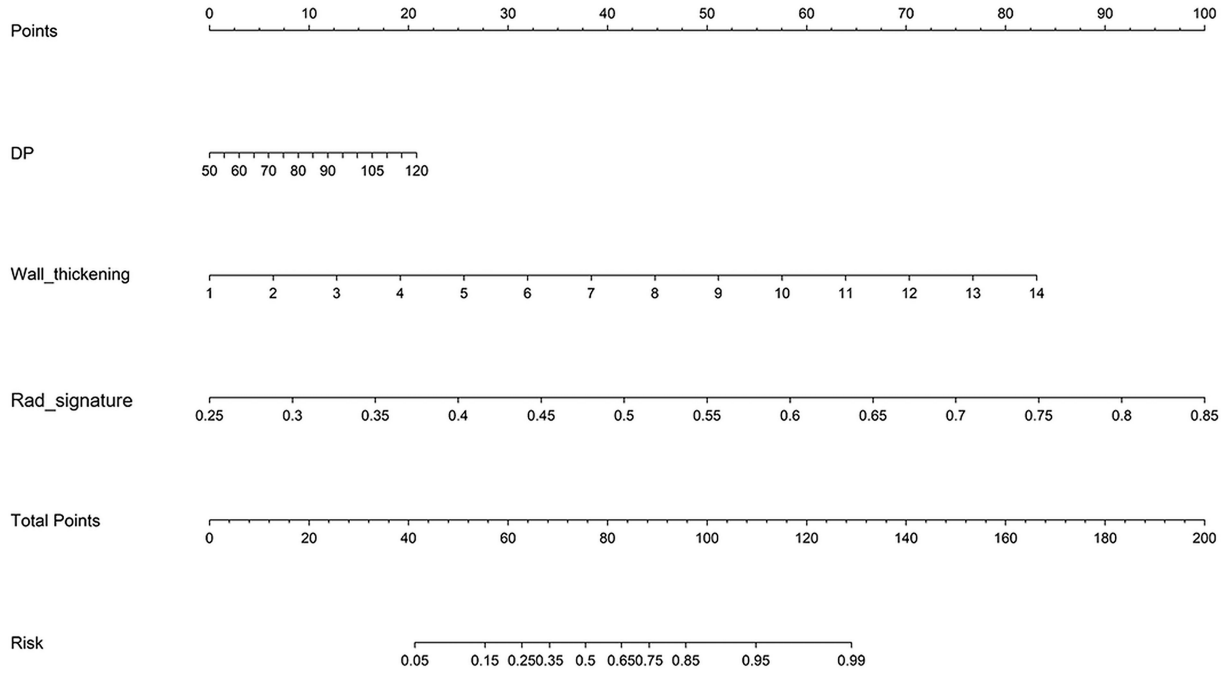


FIGURE 4
Nomogram graphic presenting six horizontal axes labeled Points, DP, Wall_thickening, Rad_signature, Total Points, and Risk, each with corresponding numerical scales arrayed for medical or statistical risk estimation purposes.

some deviation was apparent in individual probability bins, particularly in the smaller validation cohorts. These findings indicate that the models provide potentially informative risk estimates, but the probabilities should be interpreted as decision-support values rather than fixed treatment thresholds until prospective recalibration is performed (Figure 4).

Model interpretability

SHAP analysis identified wavelet_LHL_glcm_Correlation, wavelet_HLL_firstorder_Median, and logarithm_glcm_Correlation as the three most influential radiomics features. Global distribution plots showed clear differences in key texture features between the high-risk and low-risk groups. Representative-case analyses further showed that the model correctly classified both high-risk and low-risk patients while providing feature-level explanations for each prediction (Figure 5).

Discussion

The principal finding of this study is that CT-based radiomics models can effectively distinguish patients with SBO at high risk, whereas the combined model showed the most clinically useful balance of performance in external validation. Specifically, the combined model achieved AUCs of 0.942, 0.867, and 0.803 in the training, test, and external validation cohorts, respectively, and reached an external-validation accuracy of 0.833 with a sensitivity of 0.946. These data suggest that adding key clinical information to radiomics did not simply improve apparent discrimination in the development cohort, but helped preserve detection of high-risk patients when the model was applied to more heterogeneous cases. Among the candidate algorithms, LightGBM was the most stable radiomics classifier in the training and test cohorts, with AUCs of 0.934 and 0.870, indicating strong adaptability to high-dimensional radiomics features and nonlinear relationships. Importantly, the present study did not define high-risk status only by a single imaging sign or a narrow pathologic endpoint; instead, it reflected a clinically meaningful group of patients who underwent surgery or ultimately required surgical intervention after failed conservative management. From a translational perspective, this framing is relevant because the central question in daily care is often not simply whether one CT sign of ischemia is present, but whether the overall patient is likely to benefit from escalation of management.

These findings are broadly consistent with prior CT-based studies. Previous reports have shown that bowel wall thickening, decreased enhancement, mesenteric abnormalities, and ascites are closely associated with strangulation, ischemia, and failure of conservative management (9–14). Those studies established the current framework for clinical imaging assessment, but they still rely largely on subjective recognition of visible findings and therefore remain limited by interobserver variability and incomplete characterization of subtle intralesional heterogeneity. Our results suggest that radiomics may provide incremental information beyond conventional CT signs, particularly for patients whose routine imaging findings are not yet overtly alarming despite genuinely high risk. One plausible explanation is that radiomics captures spatial complexity, grayscale variation, and textural relationships that may reflect microstructural disorganization, edema, perfusion disturbance, or altered bowel wall composition before those processes become obvious on routine visual inspection. In this sense, radiomics should not be

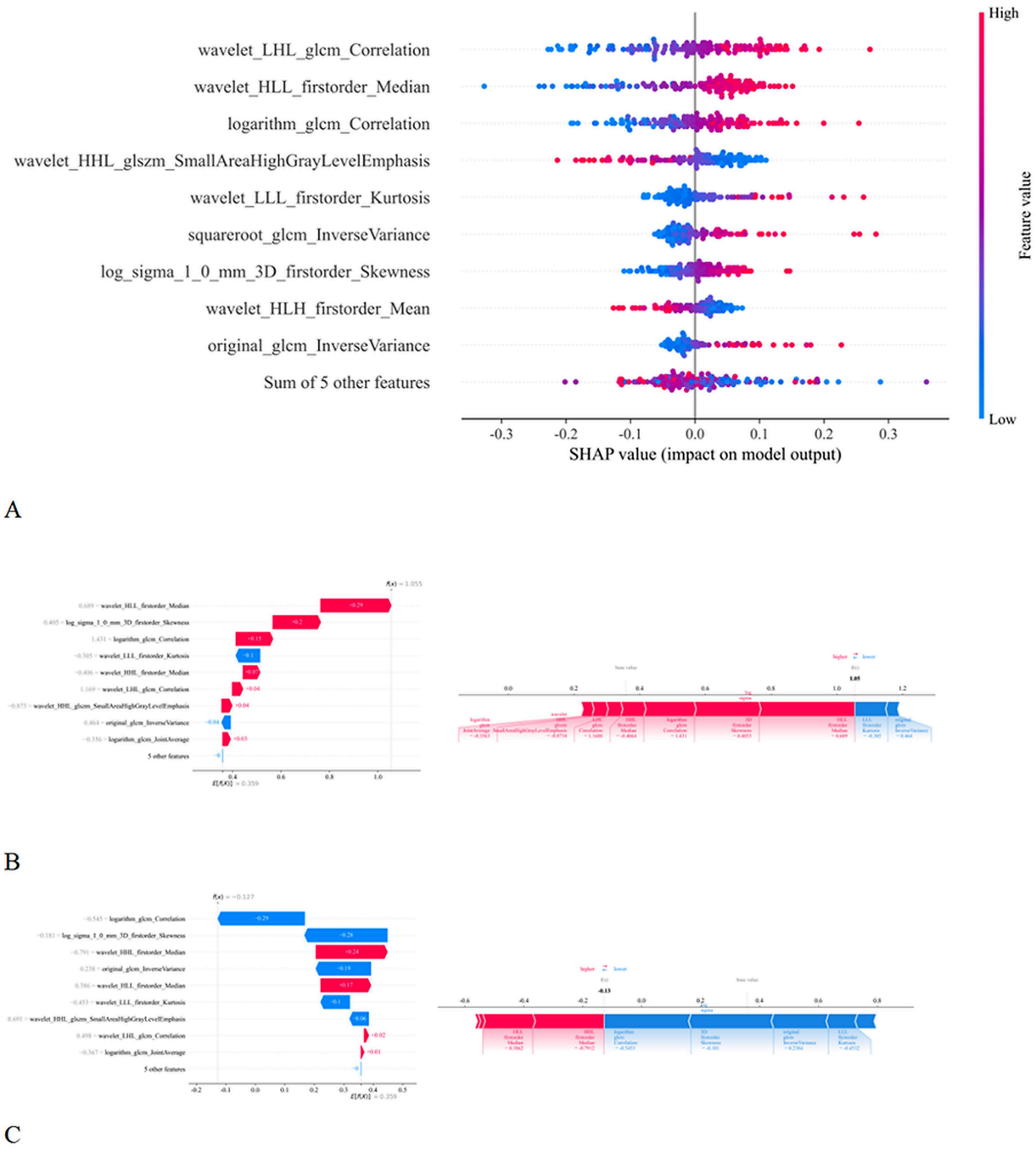


FIGURE 5
Panel A displays a SHAP summary plot of feature importance with individual feature values shown in blue to red, while panels B and C show SHAP force and bar plots for individual predictions, illustrating positive and negative feature impacts on model output.

viewed as replacing conventional abdominal CT interpretation, but as extending it with reproducible quantitative descriptors that can enrich clinical judgment in equivocal cases.

Earlier prediction tools relied mainly on clinical variables and visually assessed CT signs. In 233 consecutive patients, Schwenter et al. (23) developed a six-item clinicoradiological score for ischemia requiring bowel resection; a score of at least 3 yielded an AUC of 0.87, sensitivity of 67.7%, and specificity of 90.8%. Although clinically interpretable, that score was derived from conventional variables in a single cohort and had only moderate sensitivity. Paisant et al. (14) subsequently evaluated the Angers CT score prospectively across 15 centers and found that it was associated with failure of conservative management. However, the score remained dependent on predefined visual CT signs and was designed specifically for adhesive SBO. The present study extends these efforts by quantifying intralesional image heterogeneity, comparing multiple classifiers, and testing the models in an external cohort from separate institutions.

Machine-learning studies have addressed related but narrower endpoints. Goyal et al. (21) combined CT and clinical findings in 223 surgically evaluated patients with suspected closed-loop SBO; their internally split random-forest model achieved an AUC of

10.3389/fmed.2026.1848334

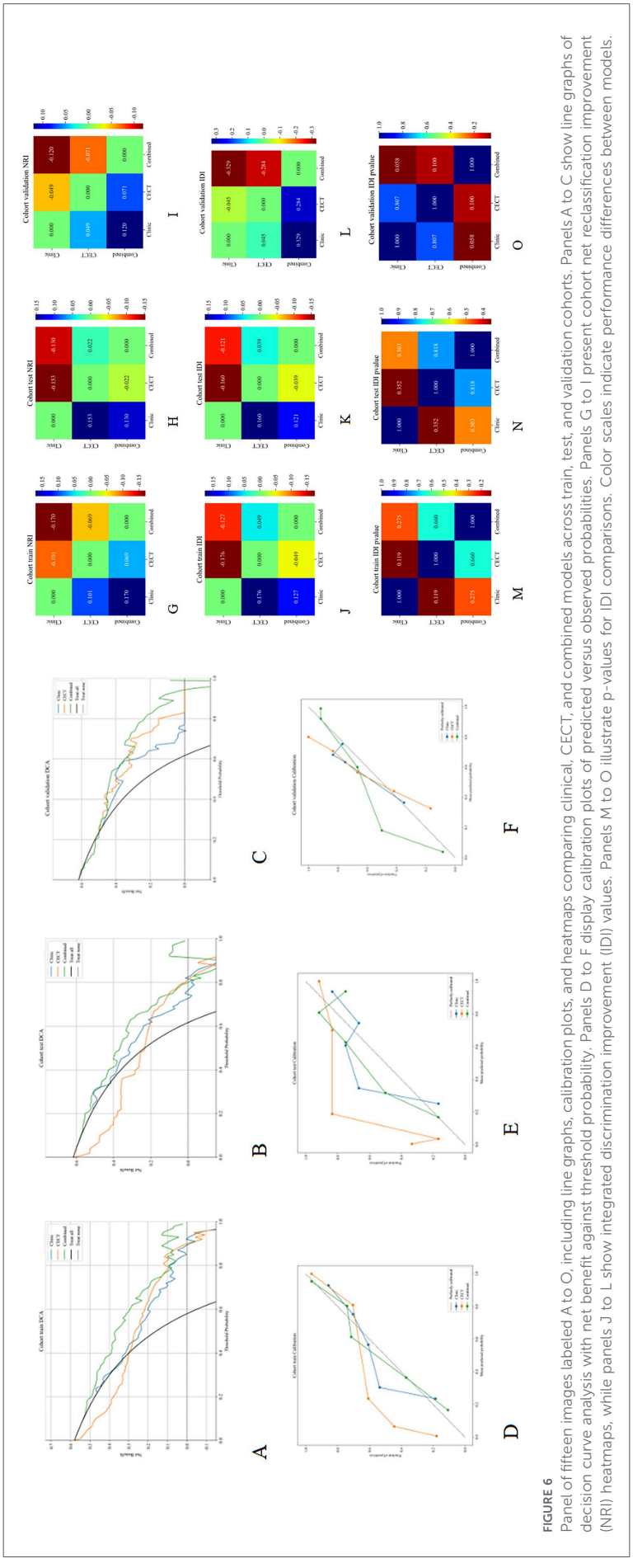


FIGURE 6
Panel of fifteen images labeled A to O, including line graphs, calibration plots, and heatmaps comparing clinical, CECT, and combined models across train, test, and validation cohorts. Panels A to C show line graphs of decision curve analysis with net benefit against threshold probability. Panels D to F display calibration plots of predicted versus observed probabilities. Panels G to I present cohort net reclassification improvement (NRI) heatmaps, while panels J to L show integrated discrimination improvement (IDI) values. Panels M to O illustrate p-values for IDI comparisons. Color scales indicate performance differences between models.

0.73, but the surgical case spectrum and absence of independent external validation limited generalizability. The prospective multicenter STRISK and NOFA scores focused on immediate surgery and failure of nonoperative treatment in adhesive SBO (26), improving clinical triage but not assessing high-dimensional CT texture. Bunino et al. (27) reported differences in entropy, uniformity, and energy between conservative and operative groups, providing preliminary evidence for SBO radiomics; however, that single-center study included only 55 patients and did not externally validate a multialgorithm model. In contrast, the present study includes multiple SBO causes, directly compares five algorithms, adds SHAP-based global and patient-level explanations, and evaluates performance in an external cohort. The variation in algorithm ranking between internal and external cohorts further shows why independent validation is necessary rather than relying on development-set performance alone.

The study (28) addressed an overlapping question but had important design differences. Wang et al. (29) analyzed 188 patients with adhesive SBO from a single hospital, all of whom underwent surgery, and used an internally split training/test design to predict bowel resection with a radiomic nomogram. The present study addresses a broader triage endpoint—surgery or failure of conservative management versus successful conservative treatment—across adhesive obstruction, internal hernia, and inguinal hernia; it compares five classifiers, evaluates clinical, radiomics, and combined models in parallel, and tests generalizability in an independent cohort from two additional hospitals. SHAP also provides global and case-level explanations.

The multicenter structure of our study also deserves emphasis because the external cohort was not merely a random hold-out subset, but a clinically different population from separate hospitals. Compared with the training cohort, the external validation cohort showed significant differences in respiratory rate, cause of obstruction, colicky pain, continuous pain, mild tenderness, rebound tenderness, decreased enhancement, and mesenteric edema, indicating meaningful intercenter heterogeneity. Against this background, some decrease in discrimination was expected when moving from development to external testing. Even so, the combined model maintained an AUC of 0.803 and achieved the highest external accuracy (0.833) and sensitivity (0.946), while the radiomics-only model had the same external AUC but lower accuracy. These findings suggest that the model captured patterns with practical generalizability rather than only center-specific noise. In other words, the value of the present study lies not in reporting an unusually high single-center AUC, but in showing that clinically useful risk stratification can be retained under more heterogeneous multicenter conditions.

Multivariable analysis provided additional clinically relevant insights that also help explain why the combined model performed better than the clinical model alone. Bowel wall thickening remained an independent predictor with an odds ratio of 1.486 (95% CI, 1.188-1.861; $p = 0.004$), which has clear radiologic-pathologic plausibility because it may reflect bowel wall edema, impaired venous return, and progressive ischemic injury. In contrast, lower diastolic blood pressure was independently associated with high-risk status, with an odds ratio of 0.955 (95% CI, 0.924-0.988; $p = 0.024$), suggesting that the transition to a high-risk state is influenced not only by local morphologic change but also by systemic perfusion status. This interpretation is supported by the model comparison results: in the external validation cohort, the clinical model achieved an AUC of 0.731, whereas the combined model reached an AUC of 0.803 with markedly higher sensitivity. Thus, the data from our study indicate that bedside physiologic information and image-derived features should be interpreted together rather than as competing sources of evidence.

Notably, the combined model did not yield a clearly higher AUC than the radiomics model in external validation, but it improved the metrics that may matter most in emergency decision-making. In the external cohort, the combined and radiomics models both yielded an AUC of 0.803, yet the combined model achieved higher accuracy (0.833) and a sensitivity of 0.946, indicating fewer missed high-risk cases. From a clinical standpoint, this distinction is important. For triage problems such as SBO, a model that misses fewer patients who will ultimately require surgery may be more useful than one that yields only a slightly higher summary discrimination metric. Our data therefore suggest that the combined model may be better suited for identifying patients who need closer observation, earlier surgical consultation, or a lower threshold for intervention, even when its AUC is not dramatically different from that of the radiomics-only approach.

Interpretability remains a central issue in the clinical translation of radiomics and machine learning (19, 20). Through SHAP analysis, we identified wavelet_LHL_glcm_Correlation, wavelet_HLL_firstorder_Median, and logarithm_glcm_Correlation as the three most influential radiomics features and visualized their contributions at both the population and individual levels. These features describe local textural correlation, grayscale-distribution characteristics, and heterogeneity and may be related to bowel wall disorganization, perfusion disturbance, or increased tissue complexity in compromised segments. However, this biologic interpretation is hypothesis-generating because no direct histopathologic or perfusion reference standard was available. Future studies correlating these radiomics features with resection histopathology, perfusion markers, and standardized CT signs are needed to confirm their clinical meaning. Accordingly, SHAP improves transparency regarding how the model reached a prediction, but it does not by itself establish the underlying tissue mechanism.

At the point of care, feature-level SHAP explanations could allow radiologists and surgeons to determine whether a high-risk prediction is driven by clinically plausible image patterns, compare the model output with conventional CT findings, and identify discordant cases that warrant additional review. This traceability may increase confidence in using the model as decision support while preserving clinician oversight rather than treating the algorithm as an autonomous recommendation.

Another point worth emphasizing is that external validation in our study also revealed meaningful differences in algorithm behavior. LightGBM was the most stable radiomics model in the development cohorts, but naive Bayes showed the best external AUC at 0.825 with an accuracy of 0.803. This shift indicates that algorithm ranking was sensitive to changes in case mix and data

distribution across centers. Rather than viewing this as a weakness, we consider it a realistic finding for medical artificial intelligence research. It suggests that no single classifier should be assumed to be universally optimal on the basis of internal performance alone and highlights the importance of comparing multiple algorithms within the same dataset. Future studies should therefore focus not only on improving discrimination, but also on calibration, threshold selection, workflow integration, and prospective evaluation of whether model-assisted care can reduce delays in recognizing patients who need urgent intervention.

Clinical implementation will require attention to workflow speed. Manual segmentation by experienced radiologists is feasible for retrospective model development but may be impractical in time-sensitive emergency settings because it adds operator time and may introduce interobserver variation. A pragmatic deployment pathway would use the model as a decision-support layer after rapid CT interpretation, with semi-automated or automated bowel-segment localization and contouring followed by radiologist review of the proposed region of interest. Prospective studies should quantify segmentation time, correction burden, failure rates, and model performance across centers before routine use.

Although the independent external cohort tested transportability across institutions with different case mix, formal subgroup stability by sex, age, and etiology remains to be quantified from patient-level predictions. Estimates in smaller etiologic strata, particularly the inguinal hernia subgroup, may be imprecise and should be interpreted as exploratory.

This study has several limitations. First, the retrospective design introduces selection bias; in particular, excluding examinations with severe artifacts, inadequate enhancement, or incomplete bowel coverage may underrepresent technically challenging real-world cases. Second, the sample size remained limited, and heterogeneity between the training and external validation cohorts may have affected model stability. Third, the workflow relied mainly on traditional machine-learning methods and manual lesion delineation rather than deep learning, automated segmentation, or longitudinal clinical variables. Fourth, the outcome was defined by the need for surgery or failure of conservative management. Although operative decisions were reviewed against prespecified clinical and imaging indications, surgeon preference, institutional protocols, bed availability, comorbidity, and patient preference may have varied across centers and introduced practice-pattern variation into outcome labeling. Fifth, the proposed biologic interpretation of the leading radiomics features was not validated against histopathology. Larger prospective multicenter studies using standardized surgical indications, explicit timing measures, prespecified image-quality criteria, and histopathologic correlation are needed to validate the model and quantify its clinical decision impact.

Conclusion

In conclusion, CT-based radiomics showed good ability to identify patients with SBO at high risk. A combined model incorporating radiomics and key clinical predictors improved external validation accuracy and sensitivity and may help guide early triage and intervention. If further validated prospectively, this type of interpretable imaging biomarker could support more standardized communication between radiologists and surgeons, improve prioritization of patients who warrant closer observation, and contribute to safer decision-making in a time-sensitive emergency setting.

Data availability statement

The data analyzed in this study is subject to the following licenses/restrictions: The data are not publicly available due to patient privacy concerns. Requests for access should be directed to the corresponding authors. Requests to access these datasets should be directed to sara_good@126.com.

Ethics statement

The studies involving humans were approved by the Ethics Committee of Quzhou People's Hospital, the Ethics Committee of Quzhou Traditional Chinese Medicine Hospital, and the Ethics Committee of Quzhou Kecheng District People's Hospital. The studies were conducted in accordance with the local legislation and institutional requirements. The ethics committee/institutional review board waived the requirement of written informed consent for participation from the participants or the participants' legal guardians/next of kin because of the retrospective design of the study. Written informed consent from the individual(s) for the publication of any potentially identifiable images or data included in this article was not required because of the retrospective design of the study.

Author contributions

GZ: Resources, Funding acquisition, Project administration, Formal analysis, Data curation, Investigation, Writing – original draft, Conceptualization, Methodology. XL: Formal analysis, Writing – original draft, Investigation, Data curation. WH: Supervision, Methodology, Investigation, Writing – original draft. XZ: Writing – original draft, Validation, Conceptualization, Visualization, Supervision. CX: Writing – original draft, Resources, Validation, Data curation. HX: Writing – original draft, Data curation, Validation, Supervision, Resources. YW: Methodology, Formal analysis, Software, Data curation, Writing – original draft. XH: Formal analysis, Writing – review & editing, Conceptualization, Methodology. DW: Writing – review & editing, Funding acquisition, Project administration, Methodology, Conceptualization, Investigation.

Funding

The author(s) declared that financial support was received for this work and/or its publication. This study was supported by the Science and Technology Project of Quzhou (2022K65, 2023K118) and the High-Level Talents Project of Zhejiang (2020-167).

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Zins M, Millet I, Taourel P. Adhesive small bowel obstruction: predictive radiology to improve patient management. *Radiology*. (2020) 296(3):480–92. doi: 10.1148/radiol.2020192234

2. Ten Broek RPG, Krielen P, Di Saverio S, Coccolini F, Biffl WL, et al. Bologna guidelines for diagnosis and management of adhesive small bowel obstruction (ASBO): 2017 update of the evidence- based guidelines from the world society of emergency surgery ASBO working group. *World J Emerg Surg*. (2018 Jun 19) 13:24. doi: 10.1186/s13017-018-0185-2

3. Paulson EK, Thompson WM. Review of small-bowel obstruction: the diagnosis and when to worry. *Radiology*. (2015) 275(2):332–42. doi: 10.1148/radiol.15131519

4. Sallinen V, Di Saverio S, Haukijärvi E, Juusela R, Wikström H, Koivukangas V, et al. Laparoscopic versus open adhesiolysis for adhesive small bowel obstruction (LASSO). *Lancet Gastroenterol Hepatol*. (2019) 4(4):278–86. doi: 10.1016/S2468-1253(19)30016-0

5. Hajibandeh S, Hajibandeh S, Panda N, Khan RMA, Bandyopadhyay SK, Dalmia S, et al. Operative versus non-operative management of adhesive small bowel obstruction: a systematic review and meta-analysis. *Int J Surg*. (2017) 45:58–66. doi: 10.1016/j.ijsu.2017.07.073

6. Peacock O, Bassett MG, Kuryba A, Walker K, Davies E, Anderson I, et al. Thirty-day mortality in patients undergoing laparotomy for small bowel obstruction. *Br J Surg*. (2018) 105(8):1006–13. doi: 10.1002/bjs.10812

7. Fung BSC, Behman R, Nguyen M-A, Nathens AB, Look Hong NJ, Pechlivanoglou P, et al. Longer trials of non-operative management for adhesive small bowel obstruction are associated with increased complications. *J Gastrointest Surg*. (2020) 24(4):890–8. doi: 10.1007/s11605-019-04156-6

8. Lau HT, Liu W, Lam V, Pang T. Early routine versus selective computed tomography for acute abdominal pain: a systematic review and meta-analysis of randomised trials. *Int J Surg*. (2022) 101(106622):106622. doi: 10.1016/j.ijsu.2022.106622

9. Millet I, Ruyer A, Alili C, Curros Doyon F, Molinari N, Pages E, et al. Adhesive small-bowel obstruction: value of CT in identifying findings associated with the effectiveness of nonsurgical treatment. *Radiology*. (2014) 273(2):425–32. doi: 10.1148/radiol.14132872

10. Chuong AM, Corno L, Beaussier H, Boulay-Coletta I, Millet I, Hodel J, et al. Assessment of bowel wall enhancement for the diagnosis of intestinal ischemia in patients with small bowel obstruction. *Radiology*. (2016) 280(1):98–107. doi: 10.1148/radiol.2016151029

11. Guerrini J, Zugna D, Poretti D, Samà L, Costa G, Mei S, et al. Adhesive small bowel obstruction: single band or matted adhesions? A predictive model based on computed tomography scan. *J Trauma Acute Care Surg*. (2021) 90(6):917–23. doi: 10.1097/TA.0000000000003182

12. de Kok BM, Toneman MK, Oei S, Westerterp M, Van Acker GJD, Van Der Pool AEM, et al. Correlation of CT findings with intra-operative outcome in closed-loop small bowel obstruction. *Eur J Radiol*. (2021) 142(109844):109844. doi: 10.1016/j.ejrad.2021.109844

13. Millet I, Taourel P, Ruyer A, Molinari N. Value of CT findings to predict surgical ischemia in small bowel obstruction: a systematic review and meta-analysis. *Eur Radiol*. (2015) 25(6):1823–35. doi: 10.1007/s00330-014-3440-2

14. Paisant A, Burgmaier J, Calame P, Loison M, Molière S, Brigand C, et al. The angers CT score is a risk factor for the failure of conservative management of adhesive small bowel obstruction. *World J Surg*. (2023) 47(4):975–84. doi: 10.1007/s00268-023-06906-9

15. Bi WL, Hosny A, Schabath MB, Giger ML, Birkbak NJ, Mehrtash A, et al. Artificial intelligence in cancer imaging: clinical challenges and applications. *CA Cancer J Clin*. (2019) 69(2):127–57. doi: 10.3322/caac.21552

16. Scapicchio C, Gabelloni M, Barucci A, Cioni D, Saba L, Neri E. A deep look into radiomics. *Radiol Med*. (2021) 126(10):1296–311. doi: 10.1007/s11547-021-01389-x

17. Song SH, Park H, Lee G, Lee HY, Sohn I, Kim HS, et al. Imaging phenotyping using radiomics to predict micropapillary pattern within lung adenocarcinoma. *J Thorac Oncol*. (2017) 12(4):624–32. doi: 10.1016/j.jtho.2016.11.2230

18. Cui Y, Zhang J, Li Z, Wei K, Lei Y, Ren J, et al. A CT-based deep learning radiomics nomogram for predicting response to neoadjuvant chemotherapy in locally advanced gastric cancer. *eClinicalMed*. (2022) 46(101348):101348. doi: 10.1016/j.eclinm.2022.101348

19. Wang Y, Lang J, Zuo JZ, Dong Y, Hu Z, Xu X, et al. The radiomic-clinical model using the SHAP method for assessing treatment response of whole-brain radiotherapy: a multicentric study. *Eur Radiol*. (2022) 32(12):8737–47. doi: 10.1007/s00330-022-08887-0

20. Yang Z, Gong J, Li J, Sun H, Pan Y, Zhao L. The gap before real clinical application of imaging-based machine-learning and radiomic models for chemoradiation outcome prediction in esophageal cancer: a systematic review and meta-analysis. *Int J Surg*. (2023) 109(8):2451–66. doi: 10.1097/JS9.0000000000000441

21. Goyal R, Mui LW, Riyahi S, Prince MR, Lee HK. Machine learning based prediction model for closed-loop small bowel obstruction using computed tomography and clinical findings. *J Comput Assist Tomogr*. (2022) 46(2):169–74. doi: 10.1097/RCT.0000000000001269

22. Huang P, Yang J, Zhao D, Ran T, Luo Y, Yang D, et al. Machine learning-based prediction of early complications following surgery for intestinal obstruction: multicenter retrospective study. *J Med Internet Res*. (2025) 27:e68354. doi: 10.2196/68354

23. Schwenter F, Poletti PA, Platon A, Perneger T, Morel P, Gervaz P. Clinicoradiological score for predicting the risk of strangulated small bowel obstruction. *Br J Surg*. (2010) 97(7):1119–25. doi: 10.1002/bjs.7037

24. Murasaki M, Nakanishi T, Kano K-, Shigemi R, Tanizaki S, Kono K, et al. Point-of-care procalcitonin may predict the need for surgical treatment in patients with small bowel obstruction. *Am J Emerg Med*. (2020) 38(5):979–82. doi: 10.1016/j.ajem.2020.02.022

25. Cho YJ, Park IS, Kim J, Cho HJ, Gwak GH, Yang KH, et al. Factors predicting the need for early surgical intervention for small bowel obstruction. *Ann Coloproctol*. (2020) 36(4):223–8. doi: 10.3393/ac.2019.09.30

26. Kupietzky A, Dodi O, Dover R, Lourie NEE, Berrebi Y, Lev-Cohain N, et al. Existing scores fail to predict bowel ischemia in patients with adhesive small bowel obstruction. *J Surg Res*. (2023 Mar) 283:416-422. doi: 10.1016/j.jss.2022.10.056

27. Friziero A, Rosso E, Zuin IS, Vallese L, Serafini S, Amico A, et al. Neutrophil to lymphocyte ratio predicts bowel ischemia in non-strangulated adhesive small bowel occlusions: a retrospective analysis from an acute care surgical service. *BMC Surg*. (2024 Jun 12) 24(1):179. doi: 10.1186/s12893-024-02476-2

28. Zein L, Calame P, Chausset C, Doussot A, Turco C, Malakhia A, et al. Preoperative prediction of inadvertent enterotomy during adhesive small bowel obstruction surgery using combination of CT features. *Eur Radiol*. (2022 Oct) 32(10):6646-6657. doi: 10.1007/s00330-022-08951-9

29. Wang Z, Zhu L, Liu S, Li D, Liu J, Zhou X, et al. Development and validation of a CT-based radiomic nomogram for predicting surgical resection risk in patients with adhesive small bowel obstruction. *BMC Med Imaging*. (2025 Feb 11) 25(1):46. doi: 10.1186/s12880-025-01575-7