

Comparison of Radiomics and Deep Learning Using Intestinal Ultrasound Images for Differentiating Inflammatory and Fibrotic Intestinal Strictures in Crohn's Disease

Runing Zhou^{1,*}, Jialin Shi^{2,*}, Jing Qin^{3,*}, Xiaowei Xue⁴, Weixun Zhou⁴, Xue Zhao², Xidong He¹, Li Ma³, Mengyuan Zhou³, Chenyi Guo⁵, Xiaoyin Bai¹, Qingli Zhu³, Hong Yang¹

¹Department of Gastroenterology, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, 100730, People's Republic of China; ²School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing, 100083, People's Republic of China; ³Department of Ultrasound, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, 100730, People's Republic of China; ⁴Department of Pathology, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, 100730, People's Republic of China; ⁵Department of Electronic Engineering, Tsinghua University, Beijing, 100084, People's Republic of China

*These authors contributed equally to this work

Correspondence: Xiaoyin Bai; Hong Yang, Department of Gastroenterology, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, People's Republic of China, Tel +86-10-69155005; +86-10-69155014, Email baixiaoyin@pumch.cn; yangh@pumch.cn

Background: Intestinal strictures represent a significant complication in Crohn's disease (CD), and accurate differentiation between inflammatory and fibrotic strictures is critical for effective treatment planning. Intestinal ultrasound (IUS) assists in diagnosing and monitoring strictures in CD, yet it faces limitations in distinguishing stricture types. This study aimed to compare radiomics-based and deep learning-based approaches for classifying bowel strictures using IUS images.

Methods: We conducted a retrospective analysis of 64 CD patients who underwent surgery between 2018 and 2023. A total of 87 ultrasound images were evaluated using radiomics and deep learning methodologies. We used 5-fold cross-validation, where in each fold, the training set was used for feature extraction and diagnostic model construction, and the test set was used to evaluate model performance. Model performances were assessed through accuracy, sensitivity, specificity, and the area under the ROC curve. Additionally, two experienced radiologists independently evaluated the nature of the strictures, and their assessments were compared with the models. Histopathological analysis via Masson's trichrome staining served as the reference standard.

Results: Results indicated the deep learning-based model had superior performance, achieving an accuracy of 83.3%, sensitivity of 88.9%, and specificity of 77.8%, outperforming the radiomics-based approach, which achieved an accuracy of 67.0%. Although expert evaluations provided the highest specificity (86.7%), the deep learning model demonstrated overall superior accuracy compared to both radiomics and expert predictions. Class Activation Mapping highlighted effective feature identification, enhancing discrimination between inflammatory and fibrotic strictures.

Conclusion: This pioneering research applies deep learning and radiomics to intestinal ultrasound for the first time to distinguish the types of bowel strictures. The real-time visualization of deep learning attention maps (class activation maps) effectively highlights discriminative pathological regions and enhances interpretability, demonstrating strong potential to support clinical decision-making in intestinal stricture classification.

Keywords: Crohn's disease, intestinal strictures, deep learning, radiomics, ultrasound imaging, artificial intelligence

Introduction

Crohn's disease (CD) is a kind of inflammatory bowel disease (IBD) characterized by transmural inflammation that can affect any segment of the gastrointestinal tract, with rising prevalence in developing countries including China.¹ Epidemiological studies reveal that 40–50% of Crohn's disease patients develop stricture-related complications within 10 years of diagnosis, rising to 70–80% after 20 years,^{2,3} while the 10-year cumulative surgical risk due to strictures reaches approximately 70%, highlighting the substantial patient and healthcare burden.⁴

While CD-related strictures can be inflammatory, fibrotic, or mixed, the accurate differentiation is critical for treatment decisions. Intestinal ultrasound (IUS) has emerged as a non-invasive tool for evaluating CD complications, with sensitivity and specificity rates of 85–95% for detecting strictures.^{5,6} IUS is a noninvasive approach with several advantages, including wide availability, convenience, and low cost, and is being increasingly promoted. Recent guidelines also support its utility in diagnosing bowel strictures.⁷ However, IUS faces limitations such as operator dependency, lack of standardized protocols, and variability in equipment, despite strain elastography potentially improving accuracy in distinguishing fibrotic from inflammatory strictures.^{8,9} These challenges highlight AI (artificial intelligence) 's potential to reduce operator bias and enhance IUS diagnostics, hence we perform this very first study to develop AI (deep learning, as automated feature extraction, and radiomics, as handcrafted features) aiding IUS-based stricture classification in CD. Radiomics captures features that are mathematically defined and can be linked to biology, while deep learning can discover complex features beyond human perception.^{10,11} The previous studies have predominantly relied on CTE or MRE findings, and there is a notable lack of research on AI models based on IUS for evaluating intestinal strictures in CD. And few study conducted in endoscopic images for stricture detecting, for its risk in leading to intestinal perforation.^{12,13} Moreover, existing studies are limited by small sample sizes and have primarily utilized AI to integrate clinical data with different ultrasound modalities for characterizing strictures, rather than employing true AI-based learning from image information to provide real-time diagnostic feedback.¹⁴

This single-center study aims to develop and validate radiomics and deep learning models to differentiate inflammatory and fibrotic strictures based on 87 IUS images from 64 CD patients, as the first study to apply two AI-based learning model in IUS images to differentiate fibrotic and inflammatory intestinal stricture.

Methods and Materials

Study Population

Patients with CD who underwent surgery and were hospitalized at Peking Union Medical College Hospital from January 1st, 2018, to December 31st, 2023, were included in the study. The specific inclusion and exclusion criteria were as follows. Inclusion criteria were: (1) CD diagnosis confirmed by Chinese IBD consensus¹⁵ and ECCO guidelines;⁵ (2) imaging evidence of strictures (luminal narrowing, bowel wall thickening, or pre-stricture dilation) via MR/CT enterography or IUS;^{3,16} and (3) age ≥ 18 years. Exclusion criteria were: (1) non-CD strictures (eg, malignancy, ischemia); (2) recent/multiple bowel resections; (3) pregnancy/lactation. Data collected included demographics (age, sex, disease duration), clinical variables (disease location, stricture number/location, prior surgeries), medication history (biologics, corticosteroids, immunosuppressants), and surgical details (stricture location/length, time from diagnosis to surgery).

Hematoxylin & Eosin (H&E) Staining and Masson's Trichrome Staining

Resected bowel specimens from CD patients are fixed in 10% neutral buffered formalin for 24 hours, embedded in paraffin, and sectioned into 4–5 μm slices. The most stenotic intestinal segment and adjacent areas are selected for staining. For H&E staining, sections are deparaffinized, rehydrated, stained with hematoxylin and eosin, dehydrated, cleared, and mounted. For Masson's trichrome staining, sections undergo similar preparation, followed by staining with Weigert's iron hematoxylin (nuclei), Biebrich scarlet-acid fuchsin (muscle/cytoplasm), and aniline blue (collagen). Slides are differentiated, dehydrated, cleared, and cover-slipped. 3D histech captured representative images, and ImageJ (Version 1.53. US National Institutes of Health, <https://imagej.net>) quantified collagen-stained tissue relative to total tissue area. The fibrosis staining area ratio was calculated for each specimen, and the median ratio was determined.

Patients were divided into two groups based on the median: those with Masson staining area above (severe fibrosis) or below the median. This binary classification aligns with methods used in prior literature.^{17,18} The diagnosing pathologist had access to clinical and imaging data necessary for standard diagnostic workflow but was blinded to the experimental ultrasound classifications. A second pathologist, performing the quantitative Masson's trichrome staining analysis, was fully blinded to all imaging data and calculated the collagen area ratio based solely on the HE-stained slides to ensure objective, research-specific assessment.

Ultrasonographic Examination and Definition of the ROI

Intestinal ultrasound (IUS) was conducted following the European Federation of Societies for Ultrasound in Medicine and Biology guidelines and Experts suggestions on the standardization of intestinal ultrasound examination and reporting for inflammatory bowel disease in China.¹⁹ A standardized, comprehensive intestinal scan was performed by one of three radiologists, each with over 10 years of experience, using a Philips iU22 (Philips Healthcare, Bothell, WA, USA) or SuperSonic Aixplorer (SuperSonic Imaging, SA, France) machine equipped with convex (C5-2) and linear (L9-3) transducers. Patients fasted for at least 8 h before US examination followed the instruction of gastroenterologists. And most of them typically follow a low-residue diet within a few days before screening or due to stricture situation, hence minimizing the interference caused by intestinal contents. A thorough scanning of the colon (from the ileocecal region to the sigmoid) and small intestine was performed with the convex transducer first. Then a detailed examination was performed by the linear transducer.

So far there is no universally accepted threshold for diagnosing intestinal stricture. The Stenosis Therapy and Anti-Fibrotic Research Consortium recommends the following ultrasound criteria for diagnosing small intestinal stricture in CD: bowel wall thickness ≥ 3 –4 mm, narrowed luminal diameter (< 1 cm), accompanied by proximal bowel loop dilation (> 2.5 cm).²⁰ Once diagnosed with intestinal stricture, the stricture segments are classified into three categories based on following ultrasonic features: fibrotic stricture, inflammatory stricture, and mixed-type. Fibrotic strictures are defined as distinct bowel wall stratification with minimal or absent vascularity, regardless of the bowel wall thickness. Inflammatory strictures are defined as loss of bowel wall stratification with long stretches of vascularity or vascularity reaching the mesentery regardless of the bowel wall thickness; or indistinct bowel wall stratification, long stretches of vascularity reaching the mesentery, regardless of the bowel wall thickness. For the clinical strategy and the characterization of US qualifications, the diagnostic value was calculated by comparing the fibrotic stricture group with the inflammatory stricture group.

Two radiologists (QJ, > 5 years' experience; ZQL, > 10 years' experience) independently reviewed the images, blinded to clinical, laboratory, and histopathological data. Disagreements were resolved by a third radiologist (WB.L, > 10 years' experience). Regions of interest (ROIs) were manually delineated on representative stricture images by a radiologist (MY.Z, 5 years' experience), blinded to histopathological findings, using ImageJ software.

Radiomics-Based Classification Method

The most representative images from each patient in the original dataset were selected, resulting in 87 images cropped to target regions. We used 5-fold cross-validation, where in each fold, the training set was used for feature extraction and diagnostic model construction, and the test set was used to evaluate model performance. The final results were averaged across the five folds to represent the overall model performance. Radiomics features are extracted using the PyRadiomics library in Python, enabling quantification of various imaging characteristics. All images were normalized to zero mean and unit variance prior to feature extraction to reduce inter-subject intensity variation. Image preprocessing followed the PyRadiomics configuration “binWidth” = 25, “force2D” = True, “interpolator” = sitk.sitkBSpline, “resampledPixelSpacing” = None. Here, a fixed gray-level bin width of 25 was used for intensity discretization. The force2D = True option ensured 2D feature extraction, consistent with the image dimensionality. B-spline interpolation was applied for any necessary image resampling to achieve smooth voxel transitions, while resampledPixelSpacing = None retained the native resolution when voxel spacing was uniform. ROIs were generated from doctor's annotation and verified using imageoperations.checkMask to ensure spatial alignment. The extracted features include first-order statistical features (eg, mean, standard deviation, etc.), shape features (eg, volume, area, etc.), texture features [such as

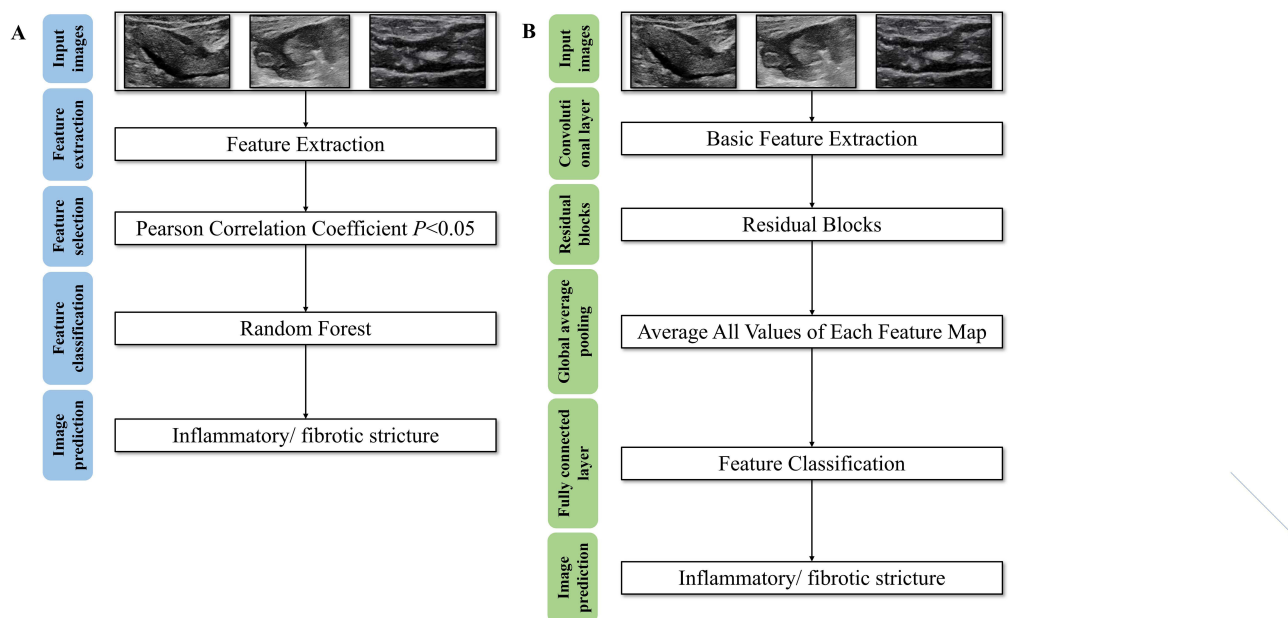


Figure 1 (A) The framework of the radiomics-based approach; (B) The framework for classifying inflammatory or fibrotic strictures based on the Resnet50 model.

grayscale covariance matrices (GLCM), grayscale four length matrices (GLRLM), and grayscale size-zone matrices (GLSZM) etc.]. Over 1100 radiomics features were calculated, with feature selection based on Pearson correlation and statistical significance ($p < 0.05$). A random forest classifier was employed to differentiate inflammatory from fibrotic strictures (Figure 1A). Model performance was assessed using accuracy, sensitivity, specificity, positive/negative predictive values, F1 score, confusion matrix, and AUC.

Deep Learning-Based Classification Model

ResNet50, a deep learning model for image recognition and classification, is a representative residual network designed to address gradient vanishing and network degradation in deep architectures. It enhances training efficiency through residual blocks, which facilitate rapid information transfer across layers. The model comprises convolutional layers, residual blocks, pooling layers, and fully connected layers. Convolutional layers extract features from input images, while residual blocks use shortcut connections to preserve and propagate information. Pooling layers reduce feature map dimensionality, mitigating overfitting and computational complexity. The fully connected layer outputs class scores, converted into probability distributions via the Softmax activation function (Figure 1B).

The ResNet50 model was subjected to 5-fold cross-validation using the same 87 images dataset used for radiomics analysis. To ensure strict test set independence, all images from the same patient were assigned to the same fold. Before training, all ultrasound images were first cropped to include only the ROI to focus on diagnostically relevant structures and minimize background noise. Subsequently, the images underwent a standardized preprocessing procedure. Each image was resized to 128*128 pixels to ensure consistent input dimensions across the dataset, and normalized to have a mean and standard deviation of 0.5 for each channel. Training employed an initial learning rate of 0.0001, adjusted periodically using step-down scheduling, with a batch size of 16 and 100 epochs. Model performance was evaluated on the test set using metrics including accuracy, sensitivity, specificity, positive/negative predictive values, F1 score, confusion matrix, and AUC.

Implementation Details Subsection

All experiments were conducted on an Ubuntu 24.04 operating system using a single Nvidia GeForce RTX 3090 GPU. The programming environment was based on Python version 3.8.19, and the deep learning framework used was PyTorch version 2.0.1.

Ethic Statement

Based on the Ethics Committee Guidelines of Peking Union Medical College Hospital for the Clinical Research Involving Human Subjects, this study is exempt from obtaining informed consent from the subjects, as such exemption does not negatively impact their rights and interests. The research utilizes identifiable human materials or data where the subjects can no longer be located, and the study does not involve personal privacy or commercial interests. Following review by the Ethics Committee (I-22PJ1092), this retrospective study has been deemed to meet the above criteria and is therefore exempt from the requirement of signed informed consent. And the study complies with the Declaration of Helsinki.

Sample Size and Statistical Analysis

The sample size was calculated based on an expected sensitivity of 80% for the diagnostic model,¹⁴ with a 95% confidence level and a confidence interval width of 0.20, yielding a minimum requirement of 61 patients. Based on this estimation for model sensitivity, we retrospectively enrolled 64 surgically confirmed CD patients from our institutional database who had undergone preoperative intestinal ultrasound. Continuous variables with a normal distribution are expressed as the mean \pm standard deviation (SD), while nonnormal variables are reported as medians (interquartile ranges [IQRs]). Categorical and discrete variables are presented as percentages. The means of two continuous normally distributed variables were compared using Student's *t*-test, and the Mann–Whitney *U*-test was applied to compare nonnormally distributed variables. The frequencies of categorical variables were compared using Pearson χ^2 or Fisher's exact test under specific conditions. $P < 0.05$ was considered to indicate statistical significance. Cohen's kappa (κ) coefficient was used to assess the agreement between the IUS findings of the two observers (Q.L.Z. and J. Q). The level of agreement was defined as poor ($\kappa < 0.20$), fair ($0.2 < \kappa \leq 0.40$), moderate ($0.4 < \kappa \leq 0.60$), good ($0.6 < \kappa \leq 0.80$) and very good ($0.8 < \kappa \leq 1.0$). All the statistical analyses were performed using SPSS (version 23.0; SPSS Inc., Chicago, IL, USA).

Results

Baseline Data for the Included Patients with CD

This study included 64 CD patients, with a median Masson's staining area of 40.10% (IQR: 35.55%–41.96%). The baseline characteristics of CD patients were presented in [Supplementary Table S1](#), grouped by Masson staining ratio, with significant differences observed between the groups (33.25% vs 47.29%, $P = 0.037$). Of 34 patients with small intestinal strictures, 24 (70.59%) were fibrotic based on Masson's staining; among 30 patients with colonic strictures, 17 (56.66%) were fibrotic. The high Masson staining group had a longer disease duration (9.65 ± 2.43 vs 7.66 ± 2.12 years, $P = 0.040$) and a longer interval from diagnosis to surgery (8.90 ± 3.12 vs 6.22 ± 1.41 years, $P = 0.047$) compared to the low staining group.

Model Performance on an Internal Independent Test Set

Radiomics-Based Method

In our experiments, we first classified intestinal stricture as fibrosis and inflammatory based on radiomics-based method. During training, using radiomics-based method, the classification has the accuracy of 67.0% (95% Confidence Interval [CI], 44.4%–88.9%), a sensitivity of 75.0% (95% CI, 40.0–100%), a specificity of 60.0% (95% CI, 28.6%–90.0%), a positive predictive value of 60.0% (95% CI, 27.3%–90.0%), a 75.0% (95% CI, 42.9%–100%) for negative predictive value, 67.0% (95% CI, 33.3%–90.0%) for F1 score, and 67.5% for AUC. We also show the importance of the features extracted by the radiomics-based approach. [Figure 2](#) shows the importance of features. Among that, the top 10 important features are `logarithm_InverseVariance`, `ShortRunLowGrayLevelEmphasis`, `gradient_SmallAreaHighGrayLevelEmphasis`, `wavelet-HLL_RunLengthNonUniformity`, `gradient_Autocorrelation`, `gradient_ShortRunLowGrayLevelEmphasis`, `gradient_RunLengthNonUniformity`, `wavelet-HLH_SmallAreaEmphasis`, `gradient_Idm`, and `wavelet-HLH_Energy`. The radiomics methods help improve the interpretability of our model.

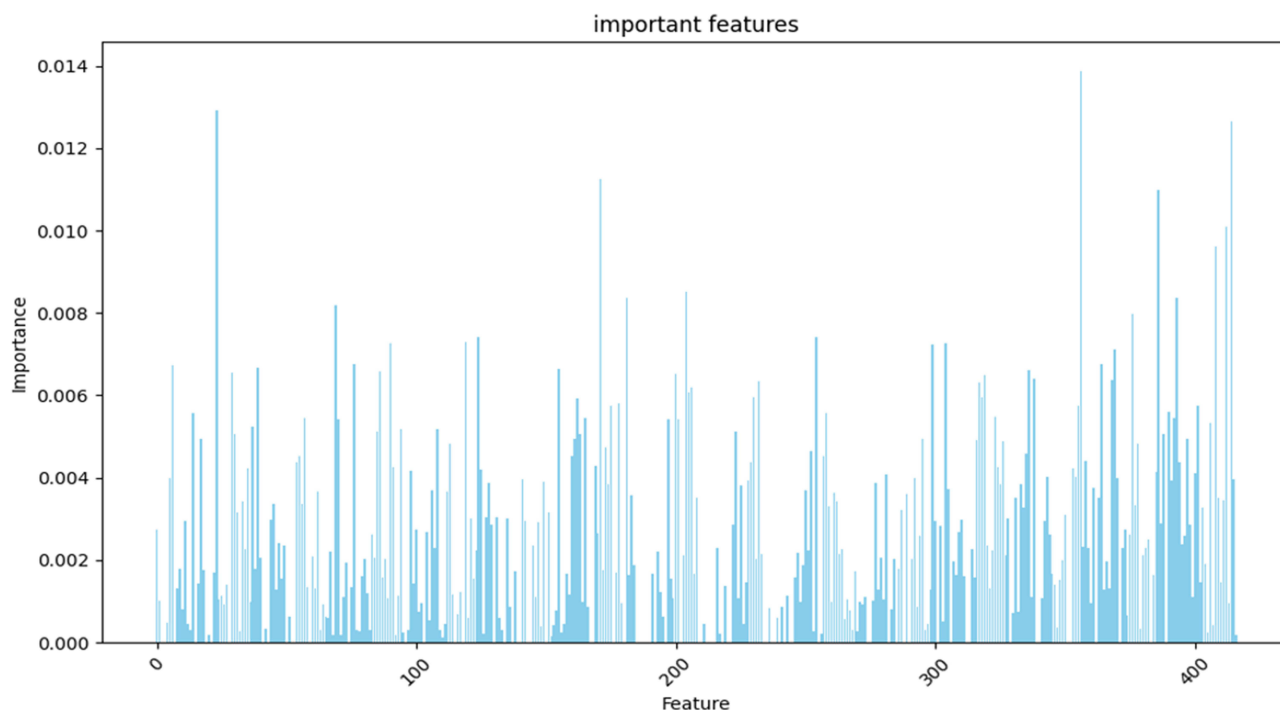


Figure 2 Importance of radiomics features.

Deep Learning-Based Method

Training with the Resnet50 model yielded the classification accuracy of 83.8% (95% CI, 66.7%–100%), sensitivity of 88.9% (95% CI, 62.5%–100%), specificity of 77.8% (95% CI, 49.9%–100%), positive predictive value of 80.0% (95% CI, 54.6%–100%), negative predictive value 87.5% (95% CI, 60.0%–100%), F1 score of 84.2% (95% CI, 61.5%–100%), and AUC of 70.0%.

Deep learning performs better compared to radiomics in our experiments. This is because radiomics relies on manual feature extraction, while deep learning uses automatic feature extraction. This can automatically extract complex and highly abstract features from data through the multi-layer neural networks, avoiding the limitations of manual intervention and showing stronger generalization ability and robustness.

We also show the confusion matrix results of both radiomics and deep learning-based methods in Figure 3. In the confusion matrices, the rows represent the true categories of the test images, the columns represent the predicted

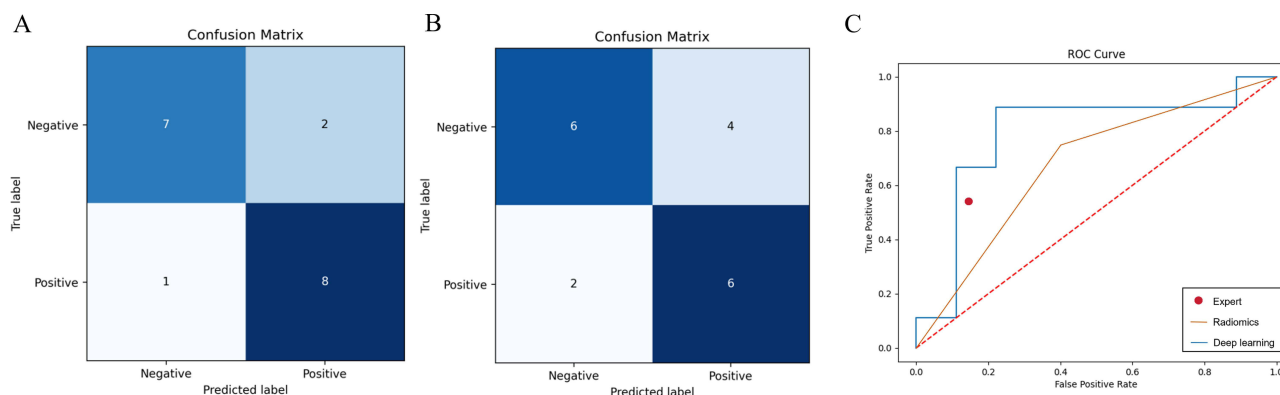


Figure 3 Testing the performance of (A) Radiomics and (B) Resnet50 models using confusion matrices (The vertical axis represents the true labels and the horizontal axis represents the model's predicted labels, with fibrotic indicated by positive and inflammatory indicated by negative); (C) testing the Performance of Resnet50 Models, Radiomics and experts'assessment based on ROC Curves.

categories of the test images (Negative for inflammatory stricture and Positive for fibrotic stricture). Obviously, the confusion matrix of the Resnet50 model (Figure 3B) shows a clear advantage on the diagonal compared with the radiomics-based method (Figure 3A). As for the ROC results, we can observe that the ROC curve for the Resnet50 model (Figure 3C) shows a trend of gradually approaching the upper left corner. This shows that the model exhibits better classification ability than the radiomics-based method (Figure 3C). The deep learning model demonstrated a significantly higher AUC compared to the radiomics model ($P = 0.018$). The deep learning model also showed a statistically significant superiority over the expert assessments ($P = 0.043$). The difference between the radiomics model and the expert assessments was not statistically significant ($P = 0.271$) (Figure 3C).

Visualization

The model's prediction results are visualized in Figure 4. Panel A displays fibrotic stricture predictions: the first three images are correctly classified, while the fourth misclassifies a fibrotic stricture as inflammatory. Panel B shows inflammatory stricture predictions: the first three are accurate, and the fourth incorrectly labels an inflammatory stricture as fibrotic.

Attentional Visualization of Deep Learning Models

To enhance interpretability, we employed class activation maps (CAMs) for attention visualization in deep learning models. CAMs, matching the original image size, assign pixel values from 0 to 1 (grayscale: 0–255), representing the contribution to the predicted output. Higher scores indicate greater sensitivity and network contribution from corresponding image regions. In our study, CAMs were generated for images in Figure 5, with heatmaps visualizing neural network features. The first row displays original images, while the second row overlays CAMs. Red areas highlight key discriminative regions, with intensity reflecting feature effectiveness.

Results demonstrate the model's ability to precisely focus on critical areas during intestinal stricture classification, effectively identifying abnormal features in the images. This underscores the model's capability in targeting relevant pathological regions for accurate classification.

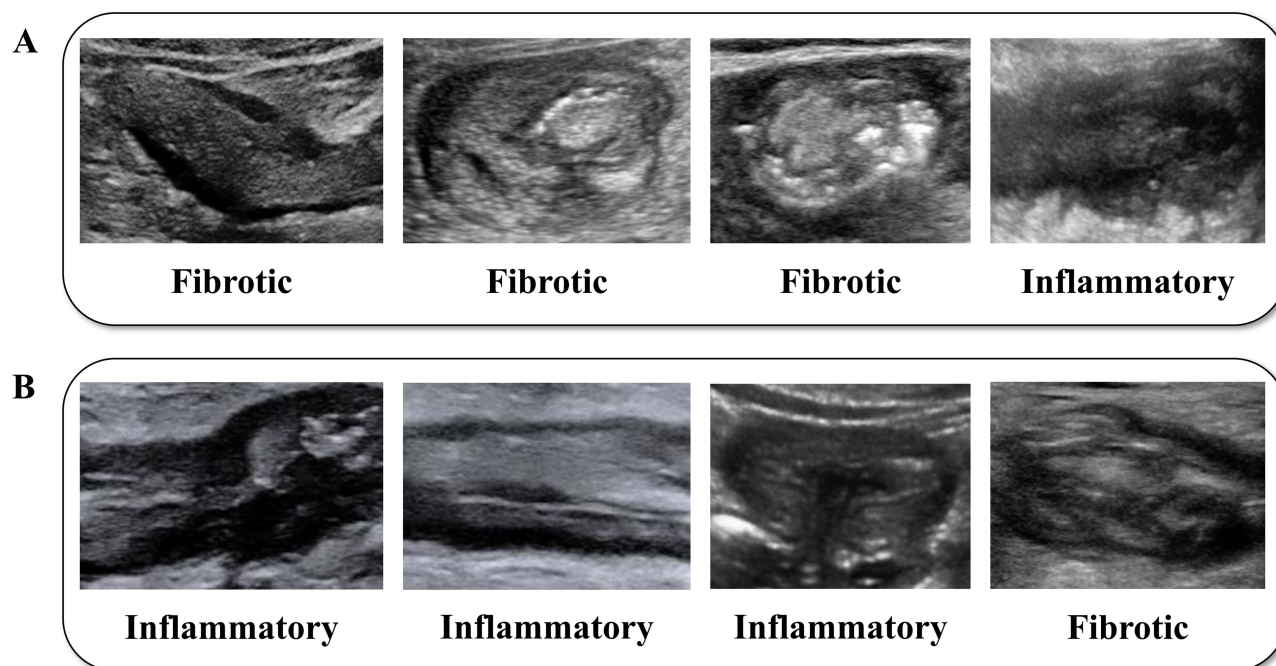


Figure 4 Visualization of attention scores for successful cases in combination with deep learning models. Representative images of patients predicted as (A) Fibrotic and (B) Inflammatory are visualized to illustrate the prediction process of the Resnet50 model.

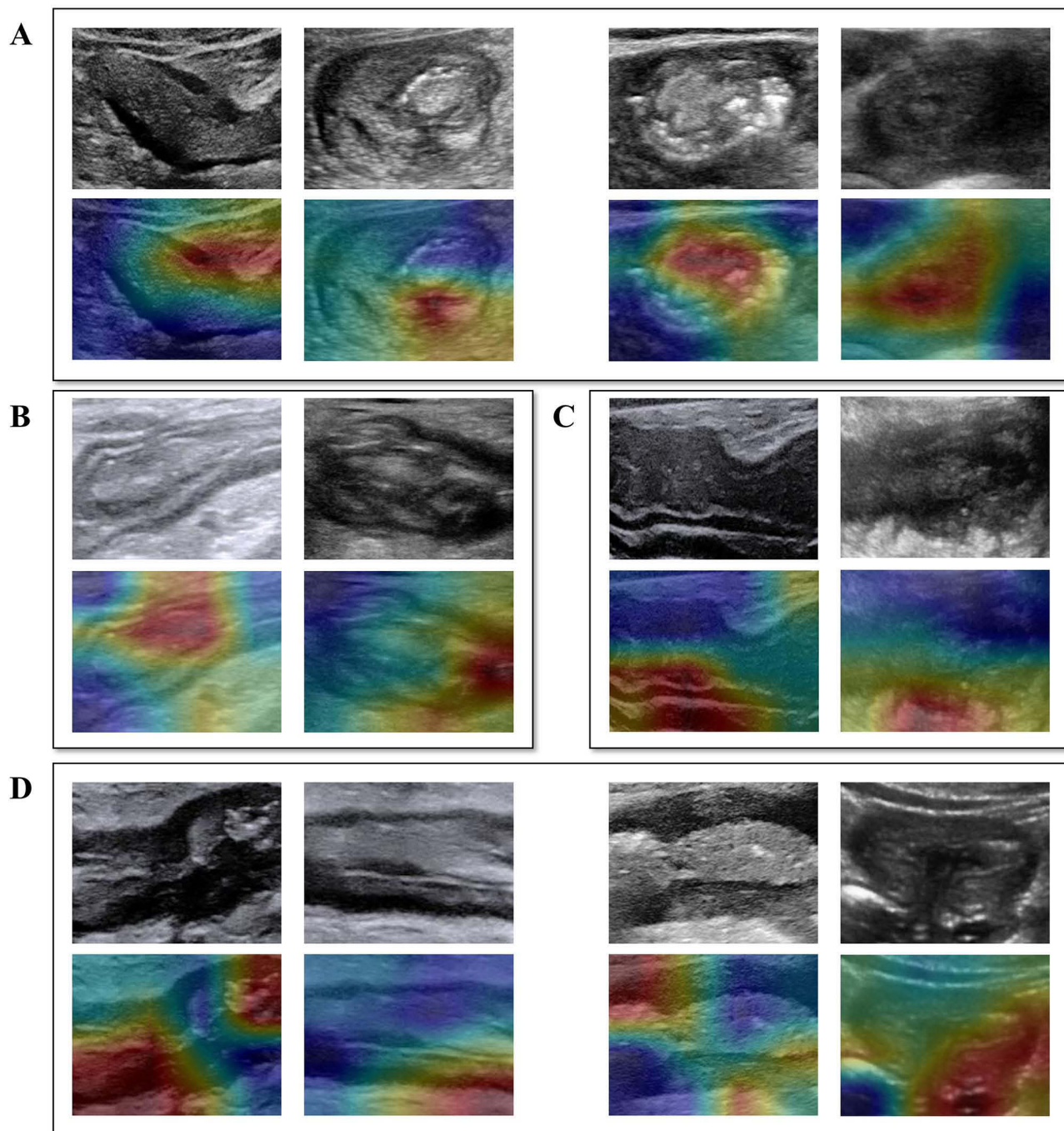


Figure 5 Resnet50 model class activation maps: **(A)** real label fibrotic, and predicted label as fibrotic; **(B)** real label as inflammatory, predicted label as fibrotic; **(C)** real label as fibrotic, predicted label as inflammatory; **(D)** real label as inflammatory, predicted label as inflammation.

Comparative Analysis of Radiomics, Resnet50 Model, and Expert Predictions

The ResNet50 model achieved the highest accuracy (83.3%), surpassing Radiomics (67.0%) and expert radiologists (73.1%) (Table 1). It also led in sensitivity (88.9% vs 75.0% for Radiomics and 54.6% for experts) and positive predictive value (PPV: 80.0% vs 75.0% for experts and 60.0% for Radiomics). Experts demonstrated the highest specificity (86.7% vs 77.8% for ResNet50 and 60.0% for Radiomics). The ResNet50 model also excelled in negative predictive value (NPV: 87.5%) and F1-score (84.2%), outperforming experts (NPV: 72.2%; F1-score: 69.7%) and Radiomics (NPV: 60.0%; F1-score: 67.0%). While ResNet50 consistently outperformed in most metrics, expert

Table 1 Comparative Analysis of Diagnostic Performance Metrics Among Radiomics, ResNet50 Model, and Experienced Expert in Classifying Intestinal Stricture

	Radiomics	Resnet50 Model	Experts
Accuracy	67.0%	83.3%	73.1%
Sensitivity	75.0%	88.9%	54.6%
Specificity	60.0%	77.8%	86.7%
Positive predictive value	60.0%	80.0%	75.0%
Negative predictive value	75.0%	87.5%	72.2%
F1-score	67.0%	84.2%	69.7%

specificity remained superior. However, expert performance in stricture classification was suboptimal overall. Inter-observer agreement between the two experts was very good (Cohen's $\kappa = 0.815$; $p < 0.001$). These findings highlight ResNet50's potential as a robust tool for medical image analysis.

Discussion

This study presents several significant findings regarding the clinical characteristic in CD patients and classification of intestinal strictures among them using AI approaches. Our deep learning-based method (Resnet50) achieved superior performance compared to the radiomics-based approach and expert predictions, with higher accuracy and better sensitivity in distinguishing inflammatory from fibrotic strictures. However, expert predictions achieved better specificity among them. Meanwhile, the CAMs visualization demonstrated that the deep learning model could effectively identify and focus on relevant pathological features, enhancing the model's interpretability and clinical applicability. AI-driven analysis into routine clinical practice holds the promise of standardizing interpretations of IUS results, thus reducing inter-observer variability.

While several AI models have been developed for differentiating Crohn's disease-related strictures, many prior studies have primarily utilized CTE or MRE. For instance, one model based on radiologist-defined strictures using automated CTE measurements achieved an accuracy of 87.6%.²¹ Another deep learning approach outperformed two radiologists (AUCs: 0.579 and 0.646; both $P < 0.05$) and was not inferior to a radiomics model (AUC = 0.813, $P < 0.05$), while requiring significantly less processing time ($P < 0.001$).²² Our model achieved comparable accuracy. Considering the operator-dependent nature of intestinal ultrasound and the relatively smaller sample size of our study compared to previous CTE/MRE-based studies, these results suggest that our model performs similarly to existing approaches.

The superior performance of deep learning over traditional radiomics highlights its potential in medical imaging analysis.^{23–25} This study is the first to compare radiomics and deep learning in differentiating CD strictures using ultrasound. Notably, expert predictions demonstrated the highest specificity, underscoring their continued importance despite AI advancements.

The successful implementation of attention visualization through CAMs represents a significant step toward interpretable AI in clinical practice. This addresses a crucial concern in medical AI applications - the "black box" nature of deep learning models. Recent work has similarly emphasized the importance of interpretable AI in clinical decision-making, showing that visualization techniques can increase physician trust and adoption of AI systems.²⁶ This automatic focusing ability not only improves the interpretability of the model but also makes the deep learning model more consistent in comparison with expert evaluation, enhancing the credibility and interpretability of the model in practical applications. To enhance generalizability, a standardized intestinal ultrasound imaging protocol should first be established to ensure consistent and effective image acquisition for AI-assisted diagnosis. In this study, the radiologists strictly followed the screening protocol recommended by Experts suggestions on the standardization of intestinal ultrasound examination and reporting for inflammatory bowel disease in China.¹⁹ And two types of ultrasound machines were applied during the study (Philips iU22 or SuperSonic Aixplorer). Both of them showed similar performance in recognizing the characteristic of fibrotic or inflammatory stricture. Also, CD patients strictly fast for 8 hours prior to

the examination followed the instruction of gastroenterologists, and most of them typically follow a low-residue diet, hence minimizing the interference caused by intestinal contents. And the radiologists adopted the unified diagnosis criteria of intestinal stricture for patients inclusion.²⁰ Based on a unified patient preparation protocol, quantitative criteria for intestinal stenosis assessment, and a standardized screening and image acquisition process, the model established herein demonstrates strong potential for generalizability following the broad implementation of these standardized examination procedures. Future validation should explicitly evaluate model performance across diverse equipment from various manufacturers and models to assess true portability. Furthermore, a prospective multi-center study specifically designed to test the model across different healthcare settings, ultrasound machines, and operator skill levels would significantly strengthen its general applicability.

The integration of ultrasound-based AI analysis into CD management could significantly impact clinical practice. Recent studies emphasize the growing role of IUS in CD monitoring, demonstrating high sensitivity, specificity, and concordance with endoscopic scores.²⁷ IUS is recommended as an ideal diagnostic tool for long-term follow-up of IBD according to both China and ECCO guidelines^{28,29} Our AI approach could enhance the utility of this non-invasive imaging modality by providing objective, quantitative assessment of stricture characteristics. Using deep neural networks can perfectly tell difference between strictures and normal mucosa (AUC = 0.989), as well as strictures and all ulcers (AUC = 0.942) based on capsule endoscopy images.³⁰ For patients intolerant to endoscopy, IUS offers a safer alternative for stricture evaluation.

The study demonstrates several notable strengths in both methodology and execution. First, it employs a sophisticated dual-approach methodology combining both radiomics and deep learning techniques, which provides a comprehensive framework for analyzing intestinal strictures in CD patients. Second, the inclusion of CAM for visualization adds a crucial layer of interpretability to the deep learning model, making the results more transparent and clinically applicable.

Despite its strengths, the study has several limitations that warrant consideration. Although this study represents the largest sample size to date for assessing intestinal stricture characteristics using intestinal ultrasound, its single-center design limits the generalizability of the findings. Additionally, the split ratio of 8:2 for training and testing sets, while common in machine learning studies, might not provide sufficient test data for robust validation given the small overall sample size. Another limitation is the lack of external validation on an independent dataset from different medical centers, which would be crucial for establishing the model's generalizability. The study also does not address the potential impact of different ultrasound equipment and operators on image quality and subsequent analysis, which could be a significant source of variability in real-world applications. In the future, as a leading center for IBD diagnosis in China, we aim to standardize and promote intestinal ultrasound practices across multiple hospitals. Future efforts will focus on expanding the deep learning model to multicenter settings, incorporating larger-scale datasets, and closely tracking patient outcomes. This will enhance the applicability of our findings and allow more patients to benefit from the research. To enhance clinical integration and model interpretability, we propose the following steps: First, the AI could be trained to recognize guideline-recommended sonographic features indicative of fibrotic or inflammatory strictures—features identifiable by sonographers but subject to interpreter experience. Second, the model's accuracy, sensitivity, and specificity should be validated through larger, multi-center studies. Third, technical integration with ultrasound systems should be pursued to enable real-time feedback during examinations. Finally, standardized imaging protocols must be widely promoted to support consistent application.

Conclusion

This pioneering study compares radiomics and deep learning for differentiating fibrotic stricture from inflammatory stricture in CD patients, highlighting the superior performance of the ResNet50 model in accuracy and diagnostic metrics. Regarding the rather small sample size and lack of multi-center data, future multi-center studies with external validation and longitudinal data are needed to assess generalizability and predictive capabilities for disease progression. Prospective studies incorporating the clinical data and IUS images, with essential follow-up, will be performed to validate clinical utility in real-world settings.

Data Sharing Statement

All data and material are shown in this manuscript.

Ethics Approval and Informed Consent

Based on the Ethics Committee Guidelines of Peking Union Medical College Hospital for the Clinical Research Involving Human Subjects, this study is exempt from obtaining informed consent from the subjects, as such exemption does not negatively impact their rights and interests. The research utilizes identifiable human materials or data where the subjects can no longer be located, and the study does not involve personal privacy or commercial interests. Following review by the Ethics Committee (I-22PJ1092), this retrospective study has been deemed to meet the above criteria and is therefore exempt from the requirement of signed informed consent. And the study complies with the Declaration of Helsinki.

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

This work was supported by National Key R&D Program of China (2023YFC2507300), Beijing Health Technology Promotion Project (BHTP P2024096, BHTPP P2024097), National High-Level Hospital Clinical Research Funding (2025-PUMCH-A-163, 2022-PUMCH-B-022, 2022-PUMCH-C-018), CAMS Innovation Fund for Medical Sciences (2024-I2M-C&T-B-004), and State Key Laboratory Special Fund (2060204).

Disclosure

The authors report no conflicts of interest in this work.

References

- Little RD, Jayawardana T, Koentgen S, et al. Pathogenesis and precision medicine for predicting response in inflammatory bowel disease: advances and future directions. *eGastroenterology*. 2024;2(1):e100006. doi:10.1136/egastro-2023-100006
- Gordon IO, Agrawal N, Willis E, et al. Fibrosis in ulcerative colitis is directly linked to severity and chronicity of mucosal inflammation. *Aliment Pharmacol Ther*. 2018;47(7):922–939. doi:10.1111/apt.14526
- Rieder F, Ma C, Hanzel J, et al. Reliability of CT enterography for describing fibrostenosing Crohn disease. *Radiology*. 2024;312(2):e233038. doi:10.1148/radiol.233038
- Thia KT, Sandborn WJ, Harmsen WS, Zinsmeister AR, Loftus EV. Risk factors associated with progression to intestinal complications of Crohn's disease in a population-based cohort. *Gastroenterology*. 2010;139(4):1147–1155. doi:10.1053/j.gastro.2010.06.070
- Gomollón F, Dignass A, Annesse V, et al. 3rd European evidence-based consensus on the diagnosis and management of Crohn's disease 2016: part 1: diagnosis and medical management. *J Crohn's Colitis*. 2017;11(1):3–25. doi:10.1093/ecco-jcc/jjw168
- Fraquelli M, Castiglione F, Calabrese E, Maconi G. Impact of intestinal ultrasound on the management of patients with inflammatory bowel disease: how to apply scientific evidence to clinical practice. *Digestive Liver Dis*. 2020;52(1):9–18. doi:10.1016/j.dld.2019.10.004
- Lu C, Rosentreter R, Parker CE, et al. International expert guidance for defining and monitoring small bowel strictures in Crohn's disease on intestinal ultrasound: a consensus statement. *Lancet Gastroenterol Hepatol*. 2024;9(12):1101–1110. doi:10.1016/S2468-1253(24)00265-6
- Chen YJ, Mao R, Li XH, et al. Real-time shear wave ultrasound elastography differentiates fibrotic from inflammatory strictures in patients with Crohn's disease. *Inflammatory Bowel Dis*. 2018;24(10):2183–2190. doi:10.1093/ibd/izy115
- Pescatori LC, Mauri G, Savarino E, Pastorelli L, Vecchi M, Sconfienza LM. Bowel sonoelastography in patients with Crohn's disease: a systematic review. *Ultrasound Med Biol*. 2018;44(2):297–302. doi:10.1016/j.ultrasmedbio.2017.10.004
- Turco S, Tiyyarattanachai T, Ebrahimkheil K, et al. Interpretable machine learning for characterization of focal liver lesions by contrast-enhanced ultrasound. *IEEE Trans Ultrason Ferroelectr Freq Control*. 2022;69(5):1670–1681. doi:10.1109/TUFFC.2022.3161719
- Whitney HM, Li H, Ji Y, Liu P, Giger ML. Comparison of breast MRI tumor classification using human-engineered radiomics, transfer learning from deep convolutional neural networks, and fusion methods. *Proceed IEEE Institute Electrical Electronics Engineers*. 2020;108(1):163–177. doi:10.1109/JPROC.2019.2950187
- Maeda Y, Dittono I, Puga-Tejada M, et al. Artificial intelligence-enabled advanced endoscopic imaging to assess deep healing in inflammatory bowel disease. *eGastroenterology*. 2024;2(3):e100090. doi:10.1136/egastro-2024-100090

13. Majtner T, Brodersen JB, Herp J, Kjeldsen J, Halling ML, Jensen MD. A deep learning framework for autonomous detection and classification of Crohn's disease lesions in the small bowel and colon with capsule endoscopy. *Endoscopy Int Open*. 2021;9(9):E1361–e1370. doi:10.1055/a-1507-4980
14. Chen YF, Liu L, Lyu B, et al. Role of artificial intelligence in Crohn's disease intestinal strictures and fibrosis. *J Digestive Dis*. 2024;25(8):476–483. doi:10.1111/1751-2980.13308
15. Inflammatory Bowel Disease Group CSoGCMA, China IBDQCCo. Chinese clinical practice guideline on the management of Crohn's disease (2023, Guangzhou). *Chin J Inflamm Bowel Dis*. 2024;08(1):2–32.
16. Bettenworth D, Baker ME, Fletcher JG, et al. A global consensus on the definitions, diagnosis and management of fibrostenosing small bowel Crohn's disease in clinical practice. *Nat Rev Gastroenterol Hepatol*. 2024;21(8):572–584. doi:10.1038/s41575-024-00935-y
17. Zhang MC, Li XH, Huang SY, et al. IVIM with fractional perfusion as a novel biomarker for detecting and grading intestinal fibrosis in Crohn's disease. *Eur Radiol*. 2019;29(6):3069–3078. doi:10.1007/s00330-018-5848-6
18. Mao H, Su P, Qiu W, Huang L, Yu H, Wang Y. The use of Masson's trichrome staining, second harmonic imaging and two-photon excited fluorescence of collagen in distinguishing intestinal tuberculosis from Crohn's disease. *Colorectal Dis*. 2016;18(12):1172–1178. doi:10.1111/codi.13400
19. Chinese Quality Control Assessment Center for Inflammatory Bowel Disease Diagnosis and Treatment IBDG, Chinese Society of Gastroenterology, Chinese Medical Association, Abdominal Ultrasound Group, Chinese Society of Ultrasound in Medicine, Chinese Medical Association. Experts suggestions on the standardization of intestinal ultrasound examination and reporting for inflammatory bowel disease in China. *Chin J Inflamm Bowel Dis*. 2024;08(2):109–115.
20. Bettenworth D, Bokemeyer A, Baker M, et al. Assessment of Crohn's disease-associated small bowel strictures and fibrosis on cross-sectional imaging: a systematic review. *Gut*. 2019;68(6):1115–1126. doi:10.1136/gutjnl-2018-318081
21. Stidham RW, Enchakalody B, Waljee AK, et al. Assessing small bowel stricturing and morphology in Crohn's disease using semi-automated image analysis. *Inflammatory Bowel Dis*. 2020;26(5):734–742. doi:10.1093/ibd/izz196
22. Meng J, Luo Z, Chen Z, et al. Intestinal fibrosis classification in patients with Crohn's disease using CT enterography-based deep learning: comparisons with radiomics and radiologists. *Eur Radiol*. 2022;32(12):8692–8705. doi:10.1007/s00330-022-08842-z
23. Song D, Zhang Z, Li W, Yuan L, Zhang W. Judgment of benign and early malignant colorectal tumors from ultrasound images with deep multi-View fusion. *Comput Methods Programs Biomed*. 2022;215:106634. doi:10.1016/j.cmpb.2022.106634
24. Bedrikovetski S, Dudi-Venkata NN, Kroon HM, et al. Artificial intelligence for pre-operative lymph node staging in colorectal cancer: a systematic review and meta-analysis. *BMC Cancer*. 2021;21(1):1058. doi:10.1186/s12885-021-08773-w
25. Bao Z, Du J, Zheng Y, Guo Q, Ji R. Deep learning or radiomics based on CT for predicting the response of gastric cancer to neoadjuvant chemotherapy: a meta-analysis and systematic review. *Front Oncol*. 2024;14:1363812. doi:10.3389/fonc.2024.1363812
26. van der Velden BHM, Kuijff HJ, Gilhuijs KGA, Viergever MA. Explainable artificial intelligence (XAI) in deep learning-based medical image analysis. *Med Image Anal*. 2022;79:102470. doi:10.1016/j.media.2022.102470
27. Madsen GR, Wilkens R, Boysen T, et al. The knowledge and skills needed to perform intestinal ultrasound for inflammatory bowel diseases-an international Delphi consensus survey. *Aliment Pharmacol Ther*. 2022;56(2):263–270. doi:10.1111/apt.16950
28. Treatment CQCACfIBDDa. Experts suggestions on the standardization of intestinal ultrasound examination and reporting for inflammatory bowel disease in China. *Chin J Inflamm Bowel Dis*. 2024;08(8):109–115.
29. Kucharzik T, Tielbeek J, Carter D, et al. ECCO-ESGAR topical review on optimizing reporting for cross-sectional imaging in inflammatory bowel disease. *J Crohn's Colitis*. 2022;16(4):523–543. doi:10.1093/ecco-jcc/jjab180
30. Klang E, Grinman A, Soffer S, et al. Automated detection of Crohn's disease intestinal strictures on capsule endoscopy images using deep neural networks. *J Crohn's Colitis*. 2021;15(5):749–756. doi:10.1093/ecco-jcc/jjaa234

Journal of Inflammation Research

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>

Dovepress
Taylor & Francis Group