Journal of Inflammation Research Dovepress

Open Access Full Text Article ORIGINAL RESEARCH

Deep-Learning, Radiomics and Clinic Based Fusion Models for Predicting Response to Infliximab in Crohn's Disease Patients: A Multicentre, Retrospective Study

Weimin Cai[1,*], Xiao Wu[1,*], Kun Guo[2], Yongxian Chen[3], Yubo Shi[4], Xinran Lin[1]

[1]Department of Gastroenterology and Hepatology, the First Affiliated Hospital of Wenzhou Medical University, Wenzhou, 325000, People's Republic of China; [2]Department of Cardiology, the First Affiliated Hospital of Wenzhou Medical University, Wenzhou, 325000, People's Republic of China; [3]Department of Chest Cancer, Xiamen Second People's Hospital, Xiamen, 36100, People's Republic of China; [4]Department of Pulmonary, Yueqing People's Hospital, Wenzhou, 325000, People's Republic of China

*These authors contributed equally to this work

Correspondence: Xinran Lin, Department of Gastroenterology and Hepatology, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou, 325000, People's Republic of China, Tel +86 18857838243, Fax +86 0576 87755312, Email lxr190910@163.com

Background: Accurate prediction of treatment response in Crohn's disease (CD) patients undergoing infliximab (IFX) therapy is essential for clinical decision-making. Our goal was to compare the performance of the clinical characteristics, radiomics and deep learning model from computed tomography enterography (CTE) for identifying individuals at high risk of IFX treatment failure.

Methods: This retrospective study enrolled 263 CD patients from three medical centers between 2017 and 2023 patients received CTE examinations within 1 month before IFX commencement. A training cohort was recruited from center 1 (n=166), while test cohort from centers 2 and 3 (n=97). The deep learning model and radiomics were constructed based on CTE images of lesion. The clinical model was developed using clinical characteristics. Two fusion methods were used to create fusion model: the feature-based early fusion model and the decision-based late fusion model. The performances of the predictive models were evaluated.

Results: The early fusion model achieved the highest area under characteristics curve (AUC) (0.85–0.91) among all patient cohorts, significantly outperforming deep learning model (AUC=0.72–0.82, p=0.06–0.03, Delong test) and radiomics model (AUC=0.72–0.78, p=0.06–0.01). Compared to early fusion model, the AUC values for the clinical and late fusion models were 0.71–0.91 (p=0.01–0.41), and 0.81–0.88 (p=0.49–0.37) in the test and training set, respectively. Moreover, the early fusion had the lowest value of Brier's score 0.15–0.12 in all patient set.

Conclusion: The early fusion model, which integrates deep learning, radiomics, and clinical data, can be utilized to predict the response to IFX treatment in CD patients and illustrated clinical decision-making utility.

Keywords: deep-learning, radiomics, Crohn's disease, infliximab, prediction

Introduction

Crohn's disease (CD), a main form of inflammatory bowel disease (IBD), is characterized as a chronic, remitting, and relapsing inflammatory disorder of the gastrointestinal tract, requiring ongoing management and surveillance.[1] Infliximab (IFX), shows efficacy in inducing and maintaining remission, also is the first anti-tumor necrosis factor (TNF) α agent. Yet, primary loss of response (PLR) to IFX is observed in 13–40% of CD patients, while 23–46% patients experience a secondary loss of response (SLR).[2,3] The reduction of response to IFX, an anti-TNF monoclonal antibody, is believed to result from inadequate drug levels, attributed to the antibody's degradation and clearance. Consequently, when a decrease in the IFX's effectiveness is suspected, a comprehensive assessment involving the evaluation of active inflammation, potential IBD complications, clinical symptoms, biochemical markers, endoscopic findings, and imaging

Received: 2 August 2024
Accepted: 15 October 2024
Published: 25 October 2024

should be undertaken. Furthermore, consistent monitoring of disease activity during IFX treatment is crucial, enabling timely adjustments to therapeutic strategies to achieve and maintain long-term remission.

For decades, the Crohn's Disease Activity Index (CDAI) has been regarded as a valuable tool for measuring the clinical activity of CD.[4–6] Nevertheless, the accurate calculation of CDAI requires over a seven-day period and the necessity for patient compliance limits its utility in daily clinical practice.[7,8] While colonoscopy offers an objective and accurate assessment of the intestinal condition, its invasiveness, high costs, and potential risks, including perforation, bleeding, and cardiovascular complications related to sedation, detract from its desirability.[9,10] Furthermore, a considerable number of CD patients are unsuitable candidates for endoscopy due to intestinal luminal strictures or a penetrating disease phenotype.[11] Additionally, the psychological and physical burdens caused by repeated invasive procedures require extended recovery times, further complicate their application for patients.[12] Both CT and magnetic resonance (MR) enterography serve as pivotal tools for assessing disease status and detecting extraintestinal complications in patients diagnosed with Crohn's disease (CD).[13,14] MR enterography, known for its provision of multi-sequence imaging data and superior soft tissue contrast resolution, plays a crucial role in evaluating disease activity and treatment response. Additionally, it aids in characterizing inflammatory or fibrotic lesions and facilitates routine clinical follow-up.[15,16] Conversely, computed tomography enterography (CTE) offers advantages such as shorter scanning times, cost-effectiveness, and high inter-reader agreement in image interpretation, thereby enhancing its availability and accessibility in clinical settings.[17] With ongoing equipment upgrades and advancements in CT reconstruction algorithms, there has been notable progress in obtaining high-quality images at reduced radiation doses. This development is particularly advantageous when imaging young patients and conducting repeated examinations over time. CTE can effectively reflect the characteristics of CD,[18–20] and there have been studies specifically investigate the correlation between the features of the intestinal wall and mesentery, and the mucosal condition under CTE.[21,22]

Recent advancements in radiomics and deep learning have shown potential use in diagnosing diseases, classifying molecular types, and forecasting treatment outcomes.[23–25] Meng et al performed research utilizing radiomics and deep learning to analyze CT scans of patients with CD.[26] Previous studies have identified relationship between the radiological features of CD and IFX response.[27–29] Moreover, recent studies indicates that imaging power radiomics can be applied to diagnose CD mucosal healing.[30,31] Based on this foundation, we posited that employing artificial intelligence (AI) for image analysis of CD might serve as a viable approach for predicting IFX response.

Medical imaging modalities like CT and MRI operate within three-dimensional (3D) spaces, and the application of two-dimensional (2D) techniques to 3D tasks can lead to loss of essential structural information. Training a 3D deep convolutional neural network (DCNN) typically requires larger datasets than 2D DCNN, presenting a substantial obstacle in medical research due to the limited available data.[32] Moreover, deep learning-derived features may be sensitive to the global translation, rotation, and scaling of images, a vulnerability not shared by radiomics features.[33] Recent studies indicate that models integrating both radiomics and deep learning outperform models relying on a single approach.[34,35]

Our study was designed to evaluate the effectiveness of 3D deep learning, radiomics, and their integrated models in the prediction of IFX response in CD patients. We implemented two strategies for model integration: early fusion based on features and late fusion based on decisions.

Material and Methods

Study Cohort

This study retrospectively collected data from 263 Crohn's disease patients across three hospitals (Figure 1). The training cohort comprised 166 patients treated at the Center 1 (Wenzhou Medical University First Affiliated Hospital) from January, 2017, to February, 2023. The external test cohort included 97 patients treated at Center 2 (Xiamen Second Hospital) and Center 3 (Yueqing People's Hospital) from March 2017 to January 2023. The three respective Hospital Ethics Committee granted approval for this study, which was carried out in compliance with the Declaration of Helsinki principles. Given the study's retrospective and observational design, the requirement for informed consent was waived. Patient confidentiality was assured by the protection of personal information in the medical record system.

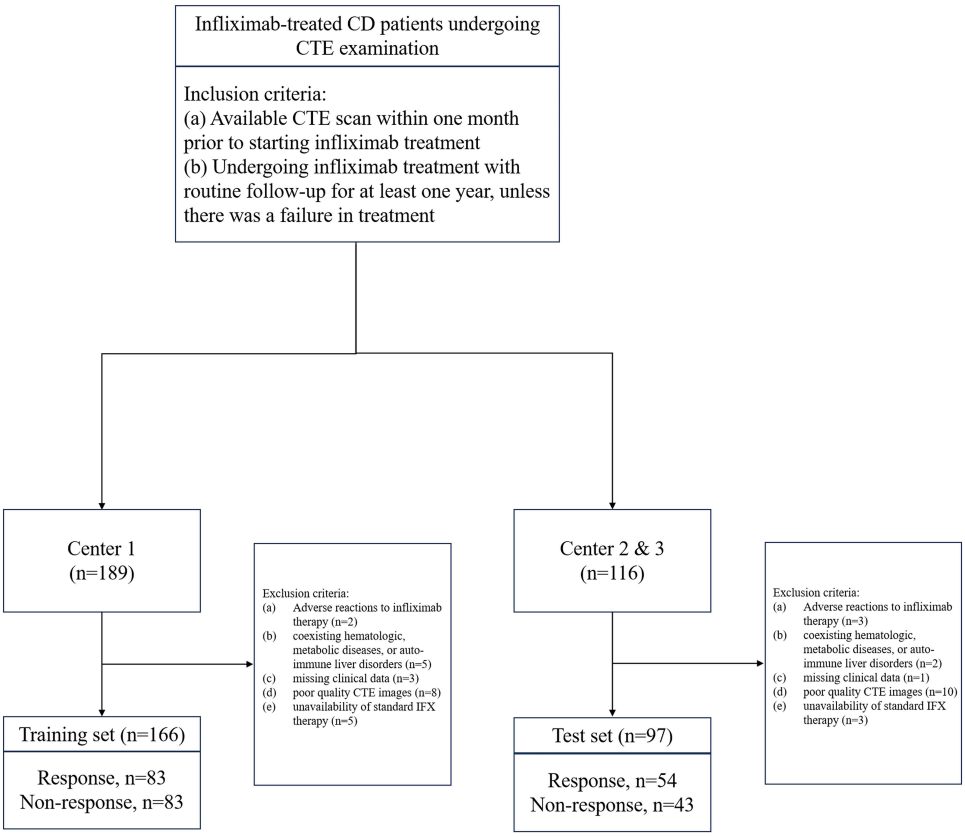

Figure 1 Flowchart diagram shows the patient selection process from three medical centers. CD, Crohn's disease; CTE, computed tomography enterography; IFX, infliximab.

The criteria for inclusion were as follows: a diagnosis of CD established according to the consensus statement of the European Crohn's and Colitis Organization;[36] available CTE scan within one month prior to starting IFX treatment; undergoing IFX treatment with routine follow-up for at least one year, or the IFX treatment failure within one year. Exclusion criteria encompassed: adverse reactions to IFX therapy; coexisting hematologic, metabolic diseases, or auto-immune liver disorders; missing clinical data; poor quality CTE images, including artifacts, overlap, noise; unavailability of standard IFX therapy. The recruitment process is detailed in a flowchart in Figure 1.

The primary outcome focused on IFX treatment failure within one year, defined as composite endpoints that encompass both PLR and SLR to IFX. PLR refers to a lack of reaction to IFX during the induction phase (0–14 weeks), while SLR is identified as a decline in clinical response (15–54 weeks) after initially responding to IFX treatment.[36–38] The outcome was assessed by an experienced multidisciplinary team at our institution, which based its determinations on clinical symptoms, laboratory indices, radiological findings, and endoscopic results. This assessment included the following components: (a) The persistence of disease-related symptoms or active inflammation, as indicated by criteria such as a CDAI exceeding 150 or showing less than a 70% decrease, or a C-reactive protein (CRP) level exceeding 5 mg/L or demonstrating less than a 50% decrease compared to baseline, was assessed during induction therapy. (b) In cases where persistent clinical remission was not achieved following induction therapy or where there was evidence of disease relapse or progression within one year post-treatment, despite an initial response to IFX, further evaluation was conducted. This evaluation encompassed criteria such as the return of CDAI levels to baseline or higher, along with active intestinal inflammation indicated by endoscopic or imaging examinations. (c) Within one year, criteria such as the necessity for surgical intervention, substitution of biological agents, escalation of corticosteroid therapy, or more than two consecutive hospitalizations due to disease progression were also considered.

CT Examination Procedure

Prior to CTE, all patients adhered to a standardized bowel preparation regimen and observed a fasting period of 12 hours. Subsequently, within one hour preceding the scan, patients orally received a 2.5% isotonic mannitol solution (400–500 mL) at 15-minute intervals. CTE was performed using either a 64-section (Philips Brilliance; Philips) or a 256-section (GE Revolution; GE Healthcare) multidetector scanner. Scanning encompassed the region from the top of the diaphragm to the symphysis pubis with patients in the supine position. Following the acquisition of nonenhanced images, a nonionic contrast agent (320 mg/mL; Ioversol, Jiangsu Hengrui Medicine Corp Ltd) was administered intravenously at a dose of 1.5–2.0 mL/kg, with an average injection rate of 3.0–4.0 mL/sec. Arterial phase images were obtained either 15 seconds after reaching an attenuation value of 100 hounsfield units (HU) in the abdominal aorta using automatic bolus-tracking or with a delay of 32–35 seconds post-contrast injection. Subsequently, portal venous phase images were acquired with a delay of 35 seconds following the arterial phase or at 70 seconds post-contrast injection. For radiomics analysis in our study, CTE portal venous phase images with a slice thickness of 5 mm were utilized, consistent with methodologies employed in prior research on radiomics in CD.

Volume-of-Interest Segmentation and Radiomics Features Extraction

Volumes of interest (VOIs) were delineated on CTE in intestinal segments reflecting the most serious inflammation for medical image analysis. This approach aligns with criteria established in previous bowel radiomics research, where active inflammation was identified through endoscopic and radiological findings.[27,29,31] Characteristics include (1) evident ulcers observable via endoscopy and/or CTE, (2) thickening of the intestinal wall and segmental mural hyperenhancement, and (3) complications such as penetration (excluding abscesses), stricture, or the presence of creeping fat around the diseased bowel segment. The segmentation process culminated with a senior abdominal radiologist's review to verify the selection of optimal VOIs. Two experienced radiologists (with the experience of 5 and 7 years, respectively), in a blinded manner (blinded to clinical characteristics and outcome), outlined each lesion's contour to capture the lesion volume using the ITK-SNAP software (version 3.8, www.itksnap.org/), excluding the intestinal lumen, guided by the diagnostic criteria for active intestinal inflammation as discerned through imaging or endoscopic outcomes.[13,39] Based on linear interpolation, images were resampled to achieve isotropic voxel dimensions of $3.0 \times 3.0 \times 3.0$ mm³. Subsequently, using Python (version 3.7.0) with the PyRadiomics library (version 3.0.1), which adheres to the Imaging Biomarker Standardization Initiative (IBSI) guidelines, a comprehensive extraction of 1116 radiomics features was performed from the VOIs within the enteric phase images of CTE. For evaluating the reliability and consistency of the delineated VOIs, a re-annotation of all patients' VOIs was conducted two months following the initial delineation. This subsequent dataset facilitated an analysis using the intraclass correlation coefficient (ICC).

Feature Selection and Model Construction

Radiomics features were standardized by z-score normalization. To address the issue of strong correlations among features (Spearman correlation coefficient of 0.9 or higher), a methodical approach involving greedy recursive feature deletion was adopted for feature filtration. This approach systematically eliminates the most redundant feature within the set iteratively until no feature exhibits a correlation coefficient above 0.9. Subsequently, features exhibiting high stability with intraobserver and interobserver ICC values surpassing 0.75 were maintained.

Further refinement of the feature set was achieved through the application of multivariate least absolute shrinkage and selection operator (LASSO) regression (LASSO parameter: Alpha:1; Family: Binomial; nfolds:10). The support vector machine (SVM) algorithm, renowned for its efficacy in delineating optimal hyperplanes for classification tasks, was employed to develop predictive models[40] (SVM parameter: Type: eps-regression; Kernel: Radial; Cost: 1; Gamma:0; Epsilon:0.1). Input to the SVM classifier comprised the training dataset, with each sample containing a set of features and an associated label denoting the response or non-response to IFX. The resultant output was a classifier model trained to predict the likelihood of non-response to IFX in patients within the test set, quantifying the probability on a scale from 0 to 1.

Clinical Model Construction

The clinical characteristics in Table 1 were used to construct the clinical model. Feature selection was performed using LASSO regression, followed by SVM classifier for classification prediction. Detailed procedures were described in the above section of "Feature Selection and Model Construction".

3D Deep Learning Model Development and Feature Extraction

In the context of 3D deep learning, the term "bounding box" specifically refers to the cubic boundary encapsulating the lesion's VOI. To standardize input dimensions, all VOI cubes were resized to 96x96x96 pixels through linear interpolation. To augment the dataset, techniques including inversion along the X, Y, and Z axes were applied. These enhanced 3D images, paired with their respective labels, served as inputs for the 3D ResNet. The training regimen spanned 300 epochs, employing an initial learning rate of 0.02, culminating in a model capable of predicting IFX response in patients. To extract the 3D deep learning features of a patient's lesion, the 3D ResNet model's penultimate average pooling layer yielded a total of 512 features.

Construction of the Fusion Model

In this study, two distinct fusion strategies were implemented to develop the fusion model.[41]

The first strategy, known as feature-level fusion or early fusion, entails the amalgamation of all features derived from various modalities into a unified feature vector. Specifically, radiomics features of the VOI were extracted utilizing PyRadiomics, and 3D

Table 1 Patient Demographics and Clinical Characteristics

Variables	Total (n = 263)	Responders (n = 134)	Non-responders (n = 129)	p
Sex, n (%)				0.983
0	62 (23.5)	32 (23.8)	30 (23.3)	
1	201 (76.5)	102 (76.2)	99 (76.6)	
Age, Median (Q1,Q3)	28 (22, 42.5)	27 (22, 38.75)	29 (23, 44)	0.312
Smoking, n (%)				0.796
0	205 (77.9)	104 (77.6)	101 (78.3)	
1	58 (22.1)	30 (22.4)	28 (21.7)	
Drinking, n (%)				0.625
0	230 (87.5)	119 (88.8)	111 (86.0)	
1	33 (12.5)	15 (11.2)	18 (14.0)	
BMI, Median (Q1,Q3)	19.15 (17.65, 21.26)	19.47 (18, 21.3)	18.83 (17.48, 20.86)	0.098
Albumin, Mean ± SD	36.16 ± 6.32	38.17 ± 5.87	34.08 ± 6.11	< 0.001
ALT, Median (Q1,Q3)	13 (9, 20)	13 (9, 20.75)	14 (8, 20)	0.937
AST, Median (Q1,Q3)	17 (13.5, 22)	17 (14, 21.75)	17 (13, 22)	0.882
Serum creatinine, Median (Q1,Q3)	66 (54, 73)	66 (54.25, 74)	65 (53, 73)	0.284
Urea Nitrogen, Median (Q1,Q3)	4.3 (3.5, 5.6)	4.55 (3.8, 5.6)	4.1 (3.2, 5.4)	0.028
Uric acid, Mean ± SD	304.87 ± 90.66	319.81 ± 84.48	289.34 ± 94.51	0.006
Total cholesterol, Median (Q1,Q3)	3.45 (3.01, 4.12)	3.55 (3.14, 4.28)	3.3 (2.95, 4.05)	0.031
Triglyceride, Median (Q1,Q3)	0.99 (0.77, 1.39)	0.96 (0.81, 1.41)	1.03 (0.74, 1.36)	0.472
High density lipoprotein, Median (Q1,Q3)	0.94 (0.81, 1.12)	1.02 (0.86, 1.2)	0.88 (0.74, 1.01)	< 0.001
Low density lipoprotein, Median (Q1,Q3)	2.01 (1.62, 2.58)	2 (1.67, 2.51)	2.01 (1.6, 2.6)	0.959
Neutrophil, Median (Q1,Q3)	4.08 (2.84, 5.8)	3.44 (2.43, 4.77)	4.99 (3.81, 6.9)	< 0.001
Monocyte, Median (Q1,Q3)	0.55 (0.4, 0.75)	0.5 (0.39, 0.63)	0.62 (0.43, 0.8)	0.003
Lymphocyte, Median (Q1,Q3)	1.41 (1.08, 1.8)	1.45 (1.1, 1.9)	1.33 (1.04, 1.72)	0.163
Hemoglobin, Median (Q1,Q3)	122 (107, 135)	126 (113.25, 136.75)	117 (102, 129)	< 0.001
Platelet, Median (Q1,Q3)	283 (219.5, 356)	263 (205, 352.5)	300 (226, 369)	0.027
PT, Median (Q1,Q3)	13.9 (13.25, 14.5)	13.7 (13.1, 14.17)	14 (13.4, 14.7)	< 0.001
CRP, Median (Q1,Q3)	12.9 (3.13, 37.35)	3.23 (1.77, 14.2)	29 (11.5, 61.4)	< 0.001
ESR, Median (Q1,Q3)	20 (7, 38)	12 (2, 24)	28 (16, 46)	< 0.001

Abbreviations: BMI, body mass index; ALT, Alanine aminotransferase; AST, Aspartate Transaminase; PT, prothrombin time; CRP, C-reactive protein; ESR, erythrocyte sedimentation rate.

deep learning features were acquired via 3D ResNet, as previously outlined. These features, in conjunction with clinical and radiological data, underwent z-score normalization. Spearman correlation, ICC analyses, and least absolute shrinkage and selection operator (LASSO) analyses were subsequently conducted to refine the feature selection. An SVM classifier was then trained to create the feature-based fusion model, designated as early fusion, with comprehensive methods detailed in the section "Feature selection and model construction".

The second strategy, decision-level fusion or late fusion, integrates the output probabilities from various models. This approach utilized a stacking ensemble learning technique to amalgamate the output probabilities derived from the three above models.[41] The trained SVM model of late fusion was evaluated in both the training and external test sets. Workflow diagram for the development of the predictive models was presented in the Figure 2.

Statistical Analyses

To determine if the sample size in our study was adequate to identify an area under the receiver operating characteristic (AUC) curve value different from 0.500, we calculated the necessary sample size using the following parameters: 80% power, a two-sided significance level of 0.05, and an alternative hypothesis that the true AUC values of Early fusion model in both the training and total test cohorts differ from the null hypothesis (AUC = 0.5). The class ratios were based on actual data from the study: 91 non-responders and 75 responders cases in the training cohort, and 44 non-responders and 53 responders cases in the total test cohort.[42,43]

The comparison of categorical variables was conducted using either the Chi-square test or the Fischer test, while the Mann–Whitney *U*-test or the independent *T*-test was utilized for continuous variables. The assessment of the predictive model's performance leveraged Receiver Operating Characteristic (ROC) curves, along with metrics such as AUC, accuracy, sensitivity, specificity, and Brier's score. The R package "glmnet" and "e1071" was used to perform LASSO analysis and SVM model construction, respectively. All statistical analyses were executed using the R programming language (version 3.4.3) and the scikit-learn library (version 0.18) in Python 3.7.

Results

Demographics and Clinical Characteristics

A total of 263 CD patients were included in this study (median age, 28 years IQR: 22–42.5 years; 201 male) treated with IFX, including 166 patients in the training set, 97 patients in the external test set. The detailed information of included patients' detailed information was summarized in Table 1. Between the non-response and response groups, there was no

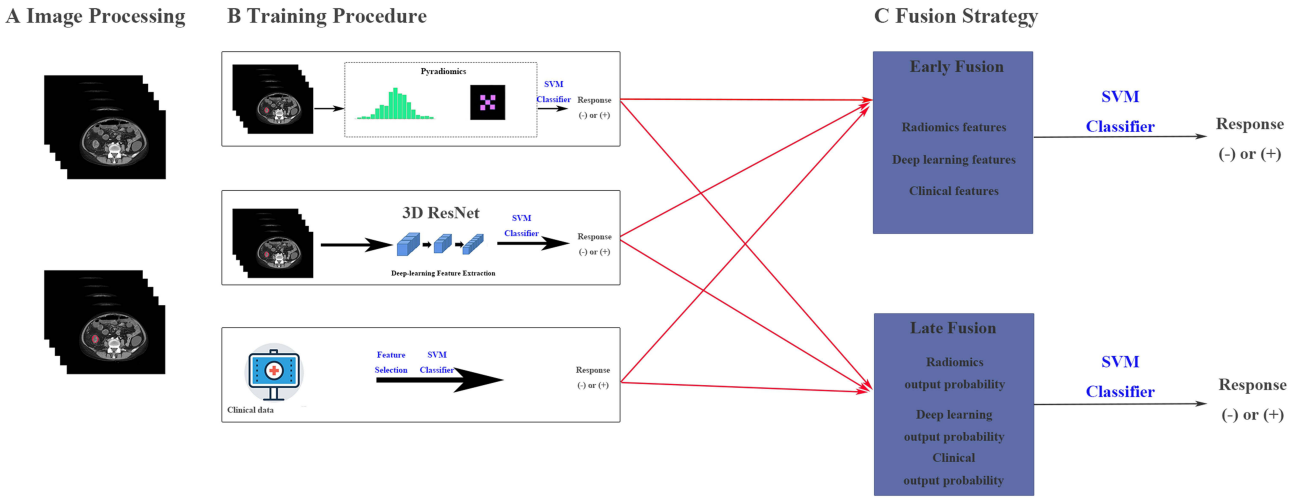

Figure 2 Workflow diagram for the development of the predictive models. Lesion segmentation and volume of interest (VOI) delineation are performed by experienced radiologists. The Radiomics model is developed using PyRadiomics. The deep learning model is developed based on 3D ResNet. The clinical data are used to construct the clinical model. For the early fusion model, the extracted features from three basic models are integrated to train an SVM classifier. For the late fusion model, the output probabilities from three basic models are used to develop a stacking model with a SVM classifier. A, the process of image processing; B, the construction of single modality models; C, the construction of fusion models. SVM, support vector machine.

significant differences in terms of sex, age, history of smoking, drinking habit, BMI, ALT, AST, serum creatinine, triglyceride, low density lipoprotein, lymphocyte count. The albumin, high density lipoprotein, neutrophil count, monocyte count, hemoglobin, platelet, prothrombin time, C-reactive protein, and erythrocyte sedimentation rate in the non-response group were significant different with those of response group.

Feature Selection and Model Development

Among the 1116 radiomic features and 512 deep learning features obtained from the CTE images, 212 radiomics features and 89 deep learning features with intraobserver and interobserver ICCs greater than 0.75 differed between the IFX treatment failure group and non-IFX treatment failure group in the training cohort ($P<0.05$). Following LASSO regression analysis, 6 radiomics features and 7 deep learning features were finally screened out. Then, we further conduct LASSO on clinical characteristics selection. Finally, 4 clinical characteristics were included into the next model construction.

Performance Analysis of the Clinical, Radiomics, Deep Learning Models and the Fusion Model

We created SVM machine learning models based on clinical, radiomics, deep-learning, and fusion model. The predictive efficiency of each model is summarized in Table 2. Figure 3A demonstrated receiver operating characteristics curves of the SVM machine learning models, with area under the curve (AUC) of 0.91 (95% CI: 0.85–0.95), 0.82 (95% CI:0.76–0.88), 0.78 (95% CI: 0.71–0.84), 0.91 (95% CI: 0.86–0.95), and 0.88 (95% CI: 0.82–0.92) for clinical, radiomics, deep-learning, early fusion, and late fusion model in the training set, respectively. The external test set cohort had AUC of 0.71 (95% CI: 0.61–0.80), 0.72 (95% CI: 0.62–0.81), 0.72 (95% CI: 0.62–0.81), 0.85 (95% CI: 0.77–0.92), and 0.81 (95% CI: 0.72–0.87) in clinical, radiomics, deep-learning, early fusion, and late fusion model, respectively, as shown in the Figure 3B. Moreover, the sensitivity and specificity of each model was: clinical model: 0.92, 0.80; deep-learning: 0.70, 0.81; radiomics: 0.59, 0.87; early fusion: 0.92, 0.81; late fusion: 0.70, 0.99 in the training cohort. And, the sensitivity and specificity of each model was: clinical model: 0.60, 0.80; deep-learning: 0.77, 0.61; radiomics: 0.60, 0.87; early fusion: 0.77, 0.85; late fusion: 0.81, 0.67 in the test cohort. Brier's score, commonly used metrics for assessing the accuracy of probabilistic prediction models, indicated that early fusion model obtained the best performance among all models (Table 2). Figure 4 also demonstrated that the early fusion model has top precision-recall curve AUC. Figure 5 displayed the decision curve analysis for the model's training and external test cohort, showing significant net gains for the early fusion based SVM model. Figure 6 presented the calibration curve of each model. Kaplan-Meier curves for IFX response stratified by early fusion are shown in Figure 7. There were significant associations (Training cohort: Log rank

Table 2 The Performance Evaluation of the Developed Models in the Training and External Test Cohorts

Model	AUC (95% CI)	p-value	Youden's index	Sensitivity	Specificity	Accuracy	PPV	NPV	Brier score
Training									
Clinical	0.91 (0.85–0.95)	0.41	0.30	0.92	0.80	0.86	0.82	0.90	0.12
Deep-learning	0.82 (0.76–0.88)	0.03	0.47	0.70	0.81	0.75	0.78	0.73	0.20
Radiomics	0.78 (0.71–0.84)	0.01	0.51	0.59	0.87	0.73	0.82	0.68	0.23
Early Fusion	0.91 (0.86–0.95)	Reference	0.36	0.92	0.81	0.86	0.83	0.91	0.12
Late Fusion	0.88 (0.82–0.92)	0.37	0.86	0.70	0.99	0.84	0.98	0.77	0.13
Test									
Clinical	0.71 (0.61–0.80)	0.01	0.45	0.60	0.80	0.71	0.70	0.72	0.23
Deep-learning	0.72 (0.62–0.81)	0.02	0.44	0.77	0.61	0.68	0.61	0.77	0.22
Radiomics	0.72 (0.62–0.81)	0.06	0.51	0.60	0.87	0.75	0.79	0.73	0.22
Early Fusion	0.85 (0.77–0.92)	Reference	0.37	0.77	0.85	0.81	0.80	0.82	0.15
Late Fusion	0.81 (0.72–0.87)	0.49	0.16	0.81	0.67	0.73	0.66	0.82	0.21

Abbreviations: AUC, area under characteristic receiver operating curve; CI, confidence interval; PPV, positive predictive value; NPV, negative predictive value.

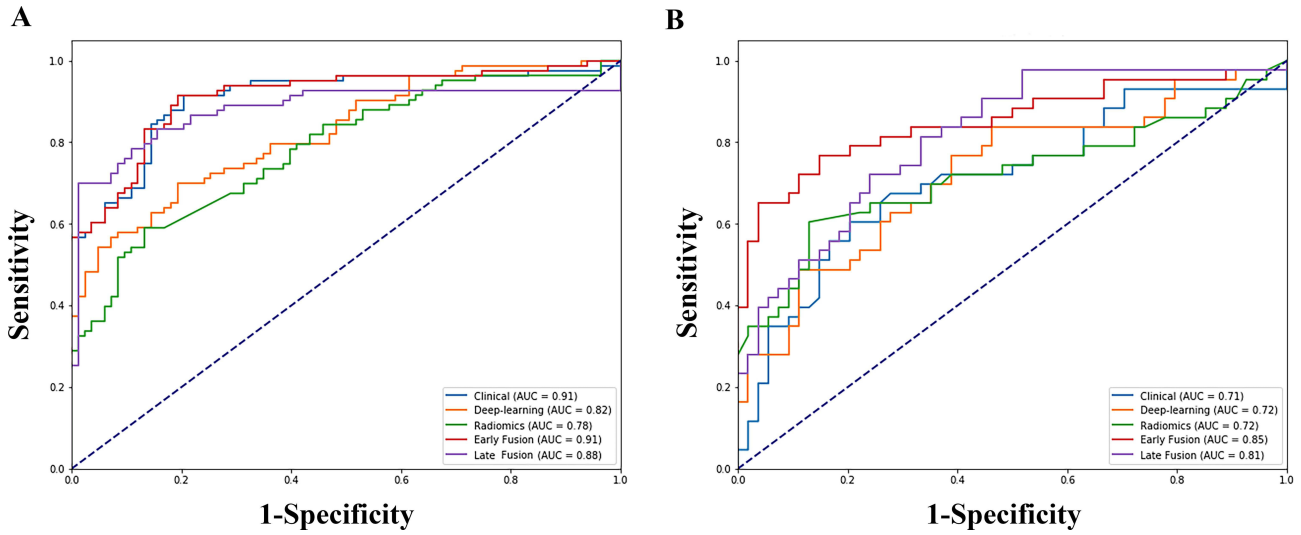

Figure 3 Prediction performance of the clinical, radiomics, deep learning, early fusion, and late fusion models with receiver operating characteristic curve analysis in the training (**A**) and test (**B**) cohorts.

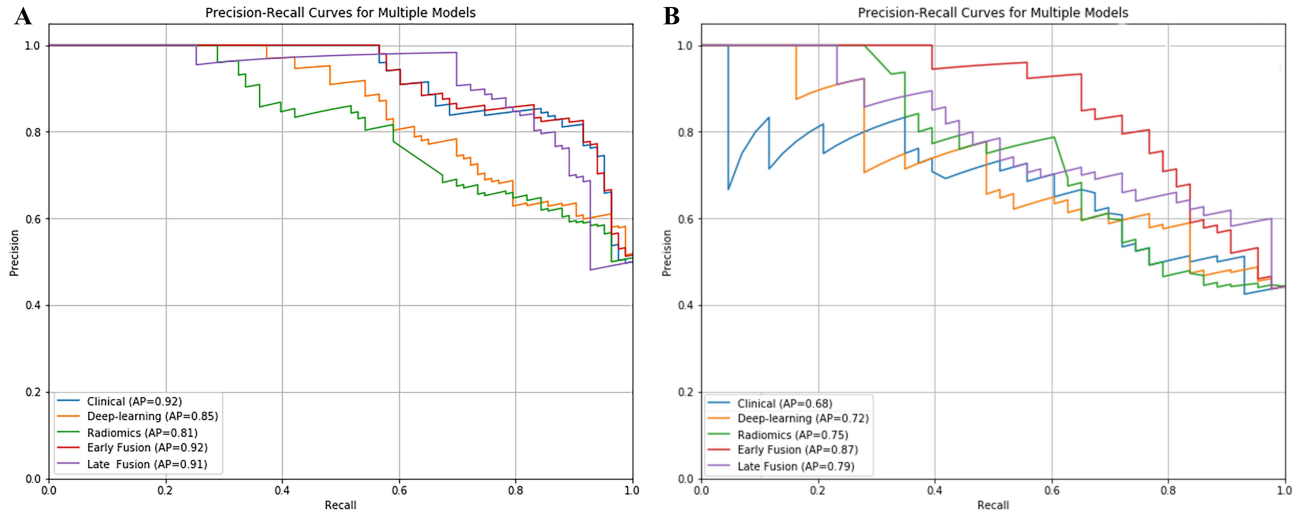

Figure 4 Prediction performance of the clinical, radiomics, deep learning, early fusion, and late fusion models with precision-recall curve analysis in the training (**A**) and test (**B**) cohorts.

test, p <0.001, hazard ratio [HR]=0.120; Test cohort: Log rank test, p <0.001, HR=0.113) among predicted responders and predicted non-responders patients stratified by early fusion model. This underscored the clinical relevance of our model in prediction of non-responsive to IFX treatment.

For the early fusion training, a sample size of 14 patients was necessary, consisting of 8 non-responders and 6 responders. To perform independent validation, a separate sample size of 19 patients was required, including 10 non-responders and 9 responders. As a result, the sample sizes used in this study—166 for the training cohort and 97 for the total test cohort—were sufficient to detect true AUC values of 0.91 and 0.85, respectively, each differing from 0.500, with 80% power.

Discussion

Currently, IFX therapy is crucial in the initial clinical management of CD patients for maintaining long-term remission. Our study developed a CTE-based multimodal fusion model to predict IFX treatment outcomes. By incorporating clinical characteristics, radiomics, deep-learning features, we crafted a comprehensive model to early identify CD patients at high

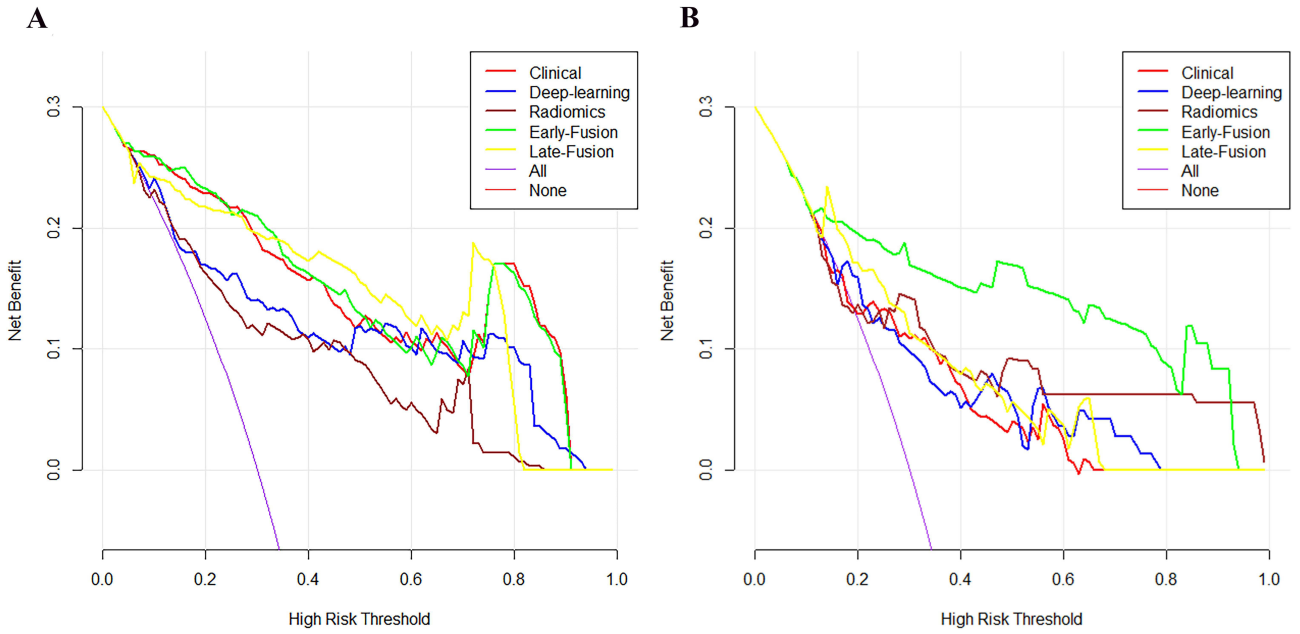

Figure 5 Decision curve analysis of the clinical, radiomics, deep learning, early fusion, and late fusion models in the training (**A**) and test (**B**) cohorts.

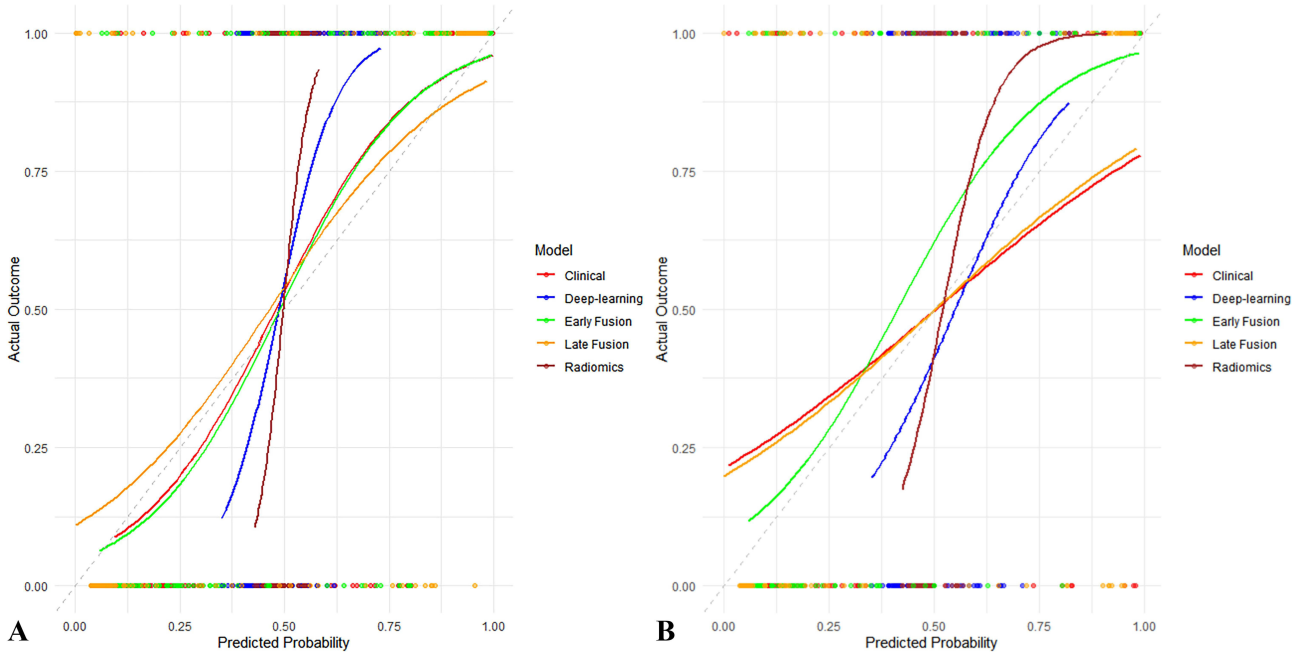

Figure 6 Calibration curve of the clinical, radiomics, deep learning, early fusion, and late fusion models in the training (**A**) and test (**B**) cohorts.

risk of IFX treatment failure. This model achieved an AUC of 0.91 (95% CI: 0.86–0.95) and 0.85 (95% CI: 0.77–0.92) in the training and test cohorts, respectively. Our findings, leveraging radiomic and deep-learning features of intestinal lesions from routine CTE images alongside clinical characteristics, could enhance personalized prediction and treatment planning for CD patients.

The prediction of response to IFX in CD patients has raised much concern for a long time. Bai et al suggested that baseline gene expression was related to PLR to IFX in CD patients.[44] Previous study by Nissila et al also reported that faecal bacterial and fungal microbiota composition could offer a predictive tool to evaluate IFX response in CD

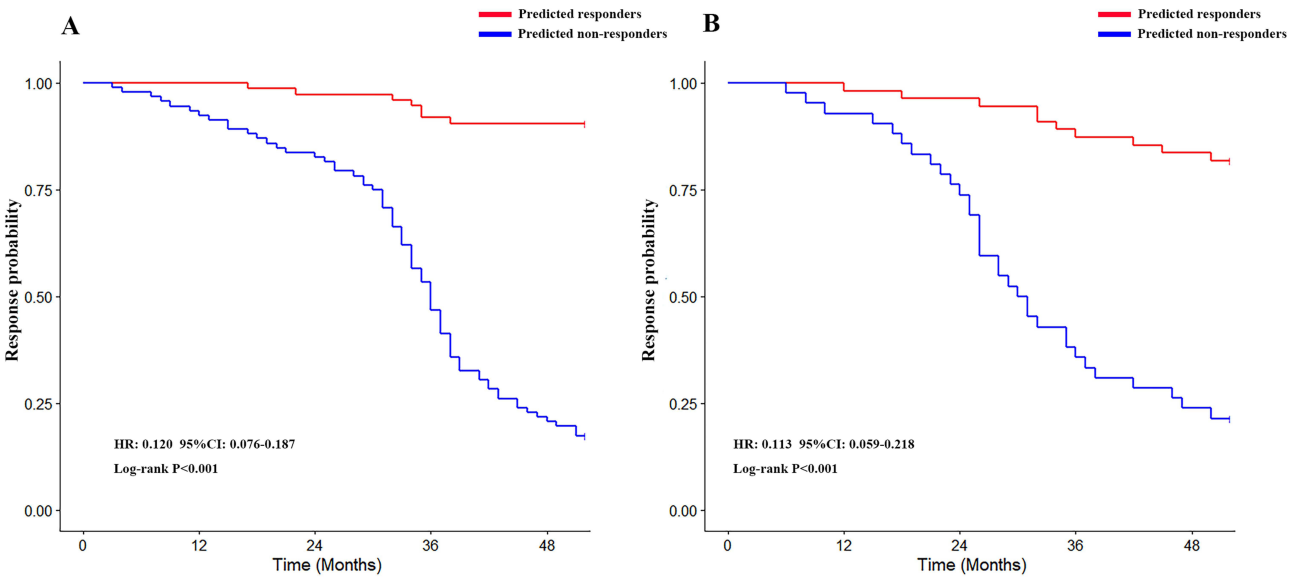

Figure 7 Kaplan-Meier survival curves for early fusion model in the training (**A**) and test (**B**) cohorts.

patients.[45] Moreover, there were several studies focused on prediction of IFX response based on clinical characteristics or radiomics data, and exhibiting moderate accuracy in predicting IFX response with AUC<0.850.[38,46,47] The present study tried to develop an integrated model based on deep-learning, radiomics, and clinical features to predict the IFX response in CD patients. Our integrated model can comprehensively analyze the routine medical imaging by deep learning and radiomics, and non-invasively predict IFX response, therefore, not only avoiding inappropriate therapy but also screening more patients suitable for IFX treatment. The findings from our study suggest that the proposed model offers additional value in the realm of personalized diagnostic approaches and treatment decision-making.

Radiomics, through the high-throughput extraction of quantitative features from medical imaging, transforms images into analyzable data for clinical decision support. This process, when combined with other significant variables, aids in constructing predictive models for diagnosis and prognosis.[48,49] Previous studies have indicated the value of radiomic features from CTE images of inflammation diseased bowel walls in predicting mucosal healing or identifying intestinal fibrosis in CD patients.[31,50] Given the common occurrence of multiple intestinal segment involvement in CD, for consistency and efficiency, our analysis focuses on the segment exhibiting the most active inflammation. Previous studies have shown a direct correlation between the thickness of the intestinal wall and disease activity in CD, with specific thickness thresholds providing accurate assessments of inflammatory activity.[51,52] Intestinal strictures in CD, characterized by varying degrees of inflammation and fibrosis, are linked to the emergence of penetrating complications.[18,53] Hence, the presence of intestinal strictures and penetrating complications in affected bowel walls are important markers of active inflammation.

Deep Learning represents a cutting-edge trend in artificial intelligence research,[54] marking significant advancements in computer vision, machine learning, and various other fields. Li et al developed and validated a deep learning-based nomogram for CT images, aimed at predicting lymph node status in gastric cancer. This model demonstrated strong discrimination ability in the test set, with an AUC of 0.821 (95% CI: 0.722–0.920).[55] However, to the best of our knowledge, there are few literatures focusing on the deep learning features from medical images in CD patients. Thus, in our study, we also separately validated the deep learning efficacy in CD.

A recent systematic review compared the performances of radiomics models, deep learning models, and multidomain fusion models in medical research.[56] The findings showed that the fusion model excelled in 63% of the studies, fell short in 25%, and matched the performance of others in 13%. Our study indicated that the effectiveness of the fusion model largely depends on the fusion methods employed. And early fusion being the predominant approach in biomedical research.[57] In a related study, Huang et al employed a CT-based Convolutional Neural Network (CNN) and clinical data

to detect pulmonary embolism, evaluating seven different fusion architectures.[58] When choosing a fusion approach for a specific medical issue, relying solely on past experience is not recommended. Instead, conducting preliminary experiments with various approaches and selecting the most optimal one is advisable.

Our study introduced a noninvasive approach using a fusion model to predict IFX response in CD patients, offering prediction support and assisting in making informed treatment choices. The model recommended alternative therapy for patients identified as non-response to IFX, therefore reducing the likelihood of complications occurrence and economic burden. Additionally, this early fusion model can aid clinicians in treatment decision by allowing for the early prediction of possible non-response, ensuring that necessary therapeutic interventions are applied without delay. Compared to previous studies focus on identifying IFX non-responders, our research not only employed potential clinical characteristics, but also tried to extract data from medical imaging based on radiomics and deep learning.[47,59] Our manuscript showed that model based on clinical characteristics, radiomics, and deep learning features using early fusion strategy could contribute to decision-making for clinicians in IFX administration management in CD patients.

It's important to acknowledge its limitations in our study. Firstly, our analysis was retrospective and prospective, which led to unavoidable selection bias. Multicenter studies are required to validate our findings. Secondly, we currently employed a manual method to delineate the VOI. In the future, we will involve leveraging artificial intelligence for automatic segmentation and measurement of the bowel, aiming to enhance the model's accuracy. Furthermore, MR enterography also plays a crucial role in assessing the disease in CD patients. Our results warrant further investigation in subsequent MR enterography studies.

In conclusion, the proposed early fusion model showed high accuracy in identifying Crohn's disease patients at high risk for IFX response. Our study introduced a noninvasive tool using standard CTE, which could help predict individual treatment responses in CD patients undergoing IFX therapy. This tool aims to enhance clinical remission rates through reasonable treatment planning.

Data Sharing Statement

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Ethics Approval and Consent to Participate

This research protocol was approved by the Medical Ethics Committee of the First affiliated hospital of Wenzhou medical university, Xiamen Second People's Hospital and Yueqing people's hospital. It was conducted according to the guidelines of the Declaration of Helsinki. All data included in this research were anonymized; thus, the need for informed consent has been waived.

Disclosure

The authors declare that they have no known-competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Spinelli A, Bonovas S, Burisch J, et al. ECCO guidelines on therapeutics in ulcerative colitis: surgical treatment. *J Crohn's Colitis*. 2022;16(2):179–189. doi:10.1093/ecco-jcc/jjab177
2. Feng J, Feng Q, Chen Y, et al. MRI-based radiomic signature identifying secondary loss of response to infliximab in Crohn's disease. *Front Nutr*. 2021;8:773040. doi:10.3389/fnut.2021.773040
3. Ding NS, Hart A, De Cruz P. Systematic review: predicting and optimising response to anti-TNF therapy in Crohn's disease - algorithm for practical management. *Aliment Pharmacol Ther*. 2016;43(1):30–51. doi:10.1111/apt.13445
4. Best WR, Becktel JM, Singleton JW. Rederived values of the eight coefficients of the Crohn's Disease Activity Index (CDAI). *Gastroenterology*. 1979;77(4 Pt 2):843–846. doi:10.1016/0016-5085(79)90384-6
5. Ding SS, Liu C, Zhang YF, et al. Contrast-enhanced ultrasound in the assessment of Crohn's disease activity: comparison with computed tomography enterography. *Radiol Med*. 2022;127(10):1068–1078. doi:10.1007/s11547-022-01535-z
6. Dane B, Sarkar S, Nazarian M, et al. Crohn disease active inflammation assessment with iodine density from dual-energy CT enterography: comparison with histopathologic analysis. *Radiology*. 2021;301(1):144–151. doi:10.1148/radiol.2021204405

7. Henao MP, Bewtra M, Osterman MT, et al. Measurement of Inflammatory Bowel Disease Symptoms: reliability of an Abbreviated Approach to Data Collection. *Inflamm Bowel Dis*. 2015;21(10):2262–2271. doi:10.1097/MIB.0000000000000496
8. Stone AA, Shiffman S, Schwartz JE, Broderick JE, Hufford MR. Patient non-compliance with paper diaries. *BMJ*. 2002;324(7347):1193–1194. doi:10.1136/bmj.324.7347.1193
9. Geldof J, Truyens M, Hanssens M, et al. Prophylactic versus endoscopy-driven treatment of Crohn's postoperative recurrence: a retrospective, multicentric European study (PORCSE study). *J Crohn's Colitis*. 2024;18(8):1202–1214. doi:10.1093/ecco-jcc/jjae011
10. Nielsen OH, Gubatan JM, Kolho KL, Streett SE, Maxwell C. Updates on the management of inflammatory bowel disease from periconception to pregnancy and lactation. *Lancet*. 2024;403(10433):1291–1303. doi:10.1016/S0140-6736(24)00052-7
11. Radmard AR, Amouei M, Torabi A, et al. MR enterography in ulcerative colitis: beyond endoscopy. *Radiographics*. 2024;44(1):e230131. doi:10.1148/rg.230131
12. Miles A, Bhatnagar G, Halligan S, et al. Magnetic resonance enterography, small bowel ultrasound and colonoscopy to diagnose and stage Crohn's disease: patient acceptability and perceived burden. *Eur Radiol*. 2019;29(3):1083–1093. doi:10.1007/s00330-018-5661-2
13. Bruining DH, Zimmermann EM, Loftus EV Jr, Sandborn WJ, Sauer CG, Strong SA. Consensus recommendations for evaluation, interpretation, and utilization of computed tomography and magnetic resonance enterography in patients with small bowel Crohn's disease. *Radiology*. 2018;286(3):776–799. doi:10.1148/radiol.2018171737
14. Ognibene NM, Basile M, Di Maurizio M, Petrillo G, De Filippi C. Features and perspectives of MR enterography for pediatric Crohn disease assessment. *Radiol Med*. 2016;121(5):362–377. doi:10.1007/s11547-015-0613-2
15. Cicero G, Alibrandi A, Blandino A, et al. DWI ratios: new indexes for Crohn's disease activity at magnetic resonance enterography? *Radiol Med*. 2023;128(1):16–26. doi:10.1007/s11547-022-01573-7
16. Napolitano M, Munari AM, Di Leo G, et al. MR enterography grading of pediatric ileocolonic Crohn disease activity based on a single bowel segment. *Radiol Med*. 2021;126(11):1396–1406. doi:10.1007/s11547-021-01409-w
17. Lu C, Rosentreter R, Delisle M, et al. Systematic review: defining, diagnosing and monitoring small bowel strictures in Crohn's disease on intestinal ultrasound. *Aliment Pharmacol Ther*. 2024;59(8):928–940. doi:10.1111/apt.17918
18. Guglielmo FF, Anupindi SA, Fletcher JG, et al. Small bowel Crohn disease at CT and MR enterography: imaging atlas and glossary of terms. *Radiographics*. 2020;40(2):354–375. doi:10.1148/rg.2020190091
19. Nardone OM, Ponsiglione A, de Sire R, et al. Impact of sarcopenia on clinical outcomes in a Cohort of caucasian active Crohn's disease patients undergoing multidetector CT-enterography. *Nutrients*. 2022;14(17):3460. doi:10.3390/nu14173460
20. Weinstein-Nakar I, Focht G, Church P, et al. Associations among mucosal and transmural healing and fecal level of calprotectin in children with Crohn's disease. *Clin Gastroenterol Hepatol*. 2018;16(7):1089–1097.e1084. doi:10.1016/j.cgh.2018.01.024
21. Sakurai T, Katsuno T, Saito K, et al. Mesenteric findings of CT enterography are well correlated with the endoscopic severity of Crohn's disease. *Eur J Radiol*. 2017;89:242–248. doi:10.1016/j.ejrad.2016.10.022
22. Li X, Wu W, Yuan Y, et al. CT energy spectral parameters of creeping fat in Crohn's disease and correlation with inflammatory activity. *Insights Imaging*. 2024;15(1):10. doi:10.1186/s13244-023-01592-6
23. Wang H, Wang L, Lee EH, et al. Decoding COVID-19 pneumonia: comparison of deep learning and radiomics CT image signatures. *Eur J Nucl Med Mol Imaging*. 2021;48(5):1478–1486. doi:10.1007/s00259-020-05075-4
24. NQK L, Kha QH, Nguyen VH, Chen YC, Cheng SJ, Chen CY. Machine learning-based radiomics signatures for EGFR and KRAS mutations prediction in non-small-cell lung cancer. *Int J Mol Sci*. 2021;22(17). doi:10.3390/ijms22179254
25. Tomita H, Kobayashi T, Takaya E, et al. Deep learning approach of diffusion-weighted imaging as an outcome predictor in laryngeal and hypopharyngeal cancer patients with radiotherapy-related curative treatment: a preliminary study. *Eur Radiol*. 2022;32(8):5353–5361. doi:10.1007/s00330-022-08630-9
26. 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
27. Song F, Ma M, Zeng S, et al. CT enterography-based radiomics combined with body composition to predict infliximab treatment failure in Crohn's disease. *Radiol Med*. 2024;129(2):175–187. doi:10.1007/s11547-023-01748-w
28. Li X, Zhang N, Hu C, et al. CT-based radiomics signature of visceral adipose tissue for prediction of disease progression in patients with Crohn's disease: a multicentre cohort study. *EClinicalMedicine*. 2023;56:101805. doi:10.1016/j.eclinm.2022.101805
29. Yueying C, Jing F, Qi F, Jun S. Infliximab response associates with radiologic findings in bio-naïve Crohn's disease. *Eur Radiol*. 2023;33(8):5247–5257. doi:10.1007/s00330-023-09542-y
30. Zhu C, Yu Y, Wang S, et al. A novel clinical radiomics nomogram to identify Crohn's disease from intestinal tuberculosis. *J Inflamm Res*. 2021;14:6511–6521. doi:10.2147/JIR.S344563
31. Zhu C, Hu J, Wang X, et al. A novel clinical radiomics nomogram at baseline to predict mucosal healing in Crohn's disease patients treated with infliximab. *Eur Radiol*. 2022;32(10):6628–6636. doi:10.1007/s00330-022-08989-9
32. Singh SP, Wang L, Gupta S, Goli H, Padmanabhan P, Gulyás B. 3D deep learning on medical images: a review. *Sensors*. 2020;20(18):5097. doi:10.3390/s20185097
33. Gong Y, Wang L, Guo R, Lazebnik S. *Multi-Scale Orderless Pooling of Deep Convolutional Activation Features. In: 2014*. Cham: Springer International Publishing; 2014:392–407.
34. Pease M, Arefan D, Barber J, et al. Outcome prediction in patients with severe traumatic brain injury using deep learning from head CT scans. *Radiology*. 2022;304(2):385–394. doi:10.1148/radiol.212181
35. Wang W, Liang H, Zhang Z, et al. Comparing three-dimensional and two-dimensional deep-learning, radiomics, and fusion models for predicting occult lymph node metastasis in laryngeal squamous cell carcinoma based on CT imaging: a multicentre, retrospective, diagnostic study. *EClinicalMedicine*. 2024;67:102385. doi:10.1016/j.eclinm.2023.102385
36. Maaser C, Sturm A, Vavricka SR, et al. ECCO-ESGAR guideline for diagnostic assessment in IBD part 1: initial diagnosis, monitoring of known IBD, detection of complications. *J Crohn's Colitis*. 2019;13(2):144–164. doi:10.1093/ecco-jcc/jjy113
37. Li L, Chen R, Zhang Y, et al. A novel model based on serum biomarkers to predict primary non-response to infliximab in Crohn's disease. *Front Immunol*. 2021;12:646673. doi:10.3389/fimmu.2021.646673

38. Kennedy NA, Heap GA, Green HD, et al. Predictors of anti-TNF treatment failure in anti-TNF-naive patients with active luminal Crohn's disease: a prospective, multicentre, cohort study. *Lancet Gastroenterol Hepatol*. 2019;4(5):341–353. doi:10.1016/S2468-1253(19)30012-3
39. Ordás I, Rimola J, Alfaro I, et al. Development and validation of a simplified magnetic resonance index of activity for Crohn's disease. *Gastroenterology*. 2019;157(2):432–439.e431. doi:10.1053/j.gastro.2019.03.051
40. Mukherjee S, Patra A, Khasawneh H, et al. Radiomics-based machine-learning models can detect pancreatic cancer on prediagnostic computed tomography scans at a substantial lead time before clinical diagnosis. *Gastroenterology*. 2022;163(5):1435–1446.e1433. doi:10.1053/j.gastro.2022.06.066
41. Wang F, Wang CL, Yi YQ, et al. Comparison and fusion prediction model for lung adenocarcinoma with micropapillary and solid pattern using clinicoradiographic, radiomics and deep learning features. *Sci Rep*. 2023;13(1):9302. doi:10.1038/s41598-023-36409-5
42. Obuchowski NA. ROC analysis. *AJR Am J Roentgenol*. 2005;184(2):364–372. doi:10.2214/ajr.184.2.01840364
43. Dercle L, Fronheiser M, Lu L, et al. Identification of non-small cell lung cancer sensitive to systemic cancer therapies using radiomics. *Clin Cancer Res*. 2020;26(9):2151–2162. doi:10.1158/1078-0432.CCR-19-2942
44. Bai BYH, Reppell M, Smaoui N, et al. Baseline expression of immune gene modules in blood is associated with primary response to anti-TNF therapy in Crohn's disease patients. *J Crohn's Colitis*. 2024;18(3):431–445. doi:10.1093/ecco-jcc/jjad166
45. Ventin-Holmberg R, Eberl A, Saqib S, et al. Bacterial and fungal profiles as markers of infliximab drug response in inflammatory bowel disease. *J Crohn's Colitis*. 2021;15(6):1019–1031. doi:10.1093/ecco-jcc/jjaa252
46. Wang Y, Luo Z, Zhou Z, et al. CT-based radiomics signature of visceral adipose tissue and bowel lesions for identifying patients with Crohn's disease resistant to infliximab. *Insights Imaging*. 2024;15(1):28. doi:10.1186/s13244-023-01581-9
47. Chen Y, Li H, Feng J, Suo S, Feng Q, Shen J. A novel radiomics nomogram for the prediction of secondary loss of response to infliximab in crohn's disease. *J Inflamm Res*. 2021;14:2731–2740. doi:10.2147/JIR.S314912
48. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology*. 2016;278(2):563–577. doi:10.1148/radiol.2015151169
49. Lambin P, Rios-Velazquez E, Leijenaar R, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer*. 2012;48(4):441–446. doi:10.1016/j.ejca.2011.11.036
50. Li X, Liang D, Meng J, et al. Development and validation of a novel computed-tomography enterography radiomic approach for characterization of intestinal fibrosis in Crohn's disease. *Gastroenterology*. 2021;160(7):2303–2316.e2311. doi:10.1053/j.gastro.2021.02.027
51. Rong C, Zhu C, He L, et al. CTE-based radiomics models can identify mucosal healing in patients with Crohn's disease. *Acad Radiol*. 2023;30(Suppl 1):S199–s206. doi:10.1016/j.acra.2023.04.022
52. Vilela EG, Torres HO, Martins FP, Ferrari Mde L, Andrade MM, Cunha AS. Evaluation of inflammatory activity in Crohn's disease and ulcerative colitis. *World J Gastroenterol*. 2012;18(9):872–881. doi:10.3748/wjg.v18.i9.872
53. 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
54. Sun Q, Lin X, Zhao Y, et al. Deep learning vs. radiomics for predicting axillary lymph node metastasis of breast cancer using ultrasound images: don't forget the peritumoral region. *Front Oncol*. 2020;10:53. doi:10.3389/fonc.2020.00053
55. Li J, Dong D, Fang M, et al. Dual-energy CT-based deep learning radiomics can improve lymph node metastasis risk prediction for gastric cancer. *Eur Radiol*. 2020;30(4):2324–2333. doi:10.1007/s00330-019-06621-x
56. Demircioğlu A. Are deep models in radiomics performing better than generic models? A systematic review. *Eur Radiol Exp*. 2023;7(1):11. doi:10.1186/s41747-023-00325-0
57. Mohsen F, Ali H, El Hajj N, Shah Z. Artificial intelligence-based methods for fusion of electronic health records and imaging data. *Sci Rep*. 2022;12(1):17981. doi:10.1038/s41598-022-22514-4
58. Huang SC, Pareek A, Zamanian R, Banerjee I, Lungren MP. Multimodal fusion with deep neural networks for leveraging CT imaging and electronic health record: a case-study in pulmonary embolism detection. *Sci Rep*. 2020;10(1):22147. doi:10.1038/s41598-020-78888-w
59. Yang T, Feng J, Yao R, Feng Q, Shen J. CT-based pancreatic radiomics predicts secondary loss of response to infliximab in biologically naïve patients with Crohn's disease. *Insights Imaging*. 2024;15(1):69. doi:10.1186/s13244-024-01637-4