Macrothink Institute™

Journal of Biology and Life Science
ISSN 2157-6076
2027, Vol. 18, No. 1

Development and Validation of a Machine Learning-Based Model for Predicting the Risk of Unplanned 30-Day Rehospitalization in Patients with Inflammatory Bowel Disease

Hong Dai

Department of Gastroenterology, Deyang People's Hospital, China.

Email: daihong0118@163.com

Yao Zeng (Corresponding author)

Department of Gastroenterology, Deyang People's Hospital, China

Lu Cao

Department of Gastroenterology, Sichuan Provincial People's Hospital, China

Received: October 10, 2026 Accepted: November 12, 2026 Published: November 17, 2026

doi:10.5296/jbls.v18i1.24163 URL: https://doi.org/10.5296/jbls.v18i1.24163

Abstract

To develop and validate a machine learning-based predictive model for the risk of unplanned readmission within 30 days among patients with inflammatory bowel disease (IBD), thereby providing a quantitative tool to aid in the early clinical identification of high-risk patients and the optimization of post-discharge management strategies, a multicenter retrospective study design was employed to collect clinical data on inpatients with IBD from three Grade A Level 3 hospitals (the top-tier in mainland China) in Deyang City and Chengdu City of Sichuan Province between October 2021 and October 2024. Using 30-day unplanned readmission as the outcome variable, we employed LASSO regression with five-fold cross-validation to select the optimal predictors and constructed three machine learning prediction models: Support Vector Machine (SVM), XGBoost, and LightGBM. The dataset was randomly split into a training set and a validation set in a 7:3 ratio, and internal validation was performed using the Bootstrap method (2,000 iterations) to compare the accuracy, sensitivity, specificity, and area under the ROC curve (AUC) of the three models. Results demonstrated that a total of 560 IBD patients were enrolled, with 78 cases (13.9%) experiencing unplanned readmission within 30 days. LASSO regression identified seven predictive factors: daily

bowel movement frequency, C-reactive protein, albumin, Barthel Index, number of medications prescribed at discharge, regular follow-up, and length of hospital stay. Among the three models, XGBoost performed the best, with an AUC of 0.842 (95% CI: 0.781–0.903) on the validation set, an accuracy of 80.4%, a sensitivity of 76.9%, and a specificity of 81.5%; LightGBM ranked second (AUC = 0.818), while SVM performed the worst (AUC = 0.763). Bootstrap internal validation demonstrated good stability of the XGBoost model (AUC = 0.836, 95% CI: 0.775–0.897). In conclusion, the XGBoost-based risk prediction model for 30-day unplanned readmission in IBD patients exhibits good predictive performance and stability, and could serve as a reference for clinical nursing risk assessment and the formulation of post-discharge follow-up strategies.

Keywords: inflammatory bowel disease, unplanned readmission, machine learning, risk prediction model, nursing risk assessment

1. Background

Inflammatory bowel disease (IBD) is a chronic, nonspecific inflammatory condition of the intestines with an unknown etiology, primarily comprising ulcerative colitis (UC) and Crohn's disease (CD) (Bruner et al., 2023). In recent years, IBD has gradually spread from Western countries to the rest of the world, and China has the highest incidence rate in Asia (Agrawal & Jess, 2022; Aniwan et al., 2022; Ng et al., 2017). It is projected that by 2025, the number of IBD patients in China will reach 1.5 million (Kaplan, 2015). IBD is characterized by recurrent flare-ups and lifelong treatment. Due to acute symptoms and complications such as obstruction, perforation, and bleeding, it severely impacts patients' quality of life and increases the societal and healthcare burden (Park & Park, 2022).

Unplanned readmission is a key indicator for assessing the quality of medical care and the effectiveness of patient management. International studies show that the unplanned readmission rate for IBD patients is 5.0% within 1 week of discharge, 14.0% within 1 month, 23.7% within 3 months, and as high as 39.2% within 1 year (Mudireddy et al., 2017). An analysis of the U.S. National Unplanned Rehospitalization Database revealed that the 30-day unplanned rehospitalization rate for Crohn's disease increased from 14.9% in 2010 to 17.6% in 2018, while that for ulcerative colitis rose from 14.3% to 15.7% (Dahiya et al., 2022). An analysis by Santiago et al. (2020) of hospitalized IBD patients in Portugal found that the 30-day readmission rate was 30.00% for Crohn's disease and 24.22% for ulcerative colitis. A systematic analysis by Nguyen et al. (2020) showed that the 30-day and 90-day unplanned readmission rates for IBD patients were 18.1% and 26.0%, respectively. Persistently high readmission rates can lead to disease exacerbation, deterioration, and even an increased risk of death (Weissman et al., 2021), while also increasing the financial burden on patients and wasting healthcare resources.

Risk prediction models can identify high-risk patients and provide a basis for personalized interventions (Mo Hangfeng et al., 2024). In recent years, machine learning algorithms have been increasingly applied in clinical prediction models, as their ability to handle high-dimensional data and nonlinear relationships surpasses that of traditional logistic regression (Riley et al., 2020). However, there is currently a lack of multicenter machine

learning studies on 30-day unplanned readmission in the Chinese IBD population. This study aims to construct and validate a risk prediction model for 30-day unplanned readmission in IBD patients based on multicenter retrospective data, thereby providing a quantitative tool for clinical care practice.

2. Research Methodology

2.1 Study Design

This study is a multicenter retrospective study. As this is a retrospective study, informed consent from patients was waived. All data were de-identified and used solely for the analysis in this study.

2.2 Study Population

This retrospective study reviewed medical records of patients hospitalized for IBD at three Grade A Level 3 hospitals in Deyang and Chengdu from October 2021 to October 2024.

Inclusion criteria: (1) Age ≥ 18 years; (2) Met the diagnostic criteria established by the IBD Working Group of the Chinese Medical Association's Branch of Gastroenterology in 2018 (Wu, 2018); (3) Complete clinical medical records from which study variables could be effectively extracted.

Exclusion Criteria: (1) Patients with malignant tumors or other severe systemic diseases; (2) Patients with outpatient visit records but no hospitalization; (3) Patients admitted for the first time for reasons unrelated to IBD or for elective procedures; (4) Pregnant women or postpartum women; (5) Patients who died upon admission; (6) Patients with missing readmission date information.

Sample Size Calculation: The model sample size was calculated based on the event-per-variable (EPV) (Riley et al., 2020). This study estimates that 6–8 study variables will be included in the model; assuming EPV = 10, the minimum number of events required is 70; According to the literature, the 30-day unplanned readmission rate for IBD patients is 14% (Mudireddy et al., 2017). Taking into account a 10% non-valid sample rate, the required sample size = 70 ÷ 14% ÷ (1–10%) ≈ 556 cases; therefore, this study plans to include a sample size of no fewer than 560 cases.

2.3 Definition of Relevant Concepts

A 30-day unplanned readmission refers to the unplanned readmission of an inpatients within 30 days of discharge due to symptoms of the same disease or a related condition. The same disease refers to a primary diagnosis of IBD at the time of readmission; symptoms of a related condition refer to a readmission triggered by an IBD flare-up, characterized by symptoms of disease exacerbation such as diarrhea, abdominal pain, bloody stools, or fever (Sbeit et al., 2021). Unplanned readmissions exclude those resulting from scheduled reasons (e.g., readmission for follow-up as directed by a physician).

2.4 Study Variables

Through a literature review and expert consultation, combined with clinical feasibility, the following candidate predictive variables were identified:

(1) Demographic characteristics: gender, age, household income, health insurance type, occupation, smoking status, alcohol consumption, and BMI.

(2) History of underlying conditions: diabetes, hypertension, coronary heart disease, congestive heart failure, chronic lung disease, and the Charlson Comorbidity Index (Charlson et al., 1987).

(3) IBD-related diagnoses and treatments: primary diagnosis, duration of IBD, number of IBD complications, medications (corticosteroids, immunosuppressants, etc.), and history of IBD surgery.

(4) Laboratory Tests: Hemoglobin, C-reactive protein (CRP), erythrocyte sedimentation rate, albumin.

(5) Clinical Examinations and Nursing Scores: Temperature/heart rate at admission, daily bowel movement frequency, endoscopy results, Barthel Index (Jiang, 2014), AHRQ-Elixhauser Index.

(6) Discharge status: length of hospital stay, pain status at discharge, number of medications prescribed at discharge, and regular follow-up.

2.5 Data Preprocessing

Statistical analysis was performed using SPSS 26.0. Continuous variables that followed a normal distribution were expressed as mean ± standard deviation ($\bar{x} \pm s$), while those that did not follow a normal distribution were expressed as median (25th percentile, 75th percentile). The Mann-Whitney U test was used for comparisons between groups. Categorical data are presented as frequencies and proportions [n (%)], and the χ^2 test was used for intergroup comparisons; when the expected frequency was <5, Fisher's exact test was used. Variables with a missing rate <20% were imputed using the Multiple Imputation by Chances (MICE) method, while variables with a missing rate ≥20% were excluded.

2.6 Feature Selection

Least Absolute Shrinkage and Selection Operator (LASSO) regression with five-fold cross-validation was used to select the optimal predictors (Garcia-Carretero et al., 2020). By introducing an L1 regularization penalty term, LASSO shrinks the coefficients of unimportant variables to zero, thereby addressing multicollinearity and high-dimensional data and improving predictive accuracy.

2.7 Model Development and Validation

Predictive models were built using Python 3.10 (scikit-learn 1.3, XGBoost 2.0, LightGBM 4.0). All study participants were randomly split into training and validation sets in a 7:3 ratio (random seed 123). Three machine learning algorithms—Support Vector Machines (SVM),

XGBoost, and LightGBM—were used to build predictive models. Internal validation was performed using the Bootstrap method (2,000 samples). Accuracy, precision, sensitivity, specificity, F1 score, and area under the ROC curve (AUC) were calculated based on the confusion matrix to compare the performance of the three models.

2.8 Statistical Analysis

Data analysis was performed using SPSS 26.0 and Python 3.10. $P < 0.05$ was considered statistically significant.

3. Study Results

3.1 Baseline Characteristics

A total of 560 IBD patients were enrolled, including 328 with UC (58.6%) and 232 with CD (41.4%). There were 78 unplanned readmissions within 30 days, resulting in a readmission rate of 13.9%. Statistically significant differences were observed between the readmission group and the non-readmission group in terms of daily bowel movement frequency, CRP, albumin, Barthel Index, number of medications prescribed at discharge, adherence to follow-up appointments, and length of hospital stay ($P < 0.05$). See Table 1.

Table 1. Comparison of Baseline Characteristics Between the Readmission Group and the Non-Readmission Group

Variable	Readmission Group (n=78)	Non-readmission Group (n=482)	Statistical Value	P-value
Age (years, ($\bar{x}$±s))	42.6±14.8	40.2±15.3	t=1.28	0.201
Male (n, %)	45(57.7)	268(55.6)	χ^2=0.12	0.729
Daily bowel movement frequency (times/day, M (IQR))	6(4~8)	3(2~5)	Z=-6.32	<0.001
CRP (mg/L, M (IQR))	28.5(15.2~45.6)	12.3(6.8~22.4)	Z=-7.15	<0.001
Albumin (g/L, ($\bar{x}$±s))	32.5±5.8	36.8±5.2	t=-6.42	<0.001
Barthel Index (points, ($\bar{x}$±s))	82.5±12.6	91.3±9.8	t=-7.05	<0.001
Number of medications prescribed at discharge (types, ($\bar{x}$±s))	6.8±2.4	5.2±2.1	t=5.89	<0.001
Regular follow-up (n, %)	42(53.8)	386(80.1)	χ^2=26.74	<0.001
Length of hospital stay (days, M (IQR))	12(8~18)	8(6~12)	Z=-5.43	<0.001

3.2 Comparison of Model Performance

The performance of the three machine learning models on the validation set is shown in Table 2. The XGBoost model performed best, with an AUC of 0.842 (95% CI: 0.781–0.903), an accuracy of 80.4%, a sensitivity of 76.9%, and a specificity of 81.5%; LightGBM ranked second (AUC = 0.818); SVM performed the worst (AUC = 0.763).

Table 2. Comparison of Validation Set Performance for the Three Machine Learning Models

Model	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC(95% CI)
SVM	73.2	69.2	74.3	0.763(0.692~0.834)
XGBoost	80.4	76.9	81.5	0.842(0.781~0.903)
LightGBM	78.6	74.4	79.6	0.818(0.754~0.882)

3.3 Feature Importance

The feature importance rankings (SHAP values) based on the XGBoost model show that the top five most influential features, in descending order: daily bowel movement frequency, CRP, albumin, Barthel Index, and regular follow-up. See Table 3.

Table 3. Feature Importance Rankings (SHAP values) for the XGBoost Model

Rank	Feature	SHAP Value (Mean)
1	Daily Bowel Movement Frequency	0.182
2	CRP	0.156
3	Albumin	0.134
4	Barthel Index	0.122
5	Regular Follow-up	0.098
6	Amount of Medication Discharged	0.076
7	Length of Hospital Stay	0.065

3.4 Internal Model Validation

Bootstrap internal validation showed that the XGBoost model had an AUC of 0.836 (95% CI: 0.775–0.897), and the calibration curve closely approximated the ideal diagonal, indicating good model stability.

4. Discussion

Based on multicenter retrospective data, this study constructed and validated a machine learning predictive model for the risk of unplanned 30-day readmission in patients with IBD. The results showed that the XGBoost model demonstrated the best predictive performance, with an AUC of 0.842 on the validation set and an AUC of 0.836 in the bootstrap internal validation, indicating that the model possesses good discriminatory ability and stability.

4.1 Predictor Analysis

This study found that daily bowel movement frequency, CRP, albumin, Barthel Index, and regular follow-up are important predictors of 30-day unplanned readmission. Daily bowel movement frequency and CRP reflect disease activity, consistent with previous studies (Mudireddy et al., 2017; Nguyen et al., 2020). Albumin levels reflect nutritional status and

the degree of inflammation; hypoalbuminemia is associated with an increased risk of readmission (Santiago et al., 2020). The Barthel Index reflects a patient's ability to perform activities of daily living; a lower score indicates poorer functional status and a higher risk of readmission (Jiang, 2014). Regular follow-up is a protective factor; patients who did not attend regular follow-up visits had a significantly increased risk of readmission, consistent with the findings of Sbeit et al. (2021). The number of medications prescribed at discharge and length of hospital stay were also important predictors, suggesting that the disease burden at discharge and the severity of the condition during hospitalization influence the risk of readmission.

4.2 Advantages of Machine Learning Models

Compared to traditional logistic regression, machine learning algorithms can automatically capture nonlinear relationships and interaction effects among variables (Riley et al., 2020). This study compared three algorithms SVM, XGBoost, and LightGBM with XGBoost performing best, likely due to its ability to handle imbalanced data and high-dimensional features. LightGBM ranked second, while SVM performed relatively poorly with small sample sizes. Future research could further explore ensemble learning strategies to enhance the models' generalization ability.

4.3 Clinical Implications

This model provides clinical nurses with a quantitative risk assessment tool to identify high-risk individuals prior to discharge, thereby enabling the development of targeted post-discharge follow-up plans and nursing interventions. For example, for patients with high daily bowel movement frequency, elevated CRP, low albumin levels, low Barthel scores, or irregular follow-up, post-discharge telephone follow-ups can be intensified, early outpatient follow-up appointments can be scheduled, and personalized dietary and medication guidance can be provided. This helps reduce the rate of unplanned readmissions, improve patients' quality of life, and conserve healthcare resources.

4.4 Limitations and Future Directions

As this study is retrospective in nature, firstly, it may be subject to selection bias and information bias. Besides, the sample was drawn from three hospitals, and its generalizability requires further validation. Furthermore, no external validation was conducted, and the model's generalizability must be confirmed by multicenter prospective studies. Future research could expand the sample size, incorporate additional potential predictors (such as fecal calprotectin and medication adherence), and explore the application of deep learning models.

5. Conclusion

Our findings highlight those clinical indicators of disease activity, nutritional status, activities of daily living, discharge medication burden, hospital stay duration and post-discharge follow-up adherence jointly determine the likelihood of short-term unplanned re-hospitalisation in IBD patients. The established XGBoost model could serve as a

quantitative risk-stratification tool for clinical front-line staff, especially clinical nurses. Before hospital discharge, clinicians and nurses can input readily-available routine clinical variables into the model to stratify patients into high-risk and low-risk groups for 30-day unplanned readmission. For high-risk individuals, targeted post-discharge management measures can be implemented, including intensified telephone follow-up, expedited outpatient revisits, individualized dietary counselling, standardized medication education, and early monitoring of bowel symptoms and inflammatory markers. Targeted risk-guided interventions are expected to help prevent IBD-related exacerbations, reduce avoidable short-term readmission events, improve patients' long-term quality of life, and optimize the allocation of limited medical and nursing resources.

The XGBoost-based model for predicting the risk of unplanned readmission within 30 days in IBD patients demonstrates good predictive performance and stability. Daily bowel movement frequency, CRP, albumin, the Barthel Index, and regular follow-up proved to be important predictive factors. This model could serve as a reference for clinical risk assessment, the development of post-discharge management and explore deep-learning or stacked-ensemble modelling strategies, so as to further improve model generalization and facilitate its eventual transformation into a usable bedside decision-aid tool for IBD clinical and nursing practice.

Acknowledgments

Not applicable.

Author contributions

Not applicable.

Funding

2025 Natural Science Research Project of Sichuan Nursing Vocational College (2025ZRY20), Deyang People's Hospital

Competing interests

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.

Informed consent

Obtained.

Ethics approval

The Publication Ethics Committee of the Macrothink Institute.

The journal's policies adhere to the Core Practices established by the Committee on Publication Ethics (COPE).

Provenance and peer review

Not commissioned; externally double-blind peer reviewed.

Data availability statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Data sharing statement

No additional data are available.

Open access

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/4.0/).

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

References

Agrawal, M., & Jess, T. (2022). Implications of the changing epidemiology of inflammatory bowel disease in a changing world. *United European Gastroenterology Journal, 10*(10), 1113–1120. https://doi.org/10.1002/ueg2.12317

Aniwan, S., Santiago, P., Loftus, E. V., & Park, S. H. (2022). The epidemiology of inflammatory bowel disease in Asia and Asian immigrants to Western countries. *United European Gastroenterology Journal, 10*(10), 1063–1076. https://doi.org/10.1002/ueg2.12350

Bruner, L. P., White, A. M., & Proksell, S. (2023). Inflammatory bowel disease. *Primary Care: Clinics in Office Practice, 50*(3), 411–427. https://doi.org/10.1016/j.pop.2023.03.009

Charlson, M. E., Pompei, P., Ales, K. L., & MacKenzie, C. R. (1987). A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *Journal of Chronic Diseases, 40*(5), 373–383. https://doi.org/10.1016/0021-9681(87)90171-8

Dahiya, D. S., Perisetti, A., Kichloo, A., Singh, A., Goyal, H., Rotundo, L., … & Inamdar, S. (2022). Increasing thirty-day readmissions of Crohn's disease and ulcerative colitis in the United States: a national dilemma. *World Journal of Gastrointestinal Pathophysiology, 13*(3), 85–95. https://doi.org/10.4291/wjgp.v13.i3.85

Garcia-Carretero, R., Vigil-Medina, L., Barquero-Perez, O., Mora-Jimenez, I., Soguero-Ruiz, C., Goya-Esteban, R., & Ramos-Lopez, J. (2020). Logistic LASSO and elastic net to characterize vitamin D deficiency in a hypertensive obese population. *Metabolic Syndrome and Related Disorders, 18*(2), 79–85. https://doi.org/10.1089/met.2019.0104

Jiang, H., Lu, H., Gu, Y., Huang, L., Ding, M., & Yu, Y. (2014). Study on correlation between Barthel self care ability score and nursing workload. *Chinese Nursing Research, 28*(34), 4238–4240. https://doi.org/10.3969/j.issn.10096493.2014.34.006

Kaplan, G. G. (2015). The global burden of IBD: from 2015 to 2025. *Nature Reviews Gastroenterology & Hepatology, 12*(12), 720–727. https://doi.org/10.1038/nrgastro.2015.150

Mo, H., Chen, Y., Han, H., Zhang, Y., Liu, Y., Zhang, M., … & Shen, J. (2024). Methods and procedures of clinical predictive model. *Chinese Journal of Evidence-Based Medicine, 24*(2), 228–236. https://doi.org/10.7507/1672-2531.202308135

Mudireddy, P., Scott, F., Feathers, A., & Lichtenstein, G. R. (2017). Inflammatory bowel disease: predictors and causes of early and late hospital readmissions. *Inflammatory Bowel Diseases, 23*(10), 1832–1839. https://doi.org/10.1097/MIB.0000000000001242

Ng, S. C., Shi, H. Y., Hamidi, N., Underwood, F. E., Tang, W., Benchimol, E. I., … & Kaplan, G. G. (2017). Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. *The Lancet, 390*(10114), 2769–2778. https://doi.org/10.1016/S0140-6736(17)32448-0

Nguyen, N. H., Koola, J., Dulai, P. S., Prokop, L. J., Sandborn, W. J., & Singh, S. (2020). Rate of risk factors for and interventions to reduce hospital readmission in patients with inflammatory bowel diseases. *Clinical Gastroenterology and Hepatology, 18*(9), 1939–1948.e7. https://doi.org/10.1016/j.cgh.2019.08.042

Park, S. H., & Park, S. H. (2022). Personalized medicine in inflammatory bowel disease: perspectives on Asia. *Journal of Gastroenterology and Hepatology, 37*(8), 1434–1445. https://doi.org/10.1111/jgh.15919

Riley, R. D., Ensor, J., Snell, K. I. E., Harrell, F. E., Jr., Martin, G. P., Reitsma, J. B., … & van Smeden, M. (2020). Calculating the sample size required for developing a clinical prediction model. *BMJ, 368*, m441. https://doi.org/10.1136/bmj.m441

Santiago, M., Magro, F., Correia, L., Portela, F., Ministro, P., Lago, P., … & Dias, C. C. (2020). Rehospitalization rates, costs, and risk factors for inflammatory bowel disease: a 16-year nationwide study. *Therapeutic Advances in Gastroenterology, 13*, 1756284820923836. https://doi.org/10.1177/1756284820923836

Sbeit, W., Khoury, T., Kadah, A., Shahin, A., Shafrir, A., Kalisky, I., … & Mari, A. (2021). Nonattendance to gastroenterologist follow-up after discharge is associated with a thirty-days re-admission in patients with inflammatory bowel disease: a multicenter study. *Minerva Medica, 112*(4), 467–473. https://doi.org/10.23736/S0026-4806.21.07442-5

Weissman, S., Sharma, S., Fung, B. M., Aziz, M., Sciarra, M., Swaminath, A., & Feuerstein, J. D. (2021). Increased mortality and healthcare costs upon hospital readmissions of ulcerative colitis flares: a large population-based cohort study. *Crohn's & Colitis 360, 3*(3), otab029. https://doi.org/10.1093/crocol/otab029

Wu, K. C., Liang, J., Ran, Z. H., Qian, J. M., Yang, H., Chen, M. H., & He, Y. (2018). Chinese consensus on diagnosis and treatment of inflammatory bowel disease (Beijing, 2018). *Chinese Journal of Practical Internal Medicine, 38*(9), 796–813. https://doi.org/10.19538/j.nk2018090106