Computational and Structural Biotechnology Journal 24 (2024) 583–592

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

journal homepage: www.elsevier.com/locate/csbj

Research article

A predictive machine-learning model for clinical decision-making in washed microbiota transplantation on ulcerative colitis

Sheng Zhang [a,b], Gaochen Lu [a,b], Weihong Wang [a,b], Qianqian Li [a,b], Rui Wang [a,b], Zulun Zhang [a,b], Xia Wu [a,b], Chenchen Liang [a,b], Yujie Liu [c], Pan Li [a,b], Quan Wen [a,b], Bota Cui [a,b], Faming Zhang [a,b,d,*,1]

[a] Department of Microbiota Medicine & Medical Center for Digestive Diseases, the Second Affiliated Hospital of Nanjing Medical University, Nanjing, China
[b] Jiangsu Engineering Research Center for Advanced Microbiota Medicine, Key Lab of Holistic Integrative Enterology, the Second Affiliated Hospital of Nanjing Medical University, Nanjing, China
[c] Department of Medicine & Therapeutics, the Chinese University of Hong Kong, Hong Kong, China
[d] National Clinical Research Center for Digestive Diseases, Xi'an, China

ARTICLE INFO

Keywords:
Fecal microbiota transplant
Machine learning
Ulcerative colitis
Clinical indicator

ABSTRACT

Background and Aim: Machine learning based on clinical data and treatment protocols for better clinical decision-making is a current research hotspot. This study aimed to build a machine learning model on washed microbiota transplantation (WMT) for ulcerative colitis (UC), providing patients and clinicians with a new evaluation system to optimize clinical decision-making.

Methods

Patients with UC who underwent WMT via mid-gut or colonic delivery route at an affiliated hospital of Nanjing Medical University from April 2013 to June 2022 were recruited. Model ensembles based on the clinical indicators were constructed by machine-learning to predict the clinical response of WMT after one month.

Results

A total of 366 patients were enrolled in this study, with 210 patients allocated for training and internal validation, and 156 patients for external validation. The low level of indirect bilirubin, activated antithrombin III, defecation frequency and cholinesterase and the elderly and high level of creatine kinase, HCO_3^- and thrombin time were related to the clinical response of WMT at one month. Besides, the voting ensembles exhibited an area under curve (AUC) of 0.769 ± 0.019 [accuracy, 0.754; F1-score, 0.845] in the internal validation; the AUC of the external validation was 0.614 ± 0.017 [accuracy, 0.801; F1-score, 0.887]. Additionally, the model was available at https://wmtpredict.streamlit.app.

Conclusions

This study pioneered the development of a machine learning model to predict the one-month clinical response of WMT on UC. The findings demonstrate the potential value of machine learning applications in the field of WMT, opening new avenues for personalized treatment strategies in gastrointestinal disorders.

Trial registration

clinical trials, NCT01790061. Registered 09 February 2013 - Retrospectively registered, https://clinicaltrials.gov/study/NCT01790061

Abbreviations: WMT, washed microbiota transplantation; UC, ulcerative colitis; IBD, inflammatory bowel disease; FMT, fecal microbiota transplantation; AEs, adverse events; BMI, body mass index; TET, transendoscopic enteral tubing; RF, Random forests; AdaBoost, Adaptive boosting; LGBM, Light gradient boosting machine; SVM, Support vector machine; MLP, Multi-layer perceptron; KNN, K-nearest neighbors; PCA, principle components algorithm; CV, cross-validated; AUC/AUROC, area under curve of receiver operating characteristic; SHAP, Shapley Additive Explanations; SD, standard deviation; P-LCR, platelet-larger cell ratio; PDW, platelet distribution width; TT, thrombin time; AT-IIIa, activated antithrombin III; GGT, γ-glutamyl transpeptidase; IB, indirect bilirubin; CHE, cholinesterase; INR, international normalized ratio; X-val, cross-validation; CRP, C-reaction protein; ESR, erythrocyte sedimentation rate; CaP, calprotectin; AChE, acetylcholinesterase; IQR, range of interquartile.

* Corresponding author at: Department of Microbiota Medicine & Medical Center for Digestive Diseases, the Second Affiliated Hospital of Nanjing Medical University, Nanjing, China.
E-mail address: fzhang@njmu.edu.cn (F. Zhang).
[1] orcid: 0000–0003-4157–1144

https://doi.org/10.1016/j.csbj.2024.08.021
Received 9 April 2024; Received in revised form 21 August 2024; Accepted 21 August 2024
Available online 24 August 2024

1. Introduction

Ulcerative colitis (UC) is a subtype of chronic inflammatory bowel disease (IBD) characterized by recurring inflammation and injury of the intestinal mucosa or submucosa, typically extending into the colon and rectum [1,2]. Increasing evidence has demonstrated the crucial role of gut microbiota in the progression of UC [3,4]. Fecal microbiota transplantation (FMT), a promising approach to reconstructing the composition and function of gut microbiota, has been effectively utilized in the treatment of recurrent *Clostridioides difficile* infection [5,6]. Additionally, FMT shows potential for other conditions related to gut dysbiosis, including UC, Crohn's disease, and other gut dysbiosis-related diseases [7].

A new method of FMT based on automatic washing and a new delivery process has been developed and termed as washed microbiota transplantation (WMT) [8] according to a consensus statement released by the FMT-Standardization Study Group in 2019[9]. WMT has been shown to have significant value in reducing the incidence of microbiota transplant-related adverse events (AEs) in UC patients, with a reduction rate exceeding 79.71 % (from 35.5 % with manual FMT to 7.2 % with WMT) [10]. This improvement is attributed to the elimination of pro-inflammatory viruses and metabolites during the washing process [8]. Our previous study demonstrated that washed preparations of fecal microbiota changed transplant-related safety, the quantitative method, and the delivery of the microbiota suspension [10].

Several clinical trials have shown that FMT or WMT are important for treating UC[11–13]. Our previously published study showed that the clinical response rate to FMT in treating UC was approximately 70 % within one month [14]. Recently, there has been an increase in the use of research paradigms based on big data in medicine and the life sciences. However, almost no models specifically focus on FMT or WMT in the context of UC. In 2022, a study performed in South Korea constructed a machine-learning model using a prospective cohort to predict the efficacy of FMT on UC. They achieved a microbial AUC score of 0.78, which decreased to 0.53 when clinical indicators were used [15].

In this study, we developed a machine learning model to determine whether patients with UC in China who present clinical symptoms prior to WMT are more likely to benefit from the model's predictive evaluations. We also explored the model's potential to assist physicians in the clinical decision-making process and reduce unnecessary medical resource expenditures. Given our team's experience and dedication to studying IBD cohorts, we aimed to provide patients and clinicians with a novel evaluation system of WMT on UC based on clinical indicators. We also established a machine learning research paradigm that can be applied not only to clinical decision-making in IBD but also to other indications for using FMT or WMT.

2. Material and methods

2.1. Study population

The study based on the clinical trials of UC using WMT (NCT01790061) was performed at the Second Affiliated Hospital of Nanjing Medical University, Nanjing, China. This study was reviewed and approved by the institutional ethical committee. Patients were recruited from April 2013 to June 2022, and the last follow-up was completed on July 25, 2022. All eligible subjects provided written informed consents prior to participation.

Inclusion criteria were: (1) patients were diagnosed as UC by a combination of typical clinical, endoscopic, and histological criteria; (2) patients did not achieve satisfactory efficacy from the previous medications [16]. Exclusion criteria were: loss to follow up at the endpoints of follow-up; a massive lack of data (missing information > 70 %); aged $<$ four years; undergoing manual FMT and accompanied with other intestinal disease and severe disease, such as malignant neoplasm, cardiopulmonary failure, serious liver and kidney disease.

The baseline characteristics of recruited patients were assessed before WMT, including age at the beginning of WMT, gender, body mass index (BMI), duration of disease, extent of disease, Mayo score, endoscopic score, delivery route, history of drug use, family history, and defecation frequency and stool traits. A series of clinical and experimental examinations before WMT were also recorded in our universal fecal microbiota bank (China Microbiota Transplantation System, CMTS).

2.2. Donor screening and WMT procedure

Healthy individuals from 6 to 24 years old were selected as candidate donors from CMTS and screened by strict exclusion criteria as described in our previous publications [17].

According to the previously reported protocols [14,16], fecal microbiota suspension was delivered into the patients' mid-gut through gastroscopic infusion under anesthesia, using a nasojejunal tube and mid-gut transendoscopic enteral tubing (TET) [18], or into the lower gut through colonic TET [18].

In order to prevent reflux of the microbiota liquid and reduce gastric acid secretion in patients with gastrointestinal motility disorders undergoing endoscopic delivery, they were administered 10 mg of intramuscular metoclopramide and an intravenous proton pump inhibitor at least one hour prior to WMT. The entire process, from the time of feces defecation to the delivery of the microbiota suspension, was completed within one hour [9]. Additionally, we selected the most suitable delivery route based on the patient's disease condition and will.

2.3. Clinical assessment and outcome

The efficacy was evaluated at 1 month after each course of WMT. The partial Mayo score was used to evaluate the clinical response of WMT. Clinical efficacy was defined as follows: a decrease of partial Mayo score of ≥ 2 and ≥ 30 % from baseline plus a decrease in the rectal bleeding subscore of ≥ 1 or an absolute rectal bleeding subscore of ≤ 1 for UC [14,19]. US National Cancer Institute Common Terminology Criteria for Adverse Events (version 5.0) was then applied to describe the intensity and relativity of AEs with WMT. AEs that occurred after WMT were recorded at hospital visits or follow-ups by telephone and were categorized as either definitely related [20].

2.4. Machine learning for prediction

Predictive models were implemented by Python (version 3.8.10, https://www.python.org/) using standard libraries that are publicly available: pandas (version 1.3.4), numpy (version 1.23.1), scikit-learn (version 0.23.1) [21], tableone (version 0.7.10)[22] and matplotlib (version 3.1.2). Logistic regression (LR), Random forests (RF), Adaptive boosting (AdaBoost), Light gradient boosting machine (LGBM), Support vector machine (SVM), Multi-layer perceptron (MLP), and Voting machine were used as classifier models for the prediction of different efficacy by using clinical information and laboratory examinations.

2.5. Variable selection and preprocess

The clinical information and laboratory examination variables were processed using the aforementioned standard libraries in Python. For variables with missing values less than 70 %, we employed the K-nearest neighbors (KNN) algorithm with default settings to fill in the missing data. Variables with missing values exceeding this threshold were dropped from the analysis.

To ensure the reliability of the model, we addressed collinearity issues by examining factors such as correlation, the coefficient of LR, the importance of RF, and significant differences. Relevant variables were carefully selected based on these considerations and included in the final model.

Samples for internal verification were selected from the previous database. To ensure an unbiased evaluation, these samples were randomly divided into a training set (80 % of the samples) and a test set (the remaining 20 %), which contributed to the balanced class proportions across the entire cohort, as well as within the training and test sets.

To prepare the variables for analysis, they were transformed and standardized. Additionally, principle components algorithm (PCA) was performed on the variables. The first three components of PCA were included in the analysis alongside the original variables. The optimal models were selected based on cross-validated (CV) results obtained through learning curve analysis. These models were further evaluated using a withheld evaluation dataset, which served as the final performance assessment for predicting the efficacy.

2.6. Evaluation and explanation of models

Prospective samples for external verification were used to evaluate the models. We assessed the models' discrimination by calculating the area under curve of receiver operating characteristic (AUROC, AUC) [23], accuracy, and confusion matrix with precision and recall. The F1-score (the harmonic mean of precision and recall) [24] and brier score (a measure of the mean squared difference between estimated risks and the actual outcomes) [25]were also calculated as a measure of model performance and calibration.

The interpretability of the model plays a crucial role in enhancing the acceptance and understanding of machine-learning predictions by clinicians and patients. To achieve this, the Shapley Additive Explanations (SHAP) summary chart was plotted. This chart helps to visualize and analyze the impact of important characteristics on the model's output, providing insights into both the overall positive or negative impact and the similarities and differences among these characteristics [26]. By utilizing the SHAP summary chart, clinicians and patients can gain a better understanding of the factors influencing the model's predictions and the relative importance of each characteristic. This promotes transparency and trust in the machine-learning model by providing clear and interpretable explanations for its decision-making process.

2.7. Statistical analysis

Statistical analyses were carried out using Python. Continuous variables were presented as mean ± standard deviation (SD) or median with quartile, whereas categorical data were reported as frequencies and proportions. The unpaired Student's t-test, Fisher's exact test, chi-square test, and Tukey HSD test were applied to analyze the differences between groups according to the type of data. P value < 0.05 was considered to be statistically significant.

3. Results

3.1. Baseline characteristics of the study participants

As shown in the flow chart (Fig. 1), 366 patients with UC who underwent WMT at our center were included for analysis. Of these, 210 were allocated for model training and internal validation, while 156 were used for external validation.

The baseline characteristics of the patients are shown in Table 1. Among the 366 participants, 58.47 % (214/366) were male, with a mean age of 40.20 years. The median disease duration was four years (IQR, 2.00–8.75). The median defecation frequency was 3.50 times per day (IQR, 2.00–6.00). The mean Mayo score was used to access the extent and severity of disease. Before WMT, the Mayo score was 7.44 (median, 8; IQR, 6–10), with 85.79 % (314/366) of the patients classified as having moderate or severe disease activity.

3.2. Patient characteristics and variables selected for internal verification

To evaluate the variables selected for use in the models, we analyzed the differences between patient data used for internal verification (Fig. S1). Patients who exhibited a positive response one month after WMT had a lower frequency of defecation ($P = 0.044$) and Mayo score ($P = 0.011$) compared to those with no clinical responses. There were no significant differences between these two groups concerning age, disease duration, or BMI.

An effective response was more commonly observed in patients with higher platelet-larger cell ratios (P-LCR, $P = 0.002$), platelet distribution

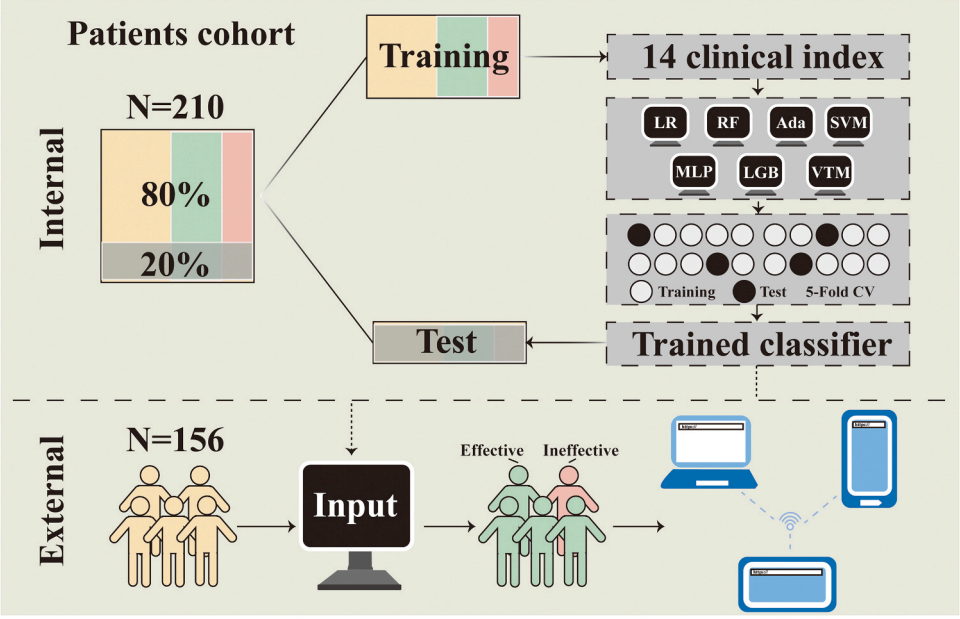


Fig. 1. Flow chart of the study. Framework for dataset partition, model training, independent validation (internal and external verification), and the online application. The individuals were divided into two groups, including the effective (green) and ineffective (red), according the efficacy at one month after WMT. LR, logistic regression; RF, random forests; Ada, adaptive boosting; LGB, light gradient boosting machine; SVM, support vector machine; MLP, multi-layer perceptron; VTM, voting machine.

Table 1
The baseline characteristics of 366 samples.

Characteristic	Result
N	366
Gender, male, n (%)	214 (58.47)
Age, years, mean ± SD	40.20 ± 15.64
Duration of disease, years, median (IQR)	4.00 (2.00 - 8.75)
BMI, kg/m^2, mean ± SD	20.72 ± 3.47
Family history, n (%)	18 (4.92)
Medication history, % (n/N)	
5-ASA	98.51 (331/366)
Steroids	42.35 (155/366)
Immunosuppressants	17.76 (65/366)
Biologic agents	7.92 (29/366)
Extent of disease, % (n/N)	
E1, proctitis	13.39 (49/366)[a]
E2, left-sided colitis	46.45 (70/366)[a]
E3, pancolitis	62.02 (227/366)[a]
Defecation frequency, n / day, median (IQR)	3.50 (2.00 - 6.00)
Mayo score before WMT, mean ± SD	7.44 ± 2.69[b]
Mayo score before WMT, median (IQR)	8.00 (6.00 - 10.00)[b]
Disease severity, % (n/N)	
Mild	10.11 (37/366)[b]
Moderate	59.56 (218/366)[b]
Severe	26.23 (96/366)[b]
Delivery route, % (n/N)	
Mid-gut	27.32 (100/366)[c]
Colonic TET	73.50 (269/366)[c]
State, %(n/N)	
Fresh	92.34 (338/366)
Frozen	0.82 (3/366)
Mixed	6.83 (25/366)

[a] 15 samples underwent WMT via mid-gut way without colonoscope ([b]), and five samples in endoscope remission underwent WMT for maintenance. [c] Three samples underwent WMT both in mid-gut way and colonic TET in the same course. 5-ASA, 5-animosalicylic acid; WMT, wash microbiota transplantation; TET, transendoscopic enteral tube; BMI, body mass index; SD, standard deviation; IQR, range of interquartile.

width (PDW, $P < 0.001$), mean platelet volume ($P = 0.002$), higher levels of HCO_3^- ($P = 0.046$), creatine kinase (CK, $P = 0.005$), longer thrombin times (TT, $P = 0.043$), and lower levels of activated antithrombin III (AT-IIIa, $P = 0.038$). The remaining laboratory indicators showed no significant differences between the two treatment response groups (Supplementary material 1).

Taking into account the coefficients of LR, the importance of RF, and the collinearity of variables (Supplementary material 2), 14 variables were selected for inclusion in the models (Table 2), including age, defecation frequency, Mayo score, P-LCR, PDW, P^+, HCO_3^-, γ-glutamyl transpeptidase (GGT), indirect bilirubin (IB), cholinesterase (CHE), CK, TT, the international normalized ratio (INR), and AT-IIIa. The darker the color, the more important the variable.

3.3. Model ensemble A and internal verification

Five-fold CV was conducted to train the models. The AUC was used to estimate the learning curve and assess model fitting. As training sizes increased, especially for the LR and SVM models, there was a noticeable trend towards convergence based on the 14 selected variables that formed ensemble A. However, a gap persisted between the training and cross-validation (X-val) lines (Fig. S2).

As shown in Figs. 2**A**, 85.71 % (6/7) of the AUC scores exceeded 0.7, and accuracy exceeded 0.73. Among all models, the RF model performed the best, achieving the highest AUC score of 0.83, underscoring its superior performance in predicting WMT treatment outcomes. Although the AdaBoost model had the lowest AUC score (0.65), it demonstrated the highest accuracy, which was on par with the SVM model.

Except for the LR model, the other six models demonstrated high recall (0.77–0.94) and F1 scores (0.83–0.89), indicating their ability to accurately identify positive instances. These models also showed low Brier loss scores, ranging from 0.14 to 0.24, indicating good calibration. Among the models in ensemble A, SVM demonstrated superior recall, F1 score, and Brier score. However, all models in this ensemble displayed high precision.

Table 2
The heatmap of characteristics of selected variables.

	P	Correlation	Coefficient	Importance
Age	0.056	0.126	0.449	0.014
Defecation frequency	0.044	−0.157	−0.527	0.013
Mayo score	0.025	−0.179	−0.725	0.017
P-LCR	0.002	0.148	0.274	0.014
PDW	< 0.001	0.156	0.689	0.011
P^+	0.260	−0.090	−0.546	0.014
$HCO3^-$	0.046	0.142	0.751	0.026
GGT	0.539	0.071	0.663	0.008
IB	0.130	−0.127	−1.050	0.018
CHE	0.134	−0.136	−1.222	0.026
CK	0.005	0.149	1.980	0.017
TT	0.043	0.157	0.089	0.018
INR	0.061	0.117	0.491	0.010
AT-IIIa	0.038	−0.137	−0.595	0.027

P-LCR, platelet-larger cell ratio; PDW, platelet distribution width; GGT, γ-glutamyl transpeptidase; IB, indirect bilirubin or unconjugated bilirubin; CHE, cholinesterase; CK, creatine kinase; TT, thrombin time; INR, international normalized ratio; AT-IIIa, activates antithrombin III. Correlation, correlation coefficient; Coefficient, coefficient of logistic regression; Importance, importance of random forest. The darker the color, the more important the variable. the *P*-value < 0.05 was considered to be statistically significant.

3.4. Internal verification of Model ensembles B and C

Clinicians may arrive at different conclusions regarding Mayo scores, particularly concerning endoscopy results and disease extent. To address this variability, we constructed Model Ensemble B to investigate whether a model excluding the Mayo score could predict the clinical response to WMT in UC patients one-month post-treatment. Additionally, Model Ensemble C was developed by applying PCA to Ensemble B for further optimization.

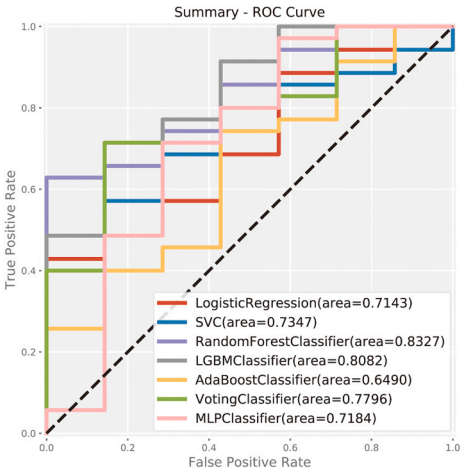
In terms of fitting, both Model Ensembles B and C exhibited trends similar to Ensemble A. The learning curves for these ensembles showed a reduction in the gaps between the lines of training and cross-validation (X-val) compared to ensemble A, which suggests that ensembles B and C achieved better fitting. The learning curve for the LR model was the same in the ensembles B and C. However, as training sizes increased, most models in Ensemble C exhibited a smoother learning curve than those of Ensemble B (Fig. S3).

Internal verification evaluations on Ensembles B and C are shown in Fig. 2B and Fig. 2C. For the AUC score and accuracy, 71.42 % (5/7) of the models in Ensemble B exceeded 0.70, while one more model did in Ensemble C. The LGBM model in Ensemble B ranked first in AUC scores, surpassing the MLP model in Ensemble C. However, the model with the highest accuracy (LGBM, 0.81) in Ensemble B was lower than in Ensemble C (RF, 0.83). In terms of recall, F1 score, and Brier score, all models except LR performed well in both ensembles. Generally, models in Ensemble B demonstrated better recall and Brier score performance, while models in Ensemble C had better in F1 scores. Notably, all models exhibited strong precision scores.

3.5. Characteristics and results of external verification

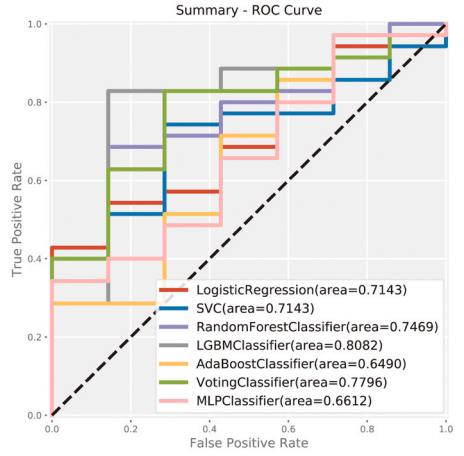
To further verify the generalizability of the ensembles, we prospectively collected data from 156 study participants for external verification. The variables from these participants are listed in Table 3. Of the 156 participants, 78.84 % (123/156) had a clinical response to treatment. The mean age was 38.73 years, and the Mayo score before WMT was 6.80 ± 2.76 (IQR, 5.00–7.00).

A



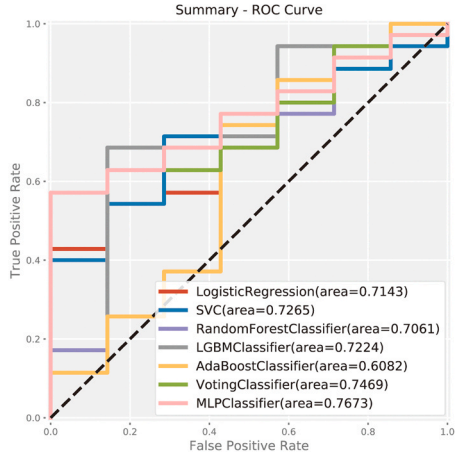
	AUC	Accuracy	Precision	Recall	F1	Brier
Logistic regression	0.7143	0.5714	0.8696	0.5714	0.6897	0.2998
Support vector classifier	0.7347	0.8095	0.8462	0.9429	0.8919	0.1390
Random forest classifier	0.8327	0.7381	0.9000	0.7714	0.8308	0.1795
AdaBoost classifier	0.6490	0.8095	0.8649	0.9143	0.8889	0.2093
LGBM classifier	0.8082	0.7619	0.9032	0.8000	0.8485	0.1844
MLP classifier	0.7184	0.7619	0.9032	0.8000	0.8485	0.2379
Voting classifier	0.7796	0.7619	0.9032	0.8000	0.8485	0.1538

B



	AUC	Accuracy	Precision	Recall	F1	Brier
Logistic regression	0.7143	0.5714	0.8696	0.5714	0.6897	0.2948
Support vector classifier	0.7143	0.7619	0.8378	0.8857	0.8611	0.1590
Random forest classifier	0.7469	0.6905	0.8929	0.7143	0.7937	0.1877
AdaBoost classifier	0.6490	0.8095	0.8649	0.9143	0.8889	0.2028
LGBM classifier	0.8082	0.8095	0.9355	0.8286	0.8788	0.1595
MLP classifier	0.6612	0.7857	0.8611	0.8857	0.8732	0.1985
Voting classifier	0.7796	0.7857	0.8824	0.8571	0.8696	0.1534

C



	AUC	Accuracy	Precision	Recall	F1	Brier
Logistic regression	0.7143	0.5714	0.8696	0.5714	0.6897	0.2948
Support vector classifier	0.7265	0.8095	0.8462	0.9429	0.8919	0.1518
Random forest classifier	0.7061	0.8333	0.8333	1.0000	0.9091	0.1297
AdaBoost classifier	0.6082	0.7619	0.8571	0.8571	0.8571	0.2181
LGBM classifier	0.7224	0.7619	0.8788	0.8286	0.8529	0.2202
MLP classifier	0.7673	0.7381	0.9000	0.7714	0.8308	0.2588
Voting classifier	0.7469	0.7143	0.8710	0.7714	0.8182	0.1673

Fig. 2. The evaluation of Model ensembles on test data of internal verification. (A) The evaluation of Model ensemble A on test data of internal verification. (B) The evaluation of Model Ensemble B on test data of internal verification. (C) The evaluation of Model Ensemble C on test data of internal verification. ROC curve, receiver operating characteristic curve; AUC, the area under curve of receiver operating characteristic; LGBM, light gradient boosting machine; MLP, multi-layer perceptron.

In terms of external verification, different results were obtained by the three model ensembles (Fig. 3). In ensemble A, 71.43 % (5/7) of the models had an AUC score of more than 0.60, with the LGBM model ranking first, with a score of 0.70. A majority of the models achieved an accuracy of over 0.75. Furthermore, all models in ensemble A performed well in terms of precision and recall. The Brier score for all models in ensemble A was below 0.25 (Fig. 3A).

In Ensemble B, only four models achieved an AUC of approximately 0.60. However, the accuracy of six models exceeded 0.76. Compared to ensemble A, the AUC of the LGBM model was lower but remained the

Table 3
The characteristic variables of 156 samples for external verification.

Characteristic	NA	Result
N	-	156
Clinical response, n (%)	-	123 (78.84)
Age, years, mean ± SD	-	38.73 ± 15.87
Defecation frequency, n / day, median (IQR)	-	3.00 (2.00 - 5.25)
Mayo score before WMT, mean ± SD	7[a]	6.80 ± 2.76
Mayo score before WMT, median (IQR)	7[a]	7.00 (5.00 - 9.00)
P-LCR, mean ± SD	2[b]	30.00 ± 10.15
PDW, mean ± SD	2[b]	12.68 ± 3.04
P^{+}, mean ± SD	2[b]	1.12 ± 0.20
HCO_3, mean ± SD	-	23.08 ± 1.99
GGT, median (IQR)	2[b]	10.00 (14.00 - 20.00)
IB, mean ± SD	-	4.91 ± 3.60
CHE, mean ± SD	-	6371.42 ± 1972.32
CK, median (IQR)	-	59.50 (33.00 – 90.50)
TT, mean ± SD	7[b]	16.81 ± 0.93
INR, mean ± SD	7[b]	1.06 ± 0.10
AT-IIIa, mean ± SD	5[b]	96.30 ± 11.88

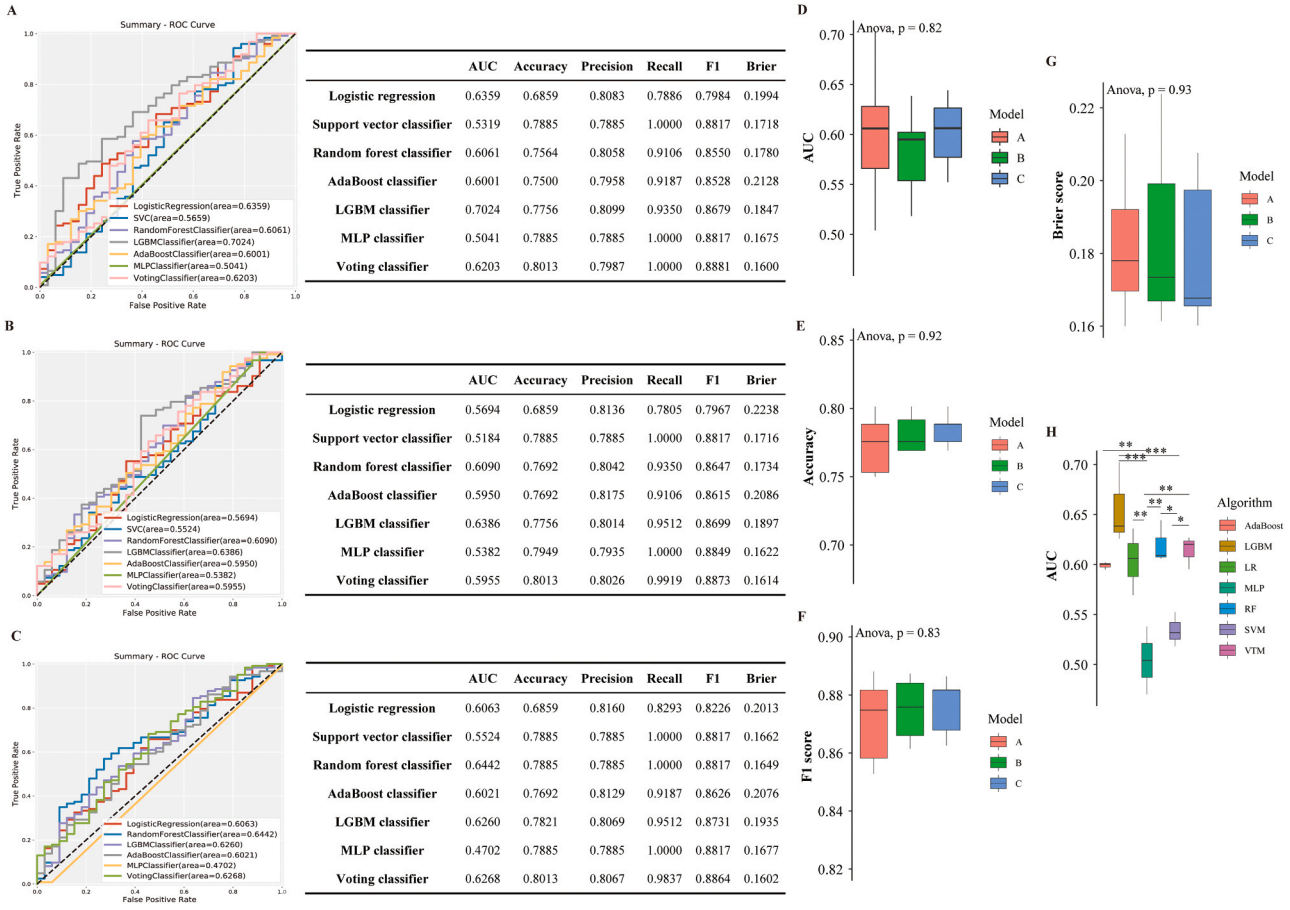
[a] Four samples underwent WMT via mid-gut way without colonoscope, and three samples in endoscope remission underwent WMT for maintenance.

[b] random missing; NA, the number of missing values; WMT, wash microbiota transplantation; P-LCR, platelet-larger cell ratio; PDW, platelet distribution width; GGT, γ-glutamyl transpeptidase; IB, indirect bilirubin or unconjugated bilirubin; CHE, cholinesterase; CK, creatine kinase; TT, thrombin time; AT-IIIa, activates antithrombin III; INR, international normalized ratio; SD, standard deviation; IQR, range of interquartile.

highest among the models in Ensemble B. Additionally, the precision, recall, and Brier scores of the models in Ensemble B were excellent (Fig. 3B).

Model Ensemble C contained five models with an AUC over 0.60, an accuracy exceeding 0.78. The RF model ranked first in terms of AUC in this ensemble. Unfortunately, the AUC of the MLP model was the lowest, falling below 0.5. However, the LR model in Ensemble C outperformed the other ensembles in terms of the remaining three evaluation indicators. Overall, the other models in Ensemble C showed good performance across the evaluated indicators (Fig. 3**C**).

There were no statistically significant differences in the AUC ($P = 0.82$), accuracy ($P = 0.92$), F1 score ($P = 0.83$), or Brier score ($P = 0.93$) between the three model ensembles. However, the AUC of Ensemble B performed worse than the others (Fig. 3D), and Ensemble C had better accuracy than the others (Fig. 3E), although these differences were not significant. In terms of the AUC, the LGBM model ranked highest, followed by algorithms based on the tree model (RF and AdaBoost), line model algorithms (LR and SVM), and the MLP algorithm (Fig. 3H). As an algorithm for integrating the results of the other six models, the AUC of the external voting machine classifier was 0.614 ± 0.017 (accuracy, 0.8013 ± 0.000; F1 score, 0.887 ± 0.001; Brier score, 0.161 ± 0.001), demonstrating good generalization and stability and good results for all the evaluation indicators among the three ensembles.



	AUC	Accuracy	Precision	Recall	F1	Brier
Logistic regression	0.6359	0.6859	0.8083	0.7886	0.7984	0.1994
Support vector classifier	0.5319	0.7885	0.7885	1.0000	0.8817	0.1718
Random forest classifier	0.6061	0.7564	0.8058	0.9106	0.8550	0.1780
AdaBoost classifier	0.6001	0.7500	0.7958	0.9187	0.8528	0.2128
LGBM classifier	0.7024	0.7756	0.8099	0.9350	0.8679	0.1847
MLP classifier	0.5041	0.7885	0.7885	1.0000	0.8817	0.1675
Voting classifier	0.6203	0.8013	0.7987	1.0000	0.8881	0.1600

	AUC	Accuracy	Precision	Recall	F1	Brier
Logistic regression	0.5694	0.6859	0.8136	0.7805	0.7967	0.2238
Support vector classifier	0.5184	0.7885	0.7885	1.0000	0.8817	0.1716
Random forest classifier	0.6090	0.7692	0.8042	0.9350	0.8647	0.1734
AdaBoost classifier	0.5950	0.7692	0.8175	0.9106	0.8615	0.2086
LGBM classifier	0.6386	0.7756	0.8014	0.9512	0.8699	0.1897
MLP classifier	0.5382	0.7949	0.7935	1.0000	0.8849	0.1622
Voting classifier	0.5955	0.8013	0.8026	0.9919	0.8873	0.1614

	AUC	Accuracy	Precision	Recall	F1	Brier
Logistic regression	0.6063	0.6859	0.8160	0.8293	0.8226	0.2013
Support vector classifier	0.5524	0.7885	0.7885	1.0000	0.8817	0.1662
Random forest classifier	0.6442	0.7885	0.7885	1.0000	0.8817	0.1649
AdaBoost classifier	0.6021	0.7692	0.8129	0.9187	0.8626	0.2076
LGBM classifier	0.6260	0.7821	0.8069	0.9512	0.8731	0.1935
MLP classifier	0.4702	0.7885	0.7885	1.0000	0.8817	0.1677
Voting classifier	0.6268	0.8013	0.8067	0.9837	0.8864	0.1602

Fig. 3. The evaluation on generalization ability of model ensembles for external verification. (A) The evaluation on Model ensemble A of external verification. (B) The evaluation on Model Ensemble B of external verification. (C) The evaluation on Model Ensemble C of external verification. (D) The comparison of AUC score in different ensembles. (E) The comparison of accuracy in different ensembles. (F) The comparison of F1-score in different ensembles. (G) The comparison of brier score in different ensembles. (H) The comparison of AUC score in different algorithms. ROC curve, receiver operating characteristic curve; AUC, the area under curve of receiver operating characteristic; LR, logistic regression; RF, random forest; LGBM, light gradient boosting machine; MLP, multi-layer perceptron; VTM, voting machine; SVM, support vector machine. * , $P < 0.05$; * *, $P < 0.01$; * ** , $P < 0.001$; P-value < 0.05 was considered as statistically significant.

3.6. Model explanations for external verification

We applied the Kernel explainer model to the external verification data generated by the three ensembles using the 'shap' Python package. The feature rankings for the VTM, an integrating algorithm, are displayed in Fig. 4. Feature importance decreases from top to bottom, with the color of the points representing the value of the corresponding variable for each sample. The colors range from blue (low value) to red (high value), and the horizontal position of each point reflects its SHAP value, indicating the variable's contribution to a clinical response to WMT one month after treatment.

The most important variables accross the three ensembles were IB, CK, defecation frequency, and CHE. In Model Ensemble C, the first and second principal components (PC1 and PC2) emerged as important features. Interestingly, PC1 was mainly influenced by CHE and defecation frequency, while PC2 was affected by P-LCR, PDW, and AT-IIIa (Supplementary Table 3), which aligns with the rankings observed in the other ensembles. Therefore, low levels of IB, AT-IIIa, CHE, along with reduced defecation frequency, and high levels of CK, HCO3- and TT, and older age may positively correlate with a clinical response to WMT one month post-treatment.

3.7. Development of a webserver platform for models

We developed a user-friendly web platform to predict the clinical response to WMT for UC patients one-month post-treatment. The platform, available at https://wmtpredict.streamlit.app/, is built using Python's Streamlit package. Input the relevant information and a simple click on the 'Submit' button in the Step one, the system activates the models to process the data. In Step Two, the platform provides predictions along with the probability of a positive or negative response to WMT based on the current patient status.

3.8. Adverse events

Thirty-eight WMT-related AEs (5.09 %, 38/746) were observed in 746 WMT procedures. Approximately 97.37 % (37/38) of AEs occurred within one week after WMT and self-resolved within a few minutes or days. Increases in the frequency of defecation (2.82 %, 21/746), fever (1.07 %, 8/746), and abdominal pain (0.67 %, 5/746) accounted for most AEs. Only one AE (myasthenia) was observed in the four weeks after WMT. This was reported in our previous study focusing on AEs[14] and categorized as possibly related to WMT. No serious WMT-related AEs were observed during the follow-up period in this study.

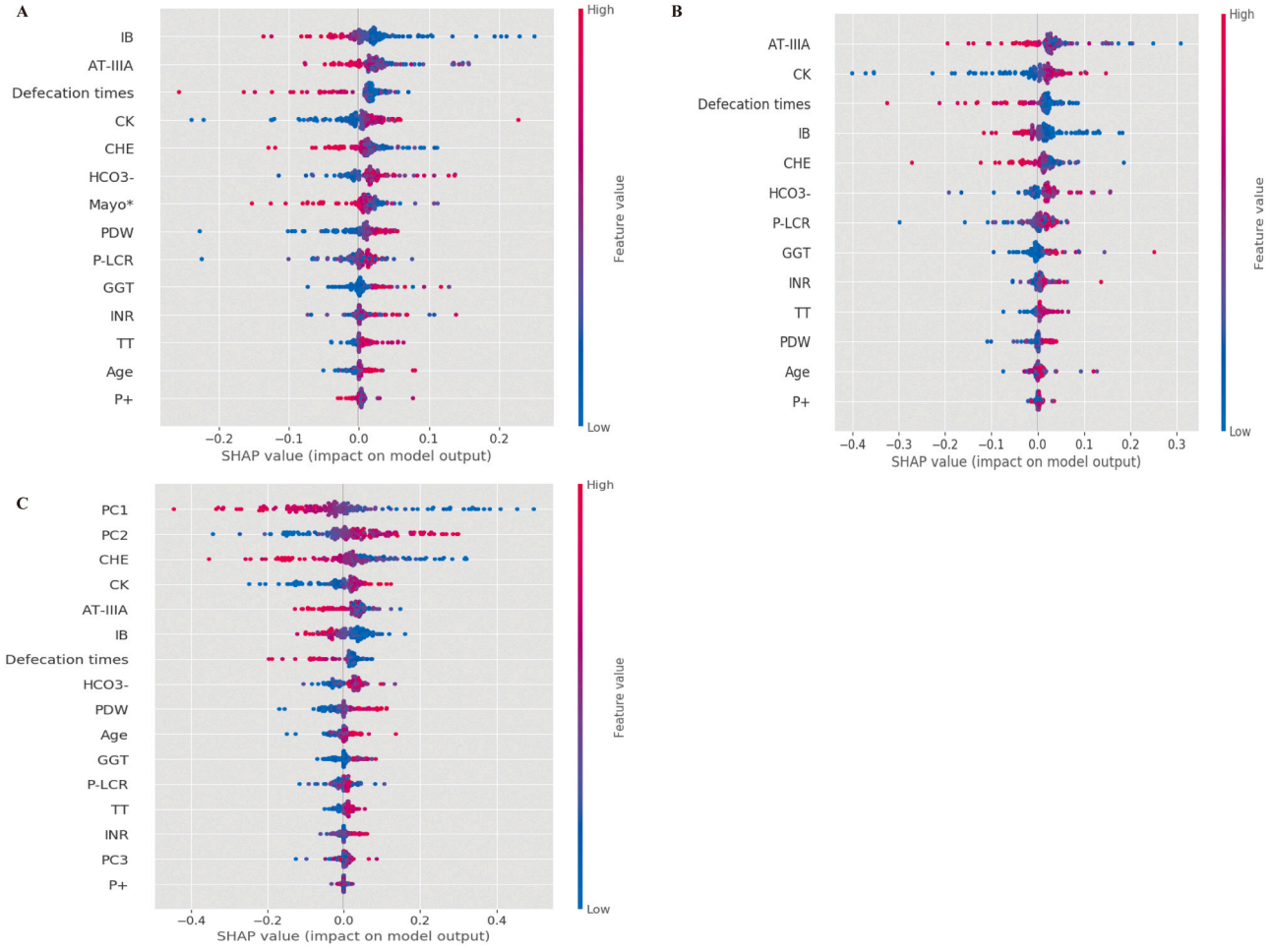


Fig. 4. The explanation and feature importance of model ensembles on external verification. (A) The summary plot of shap value on the voting machine of model ensemble A based on external verification. (B) The summary plot of shap value on the voting machine of model Ensemble B based on external verification. (C) The summary plot of shap value on the voting machine of model Ensemble C based on external verification. The importance of feature decreased from top to bottom. And the color of the points represented the value of corresponding variable every sample. The color of the points ranged from blue to red, the value ranged from low to high. The location of the point on the horizontal axis represented the shap value, which was positive to the possibility of clinical response of WMT after one month. P-LCR, platelet-larger cell ratio; PDW, platelet distribution width; GGT, γ-glutamyl transpeptidase; IB, indirect bilirubin or unconjugated bilirubin; CHE, cholinesterase; CK, creatine kinase; TT, thrombin time; AT-IIIa, activates antithrombin III; INR, international normalized ratio.

4. Discussion

A number of studies have focused on constructing artificial intelligence models to find biomarkers, microbiota (bacteriome), or metabolites using sequencing technology to predict the clinical efficacy of FMT on UC [27–29]. However, few studies have reported on predictive models based on clinical indicators for the efficacy of FMT on UC. To address this gap, we employed a machine-learning approach to construct model ensembles that investigate the effectiveness of common clinical indicators in predicting the clinical response to WMT on UC one month after treatment.

In this study, the internal verification of model ensemble A partially relied on the Mayo score, which performed slightly better than the other indicators, suggesting that baseline Mayo scores are indicative of WMT's clinical efficacy in UC. However, the Mayo score includes subjective physician evaluations, which may lead to variability in scoring based on similar or equivalent information. Additionally, the endoscopic score is limited in scope, which may hinder broader application in patient-side model integration.

Therefore, we developed Model ensembles B and C, excluding the Mayo score. While these ensembles performed slightly worse than ensemble A in internal verification, their external validation performance was either comparable or superior. Specifically, Ensemble C exhibited better generalizability and stability than ensemble A. Moreover, as the results of the evaluations manifested, the voting machine learning algorithm, which integrated the results of the other six models, provided the three ensembles with improved stability and reliability.

Previous studies have used quantifiable laboratory indicators, including C-reaction protein (CRP), erythrocyte sedimentation rate (ESR), calprotectin (CaP), and lymphocyte subset analysis, to evaluate the disease activity of IBD [30–32]. However, these features were not included in our ensemble models for various reasons. First, these indicators may have many missing values, making them difficult to effectively use in a model. Second, despite their potentially diagnostic and predictive values in IBD, they may not have significant predictive value specifically within the context of WMT for UC. For example, CaP has shown promise in the diagnosis of IBD and in predicting the relapse of quiescent IBD [33], but its role in predicting the efficacy of WMT on UC may be limited. From our previous clinical experience and relevant studies, the efficacy of WMT on UC cannot be accurately predicted by the short-term surveillance of cytokines and CRP [34]. For instance, a Lasso LR model based on white blood cells, hemoglobin, ESR, CRP, albumin, and CaP yielded an AUC of only 0.53 [15], highlighting their limited predictive abilities.

In the current ensembles, low levels of IB, AT-IIIa, defecation frequency, and CHE, along with the older age, high levels of CK, HCO_3^- and TT, were associated with a clinical response to WMT one month after treatment. Notably, defecation frequency, as a part of the Mayo score, reflected the extent of the diseases.

CHEs are specialized carboxylic ester hydrolases that break down choline esters, including acetylcholinesterase (AChE) and butyrylcholinesterase. AChE hydrolyzes acetylcholine, which regulates gut motility and mucosal responses [35]. Recently, an animal study has reported that decreasing the AChE levels can protect against tissue injury [36]. This finding may explain the negative correlation between CHE and clinical efficacy following WMT. However, this relationship is complex and warrants further investigation. Specifically, the interplay between cholinergic regulation, gut microbiota, and the host immune response in UC requires deeper exploration to better understand the underlying mechanisms.

Regarding coagulation indicators like TT and AT-IIIa, the latter serves as an antithrombin, leading to increased TT. In our ensembles, high AT-IIIa levels and decreased TTs were significant predictive features. This finding aligns with a study by Kume et al. [37], which found that patients with inactive UC had lower AT-IIIa levels and increased TTs compared to patients with active UC.

Several studies have reported that IB may be an independent protective factor against UC due to its antioxidant and anti-inflammatory effects [38–40]. Lower bilirubin levels in UC [41] are positive correlated with disease severity [42]. Interestingly, our ensembles indicated that a relatively low IB level may lead to a clinical response to WMT. The gut microbiota is involved in enterohepatic circulation, and a recent animal study reported that oral release of IB in the gut could modulate intestinal epithelium regeneration and immune response in UC [43]. Based on this, we hypothesize that WMT can regulate this circulation through microbiota, making it easier for severe UC to achieve a clinical response.

There is a notable phenomenon whereby older UC may have a higher likelihood of achieving a positive clinical response to treatment. This observation is supported by our previous study, which reported that older patients with UC were more likely to experience short-term, steroid-free clinical remission and maintain a long-term response via WMT [16]. Our ensembles also found a positive association between age and clinical response.

In this study, 73.50 % of patients received WMT by colonic TET, while few underwent WMT by mid-gut route. Our previous studies have shown no difference in efficacy and safety between the two routes [14,16]. Patients' overall satisfaction and acceptance for both routes were high, exceeding 95 % [44]. Nonetheless, colonic TET is the more recommended route for its relatively more convenient and non-disruptive nature [44]. Moreover, colonic TET is available for clinical practice and research: frequent and timely delivery of microbiota and medications, drainage and decompression for colonic perforation and ileocolic obstruction, and sampling for microbial research[45].

The incidence of AEs in the study participants was 5.09 %, and only one (0.13 %) WMT-related, long-term AE was observed. Most AEs were minimal and self-resolving, suggesting that WMT is a safe treatment for patients with UC.

This was a pilot study, and future research can build on the lessons learned. First, expanding data collection to include multiple medical institutions and clinical trials could improve the reliability and generalizability of the models. However, in clinical research, limited patient numbers and imbalanced distributions of specific disease subtypes may pose challenges to the reliability. Second, accurate and consistent data annotation is crucial, requiring a clinical expert team for this task. Third, careful feature selection and data preprocessing are essential for obtaining key information related to clinical responses following WMT. Lastly, to promote clinical acceptance and application, the interpretability of AI models is vital. Further research on interpreting and explaining machine-learning model predictions can help healthcare professionals and clinical decision-makers understand the decision-making process and the basis for predictions.

This study has several limitations. First, some data were retrospective and derived from patients' self-reports. Second, missing data led to the exclusion of some variables, which may have introduced bias. Third, although our sample size was the largest possible given the number of patients with UC who underwent FMT, the convergence and AUC results could be improved. Finally, the effects of COVID-19 should be investigated in future studies using a larger sample size.

5. Conclusions

This study utilized machine learning models on large sample sizes to predict the clinical response to WMT in UC patients one month after treatment, based on pretreatment clinical indicators. The inclusion of explainer models further enhanced the interpretability of the predictions and provided valuable insights for optimizing clinical decision-making. The findings suggest that machine learning can contribute to the field of WMT. With ongoing advancements in machine learning and the accumulation of larger datasets, integrating these models into routine clinical practice holds promise for improving patient outcomes following WMT.

Ethics approval and consent to participate

This study was reviewed and approved by the Second Affiliated Hospital of Nanjing Medical University Institutional Review Board ([2012]KY015). All eligible subjects provided written informed consents prior to participation in this study.

Authors' contributions

FZ conceived, designed and supervised the study. SZ and GL analyzed the data and wrote the manuscript. GL and ZZ handled data acquisition. PL, QW, BC and FZ made clinical diagnosis, recruited subjects and performed intervention. SZ prepared the figures. PL, GL, XW, WW, QL, YL, CL, RW and SZ contributed to text revision and discussion. All authors read and approved the final manuscript. SZ and GL contributed equally in this study.

Consent for publication

All authors declare their consent for publication. The manuscript does not contain any material that could be overtly or indirectly linked to any individual.

Funding

This study was supported by National Key Research and Development Program of China [2021YFA0717004]; China National Clinical Research Center for Digestive Diseases [2015BAI13B07] and Nanjing Medical University Fan Daiming Research Funds for Holistic Integrative Medicine.

CRediT authorship contribution statement

Sheng Zhang: Writing – original draft, Visualization, Validation, Software, Methodology. **Yujie Liu:** Writing – review & editing. **Chenchen Liang:** Writing – review & editing. **Quan Wen:** Investigation. **Pan Li:** Investigation. **Rui Wang:** Writing – review & editing. **Qianqian Li:** Writing – review & editing. **Xia Wu:** Writing – review & editing. **Zulun Zhang:** Data curation. **Bota Cui:** Investigation. **Weihong Wang:** Writing – review & editing. **Faming Zhang:** Writing – review & editing, Methodology, Investigation, Funding acquisition, Data curation, Conceptualization. **Gaochen Lu:** Wrinting – original draft, Data curation.

Declaration of Competing Interest

Faming Zhang conceived the concept of GenFMTer and trans-endoscopic tubing and the devices (FMT Medical, Nanjing, China) related to them.

Data Availability

Datasets generated for this study are included in the article. Additional data and materials underlying this article will be shared on reasonable request to the corresponding author.

Acknowledgments

We thank the scientific discussion from the construction group of microbiota medicine discipline. We gratefully thank the CMTS for the data about the WMT delivery and evaluation of efficacy and safety. We thank all the patients for their time and effort.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.csbj.2024.08.021.

References

[1] Feuerstein JD, Cheifetz AS. Ulcerative colitis: epidemiology, diagnosis, and management [J. Mayo Clin Proc 2014;89(11):1553–63.
[2] Ungaro R, Mehandru S, Allen PB, et al. Ulcerative colitis. Lancet 2017;389(10080):1756–70.
[3] Zuo T, Lu XJ, Zhang Y, et al. Gut mucosal virome alterations in ulcerative colitis. Gut 2019;68(7):1169–79.
[4] Machiels K, Joossens M, Sabino J, et al. A decrease of the butyrate-producing species Roseburia hominis and Faecalibacterium prausnitzii defines dysbiosis in patients with ulcerative colitis. Gut 2014;63(8):1275–83.
[5] Surawicz CM, Brandt LJ, Binion DG, et al. Guidelines for diagnosis, treatment, and prevention of Clostridium difficile infections. Am J Gastroenterol 2013;108(4):478–98. quiz 99.
[6] Wu X, Ai RJ, Xu J, et al. Washed microbiota transplantation for Clostridioides difficile infection: a national multicenter real-world study. J Dig Dis 2023;24(10):540–9.
[7] WANG Y, ZHANG S, BORODY TJ, et al. Encyclopedia of fecal microbiota transplantation: a review of effectiveness in the treatment of 85 diseases. Chin Med J 2022;135(16):1927–39.
[8] ZHANG T, LU G, ZHAO Z, et al. Washed microbiota transplantation vs. manual fecal microbiota transplantation: clinical findings, animal studies and in vitro screening. Protein Cell 2020;11(4):251–66.
[9] FECAL MICROBIOTA TRANSPLANTATION-STANDARDIZATION STUDY GROUP. Nanjing consensus on methodology of washed microbiota transplantation. Chin Med J (Engl) 2020;133(19):2330–2.
[10] Lu G, Wang W, Li P, et al. Washed preparation of faecal microbiota changes the transplantation related safety, quantitative method and delivery. Micro Biotechnol 2022;15(9):2439–49.
[11] Kedia S, Virmani S, Sk V, et al. Faecal microbiota transplantation with anti-inflammatory diet (FMT-AID) followed by anti-inflammatory diet alone is effective in inducing and maintaining remission over 1 year in mild to moderate ulcerative colitis: a randomised controlled trial. Gut 2022;71(12):2401–13.
[12] Costello SP, Hughes PA, Waters O, et al. Effect of fecal microbiota transplantation on 8-week remission in patients with ulcerative colitis: a randomized clinical trial. Jama 2019;321(2):156–64.
[13] Chen M, Liu XL, Zhang YJ, et al. Efficacy and safety of fecal microbiota transplantation by washed preparation in patients with moderate to severely active ulcerative colitis. J Dig Dis 2020;21(11):621–8.
[14] Ding X, Li Q, Li P, et al. Long-term safety and efficacy of fecal microbiota transplant in active ulcerative colitis. Drug Saf 2019;42(7):869–80.
[15] Kang GU, Park S, Jung Y, et al. Exploration of potential gut microbiota-derived biomarkers to predict the success of fecal microbiota transplantation in ulcerative colitis: a prospective cohort in Korea. Gut liver 2022;16(5):775–85.
[16] Liu Y, Ji X, Huang Y, et al. Older patients benefit more from sequential courses of washed microbiota transplantation than younger population with ulcerative colitis. Scand J Gastroenterol 2023:1–10.
[17] Lu G, Wang W, Li P, et al. Washed preparation of faecal microbiota changes the transplantation related safety, quantitative method and delivery. Micro Biotechnol 2022;15(9):2349–449.
[18] Wang W, Lu G, Wu X, et al. Colonic transendoscopic enteral tubing is a new pathway to microbial therapy, colonic drainage, and host-microbiota interaction research. J Clin Med 2023;12(3).
[19] Matsuoka K, Watanabe M, Ohmori T, et al. AJM300 (carotegrast methyl), an oral antagonist of α4-integrin, as induction therapy for patients with moderately active ulcerative colitis: a multicentre, randomised, double-blind, placebo-controlled, phase 3 study. Lancet Gastroenterol Hepatol 2022;7(7):648–57.
[20] Kelly CR, Kunde SS, Khoruts A. Guidance on preparing an investigational new drug application for fecal microbiota transplantation studies. Clin Gastroenterol Hepatol 2014;12(2):283–8.
[21] Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in python. J Mach Learn Res 2011;12:2825–30.
[22] Pollard TJ, Johnson AEW, Raffa JD, et al. tableone: an open source Python package for producing summary statistics for research papers. JAMIA Open 2018;1(1):26–31.
[23] Hanley JA, Mcneil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology 1983;148(3):839–43.
[24] Taha AA, Hanbury A. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. BMC Med Imaging 2015;15:29.
[25] Rufibach K. Use of Brier score to assess binary predictions. J Clin Epidemiol 2010;63(8):938–9. author reply 9.
[26] Lundberg SM, Nair B, Vavilala MS, et al. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. Nat Biomed Eng 2018;2(10):749–60.
[27] Paramsothy S, Nielsen S, Kamm MA, et al. Specific bacteria and metabolites associated with response to fecal microbiota transplantation in patients with ulcerative colitis. Gastroenterology 2019;156(5):1440–54. e2.
[28] Rees NP, Shaheen W, Quince C, et al. Systematic review of donor and recipient predictive biomarkers of response to faecal microbiota transplantation in patients with ulcerative colitis. EBioMedicine 2022;81:104088.
[29] Wu X, Li P, Wang W, et al. The underlying changes in serum metabolic profiles and efficacy prediction in patients with extensive ulcerative colitis undergoing fecal microbiota transplantation. Nutrients 2023;15(15).

[30] Sands BE. Biomarkers of inflammation in inflammatory bowel disease. Gastroenterology 2015;149(5):1275–85. e2.
[31] Cui B, Li P, Xu L, et al. Step-up fecal microbiota transplantation strategy: a pilot study for steroid-dependent ulcerative colitis. J Transl Med 2015;13:298.
[32] Dulai PS, Feagan BG, Sands BE, et al. Prognostic value of fecal calprotectin to inform treat-to-target monitoring in ulcerative colitis. Clin Gastroenterol Hepatol 2023;21(2):456–66. e7.
[33] Zand A, Stokes Z, Sharma A, et al. Artificial intelligence for inflammatory bowel diseases (IBD); accurately predicting adverse outcomes using machine learning. Dig Dis Sci 2022;67(10):4874–85.
[34] Zhang T, Cui B, Li P, et al. Short-term surveillance of cytokines and c-reactive protein cannot predict efficacy of fecal microbiota transplantation for ulcerative colitis. PLoS One 2016;11(6):e0158227.
[35] Stakenborg N, Labeeuw E, Gomez-Pinilla PJ, et al. Preoperative administration of the 5-HT4 receptor agonist prucalopride reduces intestinal inflammation and shortens postoperative ileus via cholinergic enteric neurons. Gut 2019;68(8):1406–16.
[36] Singh SP, Chand HS, Banerjee S, et al. Acetylcholinesterase inhibitor pyridostigmine bromide attenuates gut pathology and bacterial dysbiosis in a murine model of ulcerative colitis. Dig Dis Sci 2020;65(1):141–9.
[37] Kume K, Yamasaki M, Tashiro M, et al. Activations of coagulation and fibrinolysis secondary to bowel inflammation in patients with ulcerative colitis. Intern Med 2007;46(17):1323–9.
[38] Shi H, Feng Y, Jiang J, et al. Correlations between the serum bilirubin level and ulcerative colitis: a case-control study. Eur J Gastroenterol Hepatol 2019;31(8):992–7.
[39] Schieffer KM, Bruffy SM, Rauscher R, et al. Reduced total serum bilirubin levels are associated with ulcerative colitis. PLoS One 2017;12(6):e0179267.
[40] Shiraishi K, Furukawa S, Yagi S, et al. Association between serum bilirubin and mucosal healing among Japanese patients with ulcerative colitis: a cross-sectional study. Int J Colorectal Dis 2021;36(2):377–82.
[41] Zoroddu S, Di Lorenzo B, Paliogiannis P, et al. The association between bilirubin concentrations and inflammatory bowel disease: insights from a systematic review and meta-analysis. Eur J Clin Invest 2024:e14281.
[42] Zhang MH, Wang HG, Shi YT, et al. Efficacy of serum total bilirubin in predicting the severity of ulcerative colitis: a cross-sectional study. Ann Clin Lab Sci 2020;50(2):228–32.
[43] Zhuo Z, Guo K, Luo Y, et al. Targeted modulation of intestinal epithelial regeneration and immune response in ulcerative colitis using dual-targeting bilirubin nanoparticles. Theranostics 2024;14(2):528–46.
[44] Lin J, Xiong J, Jin Y, et al. Fecal microbiota transplantation through transendoscopic enteral tubing for inflammatory bowel disease: high acceptance and high satisfaction. J Gastroenterol Hepatol 2024;39(2):328–36.
[45] Zhang F, Lu G, Wang X, et al. Concept, breakthrough, and future of colonic transendoscopic enteral tubing. Chin Med J (Engl) 2024;137(6):633–5.