

RESEARCH ARTICLE



Applying machine learning to predict bowel preparation adequacy in elderly patients for colonoscopy: development and validation of a web-based prediction tool

Jianning Liu^{a*}, Wei Jiang^{a,b*}, Yahong Yu^{a*}, Jiali Gong^a, Guie Chen^a, Yuxing Yang^a, Chao Wang^a, Dalong Sun^{a,b} and Xuefeng Lu^a

^aDepartment of Gastroenterology and Hepatology, Zhongshan Hospital (Xiamen), Fudan University, Xiamen, China; ^bDepartment of Gastroenterology and Hepatology, Zhongshan Hospital, Fudan University, Shanghai, China

ABSTRACT

Background: Adequate bowel preparation is crucial for effective colonoscopy, especially in elderly patients who face a high risk of inadequate preparation. This study develops and validates a machine learning model to predict bowel preparation adequacy in elderly patients before colonoscopy.

Methods: The study adhered to the TRIPOD AI guidelines. Clinical data from 471 elderly patients collected between February and December 2023 were utilized for developing and internally validating the model, while 221 patients' data from March to June 2024 were used for external validation. The Boruta algorithm was applied for feature selection. Models including logistic regression, light gradient boosting machines, support vector machines (SVM), decision trees, random forests, and extreme gradient boosting were evaluated using metrics such as AUC, accuracy, sensitivity, and specificity. The SHAP algorithm helped rank feature importance. A web-based application was developed using the Streamlit framework to enhance clinical usability.

Results: The Boruta algorithm identified 7 key features. The SVM model excelled with an AUC of 0.895 (95% CI: 0.822–0.969), and high accuracy, sensitivity, and specificity. In external validation, the SVM model maintained robust performance with an AUC of 0.889. The SHAP algorithm further explained the contribution of each feature to model predictions.

Conclusion: The study developed an interpretable and practical machine learning model for predicting bowel preparation adequacy in elderly patients, facilitating early interventions to improve outcomes and reduce resource wastage.

KEY MESSAGES

1. This study developed a machine learning model to predict bowel preparation adequacy in elderly patients undergoing colonoscopy, notably improving prediction accuracy and aiding clinical decision-making.
2. Multiple machine learning models were used to predict bowel preparation adequacy, with the support vector machine (SVM) achieving the best performance. SHAP analysis enhanced the interpretability of the model by identifying key predictive factors, making it a reliable and transparent tool for clinical use.
3. The predictive model was integrated into a user-friendly web application, enabling healthcare providers to identify high-risk patients early and enhance the quality of bowel preparation interventions.

ARTICLE HISTORY

Received 6 November 2024

Revised 12 February 2025

Accepted 20 February 2025

KEYWORDS

Elderly patients; bowel preparation; colonoscopy; machine learning; predictive model

Introduction

Colorectal cancer, the fourth deadliest cancer globally, causes approximately 900,000 deaths annually, and this number continues to rise [1]. Colonoscopy is crucial for early screening and tumor resection [2]. Adequate bowel preparation is essential for successful colonoscopy; inadequate preparation can lead to a

high rate of missed adenomas, impact diagnostic accuracy, necessitate repeat procedures in the short term, and waste medical resources [3–5]. Despite this, inadequate bowel preparation remains a common issue. Studies show that 10%–25% of patients experience inadequate bowel preparation, and this rate is as high as 34.6% among elderly patients [4–6]. Therefore, early identification of high-risk individuals and timely

CONTACT Dalong Sun sun.dalong@zs-hospital.sh.cn Department of Gastroenterology and Hepatology, Zhongshan Hospital, Fudan University, Shanghai, China; Xuefeng Lu lu.xuefeng@zsmhshospital.com Department of Gastroenterology and Hepatology, Zhongshan Hospital (Xiamen), Fudan University, Xiamen, China.

*These authors contributed equally to this work.

Supplemental data for this article can be accessed online at <https://doi.org/10.1080/07853890.2025.2474172>.

© 2025 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

interventions are critical for patients undergoing colonoscopy.

In recent years, many studies have focused on identifying factors affecting bowel preparation quality and developing predictive models to improve preparation success rates. Gandhi et al. [7] summarized factors influencing bowel preparation quality in a meta-analysis, finding that age, constipation, and comorbid conditions were associated with higher rates of inadequate preparation. Gu et al. [8] investigated the effectiveness of different models in predicting bowel preparation failure, finding that machine learning-based methods outperformed traditional regression models.

Elderly patients, due to physiological changes and comorbid conditions, are more likely to face inadequate bowel preparation. Although several predictive models have been developed to assess the risk of preparation failure in inpatient or outpatient settings [8–11], accurate prediction of bowel preparation in elderly patients remains unmet. Given this background, our study aims to develop an easy-to-use predictive model for bowel preparation failure specifically for elderly patients. We will apply machine learning algorithms to select predictive factors and build the model, aiming to improve prediction accuracy and clinical utility. This model is intended to assist healthcare professionals in identifying high-risk patients and implementing timely interventions.

Materials and methods

Study subjects

This study included elderly patients who underwent colonoscopy examination and treatment at the Department of Gastroenterology, Zhongshan Hospital (Xiamen), Fudan University, from February to December 2023, for the development and internal validation of a predictive model. Relevant clinical data were collected. Inclusion criteria were ① Age ≥ 60 years; ② Scheduled for colonoscopy or colorectal surgery; ③ Bowel preparation regimen involved oral intake of 3 L polyethylene glycol solution. Exclusion criteria were ① Non-compliance with the bowel preparation protocol, i.e. intake of polyethylene glycol < 3 L; ② Incomplete colonoscopy due to non-preparation factors such as bowel tumors or strictures; ③ No severe cardiovascular, pulmonary, renal diseases, or cognitive impairments. This study was approved by the hospital's ethics committee (approval number: B2023-049R). As a retrospective study, informed consent was waived by the Ethics Committee.

Based on literature review results [7,12–16], this study included 21 risk factors and calculated the

sample size according to the Events Per Variable (EPV) method [17]. At least 105 cases of bowel preparation failure were required. According to previous literature, about 34.6% of elderly patients have inadequate bowel preparation. Considering a 10% sample attrition rate, the sample size for the modeling group should exceed 334 cases. Ultimately, 471 patients were included in the study.

Bowel preparation Methods

All patients received pre-examination education from a responsible nurse the day before the procedure. The education included guidance on pre-examination precautions and instructions on the use of laxatives, accompanied by written instructions. Patients were required to consume a liquid diet the day before the examination, fast after starting the laxative, and avoid drinking fluids after completing bowel preparation. All patients followed the guidelines for colonoscopy bowel preparation [18,19], using a 3 L oral polyethylene glycol solution for bowel preparation.

Assessment of bowel preparation quality

The quality of bowel preparation was assessed using the Boston Bowel Preparation Scale (BBPS) [20], which has a weighted kappa value of 0.78 and an intra-class correlation coefficient of 0.91. This scale uses a 4-point scoring system to evaluate the right, left, and transverse colon segments separately. The overall bowel preparation quality is reflected by the total score, with a score of < 6 or any segment score of < 2 considered as inadequate bowel preparation [21]. All colonoscopies were performed by well-trained and experienced endoscopists who scored the bowel preparation quality during the procedure.

Data collection and key Variable definitions

Based on the literature review, a self-designed questionnaire was used to measure the risk factors related to bowel preparation failure in the study population. The questionnaire was divided into three parts: ① General information: age, gender, Body Mass Index (BMI), daily activity level, bowel movement status, smoking history, and education level. ② Disease-related information: hypertension, diabetes, coronary artery disease, hyperlipidemia, constipation, use of calcium channel blockers (CCB), history of gastrointestinal surgery, pelvic surgery, appendectomy, cholecystectomy, and previous colonoscopy. ③ Bowel preparation information: at-home/in-hospital bowel preparation,

split/single dose, interval time to colonoscopy, characteristics of the last bowel movement, and bowel preparation quality.

The definitions of key variables are as follows: Bowel movement status refers to the patient's bowel habits over the past week, categorized into three groups: diarrhea, constipation, or normal bowel movements. The interval time to colonoscopy is defined as the time, measured in hours, between the completion of the last dose of laxative and the initiation of the colonoscopy procedure. Last bowel movement as clear liquid describes the type of rectal effluent observed during the final stage of bowel preparation, characterized by a non-brown, sediment-free liquid state. This variable reflects the most recent bowel movement after completing laxative intake and is critical for predicting bowel preparation adequacy [13].

The data were collected through the hospital's electronic medical record system and nursing assessment system, and were integrated, supplemented, and entered by two individuals after verification. Data with missing entries, non-compliance with standards, or obvious logical issues were rechecked and corrected. Since no cases with missing information were present in the final dataset, exclusion of samples due to missing values was not required.

Model selection and comparison

All data were randomly divided into a training set (70%) and a validation set (30%) to evaluate model performance and prevent overfitting. An external dataset was used for testing (external validation) to assess the generalizability of the models. The training set was utilized for model development, including feature selection and hyperparameter tuning. The internal validation set was evaluated using 1,000 bootstrap resampling iterations, during which the AUC curves for each model were calculated and plotted to demonstrate the performance distribution under different resampling conditions. This approach ensures the stability and reliability of the results. Importantly, the external test set remained completely independent throughout the entire model training and validation process, ensuring unbiased evaluation and preventing information leakage. This rigorous approach guarantees the robustness, reliability, and generalizability of the predictive models.

In this study, in addition to logistic regression (LR) [22], several machine learning models were employed to construct and compare the prediction models for bowel preparation failure in elderly patients: Light Gradient Boosting Machine (LightGBM) [23], Support

Vector Machine (SVM) [24], Decision Tree (DT) [25], Random Forest (RF) [26], and Extreme Gradient Boosting (XGBoost) (<https://xgboost.ai/>). Corresponding Receiver Operator Characteristic (ROC) curves, Decision Curve Analysis (DCA) curves, and Precision-Recall (PR) curves were plotted. Evaluation metrics for different models included prediction accuracy, sensitivity, specificity, Negative Predictive Value (NPV), Positive Predictive Value (PPV), F1 score, and Area Under the Curve (AUC).

Feature selection and model Interpretation

This study used the Boruta algorithm to optimize the number of predictors in the model. The Boruta algorithm is a feature selection method based on Random Forest, capable of identifying all variables significantly affecting the prediction outcome. It works by combining the original data with randomly shuffled "shadow" features, extending the dataset, and training a Random Forest model to evaluate the importance of each feature. The algorithm iteratively compares the importance scores of the original and shadow features, progressively eliminating unimportant features until the final classification of all features is determined [27].

SHAP (SHapley Additive exPlanations) is a tool based on Shapley values, used to interpret machine learning model outputs, providing both global and local perspectives [28]. This study used SHAP to demonstrate the association between different features and bowel preparation failure, ensuring the reliability and fairness of model interpretation, and presenting the model behavior intuitively.

External validation

The external validation was conducted using a dataset comprising elderly patients undergoing colonoscopy at the Department of Gastroenterology, Zhongshan Hospital (Xiamen), Fudan University, between March 2024 and June 2024. The inclusion and exclusion criteria were consistent with those used in the initial study.

Development of a web application

To enhance the clinical utility of the prediction model, the final prediction model was integrated into a Python web application based on the Streamlit framework. When users input the corresponding feature values, the web application outputs the prediction results and probabilities of bowel preparation failure.

Statistical methods

Data analysis and curve plotting were performed using SPSS 26.0 (<https://www.ibm.com/spss>), Python 3.11.4 (<https://www.python.org>) and R 4.4.2 (<https://www.r-project.org>). Continuous variables are presented as mean and SD. Categorical data were expressed as counts (%), and comparisons between groups were made using the χ^2 test and Fisher's exact test. Differences were considered statistically significant at $p < 0.05$.

TRIPOD AI statement

This study was reported in accordance with the TRIPOD AI guidelines [29] (Transparent Reporting of a multi-variable prediction model for Individual Prognosis Or Diagnosis – AI extension). In this study, the relevant items from the TRIPOD AI checklist were referenced and followed throughout the model development, validation, performance evaluation, and reporting processes to ensure transparency and completeness.

Result

Patient characteristics

A total of 473 patients underwent colonoscopy, with 2 patients excluded: one due to severe renal insufficiency (CKD stage 5) and another because of an intraluminal tumor that prevented complete colonoscopy. Ultimately, 471 patients were included in the study. The study design is illustrated in Figure 1.

The median patient age was 66 years (IQR 63–71), with the oldest being 87 years, and 50.32% (237/471) were male. The average BMI was $23.00 \pm 3.15 \text{ kg/m}^2$; 61.36% (289/471) had a BMI $< 24 \text{ kg/m}^2$, 33.12% (156/471) had a BMI between 24 and 27.9 kg/m^2 , and 5.52% (26/471) had a BMI $\geq 28 \text{ kg/m}^2$. Hypertension was the most common comorbidity (37.58%), followed by diabetes (20.17%), hyperlipidemia (12.74%), and coronary artery disease (6.58%); 22.08% (104/471) were on calcium channel blockers. Surgical history included appendectomy (4.03%), gastrointestinal surgery (5.52%), cholecystectomy (3.18%), and pelvic surgery (6.37%); 16.56% (78/471) were smokers. Regarding bowel preparation, 77.07% (363/471) used a split-dose regimen and 63.69% (300/471) prepared in the hospital. In terms of bowel habits, 15.50% (73/471) reported constipation, while 16.35% (77/471) experienced diarrhea; 62.42% (294/471) were undergoing their first colonoscopy. Frequently walking was reported by 81.95% (386/471); 92.14% (434/471) had clear liquid

stool during their last bowel movement. Educational levels were primary school or below for 44.80% (211/471), middle school for 26.96% (127/471), and high school or above for 28.24% (133/471). Overall, 76.86% (362/471) achieved adequate bowel preparation, while 23.14% (109/471) had inadequate preparation.

Patients were randomly divided into a training set (329/471) and a validation set (142/471) in a 7:3 ratio. General information on patients in both groups is detailed in Table 1, with no statistically significant differences observed between the groups.

Prediction model feature selection

All variables were input into the Boruta algorithm, which identified 7 significant predictors: split-dose, in-hospital bowel preparation, education, CCB, bowel movement status, activity level, and last bowel movement was clear liquid. The importance of these features was ranked as shown in Figure 2. To facilitate the identification of model features, the risk factors selected by the Boruta algorithm were assigned values, as detailed in Table 2.

Model development and performance comparison

To predict bowel preparation failure, we set bowel preparation failure as the dependent variable and used the factors identified by the Boruta algorithm as independent variables. We constructed models using logistic regression (LR), extreme gradient boosting (XGBoost), light gradient boosting machine (LightGBM), decision tree (DT), random forest (RF), and support vector machine (SVM). In the training set, the decision tree model had the highest AUC of 0.968 (95% CI: 0.948–0.988). In the validation set, the SVM model had the highest AUC of 0.895 (95% CI: 0.822–0.969) (Figure 3a). In the forest plot, the SVM model performed the best, with an AUC of 0.895 and a standard error of 0.005 (Figure 3b). DCA analysis also showed that the SVM had the highest net benefit at low decision thresholds (Figure 3c). Additionally, as shown in Figure 3d, the SVM displayed high average precision (AP of 0.853 and 0.821) in both the training and validation sets. For other performance metrics, as presented in Table 3, the XGB model had the highest sensitivity (0.766), while the SVM model achieved the highest accuracy (0.889), specificity (0.932), and F1 score (0.752).

To evaluate the impact of feature selection on model performance, multiple models were constructed using

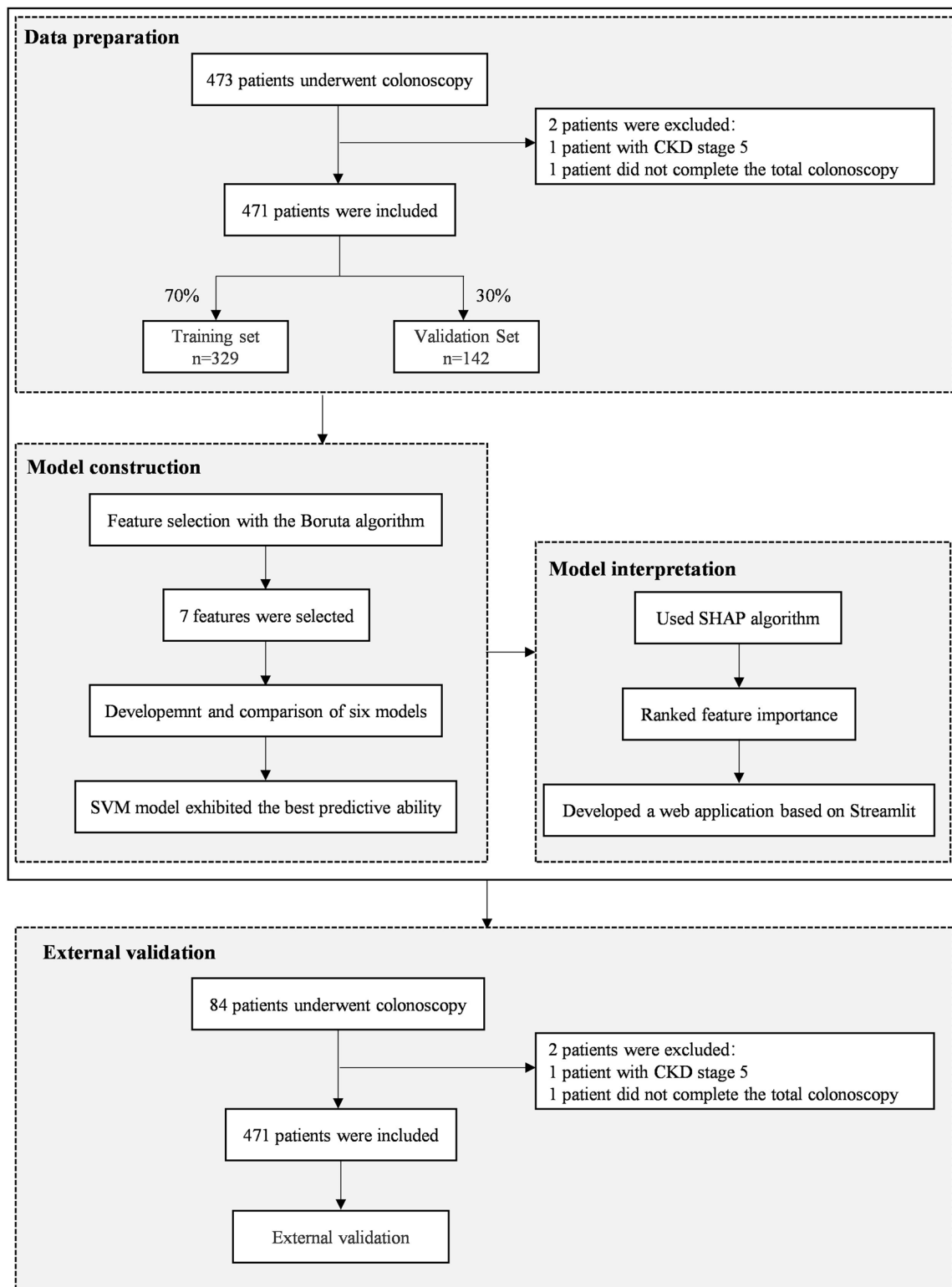


Figure 1. Research design flowchart.

all features and the top 7, 10, 13, 16, and 19 features ranked by Boruta's feature importance. The results showed that the SVM model consistently maintained the best predictive performance across all feature sets (Figure 4). Notably, when using the top 7 features

selected by Boruta, the SVM model demonstrated strong performance, second only to the model constructed with the top 13 features ($\Delta\text{AUC} = 0.018$, $p = 0.985$).

In summary, although the Delong test revealed no statistically significant differences in AUC between the

Table 1. Comparison of demographic and clinical characteristics between training set and validation set.

| Variables | Total <i>n</i> = 471 | Training Set <i>n</i> = 329 | Validation Set <i>n</i> = 142 | <i>p</i> |
|--|----------------------|-----------------------------|-------------------------------|----------|
| Age, <i>n</i> (%) | | | | 0.235 |
| <70 years | 320 (67.94) | 218 (66.26) | 102 (71.83) | |
| ≥70 years | 151 (32.06) | 111 (33.74) | 40 (28.17) | |
| Gender, <i>n</i> (%) | | | | 0.771 |
| Male | 237 (50.32) | 167 (50.76) | 70 (49.30) | |
| Female | 234 (49.68) | 162 (49.24) | 72 (50.70) | |
| BMI, <i>n</i> (%) | | | | 0.094 |
| <24 kg/m ² | 289 (61.36) | 210 (63.83) | 79 (55.63) | |
| 24–27.9 kg/m ² | 156 (33.12) | 99 (90.09) | 57 (40.14) | |
| ≥28 kg/m ² | 26 (5.52) | 20 (6.08) | 6 (4.23) | |
| HBP, <i>n</i> (%) | | | | 0.187 |
| No | 294 (62.42) | 199 (60.49) | 95 (66.90) | |
| Yes | 177 (37.58) | 130 (39.51) | 47 (33.10) | |
| DM, <i>n</i> (%) | | | | 0.112 |
| No | 376 (79.83) | 269 (81.76) | 107 (75.35) | |
| Yes | 95 (20.17) | 60 (18.24) | 35 (24.65) | |
| CAD, <i>n</i> (%) | | | | 0.139 |
| No | 440 (93.42) | 311 (94.53) | 129 (90.85) | |
| Yes | 31 (6.58) | 18 (5.47) | 13 (9.15) | |
| HPL, <i>n</i> (%) | | | | 0.979 |
| No | 411 (87.26) | 287 (87.23) | 124 (87.32) | |
| Yes | 60 (12.74) | 42 (12.77) | 18 (12.68) | |
| CCB, <i>n</i> (%) | | | | 0.195 |
| No | 367 (77.92) | 251 (76.29) | 116 (81.69) | |
| Yes | 104 (22.08) | 78 (23.71) | 26 (18.31) | |
| History of appendectomy, <i>n</i> (%) | | | | 0.710 |
| No | 452 (95.97) | 315 (95.74) | 137 (96.48) | |
| Yes | 19 (4.03) | 14 (4.26) | 5 (3.52) | |
| History of gastrointestinal surgery, <i>n</i> (%) | | | | 0.165 |
| No | 445 (94.48) | 314 (95.44) | 131 (92.25) | |
| Yes | 26 (5.52) | 15 (4.56) | 11 (7.75) | |
| History of cholecystectomy, <i>n</i> (%) | | | | 0.401 |
| No | 456 (96.82) | 320 (97.26) | 136 (95.77) | |
| Yes | 15 (3.18) | 9 (2.74) | 6 (4.23) | |
| History of pelvic surgery, <i>n</i> (%) | | | | 0.694 |
| No | 441 (93.63) | 309 (93.92) | 132 (92.96) | |
| Yes | 30 (6.37) | 20 (6.08) | 10 (7.04) | |
| Smoking, <i>n</i> (%) | | | | 0.896 |
| No | 393 (83.44) | 275 (83.59) | 118 (83.10) | |
| Yes | 78 (16.56) | 54 (16.41) | 24 (16.90) | |
| Split-dose, <i>n</i> (%) | | | | 0.893 |
| No | 108 (22.93) | 76 (23.10) | 32 (22.54) | |
| Yes | 363 (77.07) | 253 (76.90) | 110 (77.46) | |
| In-hospital bowel preparation, <i>n</i> (%) | | | | 0.908 |
| No | 171 (36.31) | 120 (36.47) | 51 (35.92) | |
| Yes | 300 (63.69) | 209 (63.53) | 91 (64.08) | |
| Bowel movement status, <i>n</i> (%) | | | | 0.551 |
| Normal | 321 (68.15) | 226 (68.69) | 95 (66.90) | |
| Diarrhea | 77 (16.35) | 50 (15.20) | 27 (19.01) | |
| Constipation | 73 (15.50) | 53 (16.11) | 20 (14.08) | |
| First colonoscopy, <i>n</i> (%) | | | | 0.073 |
| No | 177 (37.58) | 115 (34.95) | 62 (43.66) | |
| Yes | 294 (62.42) | 214 (65.05) | 80 (56.34) | |
| Activity level, <i>n</i> (%) | | | | 0.870 |
| Occasionally walking | 85 (18.05) | 60 (18.24) | 25 (17.61) | |
| Frequently walking | 386 (81.95) | 269 (81.76) | 117 (82.39) | |
| Last bowel movement was clear liquid, <i>n</i> (%) | | | | 0.288 |
| No | 37 (7.86) | 23 (6.99) | 14 (9.86) | |
| Yes | 434 (92.14) | 306 (93.01) | 128 (90.14) | |
| Interval time, <i>n</i> (%) | | | | 0.492 |
| <9h | 311 (66.03) | 214 (65.05) | 97 (68.31) | |
| ≥9h | 160 (33.97) | 115 (34.95) | 45 (31.69) | |
| Education, <i>n</i> (%) | | | | 0.777 |
| Primary school or below | 211 (44.80) | 148 (44.98) | 63 (44.37) | |
| Middle school | 127 (26.96) | 91 (27.66) | 36 (25.35) | |
| High school or above | 133 (28.24) | 90 (27.36) | 43 (30.28) | |
| Bowel preparation, <i>n</i> (%) | | | | 0.163 |
| Success | 362 (76.86) | 247 (75.08) | 115 (80.99) | |
| Failure | 109 (23.14) | 82 (24.92) | 27 (19.01) | |

Note: BMI: Body Mass Index; HBP: High Blood Pressure; DM: Diabetes Mellitus; CAD: Coronary Artery Disease; HPL: Hyperlipidemia; CCB: Calcium Channel Blocker.

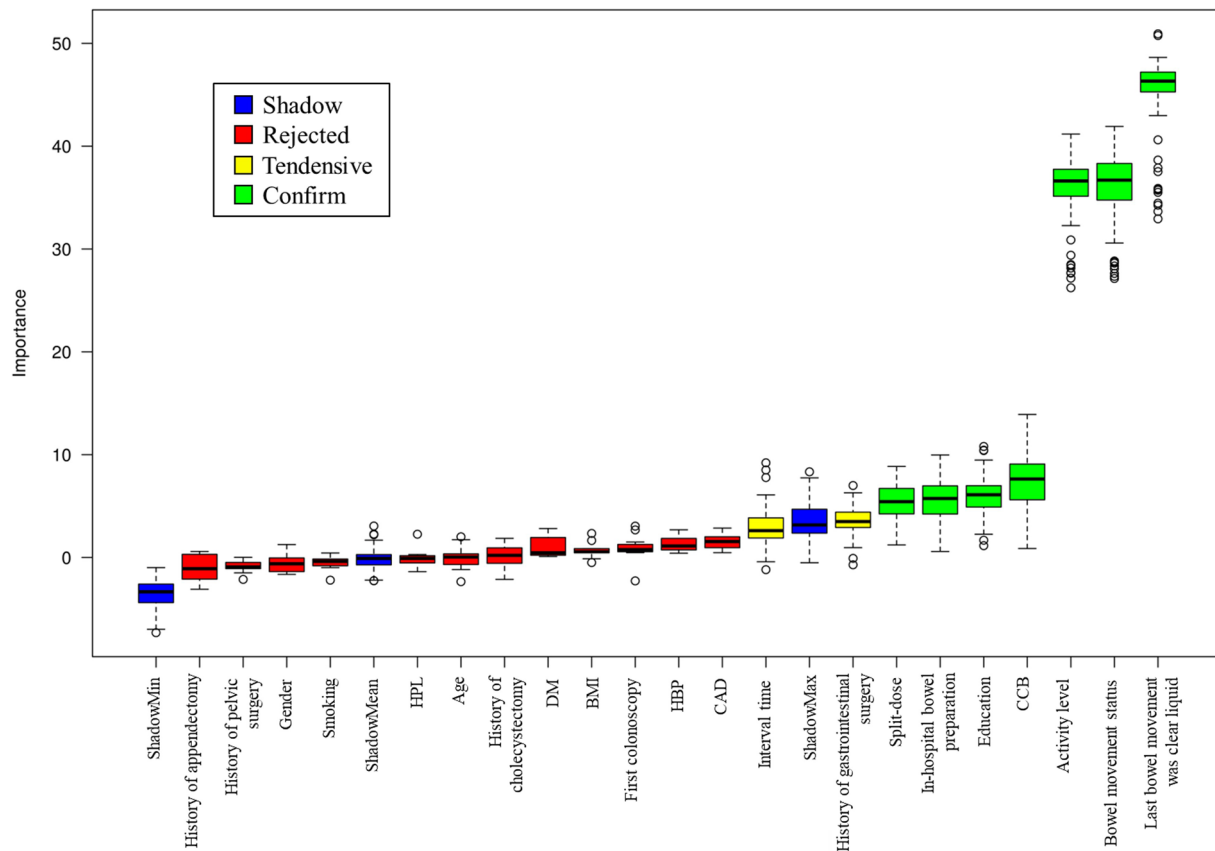


Figure 2. Importance of shadow and predictor variables selected by the boruta algorithm.

Table 2. Variable assignment and scoring table.

| Variables | Value Assignment |
|--------------------------------------|--|
| Split-dose | No = 0, Yes = 1 |
| In-hospital bowel preparation | No = 0, Yes = 1 |
| Education | Primary school or below = 1, Middle school = 2, High school or above = 3 |
| CCB | No = 0, Yes = 1 |
| Bowel movement status | Normal = 1, Diarrhea = 2, Constipation = 3 |
| Activity level | Frequently walking = 0, Occasionally walking = 1 |
| Last bowel movement was clear liquid | No = 0, Yes = 1 |

Note: CCB: Calcium Channel Blocker.

SVM model and other models (Table 4), the SVM model consistently demonstrated stable and superior performance across different feature sets. Notably, its consistent performance on multiple key metrics highlights its reliability and practical value in predicting bowel preparation failure.

Selection and external validation of the final model

To balance predictive simplicity and model performance, the final predictive model for bowel

preparation failure in elderly patients was constructed using the seven key features identified by the Boruta algorithm and implemented with an SVM approach. External validation of this model, conducted on an independent dataset, yielded an AUC of 0.889, indicating excellent predictive performance (Supplementary Figure 1). The baseline characteristics of the external validation cohort are shown in Supplementary Table 1.

To further evaluate the model's applicability to patients who had not yet initiated bowel preparation, the feature "last bowel movement as clear liquid" was excluded, and a new model was constructed. This revised model was validated externally, achieving an AUC of 0.803, which closely aligned with the internal validation AUC (0.876) (Supplementary Figure 2). These results demonstrate the robustness and consistency of the model's predictive performance across different feature sets.

The findings suggest that the final SVM model not only performs well in external validation but also maintains reliable accuracy even after excluding a key feature, highlighting its adaptability for broader clinical applications.

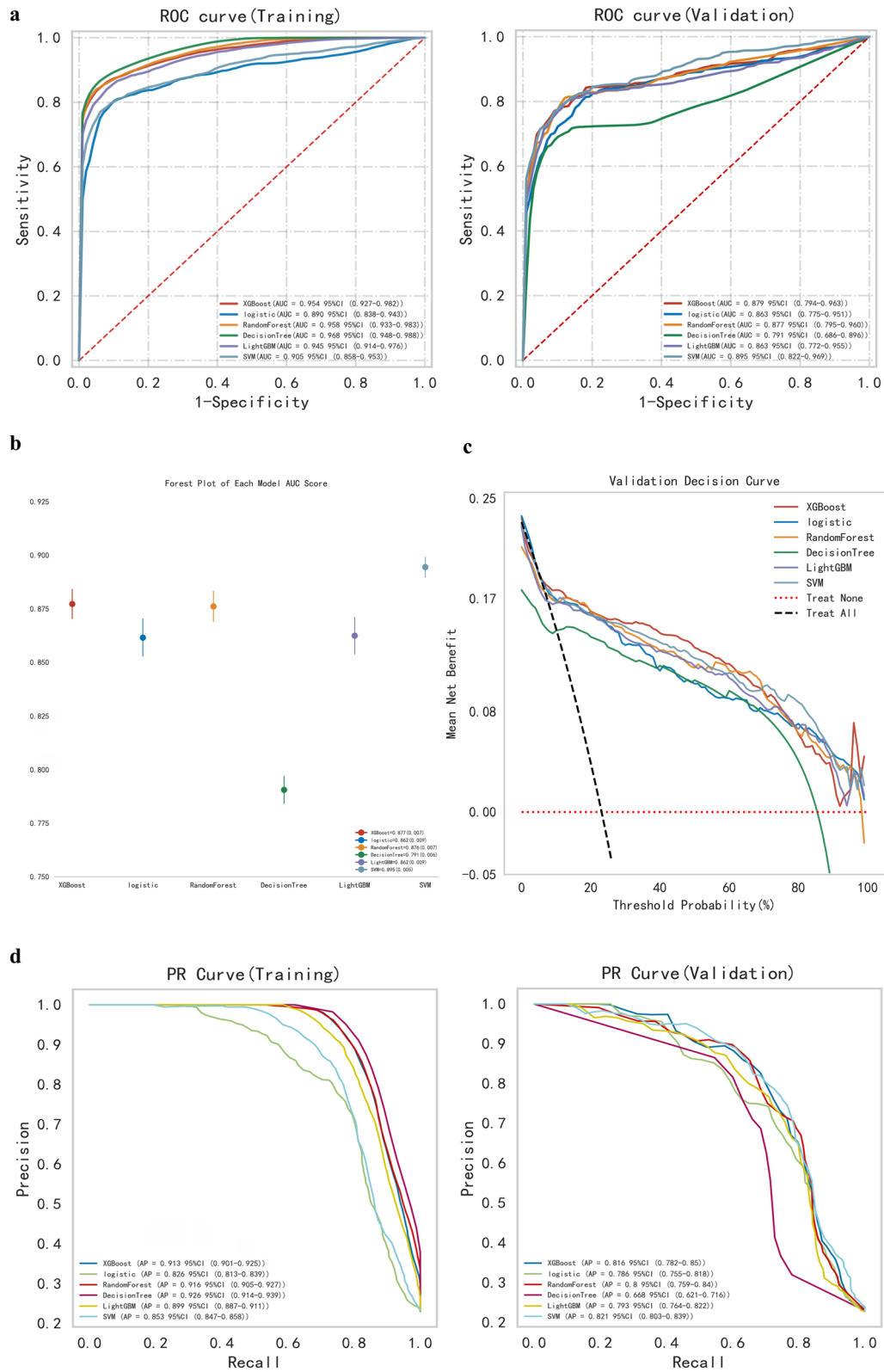


Figure 3. Model performance comparisons this figure includes multiple Sub-figures: a: ROC curves of different machine learning models, comparing training and validation performance. b: Forest plot displaying the AUC scores for each model. c: Decision Curve analysis (DCA) for different machine learning models in the validation set. d: Precision-Recall curves for training and validation sets across all models.

Table 3. Predictive performance of different models.

| Variables | AUC (95% CI) | Accuracy | Sensitivity | Specificity | PPV | NPV | F1 score |
|-----------|---------------------|----------|-------------|-------------|-------|-------|----------|
| XGBoost | 0.879 (0.794–0.963) | 0.873 | 0.766 | 0.905 | 0.723 | 0.928 | 0.738 |
| LR | 0.863 (0.775–0.951) | 0.873 | 0.702 | 0.924 | 0.757 | 0.911 | 0.722 |
| LightGBM | 0.877 (0.795–0.960) | 0.875 | 0.751 | 0.913 | 0.729 | 0.924 | 0.736 |
| RF | 0.791 (0.686–0.896) | 0.862 | 0.649 | 0.925 | 0.726 | 0.900 | 0.678 |
| DT | 0.863 (0.772–0.955) | 0.876 | 0.728 | 0.920 | 0.738 | 0.921 | 0.725 |
| SVM | 0.895 (0.822–0.969) | 0.889 | 0.739 | 0.932 | 0.771 | 0.924 | 0.752 |

Note: XGBoost: Extreme Gradient Boosting; LR: Logistic Regression; RF: Random Forest; DT: Decision Tree; LightGBM: Light Gradient Boosting Machine; SVM: Support Vector Machine.

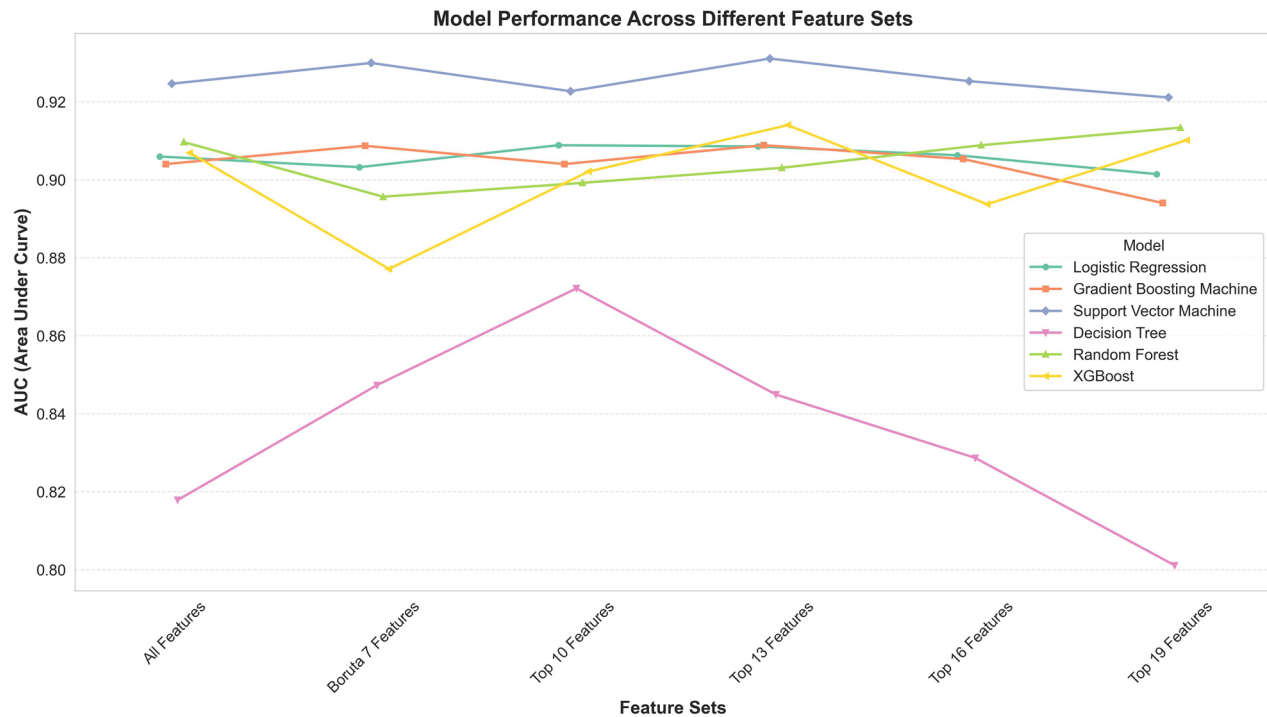


Figure 4. Model performance across different feature sets. This figure shows the AUC performance of various machine learning models across different feature sets, including all available features, the top 7 features selected by the Boruta algorithm, and the top 10, 13, 16, and 19 features ranked by importance. The figure highlights the robustness of the SVM model, which consistently achieves the highest AUC across all feature sets, and demonstrates the impact of feature selection on model performance.

Table 4. P-values from delong test comparing AUCs between different models.

| Variables | XGBoost | LR | RF | DT | LightGBM | SVM |
|-----------|---------|--------|--------|--------|----------|--------|
| XGBoost | NA | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 |
| LR | <0.001 | NA | 0.432 | 0.634 | 0.564 | 0.441 |
| RF | <0.001 | 0.432 | NA | 0.569 | 0.580 | 0.539 |
| DT | <0.001 | 0.634 | 0.569 | NA | 0.598 | 0.502 |
| LightGBM | <0.001 | 0.564 | 0.580 | 0.598 | NA | 0.528 |
| SVM | <0.001 | 0.441 | 0.539 | 0.502 | 0.528 | NA |

Note: XGBoost: Extreme Gradient Boosting; LR: Logistic Regression; RF: Random Forest; DT: Decision Tree; LightGBM: Light Gradient Boosting Machine; SVM: Support Vector Machine.

Model interpretation

To enhance the transparency and interpretability of the model, we utilized the SHAP algorithm to elucidate the model's output. SHAP offers two types of explanations: global feature-level explanations and local individual-level explanations [28]. Global

explanations reveal the overall dependencies of the model on all features. Figure 5a and b illustrates the ranking of feature importance, reflecting the average impact of each feature on the model's output. It is evident that "Bowel movement status" and "Last bowel movement was clear liquid" are the two most influential features affecting the prediction, followed closely

by “Activity level,” “Education,” and “CCB” (use of calcium channel blockers). These features play a crucial role in the decision-making process of the model.

Local explanations focus on the impact of features for individual samples, as shown in Figure 5c. In the figure, red and blue bars represent features that increase and decrease the predicted value, respectively. Specifically, the feature “CCB” (use of calcium channel blockers) being “yes” has a marked contribution to predicting failure. Other features, such as “Last bowel movement was clear liquid” and “Bowel movement

status,” also have notable influences on the model’s output.

Clinical application of a predictive model

The best predictive model identified in this study has been integrated into a web application to enhance its usability in routine clinical practice. By sequentially entering the selected seven feature values, the model automatically calculates the probability of bowel preparation failure for an individual elderly

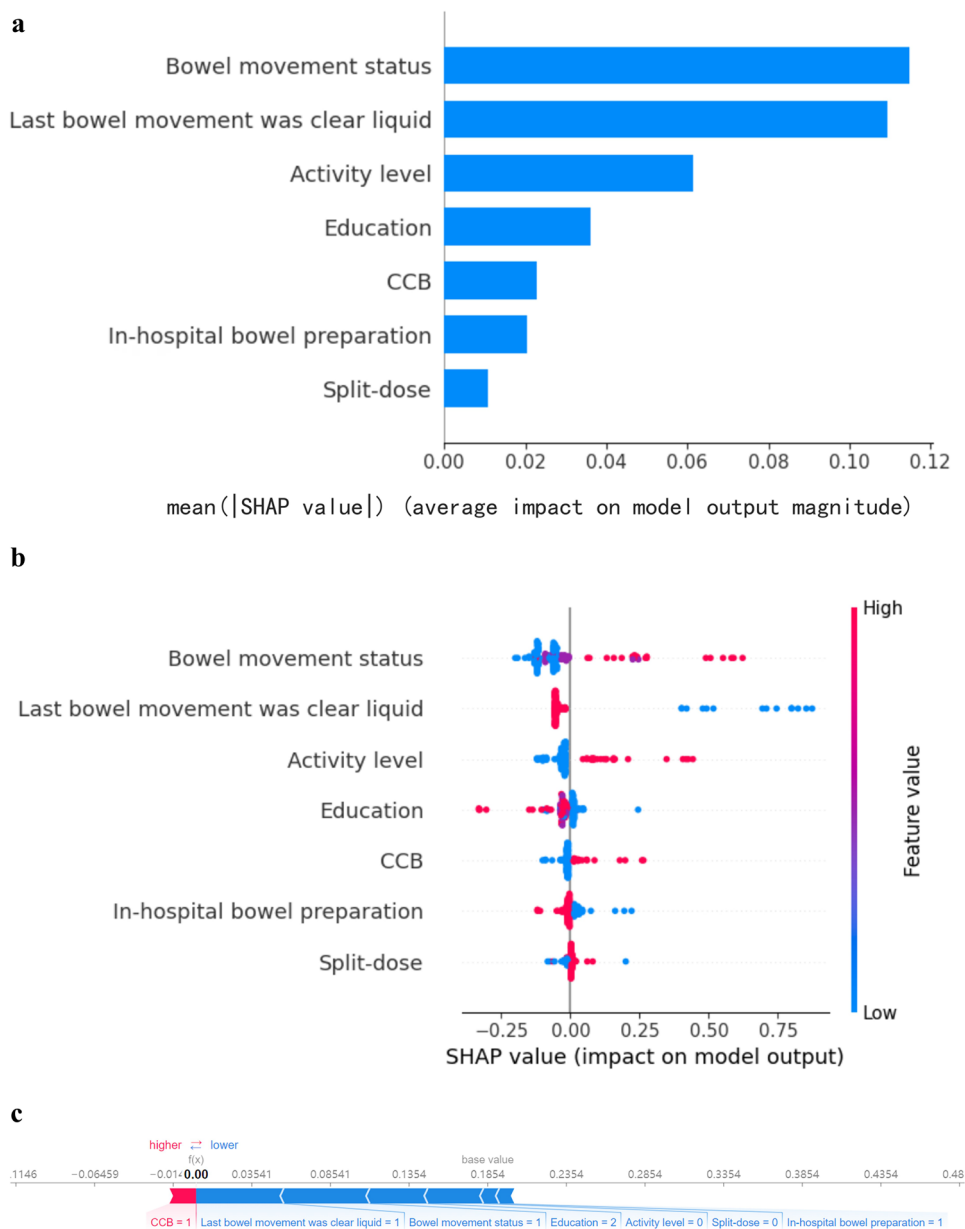


Figure 5. Feature importance and SHAP explanation plots. This figure consists of several parts demonstrating model interpretability: a: Bar Plot of Feature Importance, displaying the average SHAP values for each feature. b: SHAP Values Dot Plot, showing the impact of each feature on model output. c: Single Prediction Explanation Plot, visualizing the SHAP force plot for a single prediction.

patient and provides corresponding recommendations, as illustrated in [Supplementary Figure 3](#). This web application can be accessed online at <https://fv38cyrbas3kx3puethmad.streamlit.app/>.

Discussion

Adequate bowel preparation is essential for effective colonoscopy, and studies show that elderly patients are at high risk of inadequate preparation [30,31]. To improve the efficiency of colonoscopies and reduce the need for repeat procedures, this study developed a machine learning-based model to predict bowel preparation outcomes in elderly patients before colonoscopy.

Compared to traditional statistical regression models, machine learning captures complex nonlinear relationships between predictors, handles high-dimensional variables and their interactions, and offers precise personalized predictions [32,33]. Besides the logistic regression (LR) model, this study used various machine learning models (LightGBM, SVM, DT, RF, XGBoost) to predict the risk of bowel preparation failure in elderly patients. The results showed that, compared to the LR model, the machine learning-based models achieved higher accuracy, sensitivity, and specificity. The SVM-based model performed the best in internal validation, with an AUC of 0.895 (95% CI: 0.822–0.969), accuracy of 0.889, sensitivity of 0.739, and specificity of 0.932, notably exceeding the performance of previous studies (0.895 VS 0.80 [8], 0.895 VS 0.77 [9], 0.895 VS 0.73 [10], 0.895 VS 0.70 [11], 0.895 VS 0.63 [34], 0.895 VS 0.7835). Thus, the predictive model for bowel preparation failure in elderly patients could serve as a valuable screening tool to identify patients at high risk of failure, aiding clinical decision-making and improving bowel preparation quality in this population.

The selection of predictors is crucial in predictive model development. Previous models often used step-wise regression or empirical selection to identify variables [8–11,34–36]; we chose to use the Boruta algorithm for feature selection. Boruta is a robust feature selection method that automatically identifies and selects the most important features, reducing redundancy and enhancing model generalizability [27,37]. Our findings indicate that factors such as incremental laxative use, in-hospital bowel preparation, educational level, use of calcium channel blockers, bowel movement status, and activity level are significant predictors of bowel preparation quality in elderly patients, commonly selected as parameters in previous models [7,12]. In addition, we introduced a novel predictor “last bowel movement as clear liquid”.

Previous studies have demonstrated that the characteristics of the last bowel movement, including its color and consistency, are closely related to bowel preparation quality, as they reflect the degree of colonic cleansing [38,39]. Notably, two meta-analyses have confirmed that brown rectal effluent is a strong predictor of inadequate bowel preparation [13,40]. Our findings further support this association; incorporating last bowel movement characteristics into our model significantly improved its performance (0.803 VS 0.889, $\Delta\text{AUC} = 0.085$).

We further elucidated the decision-making process of the model using SHAP values, which not only identified the features with the greatest impact on predictions but also enhanced model transparency and interpretability. The SHAP model quantified the contribution of each feature to the predictions, crucial for explaining the “black box” of machine learning. This interpretability helps enhance clinicians’ understanding and trust in the model predictions, guiding further clinical research and interventions.

While our study primarily focuses on predicting bowel preparation outcomes, it is important to acknowledge that this approach emphasizes associations between factors and outcomes, rather than establishing causal relationships. Understanding causality, though beyond the scope of this work, is essential for advancing clinical decision-making [41]. Future research incorporating causal analysis could provide deeper insights into the mechanisms underlying bowel preparation failure, offering opportunities for targeted interventions.

Multi-model fusion, such as stacking or bagging, can enhance prediction accuracy by combining multiple algorithms but often reduces interpretability and clinical applicability. In this study, the SVM model achieved a high AUC of 0.889 in the validation set, demonstrating strong predictive performance. Given the trade-off between accuracy and simplicity, ensemble models were not employed. Future work may explore multi-model fusion to optimize performance, particularly in scenarios where accuracy is paramount.

In addition, to enhance the clinical usability of the model, we developed a web-based application using the Streamlit framework, improving its accessibility and practicality for daily use by healthcare professionals. Through an interactive web interface, clinicians can input relevant patient characteristics (e.g. education level, use of calcium channel blockers, and usual bowel movement patterns) to quickly generate individualized predictions of the risk of bowel preparation failure. For patients identified as high-risk by the model, targeted interventions can be implemented, such as administering additional doses of oral laxatives or performing

enemas, to reduce the likelihood of inadequate bowel preparation. These interventions aim to minimize the need for repeat colonoscopies, alleviate the financial burden on patients, and reduce the consumption of healthcare resources. Furthermore, the tool can also serve as a patient education resource, allowing patients to access it online and visualize their personalized risk. This feature enhances patients' understanding of and adherence to bowel preparation protocols, ultimately improving the quality of bowel preparation and increasing the success rate of colonoscopy procedures.

Our study population primarily consisted of elderly patients with chronic conditions such as hypertension, coronary artery disease, diabetes, and hyperlipidemia. These comorbidities are common among elderly populations worldwide, suggesting that our findings may be generalizable to similar cohorts. Additionally, our center implemented a standardized bowel preparation protocol, including a 3L polyethylene glycol regimen and dietary restrictions prior to laxative use. These protocols align with widely accepted clinical practices, further enhancing the potential applicability of our model in comparable clinical settings.

However, there are several limitations to this study. First, certain factors, such as the use of opioid medications and tricyclic antidepressants, reported to influence bowel preparation adequacy [8,14–16], were not present in our study population, as none of the patients were prescribed these medications. This absence may limit the applicability of our model to populations where such medications are more commonly used. Second, this was a single-center study conducted with patients from Xiamen and surrounding areas, which may restrict the model's generalizability. Third, certain variables in the study relied on self-reported data from patients, which could introduce recall bias. Lastly, although the study included single-center external validation, future research will focus on collecting multi-center data and utilizing larger sample sizes for external validation to further assess the model's stability and generalizability. Additionally, when there are significant differences between the study population and implementation setting compared to our cohort, it is essential to conduct thorough external validation and localization adjustments to avoid blindly applying the model, which may lead to inappropriate clinical decisions.

Conclusions

This study successfully developed an interpretable machine learning model to predict bowel preparation

adequacy in elderly patients before colonoscopy and created a corresponding web application. By applying this model in clinical settings, it is expected to accurately identify patients at high risk of preparation failure and enable early interventions, thereby substantially improving the success rate of bowel preparation in elderly patients and reducing medical costs.

Acknowledgements

Jiaying Liu, Wei Jiang and Yahong Yu conducted the partial data analyses and wrote the initial manuscript. Jiali Gong, Guie Chen and Yuxing Yang participated in the revision of the manuscript and the follow-up of partial data. Xuefeng Lu, Dalong Sun and Chao Wang revised the data analyses and the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

This study was approved by the Ethics Committee of Zhongshan Hospital (Xiamen), Fudan University (Approval Number: B2023-049R) and conducted in accordance with the principles of the Declaration of Helsinki. The model was developed using retrospective data collected from February to December 2023, and external validation was performed using additional data from March to June 2024. As the study was conducted within the validity period of the original ethical approval and involved only retrospective observational data, the Ethics Committee of Zhongshan Hospital (Xiamen), Fudan University waived the requirement for informed consent.

Author contributions

CRedit: **Jiaying Liu**: Conceptualization, Data curation, Formal analysis, Methodology, Project administration, Software, Validation, Visualization, Writing – original draft, Writing – review & editing; **Wei Jiang**: Formal analysis, Funding acquisition, Resources, Supervision, Writing – original draft; **Yahong Yu**: Data curation, Investigation, Supervision, Validation, Writing – original draft; **Jiali Gong**: Data curation, Formal analysis, Investigation, Supervision, Validation; **Guie Chen**: Data curation, Investigation, Validation, Visualization; **Yuxing Yang**: Data curation, Investigation, Supervision, Validation; **Chao Wang**: Data curation, Funding acquisition, Resources, Writing – review & editing; **Dalong Sun**: Data curation, Investigation, Project administration, Resources, Supervision, Writing – review & editing; **Xuefeng Lu**: Conceptualization, Data curation, Investigation, Project administration, Resources, Supervision, Visualization, Writing – original draft, Writing – review & editing.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the Xiamen municipal science and technology plans under Grant No.3502Z20214ZD1078; Fujian Provincial Natural Science Foundation Project under Grant No.2021d033; Fujian provincial medical innovation project under Grant No. 2022cxb020; and Xiamen key medical and health project under Grant No. 3502z20234006.

Data availability statement

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

References

- [1] Dekker E, Tanis PJ, Vleugels JLA, et al. Colorectal cancer. *The Lancet*. 2019;394(10207):1467–1480. doi: [10.1016/S0140-6736\(19\)32319-0](https://doi.org/10.1016/S0140-6736(19)32319-0).
- [2] Ladabaum U, Dominitz JA, Kahi C, et al. Strategies for colorectal cancer screening. *Gastroenterology*. 2020;158(2):418–432. doi: [10.1053/j.gastro.2019.06.043](https://doi.org/10.1053/j.gastro.2019.06.043).
- [3] Kluge MA, Williams JL, Wu CK, et al. Inadequate Boston Bowel Preparation Scale scores predict the risk of missed neoplasia on the next colonoscopy. *Gastrointest Endosc*. 2018;87(3):744–751. doi: [10.1016/j.gie.2017.06.012](https://doi.org/10.1016/j.gie.2017.06.012).
- [4] Millien VO, Mansour NM. Bowel preparation for colonoscopy in 2020: a look at the past, present, and future. *Curr Gastroenterol Rep*. 2020;22(6):28. doi: [10.1007/s11894-020-00764-4](https://doi.org/10.1007/s11894-020-00764-4).
- [5] Johnson DA, Barkun AN, Cohen LB, et al. Optimizing adequacy of bowel cleansing for colonoscopy: recommendations from the US multi-society task force on colorectal cancer. *Am J Gastroenterol*. 2014;109(10):1528–1545. doi: [10.1038/ajg.2014.272](https://doi.org/10.1038/ajg.2014.272).
- [6] Zhang YY, Niu M, Wu ZY, et al. The incidence of and risk factors for inadequate bowel preparation in elderly patients: a prospective observational study. *Saudi J Gastroenterol*. 2018;24(2):87–92. doi: [10.4103/sjg.SJG_426_17](https://doi.org/10.4103/sjg.SJG_426_17).
- [7] Gandhi K, Tofani C, Sokach C, et al. Patient characteristics associated with quality of colonoscopy preparation: a systematic review and meta-analysis. *Clin Gastroenterol Hepatol*. 2018;16(3):357–369 e310. doi: [10.1016/j.cgh.2017.08.016](https://doi.org/10.1016/j.cgh.2017.08.016).
- [8] Gu F, Xu J, Du L, et al. The machine learning model for predicting inadequate bowel preparation before colonoscopy: a multicenter prospective study. *Clin Transl Gastroenterol*. 2024;15(5):e00694. doi: [10.14309/ctg.0000000000000694](https://doi.org/10.14309/ctg.0000000000000694).
- [9] Dik VK, Moons LMG, Hüyük M, et al. Predicting inadequate bowel preparation for colonoscopy in participants receiving split-dose bowel preparation: development and validation of a prediction score. *Gastrointest Endosc*. 2015;81(3):665–672. doi: [10.1016/j.gie.2014.09.066](https://doi.org/10.1016/j.gie.2014.09.066).
- [10] Zhang N, Xu M, Chen X. Establishment of a risk prediction model for bowel preparation failure prior to colonoscopy. *BMC Cancer*. 2024;24(1):341. doi: [10.1186/s12885-024-12081-4](https://doi.org/10.1186/s12885-024-12081-4).
- [11] Gimeno-García AZ, Baute JL, Hernandez G, et al. Risk factors for inadequate bowel preparation: a validated predictive score. *Endoscopy*. 2017;49(6):536–543. doi: [10.1055/s-0043-101683](https://doi.org/10.1055/s-0043-101683).
- [12] Zhang Y, Wang L, Wu W, et al. Predictors of inadequate bowel preparation in older patients undergoing colonoscopy: a systematic review and meta-analysis. *Int J Nurs Stud*. 2024;149:104631. doi: [10.1016/j.ijnurstu.2023.104631](https://doi.org/10.1016/j.ijnurstu.2023.104631).
- [13] Beran A, Aboursheid T, Ali AH, et al. Risk factors for inadequate bowel preparation in colonoscopy: a comprehensive systematic review and meta-analysis. *Am J Gastroenterol*. 2024;119(12):2389–2397. doi: [10.14309/ajg.0000000000003073](https://doi.org/10.14309/ajg.0000000000003073).
- [14] Berger A, Cesbron-Métivier E, Bertrais S, et al. A predictive score of inadequate bowel preparation based on a self-administered questionnaire: PREPA-CO. *Clin Res Hepatol Gastroenterol*. 2021;45(4):101693. doi: [10.1016/j.clinre.2021.101693](https://doi.org/10.1016/j.clinre.2021.101693).
- [15] Mahmood S, Farooqui SM, Madhoun MF. Predictors of inadequate bowel preparation for colonoscopy: a systematic review and meta-analysis. *Eur J Gastroenterol Hepatol*. 2018;30(8):819–826. doi: [10.1097/MEG.0000000000001175](https://doi.org/10.1097/MEG.0000000000001175).
- [16] Yadlapati R, Johnston ER, Gregory DL, et al. Predictors of inadequate inpatient colonoscopy preparation and its association with hospital length of stay and costs. *Dig Dis Sci*. 2015;60(11):3482–3490. doi: [10.1007/s10620-015-3761-2](https://doi.org/10.1007/s10620-015-3761-2).
- [17] Vittinghoff E, McCulloch CE. Relaxing the rule of ten events per variable in logistic and Cox regression. *Am J Epidemiol*. 2007;165(6):710–718. doi: [10.1093/aje/kwk052](https://doi.org/10.1093/aje/kwk052).
- [18] [Chinese guideline for bowel preparation for colonoscopy (2019, Shanghai)]. *Zhonghua Nei Ke Za Zhi*. 2019;58(7):485–495.
- [19] Hassan C, East J, Radaelli F, et al. Bowel preparation for colonoscopy: european Society of Gastrointestinal Endoscopy (ESGE) Guideline - Update 2019. *Endoscopy*. 2019;51(8):775–794. doi: [10.1055/a-0959-0505](https://doi.org/10.1055/a-0959-0505).
- [20] Calderwood AH, Jacobson BC. Comprehensive validation of the Boston Bowel Preparation Scale. *Gastrointest Endosc*. 2010;72(4):686–692. doi: [10.1016/j.gie.2010.06.068](https://doi.org/10.1016/j.gie.2010.06.068).
- [21] Calderwood AH, Schroy PC, 3rd, Lieberman DA, et al. Boston Bowel Preparation Scale scores provide a standardized definition of adequate for describing bowel cleanliness. *Gastrointest Endosc*. 2014;80(2):269–276. doi: [10.1016/j.gie.2014.01.031](https://doi.org/10.1016/j.gie.2014.01.031).
- [22] Stoltzfus JC. Logistic regression: a brief primer. *Acad Emerg Med*. 2011;18(10):1099–1104. doi: [10.1111/j.1553-2712.2011.01185.x](https://doi.org/10.1111/j.1553-2712.2011.01185.x).
- [23] Ke G, Meng Q, Finley T, et al. LightGBM: a highly efficient gradient boosting decision tree. *Proceedings of the 31st International Conference on Neural Information Processing Systems*. Long Beach, California, USA: Curran Associates Inc.; 2017:3149–3157.
- [24] Noble WS. What is a support vector machine? *Nat Biotechnol*. 2006;24(12):1565–1567. doi: [10.1038/nbt1206-1565](https://doi.org/10.1038/nbt1206-1565).
- [25] Song YY, Lu Y. Decision tree methods: applications for classification and prediction. *Shanghai Arch Psychiatry*.

- 2015;27(2):130–135. doi: [10.11919/j.issn.1002-0829.215044](https://doi.org/10.11919/j.issn.1002-0829.215044).
- [26] Breiman L. Random forests. *Machine Learning*. 2001;45(1):5–32. doi: [10.1023/A:1010933404324](https://doi.org/10.1023/A:1010933404324).
- [27] Speiser JL, Miller ME, Tooze J, et al. A comparison of random forest variable selection methods for classification prediction modeling. *Expert Syst Appl*. 2019;134:93–101. doi: [10.1016/j.eswa.2019.05.028](https://doi.org/10.1016/j.eswa.2019.05.028).
- [28] Lundberg S, Lee S. A unified approach to interpreting model predictions. 31st conference on neural information processing systems (NIPS 2017). CA, USA; 2017.
- [29] Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ*. 2024;385:e078378. doi: [10.1136/bmj-2023-078378](https://doi.org/10.1136/bmj-2023-078378).
- [30] McNabb-Baltar J, Dorreen A, Al Dhahab H, et al. Age is the only predictor of poor bowel preparation in the hospitalized patient. *Can J Gastroenterol Hepatol*. 2016;2016:1–5. doi: [10.1155/2016/2139264](https://doi.org/10.1155/2016/2139264).
- [31] Ho SB, Hovsepian R, Gupta S. Optimal bowel cleansing for colonoscopy in the elderly patient. *Drugs Aging*. 2017;34(3):163–172. doi: [10.1007/s40266-017-0436-z](https://doi.org/10.1007/s40266-017-0436-z).
- [32] Deo RC. Machine learning in medicine. *Circulation*. 2015;132(20):1920–1930. doi: [10.1161/CIRCULATIONAHA.115.001593](https://doi.org/10.1161/CIRCULATIONAHA.115.001593).
- [33] Handelman GS, Kok HK, Chandra RV, et al. eDoctor: machine learning and the future of medicine. *J Intern Med*. 2018;284(6):603–619. doi: [10.1111/joim.12822](https://doi.org/10.1111/joim.12822).
- [34] Hassan C, Fuccio L, Bruno M, et al. A predictive model identifies patients most likely to have inadequate bowel preparation for colonoscopy. *Clin Gastroenterol Hepatol*. 2012;10(5):501–506. doi: [10.1016/j.cgh.2011.12.037](https://doi.org/10.1016/j.cgh.2011.12.037).
- [35] Fuccio L, Frazzoni L, Spada C, et al. Factors that affect adequacy of colon cleansing for colonoscopy in hospitalized patients. *Clin Gastroenterol Hepatol*. 2021;19(2):339–348.e337. doi: [10.1016/j.cgh.2020.02.055](https://doi.org/10.1016/j.cgh.2020.02.055).
- [36] Fostier R, Tziatzios G, Facciorusso A, et al. Models and scores to predict adequacy of bowel preparation before colonoscopy. *Best Pract Res Clin Gastroenterol*. 2023;67:101859. doi: [10.1016/j.bpg.2023.101859](https://doi.org/10.1016/j.bpg.2023.101859).
- [37] Kursu MB, Jankowski A, Rudnicki WR, Boruta – A system for feature selection. *Fundamenta Informaticae*. 2010;101(4):271–285. doi: [10.3233/FI-2010-288](https://doi.org/10.3233/FI-2010-288).
- [38] Fatima H, Johnson CS, Rex DK. Patients' description of rectal effluent and quality of bowel preparation at colonoscopy. *Gastrointest Endosc*. 2010;71(7):1244–1252.e1242. doi: [10.1016/j.gie.2009.11.053](https://doi.org/10.1016/j.gie.2009.11.053).
- [39] Shin SY, Ga KS, Kim IY, et al. Predictive factors for inadequate bowel preparation using low-volume polyethylene glycol (PEG) plus ascorbic acid for an outpatient colonoscopy. *Sci Rep*. 2019;9(1):19715. doi: [10.1038/s41598-019-56107-5](https://doi.org/10.1038/s41598-019-56107-5).
- [40] Feng L, Guan J, Dong R, et al. Risk factors for inadequate bowel preparation before colonoscopy: A meta-analysis. *J Evid Based Med*. 2024;17(2):341–350. doi: [10.1111/jebm.12607](https://doi.org/10.1111/jebm.12607).
- [41] Zhang Z, Jin P, Feng M, et al. Causal inference with marginal structural modeling for longitudinal data in laparoscopic surgery: A technical note. *Laparoscopic, Endoscopic and Robotic Surgery*. 2022;5(4):146–152. doi: [10.1016/j.lers.2022.10.002](https://doi.org/10.1016/j.lers.2022.10.002).