Peer Review File

Article information:

Reviewer A

Comment 1: Abstract: Data size is quite small for machine learning model, so not sure if the model was trained before on a larger dataset or if this is a training set? Stated accuracy of 81% needs a confidence interval…

Reply 1: Thank you for highlighting this concern. I want to clarify that our model was not trained on a specific dataset of a predetermined size. Given the rarity of the disease, the sample size is indeed small, and we have duly acknowledged this limitation in our documentation. I'm pleased to inform you that we have taken your feedback into account, and we've now incorporated a confidence interval for the reported accuracy of 81% (95% CI, 79.17% to 83.19%). This addition aims to provide a more nuanced understanding of the model's performance, considering the limitations associated with the dataset size. (See Page 12, line202)

Changes in the text: Using five features (standardized colon size (SCS) ratios of sigmoid colon, descending colon, transverse colon, rectum, and rectosigmoid), the model demonstrated an accuracy of 81% (95% CI, 79.17% to 83.19%).

Comment 2: Line 78: Please make sure manuscript is free of grammatical or spelling mistakes such as that one "refractorily" Line 78: "whole sentence need to be revised" …e.g., Unfortunately, up to 30% of children do not respond to 3-months of optimal laxative therapy, at which point they are considered to have refractory constipation. also where does it say that 1/3 of children will have refractory constipation in this reference?

Reply 2: Thank you for your careful review and constructive feedback. We appreciate your attention to detail. First, we have carefully reviewed the manuscript and addressed the grammatical error. Second, we have revised the sentence as follows 'Despite optimal conventional treatments, up to one-third of the children remain unresponsive for more than 3 months, at which point they are considered to have refractory constipation.' Third, Regarding the statement about 'Intractable constipation' (IC) was defined as that unresponsive to optimal conventional treatment for at least 3 months. we have provided the reference as follows 'Tabbers MM, DiLorenzo C, Berger MY, Faure C, Langendam MW, Nurko S, et al. Evaluation and treatment of functional constipation in infants and children: evidence-based recommendations from ESPGHAN and NASPGHAN. J Pediatr Gastroenterol Nutr. (2014) 58(2):258-274. doi:10.1097/MPG.0000000000000266' Thank you for the suggestion. (See Page 06, line 70)

Changes in the text: Despite optimal conventional treatments, up to one-third of the children remain unresponsive for more than three months, at which point they are considered to have refractory constipation.

Comment 3: line 122: recoverable functional constipation. I have never heard that term before…Maybe just say episodic functional constipation and refractory constipation.

Reply 3: Thank you for the suggestion, we have changed the wording from "recoverable functional constipation" to "functional constipation". (See Page 07, line 100)

Changes in the text: However, no standard has been set up for the early identification of functional

constipation and refractory constipation.

Comment 4: Line 130…Confused as to how this was a retrospective study yet patients were able to give written consent for BE…Do you mean written consent for the procedure as part of clinical care? Did they agree to obtain demographics and clinical features before your study or after your study started? If before your study started, then did you enroll patients prospectively?

Reply 4: Thank you for your thoughtful questions. We appreciate the opportunity to clarify the details that as barium enema is an invasive procedure, parental informed consent is required, along with obtaining permission for the use of imaging data. The study was reviewed and approved by the hospital's institutional review board for using medical data. Thank you for your valuable feedback.

Comment 5: Line 146: did group 1 meet ROME IV criteria? Also, which ROME criteria did you use? There is one for toilet trained and one for non toilet trained. Second group: met ROME IV criteria, so is just meeting ROME IV criteria is your definition of refractory constipation? Is this the accurate definition of refractory constipation? Patients who had persisting symptoms after 2 weeks had BE….is this a third group, or is this your entire population? I am confused by the description of study groups. Perhaps a graphical representation of study groups would help.

Reply 5: Thank you for pointing this out. We have clarified this in the methods section of the study groups (see Page 08, line 128)

Changes in the text: Due to the retrospective nature of the study, children grouping was based on the patient's medical records. Children were divided into three groups in total.

Patients with functional constipation: Children who met the Rome IV diagnostic criteria of functional constipation, and after receiving conventional therapy, symptoms gradually improved.

Patients with refractory constipation: Children who met the Rome IV diagnostic criteria of functional constipation, but showed no response to optimal conventional treatments for more than 3 months as specified in the ESPGHAN-NASPGHAN functional constipation guidelines.

Controls consisted of children who attended the clinic with symptoms of abdominal pain and bloating, after thorough examination by different tests, including barium enema, these children were last diagnosed with functional abdominal pain. These children were considered as the non-constipation group.

Comment 6: Please remove p values, meaningless in medical research. Provide us with point estimates (e.g., proportion with a 95% CI around the point estimate). E.g., line 221, duration of constipation was different, what was the difference? What is the certainty around that difference (CI)? Table 1: when conducting multiple hypothesis testing, you must adjust for multi-comparisons with a statistical correction. This is why I care more about difference in duration in months with a CI estimate.

Reply 6: Thank you for your feedback and careful review of our manuscript. We appreciate your time and valuable insights. Regarding your suggestion to remove P values, we would like to respectfully express our perspective. P values play a crucial role in statistical analysis, providing a quantitative measure of the evidence against a null hypothesis. While we acknowledge ongoing discussions within the scientific community about the interpretation and misuse of P values, they remain a widely accepted and standardized statistical tool in medical research [1-3]. In accordance with best practices, we understand your concerns and we have added 95% confidence intervals (CI)

for relevant measures in our results of Clinical Characteristics of Patient Population. And we have included 95% confidence intervals (CI) in table 1. Besides, in terms of the duration of constipation, we only conduct comparison between functional constipation group and refractory constipation group. (See Page 12, line 200)

[1]Maassel NL, Graetz E, Schneider EB, Asnes AG, Solomon DG, Leventhal JM. Hospital Admissions for Abusive Head Trauma Before and During the COVID-19 Pandemic. JAMA Pediatr. 2023;177(12):1342-1347. [2]Mir IN, Sánchez-Rosado M, Reis J, et al. Impact of fetal inflammatory response on the severity of necrotizing enterocolitis in preterm infants. Pediatr Res. doi:10.1038/s41390-023-02942-2[3]Regan AK, Moore HC, Binks MJ, et al. Maternal Pertussis Vaccination, Infant Immunization, and Risk of Pertussis. Pediatrics. 2023;152(5):e2023062664. doi:10.1542/peds.2023-062664

Changes in the text: The duration of constipation has a statistically difference between the functional group (14, 95% CI: 14.1 to 18.7) and the refractory constipation groups (22, 95% CI: 17.84 to 23.68) (*P*=0.039). And Table 1.

Comment 7: This is yet a validated tool, so not ready for clinical applications (discussion lines 290s) and you shouldn't be talking about that. Focus on next steps to validate this tool.

Reply 7: Thank you for your thoughtful review. We acknowledge that the sample size is a critical aspect of our study. To address this concern, we are currently exploring options to increase the sample size. We have added the following to the limitation (see Page 19, line 361)

Changes in the text: while our study stands as the largest of its kind, it is important to acknowledge its limitations, including a small sample size and a single-center design. Additionally, the absence of external validation is a notable constraint. We plan to extend our analysis to include other medical datasets for comprehensive validation, thereby enhancing the robustness and generalizability of our model.

Comment 8: Please remove abbreviations in table and type them out, very hard to read otherwise.

Reply 8: Thank you for pointing this out. We have type out the full names instead of the abbreviations. (See Page 31, line 608)

Comment 9: Figure 2, label the figures not with letters, but with features (Rectum). What is the error bar on top? SD? SE?

Reply 9: Thank you for bringing this up. We have labeled the Figure 2 with features and the error bar on the top is SD. And we have described below figure 2. (See Page 31, line 627)

Comment 10: In the discussion explain why ML results would be different than if you did a regression analysis with refractory constipation as your outcome?

Reply 10: Thank you for your insightful comments. We appreciate your suggestion to explain why the results from our machine learning (ML) approach might differ from those obtained through a traditional regression analysis, particularly with refractory constipation as the outcome. In our discussion, we have provided a thorough explanation of the unique strengths and potential advantages of using ML. (see Page 17, line 327)

Changes in the text: In recent years, the utilization of ML algorithms has become widespread for developing disease prediction models (36). Compared with conventional statistical methods, ML excels in accurately predicting outcomes across diverse datasets through high-level computing,

constructing algorithms for automated data-driven predictions or classifications (37). Linear regression assumes a linear relationship between the input features and the target variable, which may not capture the underlying patterns in our dataset adequately (38) … In this study, the model establishment process involved randomly dividing all patients into a training set and a test set at a ratio of 7:3. The training set was used to establish the model, and the model underwent validation on the test set subsequently. ANOVA-F was used to determine the characteristic features. SVM was used to model the data in the training set. Additionally, three-fold cross-validation was conducted to assess model stability and determine whether the model was over-fitted. In our specific case, the decision to use ML model was driven by the complex and non-linear relationships inherent in our data. Among various ML techniques, SVM stands out for its notable accuracy and efficiency, particularly in small datasets medical applications (40). SVM is a versatile algorithm that can effectively model non-linear relationships through the use of kernel functions. This flexibility allows SVM to capture intricate patterns and dependencies in the data, making it more suitable for our particular problem (41). Furthermore, research literature consistently highlights the remarkable accuracy of the SVM model in the medical diagnosis (42).

Comment 11: In the discussion, tell the reader about other AI utility that might be interesting to use, e.g., feature extraction. Why rely on humans to extract colonic diameters and standardize them?

Reply 11: Thank you for your insightful comment. We appreciate the opportunity to clarify the approach we took in our study. In the discussion section, we have provided an explanation of why we relied on standardize colon ratios rather than utilizing image analysis. (See Page 18, line 332)

Changes in the text: For image extraction tasks, Convolutional Neural Networks (CNNs) are highly effective and widely used in artificial neural networks designed specifically for image recognition and processing (39). In light of these advancements, we have made attempts to train a CNN model, but the limited sample size raised concerns about potential overfitting in the results. Human extraction of colonic diameters and standardization was chosen by the essential role of human expertise in accurately identifying and interpreting complex anatomical features, especially in children where anatomical structures can vary widely. In the future, we would extend our analysis to include other medical datasets for image extraction.

Specific comments from Reviewer B:

Comment 1: The aims could be more explicitly stated. Please, complete them.

Reply 1: Thank you for your feedback. We appreciate your suggestion to provide more explicit statements of the aims. The main objectives of our study are as follows: (see Page 08, line 111)

Changes in the text: Hence, in this study, our objective is to investigate the distinctive features of colonic anatomy in children with non-constipation, constipation, and refractory constipation. We also aim to utilize the ML model for the early identification of functional constipation and refractory constipation in children to support health practitioners in the management of refractory constipation.

Comment 2: Expand the background on pathophysiology of refractory constipation. It is too short.

Reply 2: Thank you for bringing this up. We have expanded the background on pathophysiology of refractory constipation in the introduction. We write in (see Page 06, line 79)

Changes in the text: However, for children with refractory constipation who do not respond to treatment, the situation may be substantially worse (10). This prolonged state of constipation and fecal loading can contribute to the development of colonic dilation (11). Studies indicate that dilated

segments exhibit a reduced number of high-amplitude propagating contractions and disruptions in the postprandial cyclic propagating motor pattern, causing dysfunction in colonic motility (12-14). Consequently, the dilated colon tends to reabsorb more water, resulting in the formation of hard stools and stretching of the colonic lumen (15).

Comment 3: Review more studies on radiologic features and machine learning in pediatric constipation.

Reply 3: Thank you for your inquiry regarding the existing research on radiologic features and machine learning in pediatric constipation. We have thoroughly searched PubMed for studies on radiologic features and machine learning in pediatric constipation. However, at present, there is limited research available on this specific subject.

Comment4: Provide more details on barium enema preparation and procedure.

Reply4: Thank you for bringing this up. We have added details in method section of barium enema. (See Page 09, line 152)

Changes in the text: The diagnostic testing of contrast enema was used to exclude anatomical abnormalities such as Hirschsprung's disease, an abnormally narrow, funnel-shaped rectum, and also a way to evaluate the morphology of the colon. The legal guardians of each participant provided written informed consent for taking BE. All patients underwent BE with bowel preparation before the X-ray examination. During BE, children were placed in the supine position with the assistance of their parents to ensure that they did not move during the procedure. Then, the anus was catheterized with a suitable caliber tube for inserting barium contrast (composed of barium sulfate and normal saline at a 1:5 ratio) into the rectum by using a syringe, The colon was examined in anterior, posterior, and lateral projections which allowed the pediatric radiologist to capture radiographic pictures and analyze the colonic anatomy. Measurements were performed by an experienced pediatric radiologist and cross-referenced by a pediatrician.

Comment 5: Explain how children were clinically classified into the 3 groups of the study.

Reply 5: Thank you for pointing this out. We have clarified this in the methods section of the study groups in line xx. (See Page 08, line 128)

Changes in the text: Due to the retrospective nature of the study, children grouping was based on the patient's medical records. Children were divided into three groups in total.

Patients with functional constipation: Children who met the Rome IV diagnostic criteria of functional constipation, and after receiving conventional therapy, symptoms gradually improved.

Patients with refractory constipation: Children who met the Rome IV diagnostic criteria of functional constipation, but showed no response to optimal conventional treatments for more than 3 months as specified in the ESPGHAN-NASPGHAN functional constipation guidelines.

Controls consisted of children who attended the clinic with symptoms of abdominal pain and bloating, after thorough examination by different tests, including barium enema, these children were last diagnosed with functional abdominal pain. These children were considered as the non-constipation group.

Comment 6: More information is needed on machine learning model development and validation.

Reply 6: Thank you for your feedback. We appreciate your interest in the machine learning model

development and validation process in our study. In our manuscript, we have provided a detailed account of our machine learning model development. For visual identification, we utilized Microsoft Excel and Python's Scikit-Learn library, employing a supervised machine learning script. The process involved data loading, dimensionality reduction using ANOVA F-values, and the use of a support vector machine (SVM) for classification. Moreover, we optimized various parameters, including the number of selected features, kernel type, C value, and gamma value through a grid search. The dataset underwent random division into training and test sets, allowing us to train the model and evaluate its accuracy. Additionally, we implemented k-fold cross-validation (k-fold value of 3) for a comprehensive assessment of the model's performance. The confusion matrix and decision boundaries were calculated using the parameters that yielded the best performance. Furthermore, we have reviewed several other machine learning studies, and our approach is consistent with the standard practices in the field [1-3].

[1] Tsai IJ, Shen WC, Lee CL, Wang HD, Lin CY. Machine Learning in Prediction of Bladder Cancer on Clinical Laboratory Data. Diagnostics (Basel). 2022;12(1):203. Published 2022 Jan 14. doi:10.3390/diagnostics12010203

[2] Feng M, Zhang J, Zhou X, et al. Application of an Interpretable Machine Learning Model to Predict Lymph Node Metastasis in Patients with Laryngeal Carcinoma. J Oncol. 2022;2022:6356399. Published 2022 Nov 12. doi:10.1155/2022/6356399

[3] Kim SH, Jeon ET, Yu S, et al. Interpretable machine learning for early neurological deterioration prediction in atrial fibrillation-related stroke. Sci Rep. 2021;11(1):20610. Published 2021 Oct 18. doi:10.1038/s41598-021-99920-7

Comment 7: Reduce the redundancy between tables with colon size data.

Reply 7: Thank you for your valuable feedback. We appreciate your time and effort in reviewing our work. Regarding your comment on reducing redundancy between tables with colon size data, we would like to seek clarification to ensure we address your concerns appropriately.

Could you please provide more details or specific examples to help us better understand the redundancy you are referring to? This will enable us to make targeted improvements and enhance the clarity of our tables.

The comparisons of the SCS ratios of the three groups were shown in table2. The paired-wise comparison results of functional constipation group and refractory constipation group; control group and functional constipation; control group and refractory constipation group were shown in table3. We are committed to refining our work based on constructive feedback, and your input is instrumental in this process. Thank you once again for your valuable insights, and we look forward to your response.

Comment 8: Include more visual representations of key data. It would help readers to understand the manuscript.

Reply 8: Thank you for your constructive feedback. We acknowledge the importance of visual aids in facilitating understanding. In response to your suggestion, we have provided several samples in supplementary information.

Comment 9: Clarify the order of presenting results.

Reply 9: Thank you for your feedback. We believe that our current presentation of results is clear and coherent. The order is presented as follows:

Clinical Characteristics of Patient Population:

Workflow for patient recruiting process (Figure 1).

Summary of clinical features in Table 1.

Patient categorization into control, functional constipation, and refractory constipation groups.

Statistical analysis of gender, age, and duration of constipation among the three groups.

Colon Size Ratios Analysis:

Measurement and calculation of SCS ratios of various colon segments.

Analysis of SCS ratios by gender and age, emphasizing the standardization of colon sizes.

Comparison of SCS ratios among the control, functional constipation, and refractory constipation groups.

Pairwise comparison within each group for specific colon segments.

Machine Learning Model of Constipation Classification:

Introduction of the SVM model for classifying patients.

Application of ANOVA F values to select input features.

Description of input features: gender, age, and SCS ratios of specific colon segments.

Details of the model development, including the comparison of kernel types and parameter optimization through grid search.

Presentation of model accuracy through K-fold cross-validation and confusion matrix.

Feature coefficients analysis, emphasizing the importance of SCS ratios in classification.

We hope this proposed structure aligns with your expectations.

Comment 10: Relate the key findings back to the aims that were stated in the introduction.

Reply 10: Thank you for pointing this out. According to your suggestion of comment 1, we have stated our aim explicitly and relate the key findings back to the aims in the first section of discussion. (See Page 14, line 256)

Changes in the text: In this study, we characterized the SCS ratios of colon segments in 205 children, and developed a supervised machine-learning model for earlier identification of refractory constipation in children. With an accuracy of 81%, we demonstrated that this model can effectively distinguish refractory constipation from functional constipation and non-constipation in children.

Comment 11: Discuss the limitations more thoroughly - small sample size, single center, …

Reply 11: Thank you for your thoughtful review. We acknowledge that the sample size is a critical aspect of our study. To address this concern, we are currently exploring options to increase the sample size. We have added the following to the limitation. (See Page 19, line 361)

Changes in the text: While our study stands as the largest of its kind, it is important to acknowledge its limitations, including a small sample size and a single-center design. Additionally, the absence of external validation is a notable constraint. We plan to extend our analysis to include other medical datasets for comprehensive validation, thereby enhancing the robustness and generalizability of our model.

Comment 12: Compare the findings to more prior studies on radiologic features.

Reply 12: Thank you for your inquiry regarding a comparison of our findings to prior studies on radiologic features. It is important to note that reports specifically addressing refractory constipation and colonic elongation are currently limited, we have reviewed several other radiologic features

studies, and the existing literature is discussed comprehensively within the context of our study.

Comment 13: Please, highlight the need for external validation of machine learning model.14. Discuss the potential clinical role and implications of the study.15. The conclusions about early identification/tailored management are not fully supported. Please, complete them.

Reply 13: Thank you for bringing this up. We have added details in conclusion section. (See Page 19, line 368)

Changes in the text: In this retrospective study, we described the characteristics of colonic anatomy in children with non-constipation, constipation, and refractory constipation. By building a supervised machine learning model, the three groups could be distinguished with 81% accuracy. For the future, we aim to use this model with additional medical datasets to further validate its effectiveness and expand its capabilities. The application of the machine learning model provides practical implications, which are significant for earlier identification and guiding management in children with refractory constipation. This enables clinicians to tailor management strategies, offering suggestion for further examinations or more radical treatment.

Reviewer B

Comment 1: Abstract
Please define CI in the abstract.
Reply 1: Thank you for pointing this out. We have defined CI in the abstract. (See Page 02, line 36).
Changes in the text: (95% confidence interval (CI), 79.17% to 83.19%).

Comment 2: Figure 2
a) Please define SCS in the legend.
b) Please check this letter.
Reply 2: Thank you for bringing this up. We have defined SCS in the legend and changed the letter 'ectum' to 'Rectum'. (See Page 23, line 477).
Changes in the text: Figure 2. SCS ratios of Rectum, Sigmoid colon, Descending colon, Transverse colon, Ascending colon, Rectosigmoid for the control group, functional constipation group and refractory constipation group. The error bar on the top of the bar charts represents the standard deviation (SD). SCS, standardized colon size.

Comment 3: Figure 3 and 4
Please define SVM in the legend.
Reply 3: Thank you for pointing this out. We have defined SVM in the legend. (See Page 23, line 483;485).
Changes in the text: SVM, support vector machine.

Comment 4: Figure S1
Please define SCS in the legend.
Reply 4 Thank you for bringing this up. We have defined SCS in the legend (see Page 24, line 491).
Changes in the text: SCS, standardized colon size.

Comment 5: Table 1

a) Please define CI in the table footnote.

b) Please provide the header of the first column.

Reply 5: Thank you for the suggestion. We have defined CI in the table footnote and provided the header of the first column. (See Page 22, line 455).

Changes in the text: CI, confidence interval

Comment 6: Table 2 and 3

Please provide the header of the first column.

Reply 6: Thank you for your insightful comments. We have provided the header of the first column in Table 2 and 3. (see Page 22 23, line 461 467).

Comment 7: Table S1 and S2

Please provide the header of the first column.

Reply 7: Thank you for bringing this up. We have provided the header of the first column in Table S1 and S2.

Comment 8: References/Citations.

Please add the citation for Koppen et al. at the end of the sentence.

Reply 8: Thank you for the suggestion. We have added the citation for Koppen et al. at the end of the sentence. (See Page 13, line 257).

Changes in the text: It may be due to the different control populations and imaging modalities: patients with intussusception and air contrast were used by Koppen et al (18).