




Detecting Patient Readiness for Colonoscopy Through Bowel Image Analysis: A Machine Learning Approach

Nour Kaouk¹ (✉), Lamis Amer², Tina Yaacoub¹, Youssef Bakouny¹, Chantal Hajjar¹, Flavia Khatounian¹, Joseph Amara^{1,2}, Rita Slim², Ali Mansour³ , and Cesar Yaghi^{1,2}

¹ Saint Joseph University of Beirut, Beirut, Lebanon

nour.kaouk1@net.usj.edu.lb

² Gastroenterology Service, Hotel Dieu de France, Beirut, Lebanon

³ Lab-STICC, UMR 6285, CNRS, ENSTA Bretagne, 29806 Brest, France

Abstract. Bowel preparation is a crucial step in ensuring the success and accuracy of colonoscopy procedures. Adequate bowel cleansing allows for better visualization and detection of abnormalities within the colon. In this study, we present an AI tool developed to assess the quality of bowel preparation in colonoscopy procedures. The dataset used in this study consists of 350 images of toilet bowls obtained from patients at the hospital “Hôtel Dieu de France” in Beirut, Lebanon. Their images are labeled by the professionals using the Boston scores. Our methodology involves a comprehensive pre-processing phase, encompassing detection, cropping, color adjustment, and Principal Component Analysis (PCA) on the image dataset. Subsequently, we applied different machine learning (ML) models for classification, achieving a high accuracy of 92% with Gradient Boosting. This AI-based approach exhibits great potential in enhancing the efficiency and reliability of colonoscopy evaluations, ultimately leading to improved patient outcomes and early detection of gastrointestinal disorders.

Keywords: Colonoscopy · Bowel preparation · Artificial intelligence · Classification · Machine learning

1 Introduction

Colonoscopy is of utmost importance for the screening and diagnosis of colorectal cancer, which is the third-leading cause of cancer death. The available evidence suggests that inadequate bowel preparation reduces the diagnostic yield of colorectal neoplasia and increases post-colonoscopy colorectal cancer risk [1]. Suboptimal bowel preparation has been shown to prolong the overall procedure time (e.g. increased time washing, and suctioning debris, prolonged withdrawal time), decrease the cecal intubation rate, and increase the risk of missing polyps or adenomas during the colonoscopy. It furthermore entails a shorter colonoscopy follow-up interval with a higher economic burden, therefore resulting in shorter surveillance intervals and increased costs [2]. Quality of bowel preparation has been associated with adenoma detection rates (ADR) in multiple

studies [3–5]. Similarly, a prospective multicenter randomized control trial published in 2022 included 413 patients, it reveals an ADR of 45.3% on repeat colonoscopy (on a median interval with the index colonoscopy of 28 days), an advanced ADR of 10.9%, and a serrated polyp detection rate of 14.3% [6]. A meta-analysis showed that, compared with low-quality bowel preparations, ADRs were significantly greater with intermediate (odds ratio [OR], 1.4; 95% CI, 1.1–1.8) and high-quality bowel preparations (OR, 1.4; 95% CI, 1.2–1.6) [3].

Usually, physicians rely on patients' subjective observations and descriptions to be informed about bowel preparation. Information obtained include the adherence of patients to colon cleansing guidance, and the quality of the stools in terms of color, translucency, and presence of particles. This approach often falls short in practice mainly because subjective self-assessment by patients regarding the quality of bowel preparation can be inconsistent and unreliable.

To assess the quality of bowel preparation through an advanced AI CNN model, Yan-Xing Hu et al. study [7] primarily focused on the medical approach, showcasing the potential implications and benefits of their innovative model, without delving into the technical aspects of their methodology. Previous articles in this field have predominantly emphasized the development of segmentation and classification algorithms for the detection of polyps in colonoscopy [8–10]. However, they did not explicitly address the significance of bowel preparation in the context of the colonoscopy procedure, which can be considered a potential drawback in their approaches.

To address this challenge, the proposal is to employ AI tools aimed at objectively assessing bowel preparation. This technology would allow to reduce the need for repeat procedures, saving both time and the fatigue experienced by patients. Our article aims to provide a comprehensive overview that not only highlights the medical significance of bowel preparation but also delves into the intricacies of our machine-learning model.

The rest of the paper is organized as follows to provide a clear structure and comprehensive understanding of the research. Section 2 highlights the materials and methods used to conduct this study. Section 3 focuses on the data pre-processing techniques utilized to mitigate variations in the dataset. In Sect. 4, the methodology employed for classification is described in detail. Finally, the conclusion is provided in Sect. 5, discussing the implications of the research in the field of classification.

2 Material and Methods

2.1 Image Collection and Dataset

We used a comprehensive dataset consisting of 350 images of toilet bowls obtained from patients at the hospital “Hôtel Dieu de France” in Lebanon. Patients were instructed to take pictures of the toilet bowl in the second half of their colon cleansing procedure. These images were anonymized and meticulously collected and curated to encompass a wide range of bowel preparation conditions. Figure 1 illustrates some pictures of the original dataset used in this study.

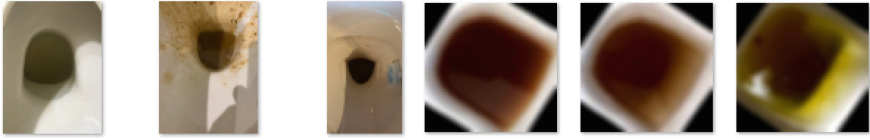


Fig. 1. Some figures of the original dataset before and after processing.

2.2 Bowel Preparation

A colonoscopy preparation begins by adjusting one's diet a few days ahead of his colonoscopy. Typically, the patient eats a low-fiber diet for two or three days, followed by a clear liquid diet on the last day. The afternoon or evening before the colonoscopy, he takes a laxative formula to purge the bowels. Studies have found that inadequate bowel preparation can lead to failed detection of cancerous lesions and are associated with an increased risk of procedural adverse events [11].

2.3 Colon Cleansing Classification

In the article, we have employed a comprehensive approach by incorporating one distinct scoring system namely the Boston scores. The Boston Bowel Preparation Scale (BBPS) is a medical scoring system used to evaluate the quality of a patient's bowel preparation before a colonoscopy. It assesses how effectively the bowel has been cleansed, with higher scores indicating better bowel cleanliness. Each segment of the colon, the right colon, transverse colon, and left colon are assigned points from 0 to 3 with regard to the cleanliness of the colon. A score of 0 includes an unprepared colon, 1 includes those in which only a portion of the mucosa of the colon segment is visible, and 2 includes those with a minor amount of residual staining and small fragments of stool present. Lastly, 3 includes those where the entire mucosa of the colon is seen well with no residual stool. The entire colon is assigned a cumulative score [12].

The dataset contains pictures of toilet bowls, which have been linked to the patients' colonoscopy tests with preparations classified into 10 different classes based on the Boston score. For classes 0, 1, 2, and 5 we have 0 images. For classes 3, 4, 6, 7, 8, and 9 we had 1, 3, 13, 25, 74, and 235 images, respectively. We split the dataset into two categories, namely 1 (good bowel preparation) for a Boston score of 9 and 0 (poor bowel preparation) corresponding to the other scores. The binarization of Boston scores helps doctors and medical professionals quickly identify the quality of preparation. Our team of doctors has validated the importance of this categorization in the medical context.

3 Pre-Processing and Data Preparation

The data pre-processing involved several techniques aimed at reducing the variations caused by different patient photographs taken with various camera phones, lighting conditions, and orientations.

Firstly, template matching approach to detect and crop the region of interest in the image dataset was applied to focus solely on the toilet and bowl area by computing

the maximum correlation value and its index. If the maximum correlation exceeds the empirical threshold of 0.6 the region of interest is cut off from the image.

Additionally, Histogram Stretching and Hamming's filter were employed to normalize the color profiles across the images, minimizing the impact of varying lighting conditions. The first algorithm expands or contracts the intensity values of the image to utilize the full dynamic range, enhancing the contrast and adjusting the color levels. The second method minimizes frequency leakage and distortion caused by sharp edges in the image. A filter size of 12 was chosen to be large enough to capture more details, smooth the image more, and reduce noise.

Finally, principal component analysis was used to orient the bowls in the same direction then, a rotation was applied with the calculated angle of rotation for each frame. The images were rotated by 45° . However, regardless of pictures orientation, the method consistently yields improved results.

Figure 1 also displays some images after the completion of all pre-processing steps carried out using MATLAB and Python.

4 Classification

Several pre-trained models, such as VGG16 and ResNet50, were tested to explore the potential use of deep learning in our specific use case. Unfortunately, a common challenge was encountered across these models: overfitting. The validation accuracy appeared to plateau, which is a known symptom of overfitting. This issue can be attributed to the relatively small sample size we have to work with, consisting of only 350 images that need to be divided into training and testing sets.

Machine learning (ML) algorithms were thus investigated. In such cases, where binary classification is appropriate, the Support Vector Machine (SVM) algorithm emerges as a good choice, as it succeeds at finding optimal decision boundaries between two classes. First, the categories are defined, and empty lists are initialized to store input and output arrays. Next, the images are resized to a 32×32 pixel size, and then the pixels are flattened into a 1D feature vector. This feature vector is used as the input to the SVM classifier. The model is trained using the training data that contains 80% of the data and its performance is evaluated on the testing data that comprises the remaining 20% of the data. The SVM classifier uses the default hyperparameter settings: The C (Regularization Parameter) is in its default value of 1. This parameter controls the trade-off between maximizing the margin and minimizing the classification error on the training data. The Kernel Type is used in its default value which is the radial basis function (RBF) kernel, which is commonly used for SVM classification. The Gamma (for RBF Kernel) parameter is 'scale', which is based on the inverse of the number of features. It controls the shape of the decision boundary.

To optimize the SVM's performance, we have considered tuning the hyperparameters by explicitly setting them and performing a hyperparameter search to find the best combination of hyperparameters for our specific dataset. To this purpose, we have used Grid Search technique in combination with cross-validation. We have obtained the best hyperparameters: {'C': 10, 'gamma': 'scale', 'kernel': 'rbf'} and achieved accuracy with best hyperparameters: 90.77%.

In our quest to find the most suitable machine learning model for our classification task, we tested four different models: Random Forest, Logistic Regression, Decision Tree, and Gradient Boosting. The Random Forest model and Logistic Regression model achieved an accuracy of approximately 87.7%, demonstrating good performance. The Decision Tree model achieved an accuracy of approximately 78%. While Decision Trees are interpretable, in this case, they had a lower accuracy compared to the other models. The Gradient Boosting model achieved the highest accuracy of approximately 92.3%.

5 Conclusion

The goal of this paper is to detect the colonoscopy preparation degree of the patient through image analysis, so after employing various pre-processing techniques to prepare our image dataset, we meticulously trained and tested our model and classified the images, achieving an accuracy of 92.3% with the Gradient Boosting model. This automated detection of bowel preparation quality addresses a medical need to help patients and physicians obtain the best quality during colonoscopy. Consequently, this innovative model will shorten the follow-up interval and reduce the economic burden, then, ensure a better medical service for the patients.

In the ongoing development of the bowel preparation assessment ML model, several areas for future work have been identified. First, there is the potential for the model to evolve into a clinical decision support system, aiding healthcare providers in real-time decision-making during colonoscopy procedures. Additionally, efforts may focus on enhancing patient education and engagement through user-friendly interfaces or mobile applications, which can positively impact patient compliance and preparation quality.

It is crucial to emphasize that the sample size of our experiment is limited, and the dataset contains a small number of images. This limitation arises from the fact that these images are obtained from real patients within a short time frame. As part of our ongoing efforts, we are committed to augmenting the number of images in our dataset over time.

References

1. Hernández G, Gimeno-García AZ, Quintero E (2019) Strategies to improve inadequate bowel preparation for colonoscopy
2. Sharma P, Burke CA, Johnson DA, Cash BD (2020) The importance of colonoscopy bowel preparation for the detection of colorectal lesions and colorectal cancer prevention.
3. Clark BT, Rustagi T, Laine L (2014) What level of bowel prep quality requires early repeat colonoscopy: Systematic review and meta-analysis of the impact of preparation quality on adenoma detection rate
4. Anderson JC, Butterly L, Robinson CM (2014) Impact of fair bowel prep on adenoma and serrated polyp detection: Data from the New Hampshire Colonoscopy Registry using a standardized preparation quality rating
5. Noh Hong S, Kyung Sung I, Hwan Kim J (2012) The effect of the bowel preparation status on the risk of missing polyp and adenoma during screening colonoscopy: A tandem colonoscopic study
6. Pantaleón Sánchez M, Gimeno Garcia A.-Z, Bernad Cabredo B (2022) Prevalence of missed lesions in patients with inadequate bowel preparation through a very early repeat colonoscopy

7. Bor Lu Y, Cun Lu S, Ning Huang Y (2022) A novel convolutional neural network model as an alternative approach to bowel preparation evaluation before colonoscopy in the COVID-19 era: A multicenter, single-blinded, randomized study
8. Wang L, Chen L, Wang X (2022) Development of a convolutional neural network-based colonoscopy image assessment model for differentiating crohn's disease and ulcerative colitis
9. Wen Y, Zhang L, Meng X (2022) Rethinking the transfer learning for FCN based polyp segmentation in colonoscopy
10. Wang Y, Feng Z, Song L (2021) Multiclassification of endoscopic colonoscopy images based on deep transfer learning
11. Patel N, Kashyap S, Mori A (2023) Bowel Preparation
12. Lai EJ, Calderwood AH, Doros G (2009) The Boston bowel preparation scale: a valid and reliable instrument for colonoscopy-oriented research