Nerthus: A Bowel Preparation Quality Video Dataset

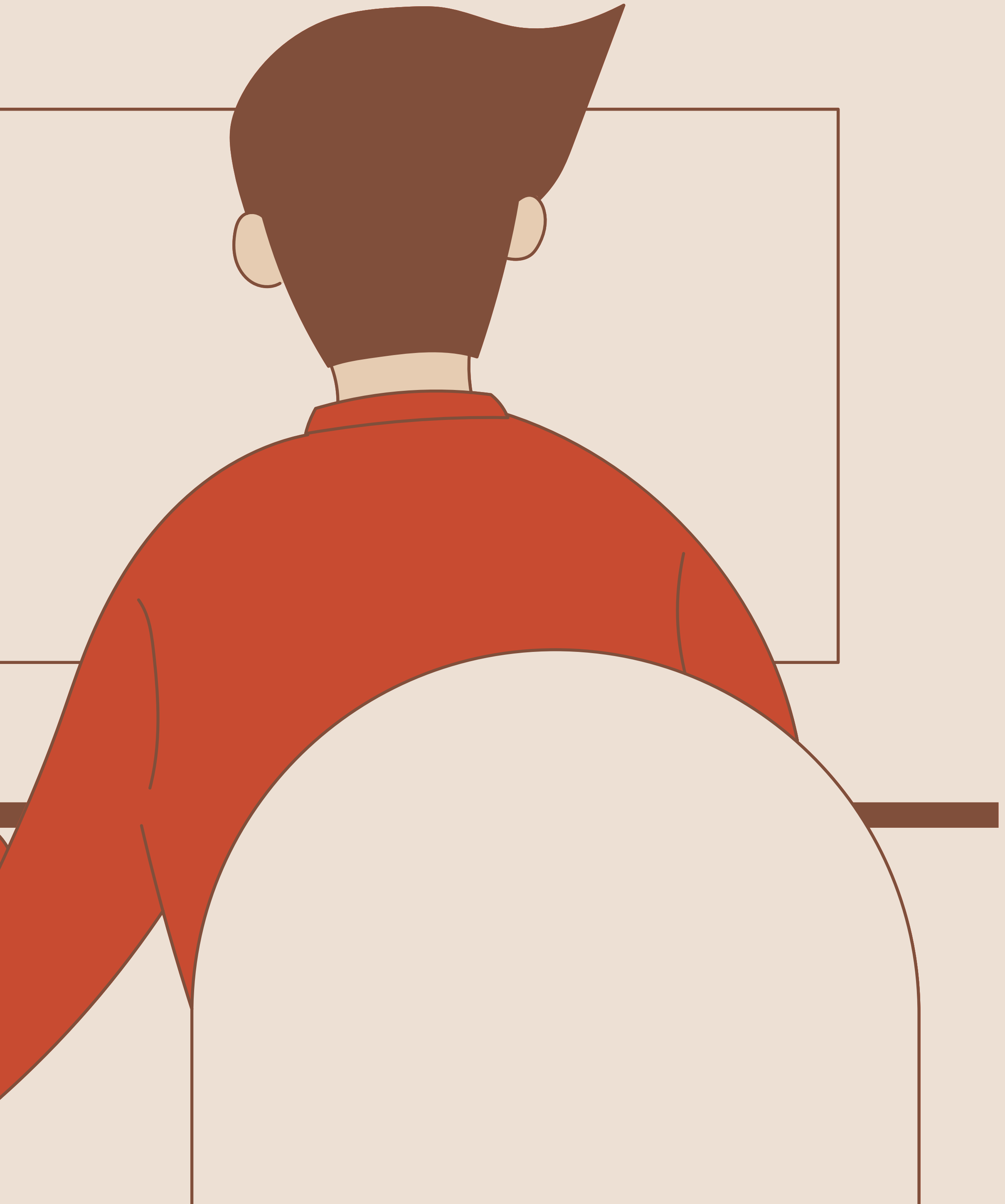
Advisor

ENG.Shady Nagy

Student

Youssef Mohamed Tawfik

Table of Contents



- Bowel preparation: Essential for a successful colonoscopy as it ensures proper visualization of the gastrointestinal (GI) tract during the examination.

- Colonoscopy significance: Used for detecting diseases such as colorectal cancer, the third most common cancer globally.

- Challenges with current assessment: Subjectivity in assessing bowel cleanliness can lead to different grading by doctors, which may impact screening and follow-up recommendations.

- Need for automation: An objective, automated system would standardize the assessment, reduce inequalities, and optimize medical resources.

- Dataset contents: The Nerthus dataset consists of 21 video clips and 5,525 frames annotated by experienced endoscopists.
- Bowel preparation quality classification: The dataset includes four classes based on bowel preparation quality, The number of frames per class varies from 500 to 2, 700, evaluated according to the Boston Bowel Preparation Scale (BBPS).
- Purpose of the dataset: To assist in developing automatic systems for evaluating bowel cleanliness and reducing human error in the process.

- Traditional classification: Previously, bowel preparation was categorized as poor, adequate, or good, but this lacked consistency.
- BBPS (Boston Bowel Preparation Scale): BBPS divides the bowel into three sections and uses a numeric scoring system (0-3) to evaluate the cleanliness of each section, providing more granularity than traditional methods.
- OBPS (Ottawa Bowel Preparation Scale): Another scale used for bowel preparation, it scores each bowel segment separately and adds a global score for the overall liquid residue in the bowel.

BBPS score points

score	Description
0	The colon section is completely blocked with solid stool, and the inner lining can't be seen.
1	Some of the inner lining of the colon is visible, but other parts are hard to see because of leftover stool or unclear liquid.
2	A small amount of stool or unclear liquid is present, but most of the colon's inner lining is visible.
3	The entire inner lining of the colon is clearly visible with no stool fragments or unclear liquid.

THIS STUDY USES BBPS, AS IT IS THE MOST WIDELY VALIDATED AND COMMONLY USED SCORING SYSTEM IN BOTH CLINICAL AND SCREENING ENVIRONMENTS TODAY [19]. THE BBPS SCALE IS DESIGNED TO EVALUATE CLEANLINESS AT WITHDRAWAL AND DOES NOT ACCOUNT FOR ANY ADDITIONAL CLEANSING PERFORMED BY THE ENDOSCOPIST, REFLECTING STANDARD COLONOSCOPY PRACTICE. THE SEGMENTAL SCORES FOR BBPS, RANGING FROM 0 (WORST) TO 3 (BEST)

- Source of data: The videos were recorded at Bærum Hospital, Norway, using high-definition colonoscopes.
- Expert annotation: The data was annotated by medical doctors, including experts from Norway, Sweden, the UK, the US, and Canada.
- Annotation process: The medical experts marked frames for bowel cleanliness levels, providing a reliable gold standard for future research and development.
- Use of ScopeGuide: The electromagnetic imaging system used in the colonoscopes helps provide accurate position information, which is essential for understanding bowel preparation quality.

- Video resolution: The dataset includes 720x576 resolution videos, with annotations sorted into folders based on BBPS scores.
- Data segmentation: The videos are segmented into shorter clips for easier use in machine learning tasks.
- Application in AI: The dataset is useful for image retrieval, machine learning, deep learning, and transfer learning applications, enabling development of automated colonoscopy assessment systems.

Key metrics

The evaluation of machine learning models should include metrics such as precision, recall, specificity, accuracy, F1 score, and frame-per-second (FPS) rate.

Precision and recall

These metrics assess how well the model detects correct bowel preparation classifications and minimizes false positives/negatives.

FPS

Important for real-time classification systems in medical settings to ensure timely assessments during colonoscopy procedures.

Classification performance

Method	PREC	REC	SPEC	ACC	MCC	F1	FPS
6 Layer CNN	0.856	0.852	0.952	0.854	0.854	0.854	42
3 Layer CNN	0.811	0.694	0.937	0.772	0.621	0.742	40
Inception v3 TFL	0.751	0.745	0.918	0.748	0.665	0.674	61
2 GF Random Forrest	0.792	0.774	0.849	0.847	0.874	0.766	310
2 GF Logistic Model Tree	0.744	0.737	0.862	0.825	0.594	0.590	200
6 GF Random Forrest	0.885	0.866	0.893	0.913	0.860	0.860	610
6 GF Logistic Model Tree	0.901	0.960	0.960	0.940	0.863	0.899	77
Baseline (JCD Random Forrest)	0.805	0.794	0.870	0.861	0.679	0.679	330
Baseline (Random/Majority)	0.240	0.489	0.512	0.652	0.000	0.000	-

In medical and computer science research, various metrics like recall, precision, specificity, accuracy, MCC, and F1 score are used with different names. Future research should include these metrics and dataset details. The best-performing method is the 6 GF LMT approach with an F1 score of 0.899. Deep learning methods performed lower due to unoptimized parameters. These results serve as useful baselines for future research and method benchmarking.

Global Features (GF)

extracte image features using the Lire software, including JCD, Tamura, Color Layout, and others. In two GF runs, we combined different features into vectors of 187 and 1186 values. The features were fused early and stored in ARFF format. Classification was done using Random Forest and Logistic Model Tree classifiers from Weka.

CNN

use Keras with TensorFlow to train two CNNs from scratch, applying ReLU, maxpooling, dropout, and Adam optimizer.

Transfer Learning

The TFL run used transfer learning with Inception v3, retraining the top dense layers for 100 epochs with RMSprop, followed by fine-tuning the top convolutional layers using the SGD optimizer at a low learning rate.

- Encouragement for research: The Nerthus dataset provides a foundation for advancing automatic bowel cleansing assessment systems, and researchers are invited to improve upon the existing methods.
- Impact of innovation: These innovations can enhance the quality of colonoscopy examinations by reducing subjectivity, saving time for doctors, and optimizing healthcare resources.

REFERENCES:

https://dl.acm.org/doi/pdf/10.1145/3083187.3083216

Thank you