Automated detection of bowel preparation adequacy with deep convolutional neural networks

Low DJ, Hong ZQ, Mukherjee A, Jugnundan S, Grover SC

Introduction: Inadequacy of bowel preparation has been shown to increase post-colonoscopy colorectal cancer. As a result, the USMTF recommends repeating colonoscopy within 1 year if bowel preparation is inadequate. However, bowel preparation documentation is poor in practice, and physician recommendations adherent to USMTF guidelines are highly variable. We present an automated computer assisted method using deep convolutional neural networks to detect adequacy of bowel preparation with the Boston Bowel Preparation Scale (BBPS).

Methods: We extracted 38523 images of colonic lumen between 2015 and 2017 from screening colonoscopy procedures. Bowel preparation scores were assessed with BBPS. Adequate bowel preparation was defined as BBPS ≥2, and inadequate bowel preparation was defined as BBPS <2. The dataset was split into 26966 images for training, 7704 for validation, and 3853 for testing. Training data was sampled with replacement from a multinomial distribution to balance subclass distributions in each batch. We developed 2 convoluted neural networks (CNN) using PyTorch with a Densenet-169 backbone pre-trained on ImageNet and fine-tuned on our data for classifying adequacy of bowel preparation (binary) and for subclassification of BBPS (multi-class). We used Adam optimiser with an initial learning rate of 3×10^{-4} and a scheduler to decay the learning rate of each parameter group by 0.1 every 7 epochs along with focal loss as our criterion for both classifiers.

Results: The overall accuracy on the test data set for BBPS subclassification was 0.91. The accuracy score for BBPS 0, 1, 2 and 3 were 0.84, 0.91, 0.85, and 0.96, respectively. The F1 score for BBPS 0, 1, 2, and 3 were 0.91, 0.92, 0.88, and 0.94, respectively. The precision score was 1.00, 0.92, 0.91, and 0.91 for BBPS 0, 1, 2, and 3, respectively. The overall accuracy on the test data set for adequacy of bowel preparation was 0.97. The accuracy score for adequacy of bowel preparation for BBPS <2 and BBPS ≥2 was 0.91 and 0.98, respectively. The F1 score for adequacy of bowel preparation for BBPS <2 and BBPS ≥2 was 0.94 and 0.98, respectively. The precision score for adequacy of bowel preparation for BBPS <2 and BBPS ≥2 was 0.96 and 0.98.

Conclusion: We present an automated computer-assisted detection method of bowel preparation with deep convolutional neural networks. The algorithm is capable of accurate classification of adequacy of bowel preparation (97%) and subclassification of bowel preparation (91%). This algorithm can be applied to automate documentation of bowel preparation, and adequacy of bowel preparation to inform the subsequent need to perform repeat colonoscopy in a timely manner. Further studies will need to be conducted to demonstrate its applicability in recorded videos and in live colonoscopy.