World Journal of Gastroenterology,

Manuscript NO: 66030

Manuscript Type: MINIREVIEWS

Title: Computer-assisted diagnostic systems to improve colonoscopy quality

RE: Response to Reviewers;

Dear Prof. Subrata Ghosh, Dr. Andrzej S Tarnawski

Editors-in-Chief,

Thank you for giving us the opportunity to submit a revised draft of our manuscript titled " Computer-assisted diagnostic systems to improve colonoscopy quality " to the World Journal of Gastroenterology. We appreciate the time and effort that you and the reviewers have dedicated to providing your valuable feedback on our manuscript. We are grateful to the editors and reviewers for their insightful comments on our paper. We have been able to incorporate changes to reflect all the comments provided by the reviewers.

Please note that we added a co-author, Dr. Yuichi Mori, to the list of authors. The information and the affiliation were included on the first page of the manuscript. We also changed the title of our manuscript to "**QUALITY ASSESSMENT OF COLONOSCOPY BY COMPUTER**". We have highlighted the changes within the manuscript.

Thank you for considering our manuscript for publication in your journal.

Sincerely,

Daniel von Renteln, M.D.

Department of Medicine, Division of Gastroenterology, Montreal University Hospital (CHUM) and Montreal University Hospital Research Center (CRCHUM), 900 Rue Saint-Denis, Montréal, QC H2X 0A9, Montreal, Canada. danielrenteln@gmail.com

Here is a point-by-point response to the reviewers' comments and concerns.

1. At times the authors mention statistical and machine learning terms, which may not be common place and familiar with the intended audience of this journal. Its is suggested that these terms are followed by a brief explanation, apt for the lay physician. These include; Deep neural network (DNN), Bag-of-features, and model overfitting.

   Author response: we added a brief explanation for each of the machine learning terms. The revised text reads as follows on:

   Page 6-7: Deep learning employs deep neural networks (DNNs), which imitates the complex interconnected neural network in the human brain. These artificial neurons are positioned in several detections and pooling layers, taking weighted data (from the precedent layer), processing it, and passing the output (processed data) to the next layer. Each layer performs as a "step of abstraction"[27], which forms a hierarchy of common features that grow in complexity throughout the layers (i.e., edge->basic shape->object->class prediction). In other words, each layer would extract useful and relevant features from a given data that would facilitate the classification of the images.

   Page7-8: Model overfitting is an error in modeling that occurs when the model is too tightly fitted to the training data and random fluctuations in the training data are learned as concepts by the model. The problem is that the fitted model does not generalize to new data due to its low bias and high variance. Overfitting can be avoided by tight monitoring of the model during the training by constantly evaluating the model performance in the training and validation data[27].

   Page 13: The bag-of-feature, inspired by the bag-of-visual-words model, is an image feature extraction method that treats image features as words by following three steps: feature detection, feature description, and codebook generation[64,65]. In this study, a histogram of visual words represented an image by a vector machine support.

2. The authors mention that the study by Klare et al, asessing the CADe system, used "colonoscopy videos". However, the researchers were employing real time CADe technology during live colonoscopies and this was blinded from the sight of the endoscopist. Prior studies have used pre-recorded videos of colonoscopies.

   Author response: Thank you. We corrected the text incorporating the correct data. The revised text reads as follows on:

   Page 11: In contrast to the results of the latter study, Klare et al[51] prospectively evaluated endoscopist performance using CADe assistance during the real-time colonoscopy procedures of 55 patients. However, the endoscopists only observed the regular monitor, and an independent investigator observed the monitor dedicated to representing the real-time outputs of the CADe system in a separate room, which was blinded from the endoscopists' sight. Therefore, the endoscopists were blinded to the real-time CADe outputs.

3. When describing the CADx system, the authors illustrate how in a trial, using nine polyp features significantly improved the system's performance compared to three features, and how this performance was comparable between CADx and experts but superior to non-experts. Please define experts and non-experts in this context.

Author response: Thank you. We added the definitions of expert and non-expert to the text. The revised text reads as follows on:

Page 12: Performance was comparable between experts (with >4 years of experience with magnification colonoscopy and NBI) and CADx, but both were significantly superior to that of non-experts (with only 1 year of colonoscopy experience and never using NBI).

4. In the paragraph following the subheading of Metadata - it is not clear what message the authors are trying to convey? Do they mean that for most studies the meta data is not available which is why it has not been assessed?

   Author response: We explained the concept of metadata and why it is critical to link the colonoscopy videos to patient-level data. The revised text reads as follows on:

   Page 16-17, under the subheading "Metadata": Most studies have failed to assess the performance and accuracy of AI models according to polyp size, polyp location, bowel preparation score, or withdrawal time[31]. Patients' information including demographic and clinical characteristics (e.g., colonoscopy indication, disease status), procedure-related quality characteristics (i.e., bowel preparation level, withdrawal time), procedure time and room, endoscopists fatigue (i.e., the procedure performed in the morning or afternoon) are the important factors that are linked with the long-term non-endoscopic outcome of interest. In other words, the detection and classification of colorectal polyps are the intermediate outcomes of the colonoscopy but the prevention of interval cancer during the surveillance period, or the evaluation of the effectiveness of medical therapy and the need for surgical treatment in patients with inflammatory bowel diseases are the ultimate goals of the colonoscopy depending on the primary indication of the procedure. As mentioned in Kudo et al[82], metadata is a critical component in establishing optimal AI platforms that can perform well in real-world practice with suboptimal conditions. For example, SSA/Ps are mainly located in the right colon, where endoscopic access and complete inspection of the mucosa are challenging[79]. Collecting a high number of colonoscopy videos with a high number of SSA/P polyps and cross-linking with patient's data would increase the accuracy and effectiveness of the colonoscopy. Future AI models must incorporate the information of the polyp size and location as well as the clinical, pre-procedural, and polyp morphological characteristics rather than focusing on the polyp images and videos alone.

5. Addition of pictures/videos of the AI technology in question would elevate the rank of the article Consider mentioning the CAD system (CADe and CADx) in the abstract as it is the meat of the article.

   Author response: We added 4 tables and 2 figures to the manuscript. The tables and figures were provided in 2 new documents (66030-table; 66030-figures). The figure legends were added to the end of the manuscript.

6. We added a short paragraph at the end of the sub-heading "Safety and cost-effectiveness" about the computer-assisted tools that obtained the regulatory approval. Here is added paragraph on **page 20**: Adaption of the newly developed AI-based techniques in routine practice, and the enhancement of endoscopists' trust in the new devices is only possible by a symbiotic relationship between academia and industry. It would facilitate obtaining regulatory approval from health authorities regarding research involving human subjects, constructing large "ground truth" data for developing AI models, and transporting knowledge and technology to ultimately access the market[95]. Several manufacturers have obtained the regulatory approvals to launch

and commercialize their AI-based colonoscopy devices around the world (Table 4); however, many of them have not provided a detailed report of their devices' performance. Further research should try to compare the performance of different AI-based systems in real-time settings by conducting prospective controlled trials with multiple intervention arms sing different commercially available AI-based colonoscopy systems. Due to the time- and cost-consuming nature of these studies, an alternative method for accelerating research is to test the "benchmarks" using the publicly available datasets[96] such as the ASU-Mayo colonoscopy video database[43], the CVC-ClinicDB database[42], the Kvasir dataset[97], the SUN-database[96], and the ETIS-Larib Polyp database. Nonetheless, these datasets contain a limited number of colonoscopy videos and images and may not reflect the true performance of an AI-based system.

7. If we added a new reference, all have been included properly in the manuscript.