March 24, 2022

Dear Dr. Andrzej S Tarnawski

I would like to resubmit our manuscript for publication in the *World Journal of Gastroenterology*, titled "Utility of a Deep Learning Model and a Clinical Model for Predicting Bleeding after Endoscopic Submucosal Dissection in Patients with Early Gastric Cancer."

My coauthors and I would like to thank you for taking the time to consider our article for publication. We appreciate the detailed suggestions provided by each reviewer. The manuscript has been rechecked, and appropriate changes have been made following the reviewers' recommendations. Changes in the revised manuscript are highlighted in red. The responses to reviewer comments have been prepared and provided below.

We thank you and the reviewers for thoughtful suggestions and insights, which have enriched the manuscript and helped us produce a better and more balanced account of the research. We hope that the revised manuscript is now suitable for publication in your journal.

I appreciate your consideration. I look forward to hearing from you.

Sincerely,

Corresponding author

Hyuk Lee, MD, PhD

Department of Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, 81 Irwon-ro, Gangnam-gu, Seoul, 06351, South Korea

Tel: +822-3410-3409, Fax: +822-3410-6983, E-mail: leehyuk@skku.edu

Co-corresponding author

Tae Jun Kim, MD

Department of Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, 81 Irwon-ro, Gangnam-gu, Seoul, 06351, South Korea

Tel: +822-3410-3409, Fax: +822-3410-6983, E-mail: taejunk91@gmail.com

RESPONSE TO COMMENTS

COMMENTS TO THE AUTHOR:

Reviewer #1 comment：

Interesting paper, looking forward to see it published

In the results: Regarding: More patients in the PEB group showed a hemoglobin drop of > 2 g/dL after ESD. The drop of Hb is an outcome for bleeding, not a risk factor for bleeding. That should be clear both in methods and results.

Response: **Thank you for your meaningful comment. We reanalyzed after excluding hemoglobin drop of > 2 g/dL and depth of invasion. We only used the factors that could be known before ESD, reflected in Method and Result section (page 7, lines 183-185; page 10, lines 259-272).**

In discussion: "Our study identified younger age, male sex, hypertension, chronic kidney disease, P2Y12RA use, anticoagulant (warfarin or DOAC) use, middle tumor location, tumor size, and a > 2-g/dL reduction in the hemoglobin level as the predictors of PEB". Why do you think that younger age was a predictor of PEB?

Response: **Thank you for your comment. Previously, several works of literature also reported that younger age was associated with bleeding after ESD. It is unclear actually why younger age was associated with PEB. Previous reports proposed that atrophic change along with aging might relate to decreasing the vascularity on the mucosal and submucosal layer. Although aging and changes in intestinal vasculature have not been**

clearly elucidated, a decrease in the volume of vasculature with aging was observed in the animals. This guess was added in the Discussion section (page 12, lines 307-312).

Please review the first sentence of the introduction: "In South Korea, the incidence of gastric cancer has high incidence,"

Response: **Thank you for your comment. We revised the sentence to "gastric cancer has a high incidence" (page 6, line 129).**

Reviewer #2 comment：

The authors proposed the utility of a deep learning model to predict post-ESD bleeding for early gastric cancer. It is very interesting but it was unclear what it is and how the readers use this in their practice. The authors should clarify what this deep learning model looks like and how it works in real practice to make readers know the real utility.

Response: **Thank you for your considerable comment. The deep learning model was reconstructed to apply to real practice by excluding the depth of invasion (submucosal invasion) and hemoglobin drop ≥2 g/dL after ESD, which cannot be known before the procedure. We added a link to calculate the deep learning and clinical models by inserting values into the formulas. Readers can easily calculate the probability of PEB through the app by inputting the patient's variables before ESD and estimate the PEB risk according to stratified categories. These corrections are added in the Method and Results section (page 7, lines 183-185; page 9, line 226; page 10, lines 259-272).**

Please clarify how different the deep learning and clinical models are (They used the same data set to develop) and how to calculate scores exactly. After I read the entire manuscript, I am still not sure.

Response: Thank you for your sincere comment. The difference between the deep learning and clinical models is the derived method. The deep learning model found the optimal hyperparameters to predict post-ESD bleeding and constructed an algorithm using all variables in the development set without external intervention. This algorithm is different from the clinical model in that it is methodologically challenging to present specific weighted factors for predicting PEB because it is a complex architecture. The clinical model used beta coefficient values of the significant factors for predicting PEB selected in multivariable logistic regression analyses. The formula is presented at the bottom of Table 2. The score as risk probability was calculated by multiplying 1000 to the obtained values in both models. We described the process of model development more accurately in the Method section. (page 8-9, lines 198-226)

To develop this model, the authors used their big data about patients' characteristics and lesions. However, they did not consider procedural factors (e.g., defect closure vs. non-closure, experienced endoscopist vs. beginner, etc.) It might affect the incidence of bleeding.

Response: Thank you for your meaningful comment. Our institution generally did not use defect closure like clipping closure or detachable snare with clips. As for the defect size, there was a missing 1,181 cases due to the change of the report form. The endoscopist's experience was not included in extracting long-term data from 2010 to 2020; considering our center is a tertiary referral center, it was assumed that experienced physicians performed most procedures. The procedure time could not be

accurately measured due to the retrospective manner. We added these as limitations in the Discussion section (page 14, lines 364-366).

Moreover, they used pathologic features like SM invasion. How can we know this before we perform ESD? Overall, I think the authors did interesting work; however, they should clarify how to use the deep learning and clinical models in real practice.

Response: **Thank you for your significant comment. As pointed out, we reanalyzed except the factors of submucosal invasion and hemoglobin drop > 2 g/dL. We derived and verified the performance of the deep learning and clinical models. First, we compared AUC. Second, the risk probability computed by deep learning and clinical models was stratified as low-, intermediate-, high-risk categories in the development set. The cutoffs were set at the 5% and 9% probabilities concerning the previous report (Hatta W et al. Gut. 2021 Mar;70(3):476-484). We proved the performance for stratifying the PEB risk of the deep learning and clinical models in the validation set. The deep learning and clinical models showed AUC 0.71 and 0.70, respectively, without significant difference (p value=0.730). In the validation set, low-, intermediate-, high-risk categories showed the actual bleeding rate of 2.2%, 3.9%, and 11.6% based on the deep learning model; 4.0%, 8.8%, and 18.2% based on the clinical model. The actual bleeding rate was slightly lower than a predicted range of intermediate-risk as ≥5% and <9% in the deep learning and was close to the upper range in the clinical model. We concluded the deep learning model showed acceptable performance considering AUC and the ability of risk stratification. An app was added to Method section to calculate easily (page 9, line 226). Based on this, physicians could give attention to the bleeding risk and carefully perform preventive hemostasis during the procedure. These can be**

use as an indicator for applying shielding methods (e.g., polyglycolic acid with fibrin glue) for high-risk patients due to the lack of randomized controlled trials about its feasibility and safety. These potential utilities of deep learning were described in the Discussion section (page 13-14 lines 359-363).

Reviewer #3 comment：

The authors have claimed that they have developed a deep learning and clinical model for the prediction of bleeding after Endoscopic Submucosal Dissection in Patients with Early Cancer. Well, the results of the manuscript are very confusing and the methods developed by the authors are clearly explained. I have mentioned some major suggestions that authors need to be incorporated for further processing.

1. The abstract of the proposed study needs to reframe and should be in a scientific language. The results mentioned in the abstract are very confusing. The authors need to clarify the results based on the parameters. The comparative results should also be there in the abstract.

Response: **Thank you for your comment. The main outcome was to construct and compare the performance of the deep learning model and clinical model for predicting bleeding risk after ESD. The performance was evaluated using two methods. First, we compared AUC for predicting bleeding after ESD. Second, we stratified the PEB risk as low-, intermediate-, and high-risk categories in the development set. The cutoff between each category was 5% and 9% referred to the previous report (Hatta W et al. Gut. 2021 Mar;70(3):476-484). The deep learning and clinical model showed acceptable performance with AUC of 0.71 and 0.70, respectively. The stratified risk of PEB**

correlated with actual bleeding rate in the validateion set. However, the intermediate-risk category with expected probabilities of ≥5% and <9% yielded a real bleeding rate of 3.9% in deep learning and 8.8% in clinical model. Reflecting on results, we revised the abstract entirely.

2. English expression needs editing and improvement. There are many typos and grammatical errors, checking the paper carefully is recommended.

Response: **Thank you for your comment. We reviewed the manuscript entirely and revised the errors.**

3. In the introduction and related work sections, the novelty of this paper w.r.t. the existing work should be stated, rather than just listing the existing work. Authors need to frame some objectives according to their novelty.

Response: **Thank you for your comment. Our novelty is first to use an artificial intelligence system for predicting bleeding after ESD. Furthermore, we constructed a deep learning model characterized by a sophisticated algorithm integrating all variables. We emphasized that in the Introduction section (page 6-7, lines 145-157).**

4. A literature review is not enough, there must have some more literature on the existing tradition and modern techniques so that authors can compare their work with the existing techniques. Create a separate method of existing techniques with their limitations.

Response: **The previous literature was limited to identifying the risk factors and did not**

suggest how to apply that clinically. The recently developed clinical model in Japan (Hatta W et al. Gut. 2021 Mar;70(3):476-484) had an advantage in a simple formula, but the interruption, heparin bridging, and replacement were analyzed as one variable, respectively. There is a need to consider how the absence of antithrombotic indications was classified. Another recently published model constructed a simple algorithm using ongoing antithrombotic agent, size, and age; among 5,080 patients, only 11 patients were classified as a high-risk group, so its usefulness is questionable (Choe YH et al. J Gastroenterol Hepatol. 2021 Aug;36(8):2217-2223.). We not only constructed a clinical model using significant factors based on the multivariable logistic regression analysis with extensive data but also derived the deep learning model using all variables and verified its performance. We assert that the ability of deep learning to predict PEB having its own algorithm including all variables proves the potential of this tool. These were added in the Discussion section (page 11, lines 299-301).

5. What preprocessing approaches have been used for preprocessing the dataset is not mentioned in the manuscript. The authors should be mentioned what preprocessing approach they have applied to the dataset.

Response: **Thanks for a critical point. We preprocessed the dataset as follows; the categorical variables were converted using one-hot encoding and the continuous variables were normalized. We added this comment to Method section (page 8, lines 202-203).**

6. The author claimed about the development of a novel deep learning approach for the

clinical data but in the paper, there is no mention of the architecture of the deep learning method. The must Clarify this thing which deep learning methods he has developed for h prediction of Bleeding after Endoscopic Submucosal Dissection in Patients with Early Gastric Cancer.

Response: **Thanks for the review. I added the description and Supplementary Figure 1 (page 8-9, lines 212-216).**

7. There are 5629 patents are taken for the prediction of after Endoscopic Submucosal Dissection in Patients with Early Gastric Cancer, I this amount of data is not enough for the deep learning model. They should mention some augmentation/bootstrapping approaches if they have applied to enhance the dataset.

Response: **Thanks for the review. We used the borderline synthetic minority over-sampling technique to generate synthetic data. Final model was trained with the 20% of synthetic data of the majority class. We described this comment in the Method section (page 8-9, lines 198-216).**

8. Authors should compare their work with some existing approaches to verify the outcomes of the proposed approach.

Response: **Previously, papers have examined whether risk probability calculated in training set using machine learning and allocated risk categories correlates with the actual rate of occurrence in the validation set. (Li D et al. Int J Cardiol. 2021 Mar 1;326:30-34. ; Na JE et al. Cancers (Basel). 2022 Feb 22;14(5):1121.) Similarly, We set the cutoff of scores between low-, intermediate-, and high-risk categories at a bleeding**

rate of 5% and 9% by referring to the previous paper in the development set and checking whether the actual bleeding rate was within the predicted range in the validation set. These are shown in Table 4.

9. In conclusion, the authors must include some comparative statistical results based on the existing techniques. Comparative result.

Response: **Thank you for your comment. We described the p value comparing the AUC between the deep learning and clinical models (Table 3).**

10. The complete manuscript needs to be rewritten in a scientific language.

Response: **Thank you for your considerable comment. We revised the manuscript after reanalysis.**

11. Highlight the changes with some font/color so that the changes can easily be traced in the revised manuscript.

Response: **Thank you for your comment. We highlight the changes as red color.**