Apr 1, 2023

Dear Editor and Reviewers:

I would like to express our heartfelt gratitude to you for spending your precious time on our manuscript entitled " ***A machine learning model for prediction of low anterior resection syndrome following laparoscopic anterior resection of rectal cancer: A multicenter study***" (ID: 84049). We thank you for providing constructive critiques, comments, and suggestions that have improved the overall quality of the paper. In the following pages, I provided detailed responses to the Reviewer's questions/comments.

Sincerely yours

Jichao Qin, M.D, Ph.D

Professor

Department of Surgery and Molecular Medical Center,

Tongji Hospital

Tongji Medical College

Huazhong University of Science and Technology

1095 Jiefang Avenue, Wuhan, Hubei 430030, China

Tel: +86-27-83665316

E-mail: jcqin@tjh.tjmu.edu.cn.

Response to individual reviewer's comments:

Reviewer #1:

Scientific Quality: Grade C (Good)
Language Quality: Grade B (Minor language polishing)
Conclusion: Major revision
Specific Comments to Authors: This manuscript aimed to develop a machine learning model using preoperative and intraoperative factors to predict major low anterior resection syndrome (LARS) following laparoscopic surgery of rectal cancer in Chinese populations. The trained random forest (RF) model had an AUC of 0.852 and a sensitivity of 0.795 and could potentially be used in the clinic to identify patients with a high risk of developing major LARS and improve their quality of life.

Answer: We greatly appreciated the reviewer's spending precious time in reviewing our study and his/her encouraging comments. We apologize for the language problems in the original manuscript, and we will be happy to edit the text further based on your helpful comments. We have carefully and thoroughly read the manuscript and tried our best to correct all the grammar mistakes and typos. Meanwhile, the revised manuscript was polished by professional English language editing company.
As the reviewer suggested, we have made some revisions, the details are as follows:

Point 1: "In the abstract, the authors claimed that based on the "decision curve analysis, the model had clinical application value." This can be discussed in the results and discussion part; what is the proposed method's theoretical and practical implication?"
Response 1: Thank the reviewer for your valuable comment. As you mentioned, the theoretical and practical implication of decision curve analysis (DCA) is important to discuss in our study. DCA provides a method to evaluate the clinical utility of a model in guiding clinical decision-making by quantifying the net benefits at some threshold probabilities[1] [2]. This can guide the selection of patients who would benefit most from the model and reduce unnecessary treatment for those who would not[3]. In our study, we conducted DCA to evaluate the applicability of the random forest (RF) model and the data indicated that the RF model had a higher net benefit than the treat-all or treat-none strategies at a wide range of threshold probabilities (internal test set: 0.15-0.75, external validation set: 0.2-0.6). In addition, we discussed the theoretical and practical implications of DCA in the discussion section of our manuscript and emphasized the advantages of this approach compared with traditional measures of diagnostic accuracy. We hope that this explanation can clarify the importance of DCA in our study and its broader implications for clinical practice. As the reviewer suggested, we removed the sentence " Decision curve analysis further demonstrated that the model had clinical application value.". Furthermore, we added significant literature on the theoretical and practical implications of decision curves in the results section (reference 28th).

Point 2: "In the abstract, please clarify how your findings or model performance compared to the state-of-the-art methods and how it can go beyond the state-of-the-art methods."

Response 2: To compare the performance of the RF with the state-of-the-art methods (the pre-operative low anterior resection syndrome score, POLARS), we provide some new data (Table 3) to show the performance of the POLARS score model in our testing set and external validation set. Table 3 indicated that the POLARS score model performs with moderate discriminative accuracy. By comparing sensitivity, specificity, positive predictive value, negative predictive value and accuracy, we found that the RF model outperforms the POLARS score model. In the abstract, we added the data in the revised manuscript to highlight the advantages of our approach (see Lines 76-79, page 4 in the revised manuscript).

Point 3: "The methodology needs to be improved by further justifying its advancement."

Response 3: Thank you for your suggestion on our research. The advances in feature selection and machine learning algorithms were described in the feature selection section and machine learning algorithms section (see Lines 180-181, page 8, and Lines 197-198, page 9, respectively, in the revised manuscript). In feature selection, when all 13 variables are used, the RF model achieves an AUC of 0.868 with sensitivity of 0.762, specificity of 0.802, positive predictive value of 0.694, negative predictive value of 0.851 and accuracy of 0.787 in the training set (Supplementary Figure 1). Notably, the performance of the RF model with all 13 variables is similar to that of 8 selected variables (Supplementary Figure 1A and Figure 2B). A smaller number of variables can reduce the complexity of a model and increase its clinical usefulness; feature selection method used in this study can help to identify the most relevant features and thus simplifying the complexity of the RF model. Regarding model selection, machine learning algorithms have been demonstrated their ability to overcome the limitations of predictors by effectively combining different types of features in a flexibility and scalability manner compared with traditional biostatistical methods [4-6].

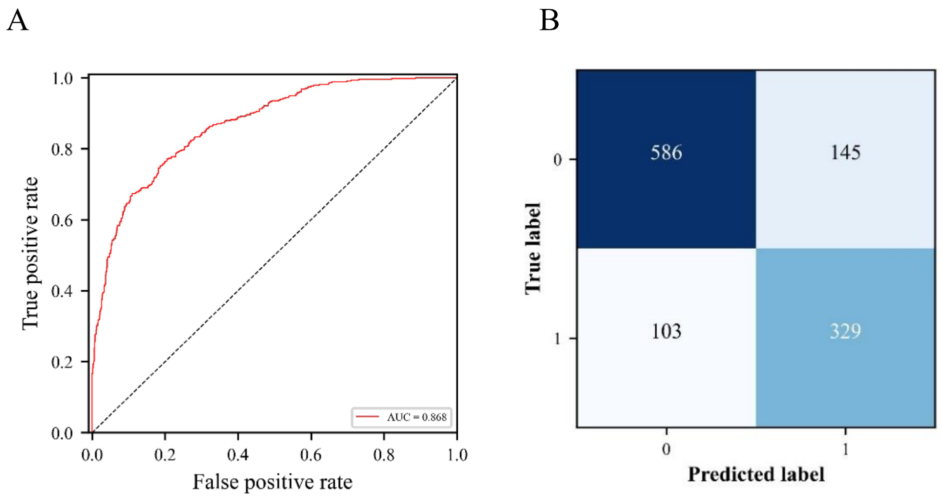

Figure 1: Performance of the RF model with all 13 variables in the training set. A: ROC of the RF model. B: Confusion matrices showing the predicted outcomes generated by the RF model. AUC, Area under receiver operating characteristic curve; RF, Random forest.

Point 4: "Figure 1 is blurry. Its quality needs to be improved."

Response 4: To further improve the quality of Figure 1, we adjusted the resolution to 1200 dpi.

Point 5: "The contribution of the paper is unclear."

Response 5: We extend our apologies for any confusion and would like to address this pertinent issue. The introduction section highlights that major LARS severely impairs quality of life, and there is a lack of an accurate model based on a multicenter study for predicting major LARS in Asian patients undergoing laparoscopic surgery. While the European population-based POLARS score model can predict LARS, its validity, when applied to other populations, particularly the Chinese population, remains uncertain. To address these challenges, this study aims to develop a machine learning model to predict major LARS based on patients from two Chinese medical centers. After external validation and comparison with the POLARS score model, the data indicated the superiority of our RF model. This study provides a new tool for predicting major LARS, which can be potentially used for rectal cancer patients to acquire early postoperative consultation and strengthen self-management to improve their quality of life. We believe that our research makes a significant contribution to the field of predicting major LARS.

Point 6: "What are the limitations of the current study? This should be clearly reported in the conclusion section."

Response 6: We agree with your comments. The limitations of the current study were reported in the revised manuscript (see Lines 362-364, page 17 in the revised manuscript).

Reviewer #2:

Point: "According to the reviewer, the work under review is interesting and well written. I do not agree with the choice of algorithms used for machine learning, in fact, I believe that other equally interesting techniques could be used; in particular NN-based classifiers."

Response: We greatly appreciated the reviewer's spending precious time in reviewing our study and his/her encouraging comments. Regarding the choice of algorithms, we believe there is no best classifier, only the most appropriate one. Random forest is an ensemble algorithm. It randomly selects different features in training samples to generate a large number of decision trees and then synthesizes the results of these decision trees to make the final classification, which is suitable for data with relatively low dimensions and high accuracy requirements. Neural networks have been proposed for a long time, but their accuracy depends on a large training set, which was originally limited by the speed of computers. Moreover, neural networks can be difficult to interpret, making it challenging to understand how the model arrives at its predictions. In summary, while we appreciate your suggestion to consider neural network-based classifiers, we believe that our choice of algorithms was the most appropriate for our research problem and our goal of providing practical and interpretable insights.

We appreciate for editors' and reviewers' warm work earnestly and hope that the correction will meet with approval. Once again, thank you very much for your comments and suggestions.

REFERENCES

1 Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making* 2006; **26**: 565-574 [PMID: 17099194 DOI:10.1177/0272989X06295361] https://doi.org/10.1177/0272989x06295361

2 Vickers AJ, van Calster B, Steyerberg EW. A simple, step-by-step guide to interpreting decision curve analysis. *Diagn Progn Res* 2019; **3**: 18 [PMID: 31592444 DOI: 10.1186/s41512-019-0064-7] https://doi.org/10.1186/s41512-019-0064-7

3 Vickers AJ, Holland F. Decision curve analysis to evaluate the clinical benefit of prediction models. *Spine J* 2021; **21**: 1643-1648 [PMID: 33676020 DOI: 10.1016/j.spinee.2021.02.024] https://doi.org/10.1016/j.spinee.2021.02.024

4 Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med* 2019; **25**: 44-56 [PMID: 30617339 DOI: 10.1038/s41591-018-0300-7] https://doi.org/10.1038/s41591-018-0300-7

5 Rajkomar A, Dean J, Kohane I. Machine Learning in Medicine. *N Engl J Med* 2019; **380**: 1347-1358 [PMID: 30943338 DOI: 10.1056/NEJMra1814259] https://doi.org/10.1056/nejmra1814259

6 Ngiam KY, Khor IW. Big data and machine learning algorithms for health-care delivery. *Lancet Oncol* 2019; **20**: e262-e273 [PMID: 31044724 DOI: 10.1016/S1470-2045(19)30149-4]https://doi.org/10.1016/s1470-2045(19)30149-4