Dear Editors and Reviewers:

Thank you for your letter and the reviewers’ comments concerning our manuscript entitled “Added-value of systemic inflammation markers in predicting postoperative pancreatic fistula following pancreaticoduodenectomy: a comparative analysis using machine learning-assisted decision-support models” (Manuscript ID: 874961). Those comments are all valuable and very helpful for revising and improving our paper, as well as the important guiding significance to our researches. We have studied comments carefully and have made correction which we hope meet with approval. The revised portion is marked in red in the paper. The main corrections in the paper and the responds to the Editor’s and reviewer’s comments are as following:

Reviewer #1:

Scientific Quality: Grade C (Good)

Language Quality: Grade B (Minor language polishing)

Conclusion: Minor revision

Specific Comments to Authors:

1. Methods: Line 131 - define ROI Results: Table 1 in the manuscript should be replaced by the supplementary table 1. Supplementary table 4: include a footnote describing what the variables with layhid are and how to interpret them. This applies to all other places in the manuscript these variables are presented or discussed.

Response: Thank you for your suggestion. We have defined ROI and described it in the methodology section, as follows:

We used the GEAW4.6 workstation to select the largest surface of the lesion in T2W1 and calculated the consistency of the extracted imaging omics features using the inter group correlation coefficient (ICC). Among them, ICC>0.75 indicates good consistency, and features with ICC>0.75 are selected for the next analysis. By calculating the correlation coefficient between each feature in the training group, the redundancy of the feature is removed. When the absolute value of the correlation coefficient between two features is greater than 0.9, one of the features is randomly removed.

Table 1 in the manuscript has been replaced by the supplementary table 1.
We have added relevant notes at the end of Supplementary Table 4, as follows:

We first provide input examples to neurons, and then transmit the signals layer by layer until the output layer results are generated; Then calculate the error of the output layer, propagate the error back to the hidden layer neurons, and adjust the connection weight and threshold based on the error of the hidden layer neurons. This iteration process is iterated until certain stop condition positions are reached.

2. Discussion: Is the incidence rate of pulmonary infection in the study 17.03%? If yes, what is 11.23% mentioned on line 258.
Response: Thank you for reminding us that we have corrected the correct incidence rate as follows:

The results of this study showed that the incidence rate of pulmonary infection after liver cancer surgery was 17.03%, consistent with previous studies.

3. Discussion: A potential limitation of the study is that it does not tell us how the features found to be important to predict pulmonary infection are clinically valuable although the authors do go on to conclude that their results are useful as such. Without a clear numerical cutoff or other guidance on how various features based on the MRI data can be used by clinicians to predict the risk of pulmonary infection, all this study can show is that the identified features are important. Further research is needed to identify specifically which features can inform clinicians. It is also not clear if the features identified in the study increase or decrease the risk of pulmonary infection.
Response: Thank you for your valuable suggestion. We agree with you that our study is that it does not tell us how the features found to be important to predict pulmonary infection. As shown in Table1 and Figure2. In this study, we delineated the ROI along the lesion boundary, covering the maximum layer of the lesion and one layer before and after the lesion. In fact, there are already differences in the quantification parameters of lung infection through ROI, such as the sum of squares (SOS), an inverse difference (IND), mean sum (MES), sum variance (SUV), sum entropy (SUE), entropy, and difference variance and entropy.

In addition, how various features based on the MRI data can be used by clinicians to predict the risk of pulmonary infection? In this study, we used least absolute shrinkage and
selection operator (LASSO) regression with minimum penalty coefficient and Pearson correlation coefficient were used to select variables that could ultimately be used to develop pulmonary infection prediction models. Without a clear numerical cutoff or other guidance in this study, because we use continuous variables and ultimately evaluate their risk based on the scores of RFM or GRLM algorithm formulas, we did not set a cutoff value. For example, in the GLRM, we can predict the risk of pulmonary infection via the “risk” score, each specific score corresponds to a specific score, allowing for precise risk assessment rather than binary risk assessment, as follows:

![Graph showing risk vs. total points]

We agree with you that further research is needed to identify specifically which features can inform clinicians. For example, as for RFM and ANNM, the application prediction model formula of RFM is as follows: \( C = \text{argmax} \left( \sum C_i \right) \), where \( C_i \) represents the type of In prediction for the \( i \)-th tree, \( C \) is the final classification result, and \( I \) is the number of trees. The formula of ANNM is as follows: \( \theta = \theta - \eta \times \nabla J(\theta) \). Among them \( \eta \) is the learning rate, so \( J(\theta) \) represents the gradient change of the loss function (i.e. \( J(0) \)). We will continuously optimize our algorithms and clinical application value.

Reviewer #2:

Scientific Quality: Grade C (Good)

Language Quality: Grade B (Minor language polishing)

Conclusion: Major revision

Specific Comments to Authors: I am really grateful to review this manuscript. In my opinion, this manuscript can be published once some revision is done successfully.

1. This study used three machine learning models and radiomic data of 505 patients for the prediction of postoperative pulmonary infection with primary hepatic carcinoma. The random forest achieved the area under the curve of 90.1% and inverse difference, the sum of squares, mean sum and sum entropy were most important predictors. I would argue that this is a great achievement. But Shapley Additive Explanations (SHAP) is gaining
immense popularity given that it provides the direction of association between the dependent variable and its major predictor. In this context, I would like to suggest the authors to derive the SHAP summary plot of the random forest.

Response: Thanks for your valuable suggestions. We agree with you that The SHAP interpretation can be used (it is model-agnostic) to compute the feature importances from the Random Forest. It is using the Shapley values from game theory to estimate how does each feature contributes to the prediction. Therefore, we have derive the SHAP summary plot of the random forest, as follows:

![SHAP summary plot of the random forest](image)

Reviewer #3:

Scientific Quality: Grade D (Fair)

Language Quality: Grade C (A great deal of language polishing)

Conclusion: Major revision

Specific Comments to Authors: see the attached microsoft office WORD file.

Authors analyzed 505 patients’ dataset so as to develop and then validate the prediction model for predicting postoperative pulmonary infection. In the work, a series of computational and quantitative analytic methods were used and mentioned in the draft of the manuscript, such as co-occurrence matrix, random forest, artificial neural network, linear regression. Finally, authors claimed that several risk factors were successfully
identified by them, and their ‘RFM’ algorithm could better predict the postoperative pulmonary infections of patients who underwent surgeries.

However, questions and issues below should be clearly explained or answered by authors.

1. In the conclusion section of abstract, authors wrote ‘depth learning’. What is “depth learning”? Do authors refer to deep learning? If not, please explain what kind of technique or technology this special term refer to.

Response: Dear reviewer, thank you very much for promptly pointing out the shortcomings in the wording of this article. We have adopted your valuable suggestions and carefully corrected vocabulary such as "deep learning" in the article. We have also consulted a large amount of literature and realized that using "deep learning" in the article is inappropriate, so we have adopted "machine learning" for correction. Thank you again for your in-depth guidance and suggestions on this article.

2. Again, in the conclusion section of abstract, authors wrote that “…, especially the RFM algorithm, which can better predict……”. Are authors sure that random forest is a kind of “depth” (or deep?) learning methods? If yes, show me where is this idea came from. Was it from your own creative definition? Or from other sources? If authors are not sure about that, authors should CAREFULLY RE-READ some introductive papers, reviews, or tutorials about machine learning, and then re-examine the correctness of that sentence. I did see several references about machine learning were cited in the manuscript, for example, the [9] and [10], but did authors really read these references?

Response: Thanks for your valuable suggestions, we have modified the conclusion section of abstract, as follows:

Postoperative pulmonary infection in patients undergoing hepatectomy may be related to risk factors such as IND, SOS, MES, SUE, SUV, energy, and entropy. The prediction model in this study based on diffusion-weighted images, especially the RFM algorithm, can better predict and estimate the risk of pulmonary infection in patients undergoing hepatectomy, providing valuable guidance for postoperative management.

In addition, we have made corrections to inappropriate expressions in the article, such as “depth” (or deep?) learning methods. We have re-examined the correctness of that sentence.
3. In the methods section of abstract, what did authors mean by writing “eleven radiomics were selected for ….”? Do authors refer to the radiomic data of eleven patients? Or I am not sure if authors were referring to eleven radiomic features? If latter is true, I seriously doubt both authors’ and the language editing service helpers’ ability in writing manuscript. It seemed that authors themselves had not checked the manuscript editions by service provider, and authors were not cared about the quality of the language editing service, as well as authors’ own manuscript, with a funny language editing certificate and a manuscript with great amounts of errors submitted to the journal.

Response: Thanks for your question and suggestions. In this study, 11 imaging omics variables were obtained through imaging data, not from 11 patients. We conducted imaging examinations on the liver and lungs of 505 patients, and obtained relevant parameters (including 11 imaging related indicators). Therefore, we have made corrections and explanations in the article to avoid being misled.

As for the quality of the language editing service, we have submitted to our paper, the Charlesworth Author Services (CAS) team (https://www.cwauthors.com.cn/) and Bullet Edits had helped us improve our language and correct grammatical errors existed in our manuscript. The CAS team confirmed that their proofreader had done a very good job of improving language and correcting grammatical errors. We also checked the manuscript again and again, and we couldn’t find grammatical errors.

4. Which hospital are authors really from or affiliated to? There were 3 different hospitals’ name in the manuscript. At title page’s author affiliation, “Jing Zhou hospital” was written. In Materials and Methods section, “Jingzhou Central Hospital” was written. In the remark section of the ethical approval paper, “Baotou Cancer Hospital” was written. Please explain why different hospital names appeared.

Response: Thank you for correcting our negligence. We have reviewed and verified the hospital name and ethical approval documents in this article to ensure that there are no errors. Once again, we deeply apologize.

5. Ethical and compliance issue: why the remark section of authors’ ethical approval paper showed “Institutional Ethics Committee of the Baotou Cancer Hospital”? Shouldn’t it be the IRB of Jing Zhou Hospital? Was this study really approved by the IRB of authors’
institution and conducted in accordance with the research ethical guideline and rules of compliance?

Response: We have reviewed and verified the hospital name and ethical approval documents in this article. As follows:

This is to certify that the design and methods of the research are in accordance with the requirements of related regulations and procedures (such as GCP, ICH-GCP) as well as the ethical principles. The IRB has approved the research to be conducted in Jingzhou hospitals.

6. It seemed that the date of ethical approval was 30 December 2022, and I believe the study plan or proposal should have been made in earlier time (Perhaps earlier than 30 January 2022? According to authors’ ethical approval document). Then why authors did not collect any informed consent from patients after the date that authors planned the study (Since or around 30 January 2022)? Was it appropriate that authors planned a study first, and then started to collect new patients’ samples/data but not letting those new-comer patients know that, “we will use your data/samples for our study”? Was such action in accordance with research compliance and ethical guideline?

Response: Thank you for your inquiry. Prior to this, we have made two applications for ethical approval of the article. However, considering the insufficient sample size for patient inclusion, we have extended it appropriately. The ethics committee has considered that it has already applied once, so it has not applied to subsequent patients (i.e. earlier than 30 January 2022). Therefore, we have conducted another ethical submission application this time. Considering all patients included and the deadline, we have applied for ethical approval again in May 2023. And we will also upload the latest ethics for review.

Thank you.

7. Again, the final day of authors’ sample/data collection was February 2023. Did authors collect patients’ informed consents after the date of ethical approval (30 December 2022)? If collected, please submit and show us. If not, explain why. Authors may also ask that institutional ethical committee whether authors should inform new patients, and why.

Response: Thanks for your question. Considering all patients included and the deadline, we have applied for ethical approval again in May 2023.

8. In the section of “Statistical methods”, are authors sure that R studio can be
downloaded from www.r-project.org? If yes, show me the snapshot of the webpage and
the download URL of R studio from www.r-project.org. If not, I seriously recommend that,
authors should be more serious about manuscript writing. Multiple low-end errors were
found in this manuscript.

Response: Thanks for your question. As for www.r-project.org, we make sure that R studio
can be downloaded from the website, as follows:

9. Are authors respecting contributors of freely available R and R studio? R language
and R studio are developed by different group of scientists and programmers, and the
basic version of these tools are freely available to public. Since authors used these tools for
data analyses, how come that not any relevant citation of R/Rstudio was found in the
manuscript?

Response: Thank you for your valuable suggestion. We have made a statement in the
acknowledgments section regarding all R package developers involved in this article. The
details are as follows:

We thank all R language and R package developers for their assistance, as well as all R
package selfless contributors used for statistical analysis and visualization in this study.

10. In section of “Predictive value of ML-based pulmonary infection prediction model”,
authors wrote that, “As illustrated in Supplementary Figure 2, RFM can distinguish
patients with pulmonary infection from those without pulmonary infection, ……”. Please tell me where is your supplementary figure 2? Did author check your manuscript and submission items?

Response: We are very sorry, but it is actually a supplement to Figure 1. We have made corrections in the article, and we apologize again.

11. Authors showed several high AUC scores, as high as 0.9x in prediction or classification performance benchmarkings. However, I don’t think authors’ results were non-bias enough, nor authors’ models developed were robust. A CRITICAL PROBLEM is the HIGH IMBALANCE in authors’ dataset. As authors wrote in the result section, in both training and validation dataset, the percentages of infected instances/patients were merely 17.xx%~16.xx%. I don’t believe such highly imbalanced training and validation dataset can generate robust prediction models nor robust results. Obviously, authors weren’t aware of this critical problem as not any imbalanced data processing methods or special data splitting strategies for imbalanced dataset was mentioned in the manuscript. Authors should re-consider the training-validation strategy to demonstrate that both your results and models have sufficient robustness.

Response: Obviously, authors weren’t aware of this critical problem as not any imbalanced data processing methods or special data splitting strategies for imbalanced dataset was mentioned in the manuscript. We really agree with your opinion, as a single center retrospective study, bias in data collection is inevitable. At the same time, in order to minimize prediction errors as much as possible, we included cross folding cross validation in our subsequent analysis and repeatedly corrected the area under the curve of ROC. The final conclusion is as follows:
Last but not least, Authors should rewrite the WHOLE manuscript in SERIOUS and CAREFUL ways.

Response: Thanks for your valuable suggestions, we have modified the whole manuscript and also We have also re polished and reviewed the manuscript. We ensure that the
manuscript meets the requirements of the magazine and reviewers within our own capabilities, striving to meet the standards.

**Editorial Office’s comments**

Authors must revise the manuscript according to the Editorial Office’s comments and suggestions, which are listed below:

(1) Science editor:

The manuscript has been peer-reviewed, and it’s ready for the first decision.

(2) Company editor-in-chief:

I have reviewed the Peer-Review Report and the full text of the manuscript, all of which have met the basic publishing requirements of the World Journal of Gastrointestinal Oncology, and the manuscript is conditionally accepted. I have sent the manuscript to the author(s) for its revision according to the Peer-Review Report, Editorial Office’s comments and the Criteria for Manuscript Revision by Authors. The quality of the English language of the manuscript does not meet the requirements of the journal. Before final acceptance, the author(s) must provide the English Language Certificate issued by a professional English language editing company. Please visit the following website for the professional English language editing companies we recommend: https://www.wjgnet.com/bpg/gerinfo/240. Before final acceptance, when revising the manuscript, the author must supplement and improve the highlights of the latest cutting-edge research results, thereby further improving the content of the manuscript. To this end, authors are advised to apply a new tool, the Reference Citation Analysis (RCA). RCA is an artificial intelligence technology-based open multidisciplinary citation analysis database. In it, upon obtaining search results from the keywords entered by the author, "Impact Index Per Article" under "Ranked by" should be selected to find the latest highlight articles, which can then be used to further improve an article under preparation/peer-review/revision. Please visit our RCA database for more information at: https://www.referencecitationanalysis.com/. Uniform presentation should be used for figures showing the same or similar contents; for example, “Figure 1 Pathological changes of atrophic gastritis after treatment. A: ...; B: ...; C: ...; D: ...; E: ...; F: ...; G: ...”. Please provide decomposable Figures (in which all components are movable and editable), organize them
into a single PowerPoint file. Please check and confirm whether the figures are original (i.e. generated de novo by the author(s) for this paper). If the picture is ‘original’, the author needs to add the following copyright information to the bottom right-hand side of the picture in PowerPoint (PPT): Copyright ©The Author(s) 2023.

We tried our best to improve the manuscript and made some changes in the manuscript. These changes will not influence the content and framework of the paper. We appreciate for Editors/Reviewers’ warm work earnestly, and hope that the correction will meet with approval.

Once again, thank you very much for your comments and suggestions.