Peer Review File

Article information: https://dx.doi.org/10.21037/jgo-24-325

Reviewer A

Comment 1: This manuscript presents a 13-gene prognostic ML signature for CRC. However, several areas need clarification and improvement.

My recommendation is for major revisions, and a substantial rewrite of all sections of the paper. While I believe a significant amount of work was done, I do not believe the authors have provided the necessary methods, background, and discussion, and that the conclusions they came to are too bold. I believe there is *something* to present with the work they have done but they have not presented it in adequate detail. It is clear the authors have done work, and that there are results that can be presented that will be worthy of publication, however, these must be presented properly and the author must not overstate the claims. I believe the authors would benefit from having a statistician review their results and figures and assist them with proper scientific quantitative reporting of results during revisions.

Reply 1: We thank the reviewer for this thoughtful review and comment. In this revision, we have carefully modified the manuscript according to the suggestions and marked the revision content in red color.

Comment 2: The abstract needs work, the background section is very short. The methods section does not mention the specific techniques used.

Reply 2: In this revision, we have expanded the background section and updated the methods section to include specific details about the techniques used in our study.

Changes in the text: Page 2, Lines 21-33, 37-40.

Comment 3: Page 2 bullet points spacing is inconsistent

Reply 3: In this revision, we have modified the bullet points spacing on page 2 to ensure consistent spacing throughout the manuscript.

Changes in the text: Page 3, Lines 56-69.

Comment 4: The statement in “What is known and what is new” should possibly have a citation attached if within the format of the journal

Reply 4: Thank you for your suggestion. We have reviewed the requirements of the journal and found that citations are not required in the "What is known and what is new" section.

Comment 5: R package names are traditionally italicized within text e.g. survminer, survival, compareC, rms, nomogramEx, ggDCA, timeROC, ESTIMATE, CIBERSORT

Reply 5: In this revision, we carefully reviewed our paper and made sure that R package names such as survminer, survival, compareC, rms, nomogramEx, and timeROC are italicized within the text. However, as ESTIMATE and CIBERSORT are algorithm names, they were not italicized.

Changes in the text: Pages 6-9, Lines 150, 156, 160, 166, 168, 174, 177, 180, 207

Comment 6: R packages should all be cited within the line where they are introduced, many don’t appear to be cited at all.

Reply 6: We apologize for the oversights in the previous version of the manuscript. In this revision, we have carefully modified our manuscript and ensured that all R packages are cited within the line where they are introduced.

Changes in the text: Pages 6-9 , Lines 150, 157, 160, 167, 168, 174, 177, 180, 208

Comment 7: LASSO is an acronym please introduce it and provide background as to why you chose

it

Reply 7: We have introduced LASSO as an acronym for Least Absolute Shrinkage and Selection Operator, which is a linear regression method that adds a penalty term to the sum of squared residuals to force some coefficients to be exactly zero. The combination of LASSO Cox regression and gradient boosting machine (GBM) was selected as the strategy to construct the prognostic signature, as it obtained the highest C-index value among the 97 combinations.

Changes in the text: Page 14, Lines 346-349.

Comment 8: Line 139/140 – elaborate, "various" is vague, how were these chosen? Why did you choose this approach?

Reply 8: Yes, we agree that "various" is vague. In this revision, we have modified the sentence as "To construct a reliable prognostic signature, we applied a comprehensive machine-learning survival framework to the 22 prognostic genes in the GSE39582 training dataset. The framework consisted of 97 algorithm combinations, which were used to develop corresponding models". To build a prognostic signature with optimal predictive ability, we combined 9 machine learning algorithms in pairs, resulting in 97 combinations. This approach was taken as each algorithm has its own strengths and weaknesses, and combining them enabled us to leverage their respective advantages and minimize their limitations.

Changes in the text: Pages 10, Lines 227-230.

Comment 9: Lines 153-155 "robust" is the wrong choice of word here, as robust should be based on comparison to existing models, consistency across datasets, clinical or practical relevance of the features. The AUC values shown are 0.77-0.79 which are generally classified as 'fair performance' (above 0.8 being 'good', above 0.9 being 'excellent'). Line 157 is a better use of the word robust.

Reply 9: We apologize for the inappropriate wording. In this revision, we have modified "robust" into "relatively good".

Changes in the text: Page 10, Line 251.

Comment 10: Model performance metrics beyond AUC should be stated such as sensitivity, specificity, F1 score - these can be calculated even for LASSO

Reply 10: Thanks for your kind suggestion. In ROC curve analysis, the x-axis represents 1-specificity, which is the false positive rate (FPR), and the y-axis represents sensitivity, which is the true positive rate (TPR). The AUC (Area Under the Curve) value in ROC curve analysis of prognostic models, summarizes the overall performance of the model in distinguishing between positive and negative cases and is an important indicator of the predictive accuracy of the prognostic model. Therefore, time-dependent ROC curve analysis is the most commonly used method for evaluating prognostic model performance. While we agree that sensitivity, specificity, and F1 score metrics can provide a more comprehensive evaluation of model performance, we encountered some difficulties in calculating these metrics for our 13-gene prognostic signature constructed by the combination of LASSO and GBM algorithms. We will continue to explore suitable methods for calculating these metrics in future studies.

Comment 11: Cross-validation strategy is not mentioned – this is major

Reply 11: Thanks for your kind reminding. Ten-fold cross-validation was performed as part of the LASSO and GBM algorithms, and we have included a description of this in the revised manuscript.

Changes in the text: Page 14, Lines 352-356.

Comment 12: Line 159 what is the optimal cutoff point? I only see it introduced in line 93 but it is not described in detail anywhere, how is it calculated? What does it represent?

Reply 12: In this revision, we have further introduced the optimal cutoff point as follows:

The "surv_cutpoint" function in "survminer" was used to determine the optimal cutoff point to divided CRC patients into the high- and low-risk groups. This cutoff point corresponds to the risk score value that can maximize the difference in overall survival time between the two groups.

Changes in the text: Page 6-7, Lines 150-153.

Comment 13: Throughout the paper I am seeing explanations of what you did, but limited explanation of why you did it

Reply 13: In the revised manuscript, we have provided more extensive explanations of the rationale behind our methods and analyses.

Changes in the text: Page 9, Lines 217-219; Page 10, Lines 227-229; 243-245; Page 13, Lines 310-311.

Comment 14: Lines 201-202 finding TP53 and KRAS had the highest mutation frequency in a cancer sample is not a novel or exciting discovery, these are some of the most common genes known to cancer if not the most common.

Reply 14: We agree that the highest mutation frequencies of TP53 and KRAS in our cancer samples are not a new or surprising finding, as these genes are known to be among the most frequently mutated genes in cancer. Through tumor burden analysis, although there were differences in the type and frequency of mutations in certain genes between the low-risk and low-risk groups, we found no significant difference of the mutation frequency of the signature-related genes between the two groups. Therefore, in this revision, we have deleted this content in the manuscript.

Comment 15: 200-206 I would recommend to explain why you did the tumour mutation burden study, and explain why such studies are not useful and why your machine learning approach has greater potential.

Reply 15: Tumor mutation burden is a measure of the number of somatic mutations present in the tumor genome, and it has been shown to be associated with response to immunotherapy and overall survival in some cancer types. Therefore, including tumor mutation burden analysis in prognostic modeling studies can provide valuable insights into the underlying biology of the tumor and improve the accuracy of outcome prediction. Accordingly, we have performed tumor mutation burden analysis and focused on the differences between high- and low-risk groups. However, we found no significant difference of the mutation frequency of the signature-related genes between the two groups. Therefore, in this revision, we have deleted this content in the manuscript.

Comment 16: 207-214 MAJOR: correlation does not equal causation. This needs to be specified as a PRELIMINARY analysis. You do not report the statistical significance of the claims or the effect size. There are no quantitative values mentioned. The figure is a boxplot with no p-values. I am not convinced by the claim you have made based on the available data.

Reply 16: Yes, we agree that correlation does not equal causation. In the original corresponding Figure 7, the statistical significance was indicated by asterisks in the boxplot. In this revision, we have specified this step as a preliminary analysis and used p-values to the significance of the difference in drug IC50 values between the high- and low-risk groups in the boxplots.

Changes in the text: Page 13, line 311; Figure 7.

Comment 17: Line 246 – 247 – you are citing papers that had models with 0.656 0.642 and 0.698 AUCs – these are VERY low AUCs, generally for a predictive model we want to see minimum 0.8.

Reply 17: Yes, for a predictive model, we typically expect to see an AUC value of at least 0.8. However, some published prognostic signatures have reported AUC values lower than 0.8, even as

low as 0.7. For instance, the AUC values reported in the papers we cited are relatively low for predictive models. Therefore, in this revision, we have removed these examples and discussed our model in a different way.

Changes in the text: Pages 14-15.

Comment 18: Line 273-275 I think the claims are too bold for the provided proof.

Reply 18: We acknowledge that the claims made in lines 273-275 may be too bold given the level of evidence provided in the manuscript. We have revised the manuscript to present our findings in a more cautious and nuanced manner.

Changes in the text: Page 16, Lines 394-399.

Comment 19: Why didn't you look at functional analysis? Gene enrichment/ontology? I think this would be stronger than your attempt to use correlation to link to treatment response

Reply 19: Thank you for your suggestion. We agree that functional analysis, such as gene enrichment and ontology analysis, could provide valuable insights into the biological mechanisms underlying the signature. However, as a prognostic model, we are primarily interested in its ability to predict patient survival and treatment response. Therefore, we conducted drug sensitivity analysis to evaluate its performance in predicting treatment response.

Comment 20: Discussion over all is weak – for example table S1 should not be a supplement but should be included, and discussion on how the data was

Reply 20: We agree that the discussion in our manuscript could be strengthened, and we have made several revisions to address the issues you have raised. We have included Table S1 in the main body of the manuscript and provided a more detailed explanation of how the data was preprocessed and analyzed.

Changes in the text: Page 5, Line 122; Pages 14-16, Lines 331-392.

Comment 21: I don't think I could replicate the study by what you have provided. Data availability – provide direct links to the data. Code is not provided. 13-gene prognostic signature is the title of the paper but the code for the signature is not provided – how is it supposed to be implemented by others? Was the goal of this paper really to produce a signature or is it exploratory biomarker analysis? Recommend the author review other papers from highly reputable journals that report ML signatures to see how this is more commonly approached.

Reply 21: We acknowledge that we could have provided more information on data availability and code for the 13-gene prognostic signature. In this revision, we have signed the Data Sharing Statement that we can provide the data and make our code available to ensure that our study can be replicated by others. The link will be provided by the editor.

The goal of this paper is to develop a prognostic gene signature for predicting patient survival and treatment response in cancer. Our approach involves using machine learning algorithms to identify a set of genes that are highly correlated with patient outcomes and then constructing a signature based on these genes. The approach was adapted from a high-quality paper published in Nature Communications [Liu Z, Liu L, Weng S, et al. Machine learning-based integration develops an immune-derived lncRNA signature for improving outcomes in colorectal cancer. Nat Commun 2022;13:816.].

Changes in the text: Page 6, Line 130-143; Page 16, Line 406.

Comment 22: There is no mention of how you controlled for overfitting – yes you mentioned the word overfitting in line 236, but if you are creating multiple LASSO models (you made 97!) there is still a strong possibility to overfit even though LASSO is 'lower risk' of overfitting.

Reply 22: We acknowledge the importance of controlling for overfitting in our LASSO models, and we have taken several steps to do so. We have used cross-validation to evaluate the performance of our models and to select the optimal hyperparameters, which can help to prevent overfitting. We have also used a penalty parameter to shrink the coefficients towards zero and to reduce the risk of overfitting. However, we acknowledge that there is still a possibility of overfitting, given that we created multiple LASSO models. In the revised manuscript, we have provided a more detailed explanation of how we controlled for overfitting and the limitations of this approach. **Changes in the text:** Page 14, Lines 352-356.

Comment 23: Your study would be better supported by a proper comparative analysis

Reply 23: We have compared our prognostic model with other 40 published models and discussed the results.

Reviewer B

1. Please check all abbreviations in the abstract, highlight box, and the main text. Abbreviated terms should be full when they first appear.

E.g.,

LASSO, GBM, CRC in Highlight box

ROC in **Introduction**

Reply: Thank you for your suggestion. In this revision, we have checked all abbreviations in the abstract, highlight box, and the main text, and supplemented the missing abbreviations.

Changes in the text: Page 3, Line 65 and 68; Page 5, line 112.

2. Table 1

Please revise this typo.

unknow
T stage, n (%)
T0
T1
T2
T3
T4
Tis
unknow
N stage, n (%)
N+
N0
N1
N2
N3
unknow
M stage, n(%)
M0
M1
MX
unknow

Reply: We apologize for the spelling mistakes. In this revision, we have modified "unknow" into "unknown" in the Table 1.

3. ALL abbreviations used in each table/figure or table/figure description should be defined in a footnote below the corresponding table/figure. Please check all figures/tables and provide them correspondingly.

E.g.,

CRC in Table 1

CRC in Table S1

TCGA in Figure 1

OS, CRC, TCGA, LASSO, KM, ROC in Figure 2

OS, CRC, KM, TCGA, ROC, in Figure 3

OS, CRC, ROC in Figure 4

KM, OS in Figure 5

COAD, READ in Figure 8

Reply: Thank you for your suggestion. In this revision, we have added abbreviations used in each table/figure or table/figure description.

Changes in the text: Page 20, Line 539; Page 21, line 546; Page 23, line 556-559; Page 25, line 571-572; Page 26, line 579; Page 27, line 588; Page 30, line 628-629.

4. Since there is only one part in Figure 7, please revise Figure 7A to Figure 7

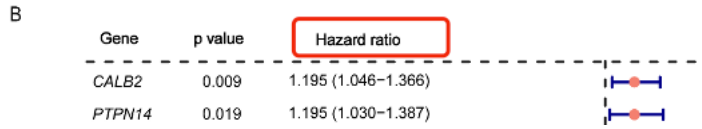
320 inhibit half of the measured biological activity. As shown in Figure 7A, the IC50

Reply: Thank you for your suggestion. In this revision, we have modified Figure 7A to Figure 7.

Changes in the text: Page 13, Line 322.

5. Figure 1B

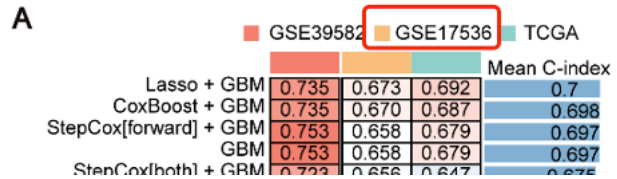
Please add "(95% CI)" after "Hazard ratio".



Reply: Thank you for your suggestion. In this revision, we have added "(95% CI)" after "Hazard ratio" in Figure 1B.

6. Figure 2A

GSE17535 or GSE17536? Which one is correct? Please check and confirm.



546 CRC patients in the GSE39582, GSE17535 and TCGA datasets. (B and C) LASSO Cox regression

Reply: We apologize for the spelling mistake. The accession number GSE17536 is correct. In this revision, we have modified "GSE17535" into "GSE17536".

Changes in the text: Page 23, Line 550.

7. As for the special symbols, please explain their meaning in the legend.

E.g.:

Fig 3: *, **, ***, ****

Fig 4: *, ***

Fig 6A: *, **, ***

Fig 8: *

Reply: Thank you for your suggestion. In this revision, we have explained the meaning of those special symbols in the legend of Figure 3, 4, 6 and 8.

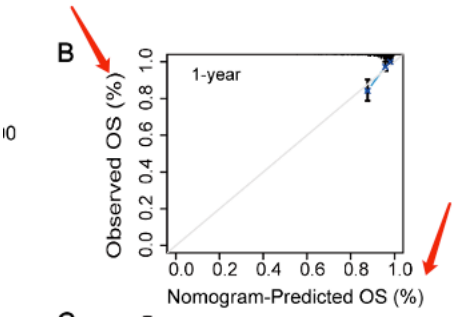
8. Figure 4B-4D

The correct format for the x-axis and y-axis should be one of the following, please revise.

a) If the description is with %, the numbers usually should be 0-100.

b) If the description is without %, the numbers usually should be 0-1.0. There's no need to indicate % in both description and data.

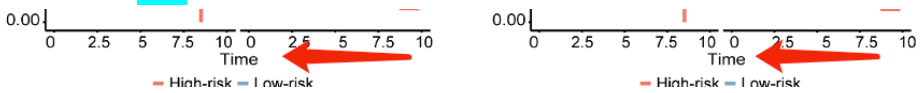
With the symbol %, the value of the data 0-1.0 will change a lot, e.g., 0.4% = 0.004, please check if the value of 0.00X in this figure is accurate.



Reply: Thank you for your suggestion. In the revised Figure 4, we have modified the titles of x-axis and y-axis into "Nomogram-predicted OS" and "Observed OS", respectively.

9. Figure 5B

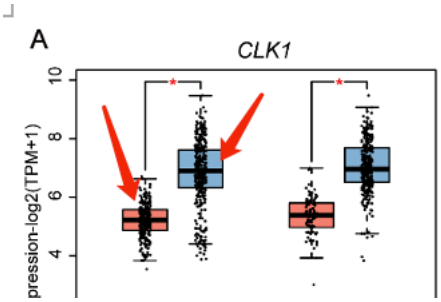
Please provide the unit for "Time".



Reply: Thank you for your suggestion. In this revision, we have added the unit for "Time" in Figure 5B.

10. Figure 8A

Please explain the meaning of red and blue in the legend/Figure.



Reply: Thank you for your suggestion. In this revision, we have added the legend of red and blue in the Figure 8A and 8B.