

Peer Review File

Article information: <https://dx.doi.org/10.21037/jmai-22-98>

Reviewer A	Response
1. All analyses were based on one dataset only without external validation data. Thus, this part should be improved, especially recently there have been a lot of public datasets for GBM patients.	Although there are a lot of public datasets of GBM patients, several datasets are limited clinical profiles, more genomic profiles for molecular research, or lack of consistent acquisition protocol and data collection. Therefore, we performed only our dataset in the present study.
2. The methodology is not novel. So, I'm wondering about the limited contribution of this study.	Because we selected features associated with prognosis by statistical analysis, multicenter trials (for the number of patients) or meta-analysis (for data pooling) will be improved the feature selection processes by statistical method. So, I revised this according to the reviewer's comment. (Lines 230-232)
3. When comparing the performance results among methods/models, the authors should conduct some statistical tests to see significant differences.	I revised it according to the reviewer's comments. (Line 166-168)
4. Uncertainties of models should be reported.	I added 95%CI in Table 3 according to the reviewer's comment. (Line 166-168)
5. Nomogram model should be added.	I added a nomogram in Figure 1 according to the reviewer's comment and revised the manuscript. (Lines 92-94 and 154-155)
6. How did the authors perform hyperparameter tuning of their models?	I revised it according to the reviewer's comments. (Line 106-110).
7. More references on machine learning-based glioma prediction should be added to attract a broader readership i.e., PMID: 35767281, PMID: 33735760.	I revised citations according to the reviewer's comments. (Reference 18,19)
8. The authors are suggested to conduct cross-validation during their training process.	I reported cross-validation in the revised manuscript according to the reviewer's comment. (Line 106-107).

9. Quality of the figures should be improved.	I converted our figure to 300 dpi according to the reviewer's comment.(Figure 1-4)
10. KM curves look strange. Normally, it displays 2 groups of patients (i.e., low-risk and high-risk) with p-values to show the significance.	The KM curves in Figure 2A are built from the survival data of both train and test datasets, and I added the p-value of the log-rank test to show insignificantly different prognoses between datasets. Additionally, KM curves in figure 2B are built from survival functions among various methods using a train data set. So, I revised the figure legend. (Figure 2)
11. Table 1 should contain some statistical tests and p-values.	I revised table 1 according to the reviewer's comment. (Table1)

Reviewer B	Response
1. Regarding the TRIPOD Checklist, here are some suggestions.	I send the TRIPOD checklist as attached file.
Major	
(1) From the title, this manuscript focuses on the comparison between Cox Regression and Machine Learning models. However, we fail to see the relevant descriptions in the Abstract, the authors just specify the results of various time-to-event ML models. It should be highlighted in the results and conclusion.	I revised it (abstract).
(2) The Method is incomplete. The development process of various models (Cox regression model, parametric survival model, and ML model) should be stated in detail, including the potential predictors' selection and measurement (Only stating the final predictive variables in the Results is insufficient).	I revised it (line 103-105, 110-118).
Minor	
(3) Please state explicitly in the title whether the study is a development model study, validation model study, or both. The same applies to the last sentence of the introduction.	I revised it (line 61-62).
(4) The keywords should be below the Abstract and usually include 3-5 words.	I revised it
(5) Please add the participants' source and main outcome measures in the Abstract-Methods.	I revised it (Abstract, method line 66-69,86-90).
(6) "The multilayer perceptron had the highest value of Harrell's concordance index,	I edited it (abstract).

<p>whereas the random survival forest model had the lowest root mean square error and mean absolute error for the predicted number of patients at risk over time". Specific values need to be given. In addition, why not report the predictive performance of Cox hazard regression and parametric survival models? It should be added to the results.</p>	
<p>(7) "Katzman et al. trained a neural network to generate individualized therapy suggestions that would extend the survival period of a group of patients." I assume this sentence is linked with the following content- "Because personalized prognostication is..." If so, please introduce personalized prognostication separately to avoid misunderstanding.</p>	<p>I edited it (line 52-55).</p>
<p>(8) "All newly diagnosed glioblastoma patients were hospitalized between 2007 and 2021 in a tertiary center in Southern Thailand". The dates should be specific to the month. And, in prognostic modeling studies, the duration of follow-up is also critical.</p>	<p>I edited it (line 66-68).</p>
<p>(9) Except for exclusion criteria, detailed inclusion criteria of participants should also be provided. Moreover, it's highly recommended to draw a flow chart to clearly show the selection process, and the eligibility criteria should also include. The flow chart is better placed in the results part.</p>	<p>I added the workflow of study as figure 1.</p>
<p>(10) "The ML was performed using the Python program version 3.8.7 with the PySurvival package26 (Python Software Foundation, USA)". It's suggested to provide full information or R code for reproducibility.</p>	<p>I added source code as supplement.</p>
<p>(11) If applicable, authors are advised to describe how the variables are handled in the analysis, including the method of defining cut-point values for converting continuous variables into categorical variables et al.</p>	<p>I edited it (line 84-86).</p>
<p>(12) For authors' kind reference, give definitions to some important outcome measures, such as overall survival, survival probability and median survival time.</p>	<p>I edited it (line 86-91).</p>
<p>(13) Please move the statistical analysis to the end of the Methods part. After this sentence-". In both the univariate and multivariable analyses,... select the final predictive model", please add that candidate</p>	<p>I edited it (line 96-97).</p>

<p>risk factors with $P < 0.10$ from the univariate regression analysis were entered into the multivariable regression model.</p>	
<p>(14) "Following the data split, 179 patients were used for the development of the ML model, and the remaining 56 cases were used to assess the model's performance". The patients' amount is inconsistent with the previous description (208).</p>	<p>I edited it (line 150-153).</p>
<p>(15) Please add the reference to support this sentence-" A population-based trial discovered that the median overall survival for TMZ with radiation and radiotherapy alone was 16.2 months and 9.0 months, respectively".</p>	<p>I edited it (line 216).</p>
<p>(16) Participants with missing data have been excluded from any analyses, at least the biased estimates of the model's predictive performance caused by this selection bias need to be discussed.</p>	<p>I edited it (line 260-262).</p>
<p>(17) Tables & Figures (a) In Table 1, please avoid reporting $P = 1$, $P > 0.99$ is fine. (b) Please add a detailed figure legend to explain Figure 1. (c) From Fig 3, the predicted graph of the Cox and CSF seems to be close to the actual graph as well. Why only report MLP and RSF?</p>	<p>I edited table1 and figures. I edited it (line 194-195).</p>
<p>(18) According to the author's instruction (https://cdn.amegroups.cn/static/public/2.1-Structure%20of%20Original%20Articles-template-V2022.11.4.docx?v=1679475672441) (a) It's strongly recommended to add a highlight box to highly summarize the key findings/recommendations, innovation, and potential implications of the study. (b) Introduction should be restructured into three parts: a) Background, b) Rationale and knowledge gap, c) Objective. (c) Similarly, the discussion is structured in five parts: a) Key Findings, b) Strengths and limitations, c) Comparison with similar researches, d) Explanations of findings, e) Implications and actions needed.</p>	<p>I added the highlighted file.</p>