
Peer Review File

Article information: <https://dx.doi.org/10.21037/jtd-23-1058>

Reviewer A

The utilization of the Cloud-LSSVM model appears to be a novel and promising method in dealing with uncertain problems in prognostic modeling.

The comprehensive analysis of large sample of patients provides a robust foundation for the findings.

Overall, this article presents valuable insights into utilizing advanced computational methods for prognostic modeling in ESCC. I believe this study will make a significant contribution to the field.

Response: Thank you for your positive feedback. We are submitting a revised manuscript to address these concerns. Detailed point-by-point responses to these concerns are provided hereinunder.

Comment #1

*119 The T, N, and M of all patients were staged according to the seventh edition of
120 The American Joint Committee on Cancer (AJCC) esophageal cancer staging protocol.*

Regarding the staging of esophageal cancer, the authors mentioned using the seventh edition of The American Joint Committee on Cancer (AJCC) esophageal cancer staging protocol. However, considering that the current standard is the eighth edition. I would recommend considering the utilization of the more recent edition.

Response: We appreciate the reviewer's careful review of our manuscript. We have modified our text as advised (see Page 5, line 168)

“The T, N, and M of all patients were staged according to the eighth edition of The American Joint Committee on Cancer (AJCC) esophageal cancer staging protocol.”

Comment #2

*144 The factors affecting prognosis are generally uncertain, and the probability theory and
145 fuzzy mathematics are insufficient to deal with the uncertain problems.*

Regarding this sentence, the word 'uncertain' is repeated twice in a similar structure. I would suggest considering variations in terminology to avoid redundancy.

Response: We appreciate the reviewer's careful review of our manuscript. We have modified our text as advised (see Page 5, line 98)

“...fuzzy mathematics are insufficient to deal with the this problem.”

Reviewer B

- 1) First of all, the authors need to indicate the development and validation of a prognosis prediction model by using LSSVM and comparisons between models based on LSSVM and other algorithms in the title. My further major concern regarding this study is the poor predictive accuracy of LSSVM, as indicated by the C-index, which are lower than 0.75. Because of this, I cannot see the necessity of the current study.
- 2) Second, the abstract needs some revisions. The background did not indicate the strengths of LSSVM in comparisons to other algorithms and why LSSVM-based prediction model is able to accurately predict the prognosis. The methods need to describe the inclusion of subjects, the assessment of potential predictors, follow up procedures, measurement of prognosis outcomes, and the generation of training and validation samples. The results need to briefly summarize the clinical characteristics of the two samples, and the survival rates in the whole samples. The conclusion should be more detailed since the comparative results of these algorithms may be sample-dependent.
- 3) Third, in the introduction of the main text needs a detailed review of known prognosis prediction models, have comments in their algorithms and predictive accuracy, analyze the limitations of algorithms used by prior studies, and explain why LSSVM is potentially accurate.
- 4) Fourth, in the methodology of the main text, please indicate the clinical research design of this study, describe how the SEER patients were followed up, the recruitment of the clinical sample and its follow up procedures, the generation of training and validation samples, and the prognosis outcomes to be predicted. In statistics, please indicate the threshold values of C-index for a good prediction model.
- 5) Fifth, several potentially related papers should be reviewed and cited:
 1. Deng J, Weng X, Chen W, Zhang J, Ma L, Zhao K. A nomogram and risk classification model predicts prognosis in Chinese esophageal squamous cell carcinoma patients. *Transl Cancer Res* 2022;11(9):3128-3140. doi: 10.21037/tcr-22-915.
 2. Lian L, Teng SB, Xia YY, Shen XM, Zheng Y, Han SG, Wang WJ, Xu XF, Zhou C. Development and verification of a hypoxia- and immune-associated prognosis signature for esophageal squamous cell carcinoma. *J Gastrointest Oncol* 2022;13(2):462-477. doi: 10.21037/jgo-22-69.
 3. Lin Y, Tang M, Liu Y, Jiang M, He S, Zeng D, Cui MY. A narrative review on machine learning in diagnosis and prognosis prediction for tongue squamous cell carcinoma. *Transl Cancer Res* 2022;11(12):4409-4415. doi: 10.21037/tcr-22-1669.

4. Liu C, Han J, Han D, Huang W, Li B. A new risk score model based on lactate dehydrogenase for predicting prognosis in esophageal squamous cell carcinoma treated with chemoradiotherapy. *J Thorac Dis* 2023;15(4):2116-2128. doi: 10.21037/jtd-23-388.

Response. Thank you for your positive feedback. We are submitting a revised manuscript to address these concerns. Detailed point-by-point responses to these concerns are provided hereinunder.

R1. Thank you for your positive feedback. The c-index values of SEER data and Linzhou database showed that the accuracy of Cloud-LSSVM model was higher than that of differentiation grade, random forest and nomogram, which indicated the accuracy of our model. The sample in this paper was drawn from the SEER database, which has only clinical data and inherently low c-index values. If we replace the database, c-index will increase to 0.79.

R2. Thank you for your positive feedback. We have modified our text as advised (see Page 2-3, lines 62-68 and 73-79 and 91-93)

“We studied 4,771 ESCC patients(training samples) from the Surveillance, Epidemiology, and End Results (SEER) database and 635 ESCC patients(validation samples) from the Henan Provincial Center for Disease Control and Prevention (HCDC) database, with the same exclusion criteria and inclusion criteria for both databases, and obtained permission to obtain a research data file in the SEER database from the National Cancer Institute.”

“The overall median survival time of the SEER database was 14 months (HCDC samples was 46 months), the mean survival time was 26.5 months (HCDC samples was 36.8 months), and the 3-year survival rate was 65.8%. This is because most of the patients with Henan samples are early ESCC, and most of the seer patients are T3 and T4 people.”

“Due to the difference of clans between training samples and test samples, the accuracy of prediction is generally not high, but the accuracy of Cloud-LSSVM model is much higher than other models.”

R3. Thank you for your positive feedback. We have modified our text as advised (see Page 4, lines 116-122).

“Machine learning is widely used in cancer research, mainly in cancer diagnosis, image recognition, prognosis prediction and so on. While the results are encouraging, machine learning has its limitations, such as being less sensitive to missing data and simpler algorithms, which are usually used for text classification.(15) The amount of data that

neural networks need to process is too large to handle multi-dimensional data. Although the nomogram is simple and easy to use, it has defects in processing continuous variables. (16-18)”

References

15. Lin Y, Tang M, Liu Y, Jiang M, He S, Zeng D, Cui MY. A narrative review on machine learning in diagnosis and prognosis prediction for tongue squamous cell carcinoma. *Transl Cancer Res* 2022;11(12):4409-4415.
16. Deng J, Weng X, Chen W, Zhang J, Ma L, Zhao K. A nomogram and risk classification model predicts prognosis in Chinese esophageal squamous cell carcinoma patients. *Transl Cancer Res* 2022;11:3128-3140.
17. Lian L, Teng SB, Xia YY, Shen XM, Zheng Y, Han SG, Wang WJ, Xu XF, Zhou C. Development and verification of a hypoxia- and immune-associated prognosis signature for esophageal squamous cell carcinoma. *J Gastrointest Oncol* 2022;13(2):462-477.
18. Liu C, Han J, Han D, Huang W, Li B. A new risk score model based on lactate dehydrogenase for predicting prognosis in esophageal squamous cell carcinoma treated with chemoradiotherapy. *J Thorac Dis* 2023;15(4):2116-2128.

R4. Thank you for your positive feedback. We have modified our text as advised (see Page 5, lines 163-168 and 18-189)

“The overall survival estimate registered in the SEER database is the “cause-specific classification of death”, and stratified “dead (attributable to this cancer dx)” or “alive or dead of other cause”. Survival time was calculated from the diagnosis date to the date of death or last contact. The last contact, or the cut-off date of the study, was December 31, 2019, which was the last date of update on the follow-up time. ”

“The C-index between the predicted probability and the actual outcome was calculated to judge the prognostic accuracy of the model. In this paper, C-index less than 0.65 was considered as low accuracy, and C-index more than 0.65 was considered as high accuracy.”

R5. Thank you for your positive feedback. We have modified our text as advised (see Page 13)

References

15. Lin Y, Tang M, Liu Y, Jiang M, He S, Zeng D, Cui MY. A narrative review on machine learning in diagnosis and prognosis prediction for tongue squamous cell carcinoma. *Transl Cancer Res* 2022;11(12):4409-4415.
16. Deng J, Weng X, Chen W, Zhang J, Ma L, Zhao K. A nomogram and risk classification model predicts prognosis in Chinese esophageal squamous cell carcinoma patients. *Transl Cancer Res* 2022;11:3128-3140.
17. Lian L, Teng SB, Xia YY, Shen XM, Zheng Y, Han SG, Wang WJ, Xu XF, Zhou C. Development and verification of a hypoxia- and immune-associated prognosis signature for esophageal

squamous cell carcinoma. *J Gastrointest Oncol* 2022;13(2):462-477.

18. Liu C, Han J, Han D, Huang W, Li B. A new risk score model based on lactate dehydrogenase for predicting prognosis in esophageal squamous cell carcinoma treated with chemoradiotherapy. *J Thorac Dis* 2023;15(4):2116-2128.