

Respond to reviewers' comments

Reviewer #1:

- 1. Authors constructed a prediction model of postoperative delirium and verified it using SMOTE. There are like typos in some key statements, rendering the conclusions contradictory to the date presented. RESULTS of Abstract indicated that P1 outperformed P2. In Model Comparisons of RESULTS, however, it was said that "there was no significant difference in performance". Yet again in Research Results of ARTICLE HIGHLIGHTS, it was stated that the model based on SMOTE outperformed traditional methods. Minor comments: 1) "P2" was not defined in Abstract. 2) "ICD-10 coding classification" -> "ICD-10; maybe provide a reference to ICD-10? 3) "willing participation" -> "willing to participate"? 4) Define abbreviations ICU, BMI, and VAS. 5) Is this manuscript a research article or a review? If it's a research article, why does literature search appear in "Data Collection"? 6) "nebulization"->"nebulizer"? 7) "technological roadmap" -> "flowchart"? 8) For the definitions of P1 and P2, please indicate units when applicable, e.g., for "Surgery duration". 9) Why were the p values so large for the correlation ($p=0.784$ for P1 and $p=0.990$ for P2)? What test were those p values for? 10) "1/019"->"1.019"? 11) "index scores" -> "index" 12) "can further mitigate" -> "may further mitigate"? 13) "the SMOTE oversampling (technique)" -> "SMOTE" 14) Ref. [28] missing journal name 15) Figure 1: Define abbreviations SMOTE&ROC in the caption 16) Figure 1: What is "Frequency of risk prediction models"?**

Answer:

- 1. Regarding Typos and Date Contradictions in Conclusions:** First, we appreciate the reviewer's attention to typos in our manuscript. We have meticulously reviewed and corrected all identified typographical errors throughout the article. As for the contradiction between the conclusion and the date mentioned, we have re-examined the relevant sections and made necessary corrections to ensure consistency and accuracy.
- 2. About the Comparison of P1 and P2 in the Abstract:** We acknowledge the potential misunderstanding in the abstract's description comparing P1 and P2. In the detailed sections of the article, we mentioned that there were no significant performance differences between P1 and P2. This misunderstanding might have arisen from the abstract's condensed expression. To avoid confusion, we have rewritten the abstract to more accurately reflect the comparison results between the two models.
- 3. On the Undefined Issue of "P2":** We thank the reviewer for pointing out this oversight. In the abstract of the paper, "P2" refers to the SMOTE-based logistic early warning model (P2).
- 4. Regarding the "ICD-10 Coding Classification" Suggestion:** Your suggestion to provide references for ICD-10 is very reasonable. The International Classification of Diseases, 10th Revision (ICD-10), is an international standard used for diagnosis and health condition classification. In our paper, we briefly introduce the basic concept of ICD-10 and provide relevant references or links for readers to further understand this

standard.

5. On the Expression “Willing to Participate”: This might have been a typographical error. Consistency in terminology throughout the text is essential. We have already made this correction in the main text.

6. About the Definitions of Abbreviations ICU, BMI, and VAS: Thank you very much for your reminder. ICU stands for “Intensive Care Unit,” BMI for “Body Mass Index,” and VAS for “Visual Analogue Scale.”

7. On the Paper Type and “Data Collection” Section in Literature Retrieval: This paper is an original article. The “Data Collection” section aims to express that no published papers were found through retrieval. It also seeks potential factors for further analysis and research through literature retrieval. To avoid misunderstanding by the readers, it has been removed.

8. Thank you for the reviewer's suggestion. Based on your feedback, I have revisited Figure 1 and confirmed that this figure indeed more closely aligns with the nature of a flowchart. Therefore, I have changed the legend from ‘Technological Roadmap’ to ‘Flowchart’ to ensure the figure more accurately reflects its content and purpose. Thank you for your guidance, which helps improve the clarity and accuracy of the paper.

9. Thank you for the reviewer’s suggestion regarding the specification of units. I fully agree on the importance of accurately expressing measurement data and results in scientific research. Therefore, I have reviewed and updated all relevant data points in the paper to ensure each measurement is appropriately annotated with its unit.

10. The large p-value issue mentioned by the reviewer actually reflects a common misunderstanding in statistical analysis. In statistics, the p-value is an indicator used to determine whether there is a significant difference between observed results and hypotheses. Generally, a p-value less than 0.05 is considered statistically significant, indicating a low probability of the results occurring by chance, thereby supporting the research hypothesis. In the main text, the p-values were 0.784 and 0.990, respectively, values far higher than 0.05, usually indicating no statistical significance. However, the main text also mentioned the values of R^2 (coefficient of determination), which were 0.349 and 0.355, respectively. R^2 is an indicator of model fit, showing the percentage of variance in the dependent variable that is explained by the independent variables. In some cases, even if the p-value is not significant, R^2 can still provide important information about the fit of the model. We thank you for inquiring about the statistical methods used in our paper. To ensure transparency and reproducibility of the analysis, we have detailed every statistical technique used in the methodology section of the paper.

11. Due to a mistake, "1.019" was written as "1/019," and it has been corrected in the main text. "Index scores" was changed to "index." "Can further mitigate" was altered to "may further mitigate." "The SMOTE oversampling (technique)" was changed to "SMOTE." Ref. [28] has been corrected to include the journal part. The abbreviations in Figure 1's title have been defined. The term "Frequency of risk prediction models" might be somewhat ambiguous, as it is not a common term in statistical or data analysis. In the context of risk prediction models, it refers to the frequency at which the model predicts an event (such as postoperative complications, disease recurrence, etc.) to

occur. For example, in medical research, a risk prediction model might be used to estimate the probability of a patient experiencing a certain complication within a specified time frame.