Peer Review File

Article Information: https://dx.doi.org/10.21037/jtd-24-1362

**Reviewer A:**

**RC1:** The title needs to clearly indicate the development and validation of the prediction models.

Reply 1): We have modified the title as advised.

Changes in the text: Page 1, line 3

**RC2:** In the abstract, the authors did not explain why the ML is potentially accurate for the prediction in the background, the methods did not describe how the training and validation samples were generated and how the potential predictors were identified, as well as how the hemorrhea and major cardiovascular adverse events were diagnosed, the results need to describe the training and validation samples and the accuracy parameters in both samples, and the conclusion needs comments for the clinical implications of the findings.

Reply 2): We have modified our text as advised.

In the 'Abstract' and 'Introduction', we added some statements and described the advantages of machine learning in reducing model errors.

Changes in the text: Page 2, line 35-41; Page 4, line 101-123.

In the two 'Methods', the training set and verification set were generated by overall randomization in proportion. The diagnosis of adverse events was made by experienced cardiologists during outpatient/inpatient follow-up, and the model did not directly diagnose adverse events, but only suggested the risk and the underlying risk features. The the potential predictors were identified by SHapley Additive exPlanation. SHAP showed the overall predictive feature ranking of the adverse events (Figure 4B/5B/6B) and the specific risk feature ranking belonging to a particular individual(Figure 4C/5C/6C). The 'Methods' part has been rearranged.

The baseline characteristics of the training and validation sets are shown in Table 3(appendix 1).

Changes in the text: Page 2, line 48-49/56-57; Page 5-6, line 153-182; Page 8, line 235-236[Table 3(appendix 1)].

The statements in the two 'Conclusions' have been adjusted.

Changes in the text: Page 3, line 71-74; Page 12, line 378-384.

**RC3:** In the introduction of the main text, the authors need to analyze the challenges in the prediction of hemorrhea and major cardiovascular adverse events, the limitations of available

prediction models, and why ML algorithm is potentially accurate.
Reply 3): We have modified our text as advised.
Changes in the text: Page 4, line 101-123.

**RC4:** In the methodology, I suggest the authors to do sensitivity analysis, that is to use one center data to develop the model and the other center data were used as the validation sample. The authors need to report the threshold value of AUC for a good prediction model.
Reply 4): The two centers we selected were two affiliated hospitals of Nanjing Medical University. The populations were all from Jiangsu Province, with similar diets and lifestyles, the same testing methods, diagnostic criteria, diagnostic, inclusion and exclusion criteria, so we merged the data of the two centers, and every data was independent and non-repetitive. We randomly allocated the merged data to the training set and validation set in proportion. Although the performance and generalization ability were weaker than the method described by the reviewer, the sample size during training and validation was guaranteed. Dong J et al. also trained the data with a larger sample size by merging the data of the entire cohort in their study(Dong J, Feng T, Thapa-Chhetry Bet al. Machine learning model for early prediction of acute kidney injury (AKI) in pediatric critical care. Crit Care. 2021;25:288.). In future studies, we will conduct multicenter prospective cohort studies in conjunction with other centers to verify and adjust our model.
In the 'Limitations', it has been stated.
Changes in the text: Page 1, line 350-359.

**RC5:** Finally, please consider to cite several related papers: 1. Liu J, Zhang Q, Liu Z, Wang X, Gong Y, Fan F, Zhang B, Jia J, Zhang Y, Liu Y, Zheng B, Li J, Huo Y. Microvascular reperfusion of fibrinolysis followed by percutaneous coronary intervention versus primary percutaneous coronary intervention for ST-segment-elevation acute myocardial infarction. Quant Imaging Med Surg 2024;14(1):765-776. doi: 10.21037/qims-23-666. 2. Du M, Ye X, Li D, Yang C, Dai R. Development of a prediction model for exercise tolerance decline in the exercise assessment of patients with acute myocardial infarction undergoing percutaneous coronary intervention revascularization in the acute phase. J Thorac Dis 2023;15(8):4486-4496. doi: 10.21037/jtd-23-554. 3. Xiao Z, Riletu A, Yan X, Meng Q, Zhang W, Zhang N, Ma C, Guo X, Han J, Nie H, Deng H, Liu J, Chen J, Dong Y, Liu T. Association of serum cystatin C level and major adverse cardiovascular events in patients with percutaneous coronary intervention. Cardiovasc Diagn Ther 2024;14(4):621-629. doi: 10.21037/cdt-23-482. 4. Ranasinghe S, Tjoe B, Shufelt C, Wei J, Lauzon M, Luu J, Asif A, Lewis J, Pepine CJ, Shaw LJ, Handberg E, Merz CNB. Association of abnormal electrocardiography response on dobutamine stress echocardiogram with longer-term major adverse cardiovascular events in women with symptoms of ischemic heart disease. Cardiovasc Diagn Ther 2023;13(6):948-955. doi: 10.21037/cdt-23-296.

Reply 5): We have modified our text as advised. The above articles have been added to the 'References' after reading, thanks for the reviewers' suggestion.
Changes in the text: Page 4, line 101; Page 14, line 430-444.
At last, thank you for your careful evaluation of this manuscript.

**Reviewer B**

The manuscript titled "Prognostic prediction modeling study in acute myocardial infarction patients after percutaneous coronary intervention: hemorrhea and major cardiovascular adverse events" presents a comprehensive study aimed at developing and validating a machine learning model for predicting the risk of hemorrhea and major adverse cardiovascular events (MACE) in patients undergoing percutaneous coronary intervention (PCI).

Introduction
- The introduction could benefit from a more detailed discussion on how machine learning has been previously applied in similar contexts, providing a gap that this study aims to fill.
Reply: As of now, machine learning has gradually been applied to the prediction of adverse events after PCI. Previous studies researches paid little attention on patients' medication information while paying too much attention on the past history, laboratory data, ignoring the effects of medication on prognosis. In addition, few research focused on individualized risk factor identification. The order of the relevant paragraphs and some expressions have been adjusted.
Changes in the text: Page 4, line 119-123.

Method
- The section could improve by providing rationales for the choice of machine learning algorithms and why others were not considered.
Reply: This study aims to build a predictive model and explore the advantages and disadvantages of logistic regression and other machine learning models. Logistic regression model is chosen to explore the linear relationship, while other machine learning models can explore the nonlinear relationship. The relevant expressions have been added in the '**Statistical analysis'** section, and the order of '**Statistical analysis**' and '**Feature selection**' has been adjusted.
Changes in the text: Page 7, line 207-211.

- The handling of missing data and the decision to exclude features with more than 80% missing values could be better justified.

Reply: Relevant expressions have been added referring to the reviewer's suggestion below.
Changes in the text: Page 12, line 384-387.

Results

- The results section would benefit from a comparison of the performance of the XGBoost model with that of logistic regression in a more detailed manner.

Reply: Indicators such as accuracy, sensitivity, specificity were included in Table 2, and the net reclassification index (NRI, at the baseline level of logistic regression) was also involved. The results suggested that the machine learning models were all superior than LR, and XGBoost was synthetically optimal in the machine learning models. Baseline for NRI comparison was added.
Changes in the text: Page 8, line 224-225.

- It lacks a discussion on the clinical implications of the predictive features identified by the model.

Reply: In this study, three different outcomes were predicted, each outcome had different risk features, and SHAP had given many features in the results. The discussion of their (at the top of the lists) clinical implications was in the '**Discussion**' section.
Changes in the text: None.

Discussion

- The discussion could be enhanced by suggesting specific ways in which the model could be integrated into clinical practice.

Reply: We adjusted the position of the following paragraph '*The two centers we selected were two affiliated hospitals of Nanjing Medical University. The populations were all from Jiangsu Province, with similar diets and lifestyles, the same testing methods, diagnostic criteria, diagnostic, inclusion and exclusion criteria, so we merged the data of the two centers, and every data was independent and non-repetitive. We randomly allocated the merged data to the training set and validation set in proportion. Although the performance and generalization ability were weakened, the sample size during training and validation was guaranteed. Dong J et al. also trained the data with a larger sample size by merging the data of the entire cohort in their study(~~33~~32).*'(The citation order has already been changed.)
Another application, early clinical warning, is suggested in this citation. Relevant expressions has been added.
Changes in the text: Page 11-12, line 352-364, **line 360-364**.

- It does not sufficiently explore the implications of the findings for future research or how the model might be improved or tested in different settings.

Reply: Validation of multi-centers, cross-regions, and cross-populations, adjustment of features and model complexity has been described in the '**Limitations**' section. Description of the ensemble learning in '**Discussion**' section has been added.
Changes in the text: Page 12, line 365-373; Page 13, line 403-411.

Overall Comment:

1. Model Generalizability: The paper could address the limitation regarding the generalizability of the machine learning models. Since the data comes from specific medical centers within a defined time frame, the models might not perform as well with data from different geographic locations, healthcare systems, or patient demographics. An example improvement could be: "The predictive models were developed using data from XXX Hospital and XXX University Medical Center, which may limit their applicability to different healthcare settings or populations with varied demographic characteristics. Future work should focus on validating and potentially adjusting these models with data from a broader range of settings."
Reply: As suggested, we have adjusted the expression of point VII in 'Limitation'.
Changes in the text: Page 13, line 403-407.

2. Feature Selection and Model Complexity: The paper does well to describe the predictors used in the model but could benefit from a discussion on the limitations related to feature selection and model complexity. For instance, the inclusion or exclusion of certain predictors might have implications for the model's interpretability and clinical utility. An example addition could be: "The selection of 53 clinical features for model training was based on availability and presumed relevance to the outcomes of interest. However, this approach may have overlooked potentially influential predictors not included in our dataset. Additionally, the complexity of our models, particularly with a large number of features, might limit their practical application in clinical settings where simpler models may be preferred for ease of use."
Reply: First, we added limitations to the application of the online calculator in clinical settings regarding the complexity of the model in '**Discussion**' section. Second, we added the influence of clinical feature selection on model prediction in '**Limitation**' section(point I).
Changes in the text: Page 10, line 319-322; Page 12, line 388-391.

3. Impact of Missing Data on Model Performance: While the paper mentions handling missing data through deletion or imputation for features with less than 80% missing values, it could elaborate on how this approach might have impacted the model's performance and predictions. An example elaboration might be: "Our approach to managing missing data, including the deletion of features with high levels of missingness and imputation for others,

might have introduced biases or inaccuracies in the models. The extent to which these methods influenced model performance and predictions remains an area for further investigation."

Reply: Further studies on the treatment of missing values and thresholds need to be further confirmed in the future, and relevant expressions have been added referring to the reviewer's suggestion.

Changes in the text: Page 12, line 384-387.

4. Temporal Changes and Model Relevance: The study covers data from January 2007 to January 2022. The paper could discuss how changes in medical practices, technologies, and patient characteristics over this period might affect the models' relevance and accuracy over time. Suggested addition: "The span of our dataset across fifteen years incorporates a period of significant evolution in medical practices and technologies related to PCI and AMI treatment. As such, the predictive performance of our models may vary when applied to more recent patient cohorts or future cases, necessitating periodic reevaluation and updating of the models to maintain their accuracy and relevance."

Reply: The impact of the time span on the forecast is indeed not taken into account, and multi-center external validation, subgroup analysis by time period, is needed in the future. Relevant expressions have been added referring to reviewer's suggestion.

Changes in the text: Page 13, line 407-411.