To Reviewer #2

General comments: *The authors presented a comparison of bagging, boosting and stacking ensemble methods to evaluate the landslide susceptibility mapping in the Three Gorges Reservoir area of China. Although the manuscript is well written, but the presented methods and presentation of data alacks novelty. One can easily guess that which models work better in the start without further reading the contents, as these has been done in several previous studies. From a readers viewpoint, I would like to see real discussion of science; however which is missing in this paper.*

Response: We thank you for your recommendation and valuable comments, which have ultimately improved this manuscript. We greatly appreciate your extensive and thoughtful review of our manuscript. According to your comments, we have made point-by-point corrections, which we hope will meet your approval.

Point-by-point responses to your detailed comments:

1.Comment*: For example, the authors analysed 25 factors (and surprisingly to me this do not find a collinearity problem); how these factors contribute to the landslides in TGD area?. I also see that feature importance is signifcantly high for Altitude. Some other important factors do not contribute at to the model as well such as the slope. Explanation of that adds to discussion.*

Response: Thank you for your careful insights. Through the feature importance analysis, the paper concludes that altitude, terrain surface texture (TST), distance to residents, distance to rivers and land use are five important factors that affect the occurrence of landslides. Statistical landslides are mainly distributed in which range of each factor and how the factor causes the landslide to occur. Due to the large number of evaluation factors, only the evaluation factors that have a greater impact on the occurrence of landslides are analyzed.

2.Comment*: Also, authors have computed the results of zoning in Table 3,4 and 5. What is their meaning to a reader ?*

Response: Thank you for your careful insights. For readers, from Tables 3, 4, and 5, it can be seen that the number of grids, the number of landslide grids, and the frequency ratio of the three models at each susceptibility zoning level at different grid sizes. A high frequency ratio means that more landslide grids are divided under this susceptibility level, and a high frequency ratio of the extremely high-prone areas and high-prone areas indicates that the model predicts better.

3.Comment*: Additionally, authors hsould validate the model not in the training site. They should have choose the adjoining catachments to check whether the result still hold valid (Like AUC of 0.95).*

Response: We thank you for your valuable comments. The model accuracy rate in Table 6 includes the accuracy rate of the training data and the accuracy rate of the test data, and the test data does not participate in the model training.

4.Comment*: Again, the comparisojn of 30-60-90 m grid size is inappropriate, as these are again known from several past works.*

Response: Thank you for your careful insights. It is impossible to know which evaluation unit is suitable for landslide susceptibility evaluation in this area without comparing the evaluation units of different sizes. After comparing the 30-60-90 m grids, we found that with the increase of the grid size, the phenomenon of model overfitting will be more serious, so we chose the 30m grid for better prediction results.

Thank you very much for your insightful and detailed comments.