General comments

The manuscript focuses on changes in national-scale rainfall-induced landslide susceptibility in New Zealand under the background of climate change. The topic has practical significance and is also within the scope of NHESS regarding natural hazards, climate change impacts, and risk assessment. Based on landslide data triggered by Cyclone Gabrielle, the author developed a LightGBM model and combined topographic, geological, soil, vegetation, and rainfall factors to predict landslide susceptibility under current and future SSP scenarios. Overall, the research idea is relatively clear, and the framework also has certain application value. However, the current manuscript still has several key methodological issues. The landslide inventory is mainly constructed based on NDVI-derived bare-ground change and a slope threshold, which introduces certain uncertainty in sample classification. The model training mainly relies on samples from the single Cyclone Gabrielle event and a limited region, but the model is then extended to the national scale and to different future storm scenarios, while its generalization ability has not been sufficiently validated. Important influencing factors such as antecedent soil moisture and aspect were not included. The resampling of HIRDS rainfall data from 2000 m to 25 m may also introduce scale mismatch and spatial bias. In addition, although the ROC-AUC is high, the PR-AUC for the landslide class is relatively low, and at the 5% threshold, false positives are clearly more numerous than true positives, indicating that the model has a clear tendency toward over-prediction. Therefore, I believe that the current results are not sufficient to support the main conclusions of the manuscript regarding changes in landslide susceptibility at the national scale and under future climate scenarios. The author is advised to consider resubmission only after further improving the landslide inventory, adding validation using independent events or independent regions, conducting sensitivity analyses of thresholds and prior probabilities, and fully discussing the uncertainty of future rainfall data. For the above reasons, I recommend that the current manuscript not be accepted.

Specific comments

1. The rationale for the construction of the landslide inventory and the selection of thresholds needs to be further explained. In the manuscript, Sentinel-2 images before and after Cyclone Gabrielle are used to identify bare-ground changes through NDVI changes, and then a slope >10° rule is applied to distinguish landslides from flood deposits and other areas. This approach is reasonable to some extent, but NDVI essentially identifies bare-ground change rather than landslides themselves. At the same time, the slope >10° threshold is an empirical threshold, which may affect which areas are ultimately defined as landslide samples. The author is advised to explain the basis for selecting the NDVI threshold and the slope threshold, and to further discuss whether different slope thresholds would lead to obvious changes in the number of landslide samples, their spatial distribution, and the model results.

2. The author combined the landslide data identified in this study with the landslide inventory from Dragonfly Data Science, and only used pixels identified as landslides in both datasets as positive samples. Areas identified by only one of the two datasets were classified as "possible landslides" and excluded from both training and testing. This approach can reduce misclassification, but it may also make the training samples more biased toward relatively obvious and larger landslides, causing the model to be more capable of identifying large areas of bare-ground change rather than necessarily representing all rainfall-induced landslides. The author is advised to further explain whether this treatment affects the completeness of the landslide samples, and whether excluding the "possible landslide" areas changes the spatial representativeness of the non-landslide samples.

3. Regarding the representativeness of the training area and the issue of national-scale extrapolation, the author is advised to provide further justification. The model is mainly trained using landslide samples from the area affected by Cyclone Gabrielle, but it is ultimately applied to the whole of New Zealand. Considering that different regions may vary in terms of topography, geology, land cover, and rainfall characteristics, whether a model trained only on this event and this region is sufficient to support national-scale prediction needs to be more fully explained.

4. The model training samples come from the single event of Cyclone Gabrielle. This event has its own rainfall path, duration, storm direction, antecedent wetness conditions, and regional spatial characteristics. Therefore, whether the model can be generalized to other rainfall events still needs further discussion. If there is no validation using other independent rainfall events, the author is advised to be more cautious when describing the applicability and generalization ability of the model.

5. Rainfall-induced landslides are usually affected not only by event rainfall, but also closely related to pre-storm soil moisture conditions. The author is advised to further discuss the possible impacts of the absence of antecedent soil moisture, and to consider whether antecedent rainfall indices, soil moisture reanalysis data, or other data could be used for supplementary testing or sensitivity analysis.

6. The author did not include the aspect variable, on the grounds that the influence of aspect on landslides may depend on storm direction, while future storm direction is unknown. This explanation is understandable, but for the specific training event of Cyclone Gabrielle, aspect and storm direction may still have influenced the landslide distribution. The fact that storm direction cannot be determined in future scenarios does not mean that aspect is completely unimportant in the current event-based modelling. The author is advised to further explain the possible influence of excluding the aspect variable on model interpretation and prediction results.

7. Regarding the accuracy and applicability of the future precipitation data itself, the author is advised to provide further explanation. Precipitation has significant spatial heterogeneity, and whether HIRDS data are reliable in representing the spatial differences in future precipitation needs to be clarified. In addition, the original resolution is 2000 m, and it is resampled to 25 m; this process may further introduce spatial bias. The author is advised to discuss whether different resampling methods, such as nearest neighbour, cubic convolution, and inverse distance weighting, would

have a significant impact on the landslide susceptibility prediction results.

8. The model results show an increase in landslide susceptibility under future warming scenarios, and this direction is reasonable. However, the magnitude of the increase may be affected by factors such as the probability threshold and the prior probability setting. The author is advised to further explain whether the magnitude of the future increase in landslide susceptibility remains stable under different thresholds or different prior probabilities.

9. Although the ROC-AUC is high, the PR-AUC for the landslide class is relatively low, and at the 5% threshold, false positives far outnumber true positives. The author is advised to make this point clearer in the interpretation of the results, and to further discuss the possible effects of this over-prediction on the landslide susceptibility maps, the estimated affected area, and the exposure assessment of buildings and roads.