General comments:

The problem of landslide volume estimation has been a focus for the community for quite some time, through methods such as area-volume scaling, geometrical modelling, numerical simulations, and more. This parameter is crucial as it helps gauge the magnitude of landslides, particularly at regional scales. Most highly accurate methods, like numerical simulations, often struggle at the regional scale. This manuscript offers a valuable reflection of data-driven modelling for delivering robust regional-scale analyses of landslide masses. Kudos to the authors for this interesting research, which has significant implications for hazard prediction and modelling. However, there are some major comments and curiosities I have. I believe the study is promising and of great interest to the landslide community, but it requires further work. The English language writing can be improved, especially in the Introduction. Some sentences read awkwardly and are hard to follow. Sentence phrasing must be improved to make the manuscript clearer, particularly for non-native English readers.

Specific major comments:

1. The Introduction needs to be revisited for editing in both grammar and phrasing of the language. Moreover, the motivation for the importance of volume quantification appears to be a bit lacklustre. I do not see a geomorphological connection as to why volume estimates are important to understand process mechanism and kinematics. Although, the manuscript does not explore said mechanism and kinematics expressions, however, to build a succinct story, a logical connection between the geomorphology and the surface failure should, in my opinion, be expressed to highlight why volume estimations are important as it directly feeds into the story of hazard prediction moving forward.

2. Are the training and testing datasets split randomly with keeping the training data fixed or is the split performed geographically? It would be interesting to see a geographically split dataset to see how well the model(s) perform due to apparent differences in the geological and environmental conditions across the study area.

3. One of my main concerns, or rather my curiosity, is regarding the data set itself. The volume information, along with the inventory, is particularly noteworthy in this case, as most inventories lack volume data. Keeping this in mind, how do the authors think about the application of such methods in other areas? Now, the authors have created a method that works pretty well within the given region. Instead of finding other regions (which might be difficult and time-consuming) could the authors simply use the model and predict volumes on similar nearby regions where the volumes are not calculated? This could serve as a simple prediction example demonstrating the method's application, without requiring extensive investigation. This approach is important as it helps the authors extend beyond a simple ‘exercise’ of the method, since it is currently applied only in the study area. Moreover, this would make the claim in Conclusion, Lines 346-349 more credible.

4. My biggest concern is related to the soil-depth. Now, it is impossible to imagine the calculation of volumes without the depth of the material that has failed as that is the 3rd

dimension required for volume calculations. It appears that the soil depth was 'removed' after feature importance analysis for the best performing EGB model. Sure, the depth information might not have been that important in this example of model training for this region, but I would argue that in other regions, particularly if the region contains multiple deep-seated landslides and the failure surface runs deep until the bedrock. I am just not convinced that removing soil depth makes sense, as geomorphologically, depth (which also relates to soil composition) is very important for accurate volume estimation and calculation.

5. Another question is pertaining to the type of failure movement. The inventory contains multitude of information but what about the movement types of the landslides? What types of landslides are considered in the inventory? Because clearly shallow and deep landslides would require separate treatments when looking at volume predictions because the material composition, material type, and material depths would be tremendously different. Do the authors combine these landslides together? What is the proportion of these landslide types? Also, are there prevalent debris flows, because volumes of debris flows is another story altogether since entrained volumes due to channelization are different than surface failure volumes. I see that the Discussion can be improved a lot by addressing and discussing these topics and limitations.

6. The Discussion section is oriented quite too much on the aspects of the different models, conditioning factors, and their roles in the prediction of the volumes. As I mentioned in my previous comment, not much is discussed on the practical questions of scalability, different modes of movements, soil depths, runout volumes of entrained materials etc. These are essential topics as the direct counterpart of statistical models, i.e., numerical models tend to answer these questions. So, a comparison with the literature in that order is missing which I believe would add new levels of arguments to put forward by the authors and cement why their method works well despite lacking/following physical laws.

7. In Table 1, under Geomorphology, the feature "erosion" is presented. Now, erosion itself can be referred to the volume, which is the main variable that the authors are trying to estimate. So, how is this variable used in the training regime? Or is this erosion feature different than the output of volume? Also, there are summary statistics of the erosion under Table 2. Why is that? My concern is that the authors are not clear as to what 'erosion' refers to in the data-driven model construct. If it is in fact similar to volumes, then the predictor variable and output variables are more or less the same. This needs further in-dept clarification.

8. Table 1: Descriptions should be written properly for each feature/variable. At the moment, the descriptions read more like a summary of the sub-groups, written altogether. Please provide descriptions individually for each feature properly. For example, Slope angle, slope aspect, and slope length are all written in one statement. Make them three individual statements to make it clearer to understand. Also, the descriptions are not clear enough. For

example, “There exists an established relationship between the slope morphology and volume landslide due to rainfall”. This is not a description. It is a reasoning to justify a claim. Please provide appropriate descriptions.

9. Lines 311-312: It would be nice explain why the random forest works well with smaller volumes. The connection between the machine learning predictions and the scale of the estimated volumes should be explained more intricately to provide a grounded understanding. Does the EGB model predict larger volumes more accurately than the rest, like Random Forest? If so, then why? Please explain these aspects.

Minor comments:

1. Line 31: “high”, should be “height”.
2. Line 36: “resulting volume of landslides”. Change this to “resulting surface failure”.
3. Line 38: “fragilize”. Not sure if such a word is used commonly to express the weakening of slopes. I'd rather opt for 'weaken'.
4. Similar English issues are found in Section 2 (Study area). Please address the language issues.
5. Figure 2. Font size of plot (b) is different than the rest, and also stretched. Please make all font sizes uniform.
6. Line 111: Replace ‘joined’ with ‘combined’.
7. Line 128: “flown away”? I am not sure if using this term is accurate. Generally, we refer to them as “removed material” from the surface. Can you please double-check this?
8. Is the slope angle the average angle of the terrain where the landslide was located or is the angle of reach? In my opinion, the angle of reach would make more sense as landslides that are closer to each other will exhibit different angles of reach but the same adjacent landslides would bear the same average slope angle as you are averaging based on the terrain. Please make it clear as to which one you have considered and why.
9. Line 136: What do you mean by ‘composing material’? This is not clear.
10. Lines 140-142: Please check the English grammar here. The sentence can be improved a lot.
11. Line 341: Change to “Among the tested models,”
12. Conclusion- Line 349: Change from “can be a better tool” to “can be a good tool”.