Review of Chou et al., Prediction Deep-Seated Landslide Displacements in Mountains through the Integration of Convolutional Neural Networks and Age of Exploration-Inspired Optimizer

Summary and Recommendations

This paper by Chou et al. describes an effort to test the sensitivity of various machine learning models on forecasting deep-seated landslide displacement over single-day and weeklong timescales. The authors utilize two sets of extensometer data that record landslide displacement at Lushan Mountain in Taiwan over a period from 2009-2017, along with four records of groundwater well data and satellite-derived temperature and humidity data. Over this time, the extensometer data record multiple pulses of movement that appear to correspond to peaks in groundwater levels, suggesting a connection to pore-water pressure increases via rising water tables. The authors employ their record of time series data to train a bevy of various AI models, and then from the two top-performing models fine-tine their hyperparameters using a newly released optimization algorithm (the Age of Exploration-Inspired Optimizer, or AEIO). The authors find that: 1) many models perform well in forecasting landslide displacement although there are tradeoffs between accuracy and computation time (impressively low errors from ~4-7% in the best cases); and 2) the AEIO algorithm successfully reduces uncertainty in their top models.

Overall, the authors present a clear description of the AI models used in the analysis and show convincingly that for the study monitoring sites machine learning algorithms can indeed be used to accurately forecast landslide displacement, even at the multi-day time scale. Showing that these methods yield a ~5% error on a seven-day forecast of landslide displacement is highly impressive and has obvious societal relevance. The AEIO method (complete with a very fanciful Fig. 8) does appear to work well in reducing the prediction uncertainty for the top-performing models. Therefore, I think the paper succeeds in showing the practical utility of applying optimized AI-based methods to this type of extensometer data and the benefits of running an optimization scheme on improving model performance. As presented, however, the manuscript feels somewhat lopsided as there is comparatively little information about the landslide itself and any in-depth analysis on connections from the model(s) to the results. For example, how much does the choice of input parameters impact performance? Are four groundwater datasets necessary, or would one suffice? Does including humidity data actually help improve model results, or is it extraneous? These are the types of questions worth discussing that may help yield more insight and understanding that may expand the utility of these results beyond the authors' study data (and thus would be of increased relevance to the global NHESS readership). Beyond these primary concerns, there are a number of smaller line-by-line technical and editorial comments I provide below that warrant addressing by the authors. If the authors can address these comments, I think this manuscript will make a useful contribution to NHESS.

Line-by-Line Comments

1: I'm not sure the phrase "in Mountains" is necessary here.

2: I believe the word "an" should perceive "Age of Exploration-Inspired Optimizer"

9: Nothing is done in this manuscript to show that deep-seated landslides are becoming increasingly frequent due to changing climate patterns. Is there a reference the author can provide that shows this in order to justify its presence in the abstract? This is certainly a nuanced topic as projected climatic changes may impact different areas (and thus landslide-triggering potential) differently across the globe, and therefore it is difficult to make these blanket statements.

11: insert "by" after "displacements"

25: There are certainly more than 378 landslides recorded worldwide between 1997 and 2017. Is this a specific subset of slides from this study? If so, a little more context needs to be provided here on what this number represents.

35: The 10 m threshold for defining a deep-seated landslide seems arbitrary. Dou et al. (2015) use 10 m as an example in their example sketch (their Fig. 5), but they do not reference this as a specific genetic guideline. Please use a more appropriate definition here.

41-42: This sentence feels out of place here since the paragraph is just discussing background. It would fit better in the final paragraph of this section outlining the goals of the specific study (i.e., lines 63-76)

54: editorial suggestion – can remove "In contemporary times"

55: CNN has not been defined before the introduction of this acronym

64: The term "predict deep-seated landslides" sounds vague. Predicting incipient failure? Reactivation of an already-established failure? Please specify.

65-66: Please list references of pre-existing work that you are referencing here

67: Impressive! At what depth is the failure plane for each of these extensometers?

88-92: This section feels quite under-referenced, as there are numerous theoretical and observational examples of groundwater impacts on deep-seated landslide failure.

93-94: This is another purely editorial comment, but the citation style presented here could be more succinct. For example, "Similarly, Preisig (2020) developed…" rather than "Similarly, Presig developed a groundwater prediction… (Presig, 2020)." This same style is utilized throughout the manuscript

103-105: In what way did Lin et al. "somewhat overlook" the importance of hydrological conditions in landslide formation here? Please be specific.

110 (Figure 1). Where is the actual landslide here? Below the diagram? I find the arrow below the right diagram very confusing and vague. A schematic failure plane perhaps informed by the borehole data would be useful for clarifying what it is the authors are trying to illustrate here.

122: Numerical models can simulate many scales, not just the laboratory scale. Please fix.

125: Does this Mufundirwa et al. reference also utilize a numerical model? If not, this paragraph should perhaps speak to both laboratory and numerical studies.

130: editorial – can delete "Meanwhile," here

135-136: What are "micro-units" here?

140-142: The previous paragraphs have not demonstrated that these "conventional methods have shown limited success in handling big data…" More information needs to be provided in this or the previous paragraphs to provide justification for this argument.

154: Is there any discussion on why this model was the most successful?

163-164: Please define what the term "feature engineering" is here

166: these parameters (topographic slope and soil parameters) don't necessarily have to be one-dimensional. Topography can be 2-D and soil parameters can be 3-D (and perhaps even time-dependent).

168-169: from my limited understanding of AI-based models, most are black boxes and therefore disentangling physical processes can be difficult. I thought this was the domain of physics-informed neural networks?

184: "predicting landslide displacement" would be more specific here

Section 3.1 (Lines 218-277): This part confused me at first because CNN's deal with imagery and you are using time series vectors. It is later clarified in the paper that the time series data are converted to images for use with the models, but it would be worth stating something up front that vector data can also be utilized in this construct with the proper transformation.

250 (Fig. 3): the 3x3 kernel illustrated here is mislabeled as 2x2

292: It's not clear here why RNNS are well-suited to learning time series with short-term dependencies. Please clarify.

318-322 (Performance Metrics): If you are assigning a separate section to performance metrics, it would be good to describe what each one is and the benefits and drawbacks for each metric.

328: What exactly is a particle in this instance? Some context is needed here.

337 (Fig. 8): The red arcuate arrows that link the positional strategies appears to suggest that once one strategy is selected, the explorer goes from one strategy to the next when in fact they return to the middle after each time step (correct?). If that is the case, then the arrows should arc back down to the central location to reflect the decision-making process that occurs with each positional change.

361-372: These two steps need to be elaborated on a little bit more, as it is presented somewhat confusingly and the equations for (8) and (9) are identical.

388 (Fig. 10): Much more information is needed in the figure caption here, as the current captions are essentially vacant. Additionally, the map in (a) is missing crucial information such as latitude and longitude graticules, and contains extraneous information (e.g., random text and other symbols that are not defined). With regard to (b), was the landslide failure plane identified with these cores? Or is the failure plane depth only known in the extensometer boreholes? Please provide more information here or elsewhere in the manuscript.

407: Please cite the previous research here

407-412: I think the term “cleavage” is misused here. Do the authors mean “fracture”? Typically, cleavage refers to the tendency of a mineral to break along planes defined by crystal lattice structure and are typically not seen at the scale of an entire hillslope. Lastly, it would be worth putting these observation zones on the map of the landslide for reference.

426: How was the rainfall data measured? Via a local rain gauge? If so, can put it on the map as well.

438-442 (Figs. 12-14): It is very difficult to compare the time series data as all the axes are scaled differently. I strongly recommend making one three-panel figure that is aligned in the time dimension instead of three separate figures. I would also recommend putting the known storms from Table 2 as vertical bands on each subplot. This will really help unify the datasets and make it much easier for readers to discern how precipitation, groundwater levels, and landslide displacement are aligned.

446-447: Should be “June” instead of “August”, otherwise the groundwater will be responding to a future event!

457-458: The groundwater levels that are apparently driving displacement here are 10s of meters below the ground surface (e.g., Fig. 14). Which impacts on soil structure by thermal processes are you referring to? Do thermal effects at this depth contribute to landsliding? Please provide context and references here to back up this statement or otherwise remove.

459-461: Please describe more the data used here? For example, is it daily data? What is the grid size? What is the measurement source?

488: Indeed! Having forecast data a week in advance would be extremely beneficial.

499: Specify process to be modeled (i.e., landslide displacement)

502: editorial comment – shouldn’t end sentence with “…”

509 (Fig. 16): Very helpful flow chart

545: It would be helpful to have a figure showing a subset of the models plotted alongside the displacement data so readers could see how the differences in MAPE are actually reflected in the time series predictions

568: change “landslides” to “landslide displacement” or something similar

660-664: This is a nice motivating paragraph that belongs in the introduction and would help provide context for the study.

668-668: Are these models not considered “conventional”? If not, why not? Could also be specified earlier on in the manuscript.

678: Here again would be a great place to delve into the “why” a little bit more. Any thoughts why a certain class of models outperforms the others? This discussion section is quite short relative to the rest of this paper, and there are a lot of aspects to potentially discuss. Does withholding certain parameters (e.g., temperature, humidity, or both) impact the results substantially? If so, why might that be the case? Since so much work has been done to get to this stage of predictive success, a small amount of additional work may help elucidate the role of specific processes in aiding the predictability of landslide displacement in this context that could be useful for the broad readership of NHESS.