RC3: Comment on EGUsphere 2024-2374 (Karsten Müller, 08 Nov 2024)

The authors evaluate the performance and interpretability of a random forest model trained to predict avalanche danger levels under dry-snow conditions. Their machine learning model aligns well with human forecaster assessments when used operationally alongside them, providing a higher temporal resolution in danger level assessments. However, the model underperforms in situations where persistent weak layers dictate the avalanche danger. A challenge in evaluating complex machine learning models, often referred to as "black- box" models, is discerning the rationale behind the model's outputs. The authors employ the SHAP (Shapley Additive Explanations) method to identify and explain which input parameters most strongly influence the model's predictions. By understanding the model's decision-making process, avalanche forecasters can better trust and validate its assessments. This interpretability allows forecasters to compare the model's analysis with their own, offering insight into potential discrepancies and prompting consideration of overlooked factors in their own assessments. Comprehensible model results can also be a valuable tool for forecaster training to showcase which parameters to consider given certain avalanche conditions. This work showcases how to increase the value of machine learning models for operational avalanche forecasting and similar operations. This study demonstrates how enhancing model interpretability can enhance the practical value of machine learning models in operational avalanche forecasting or similar operations. I recommend to publish this paper after considering my points listed in my detailed comments below. Kind regards, Karsten Müller.

Thank you very much for your review and constructive feedback on our manuscript. We greatly appreciate your positive remarks regarding the potential practical value of our study for avalanche forecasting. We will carefully address all the points outlined in your detailed comments to improve the manuscript. Thank you for your evaluation and recommendation to publish our work. Please find below the reply to your comments.

Kind regards,

Cristina Pérez-Guillén

Detailed comments

Paper title: Consider replacing "explainability" by "interpretability".

Thank you very much for your suggestion. However, we think that the term "explainability" is more appropriate in this context.

Line 10: "...though it decreased the performance during periods..." - the performance of what?

We will rephrase the sentence to improve clarity.

Line 111: remove "the" before 60% and "s" in predicts and remove "of" after danger level.

Thank you very much, we will correct this.

Line 136: "...with only one avalanche forecaster having access to the predictions." Do you mean "only one that is currently on duty" or "only one forecast from the entire forecasting group not necessarily participating in the assessment"? Please clarity.

Only Frank Techel from the avalanche warning team forecasters could access the model's predictions. We will rephrase this.

Table 1: the abbreviation, $\Delta D_{bu,a}$ is only explained in appendix. Please explain it here if you use it in the table caption.

We will add the explanation in Tables 1 and B1.

Line 244: "old snow problem (OS)" - consider replacing the term "old snow" by the EAWS standard name for the

avalanche problem "persistent weak layer (PWL)".

Thank you very much for your suggestion. We will replace "old snow" by "persistent weak layer (PWL)" in the new version.

Figure 6: Please consider using the same scales in plots a and b and c and d. Or add a note in the caption that different scales are used. "old snow" see comment to line 244.

We prefer using two scales for each avalanche problem to visualize the differences between regions better. We will add in the caption that different scales are being used.

Line 327 and following paragraph: It is not clear to me what is meant here - please consider rephrasing.

Thank you very much. We will rephrase it for clarity.

Line 339: Swap words "slope" and "virtual": ...for four virtual slope aspects...

Thank you very much. We will do it.

Line 352: Please state more clearly what do you mean by "both types"?

We refer to 'nowcast' and 'forecast' predictions with both types. We will rephrase it.

Line 371: put "Schweizer et al., 2020" at all in parenthesis.

Thank you very much. We will do it.

Line 373: "rapid increases and danger levels correlating with forecast in the bulletin. . . " Replace "forecast" with "danger level" or just leave it out.

We will do it.

Line 374: again, consider replacing OS problem by PWL problem.

Thank you for the suggestion. We will change it.

Line 405: have you considered using "minimum and maximum temperature or temperature difference during the day" or "strongest temperature change over six hours" as an input parameter to capture rapid temperature changes?

In the initial stages of model development, we tested the model by incorporating maximum, minimum, daily range, and differences for all meteorological features. However, as this did not improve performance, we decided to reduce the number of features and correlations by including only mean values. Nevertheless, features that capture short-term temperature changes could be valuable for testing in future model developments.

Line 431: please check this sentence. Is there a comma missing after stability indices?

We will modify it.

Line 577: remove reference to Techel et al. ... discussion.

We will remove it.