

The Cryosphere Discuss., referee comment RC2  
<https://doi.org/10.5194/tc-2022-80-RC2>, 2022  
© Author(s) 2022. This work is distributed under  
the Creative Commons Attribution 4.0 License.

## Comment on tc-2022-80

Ron Simenhois (Referee)

---

Referee comment on "Automated avalanche mapping from SPOT 6/7 satellite imagery with deep learning: results, evaluation, potential and limitations" by Elisabeth D. Hafner et al., The Cryosphere Discuss., <https://doi.org/10.5194/tc-2022-80-RC2>, 2022

---

Below are more specific comments for the manuscript.

Line 125: You mention the patches' size in several places in later parts of the manuscript. Maybe mention it here (pixel and ground cover) instead (except in 4.2) for better clarity.

Lines 127-133: Your method description of balancing the data is unclear. If I understand correctly, you under-sampled non-avalanche areas by considering only patches centered around an avalanche for 95% of the images. Is this correct? Please make the description clear.

Line 129: You use "First" where is second?

Line 146: Did you try the IoU loss function to reduce the effect of the unbalanced dataset?

Lines 150-151: did transforming shadow pixels' negative values  $v \rightarrow (-3 \rightarrow v^2)$  help? I don't see any reference to it in other places in the manuscript.

4.1 Results and generalization ability: This section should be reorganized for better clarity. You initially present your detection results on the test set and then introduce the previously unseen test set concept. I suggest that you swap the first and second paragraphs.

Table 2: some of the fields are in bold font. Why? Am I missing something?

Figure 6: For clarity, I recommend replacing "Avalanche Score" with the commonly used term: "model confidence"

Line 232: Do you think that the experts' difficulty with low light areas may cause lower quality data in shadowy areas and, as a result, reduce the model performance in these areas?

Table 4: Is this pixel-level comparison or detection of an avalanche?

Figure 8: The color scale here is unclear. What do the 1 to 5 values mean?

4.4 Limitations of this study section: Is this it? How about the experts' calcification

variation in dark areas that suggests that the data quality may be unclear?

Line 285: The F1-scores seem low... Is this increase in an F1-score? It doesn't add up otherwise... Please make it clear.