Review of "Autoencoder-based feature extraction for the automatic detection of snow avalanches in seismic data"

January 13, 2025

General comments

The manuscript by Simeon et al. presents a novel method to automatically detect snow avalanches in seismic data collected at a test site above Davos, Switzerland. Specifically, the performance of three different algorithms is assessed: seismic attributes, temporal autoencoder, and spectral autoencoder. Based on this, they find that the inclusion of features from the frequency domain improves model performance and unsupervised autoencoders show potential as an alternative to the standard expert-based seismic attributes classification.

I believe the science behind this study is sound and aligns with the focus of Geoscientific Model Development. However, I suggest some restructuring and clarification prior to publication. In particular, the Discussion section is quite long and, as pointed out by another referee, the manuscript still contains some repetitive information. Considering Provost et al. (2017) found higher true positive rates for non-windowed signals, I believe determining the effect of different window lengths would add value to the manuscript. At least the potential effects of choosing a different window length should be discussed. I recommend the authors also take my specific comments listed below into account.

Specific comments

L41 – *radius of several kilometres*: This becomes clearer later on in the manuscript, but I would suggest being precise from the very beginning.

L46 – *other types of mass movements*: Consider adding a few examples.

L47 – *other seismic sources* "such as earthquakes"?

L71: You describe what an encoder is, but not a decoder.

L80: Why did you use 10 s windows and not, e.g., 5 s, 20 s, or maximum length of picked avalanche events?

L91: What are the advantages of a *star-like* pattern compared to others?

L127 – *non-background noise signals*: having 912 non-background noise signals that then split into avalanche and noise events is a confusing terminology. Consider using a different term for *non-background noise signals*.

L139 – *Fig. 2b*: The labels of the panels are missing in Fig. 2.

L148 – *scores exceeded 1.5*: Consider changing to *was at least 2.0*. Readers just quickly skimming the manuscript might misinterpret it as ≥ 1.5

L167 – *even*: change to *roughly even*.

L167 to 169: You differentiate between dry and wet avalanches here but never address it in the results.

L173: Sect. 4.3 referenced before 4.2.

Fig 3: Could add brackets to the y-label clearly indicating which are training and test folds.

Fig 4: Only the encoder is shown. Is the decoder part of the autoencoder not used?

L194: Why is the decoder discarded?

L264: Could include these results as supplementary material.

L280 – *Gini information criterion*: Missing citation.

Fig. 7: Consider adding a colour bar.

Sect. 5.1: I recommend integrating Sect. C1 into the main text. Otherwise, it is difficult to interpret the results.

Fig. 8: It is not clear what the x- and y-labels are.

Fig. 10 – *up/down*: Change to *top/bottom*.

L331 and 385: Sect. 6.4 referenced before 6.3. I recommend to first discuss the missed avalanche windows and false alarms, and then the applicability to early warning systems.

L367: Same as the first sentence of the paragraph.

L378: *dotted blue line* in Fig. 11.

L409 to 411: It would be interesting to see how the issue of identifying avalanche onset evolves under different signal window lengths (e.g., trading slightly later detection for overall better performance).

L441: What could these unknown sources be?

L445: Compared to previous studies, why can the models discussed here better differentiate between avalanches and earthquakes?

L466: How do the approaches of Bessason et al. (2007) and Hammer et al. (2017) compare the this study?

L504: I recommend ending with your key takeaway.

I hope the authors find my comments helpful.

Sincerely, Kevin Hank