The authors have applied deep learning autoencoder models for the automatic and unsupervised extraction of features from seismic records. These extracted features were then used in classifiers to identify snow avalanches. This study presents a novel and relevant approach to enhance machine learning predictions, which could be useful not only for identifying snow avalanches but also for detecting other types of natural events. The overall methodology is well-defined, and the manuscript is well-written and easy to follow. I recommend the publication of this manuscript after the following issues are addressed

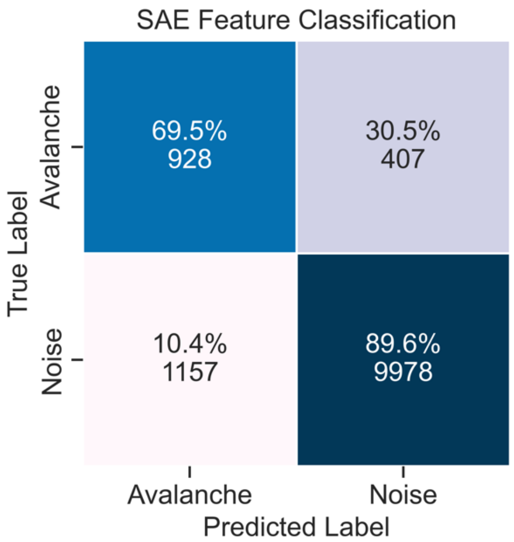
Major:

1) Given that the models tend to miss the onset of an avalanche, the authors should have included a scenario where only verified avalanches were used, excluding non-verified ones during the training of the autoencoders. While I am not suggesting that this must be incorporated in the revised version, as this conclusion emerged only after the study was completed, it is still worth mentioning.

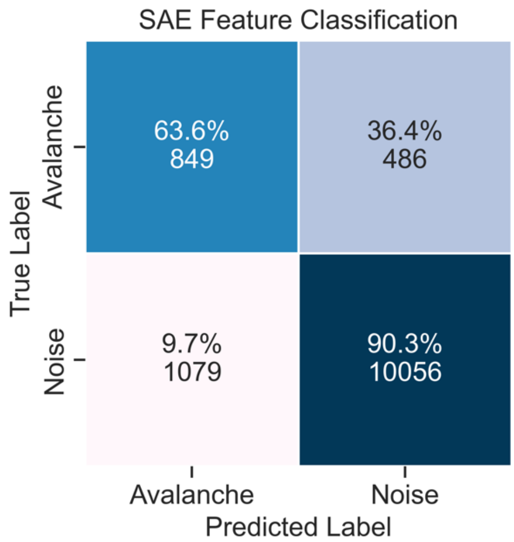
We would not expect the tendency to miss the onset of avalanches to be reduced when using only verified avalanches. All models tend to misclassify avalanche onsets as well as avalanche signals with a low signal-to-noise ratio. This suggests that the reason for missed onsets is rather found in the nature of mass movement signals and seismic recording. Avalanches are variable, moving sources of seismic energy, which attenuate significantly with distance. When an avalanche releases, the generation of seismic energy is typically low but increases as the flow moves downward due to the entrainment of mass and acceleration. The avalanche descent causes erosion processes, impacts with the terrain and the snow cover, and a final mass deposition, all of which are sources of seismic energy (Pérez-Guillén et al., 2016). Additionally, we expect an increase in seismic amplitudes over time due to a reduction in the source-receiver distance, as all avalanches approach the seismic array with time at our test site. Moreover, fully verified avalanches in our study are avalanches that were detected by the Doppler radar and/or verified with camera images. Some of them are small and thus, the signal-to-noise ratio is low. Installing a sensor in the avalanche release area would allow for recording the onset of avalanches with a higher signal-to-noise ratio, thus improving the performance of a model trained with this data. However, such a configuration would be limited to recording the onset of avalanches in the specific path where the sensor is installed, but not in all the avalanche paths of Dischma (Fig. 1), which can originate from different slope aspects and elevations.

Finally, unsupervised autoencoders are entirely independent of any class labels or information. Thus, by considering only verified avalanches, we would not reduce class ambiguity from the autoencoder's perspective but the dataset size and with it, valuable information might be lost. Nevertheless, we followed the reviewer's suggestion and retrained the spectral autoencoder and random forest model on only verified avalanches, i.e. avalanches that reached an expert score of 3. The comparison of both approaches is shown in the following figure. We indeed observe no improvement in the number of detected avalanche windows but a reduction.

Original results



Verified avalanches only



In conclusion, we are aware of this limitation and its significance for a potential early-warning system. For future studies aimed at developing an early-warning model, we would suggest examining the avalanche onsets in more detail and developing specialized models based on only these windows.
We will include parts of this reasoning and outlook in the final version of the manuscript.

2) How were the machine learning algorithms implemented, including details such as programming languages and libraries used?

The code is predominately written in Python using the PyTorch library for the autoencoder models, the random forest implementation of the Scikit-learn library, the Pandas library for handling the data and more standard Python libraries such as NumPy and SciPy. We will include this specification in the main text of the final manuscript under code and data availability.

Minor:

Line 218: "The best model from the cross-validation procedure (Table F2) was composed of convolutions with kernel size 20 (or 0.1 s) and stride 10. " Was the MSE (Mean Squared Error) the primary metric used for classification?

The mean squared error was used to develop, more specifically to train the autoencoders and optimize their reconstruction. Since autoencoders aim at reconstructing a given input signal, they require a reconstruction loss for training. The primary metric used to evaluate the classification was the avalanche class f1-score (line 286 in the preprint).

Line 228: "As an activation function, we use the leaky rectified linear unit (leaky ReLU; (Xu et al., 2015))". Was the activation function unchanged during the hyperparameter optimization process?

No, it was not. We included both the leaky ReLU and the Tanh activation function in the autoencoder optimization process.
We will clarify this in the revised manuscript.

Line 318: Please replace "This for" by "For this"

Yes, we will.

References:

Pérez-Guillén, C., Sovilla, B., Suriñach, E., Tapia, M. and Köhler, A., 2016. Deducing avalanche size and flow regimes from seismic measurements. *Cold Regions Science and Technology*, *121*, pp.25-41.