

Earth Surf. Dynam. Discuss., referee comment RC1
<https://doi.org/10.5194/esurf-2022-15-RC1>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on esurf-2022-15

Anonymous Referee #1

Referee comment on "Automated classification of seismic signals recorded on the Åknes rock slope, Western Norway, using a convolutional neural network" by Nadège Langet and Fred Marcus John Silverberg, Earth Surf. Dynam. Discuss.,
<https://doi.org/10.5194/esurf-2022-15-RC1>, 2022

Dear Editors,

Dear Authors,

The authors propose a comprehensive paper on the implementation of a Convolutional Neural Network-based supervised classification algorithm for the classification of microseismic sources on the Aknes landslide. After demonstrating the ability of the machine learning algorithm to correctly identify sources on a limited dataset, the authors use it to produce a new catalog of 60,000 events. The catalog analysis shows complex links between microseismic activity and potential environmental forcings.

This paper is very interesting because it proposes for the first time to use Convolutional Neural Network (CNN) on seismic signals recorded on unstable slopes. The analysis of the catalog produced is very exhaustive, relevant and brings very interesting elements to better understand the link between the micro-seismicity endogenous to a landslide and the external forcings, and the associated mechanisms. The article is very well written, easy to follow and understand, and the figures are of excellent quality.

Overall I find this article almost publishable as is. I nevertheless have some minor comments and questions, especially on the Machine Learning part. I detail my comments below:

L202-205 : I think the training set should be better described. In any implementation of supervised classification algorithm it is very important to know the exact number of events used to train and to test the algorithm, as those can greatly influence the performance of

the algorithm. Then the description here is confusing : first you say that « most classes constitute around 12% of the training set », but then that this is not the case for the HF SQ and RE classes. I would suggest adding two columns in table 1 with the number of events in the training and in the testing sets for each class.

L216-219 : Some previous studies also proposed features computed on the spectrograms, and they usually are amongst the best features for the classification (e.g. Provost et al., 2017 ; Hibert et al., 2017 ; 2019 ; Maggi et al., 2017 ; Wenner et al., 2021). This should be mentioned here. You propose a new approach based on the spectrogram but you are not the first to propose to use this data transformation and this should be acknowledged in your description here.

L222-227 : I don't understand the drawbacks described here. Features computation is not more complicated than the computation you do to transform your data into spectrograms. I understand that you need some arguments to tell the readers why you choose to work with spectrograms, but I don't think the arguments you propose here are valid. One could argue that your approach based on spectrograms is more susceptible to computation parameters, as the resulting spectrogram is completely controlled by the way you compute it (e.g. what is the influence of the spectral transformation you use? Between DFT, wavelets, Z-Transform, etc ? What about the window shape ? the window length ? the overlap ?). What would be the best option is the possibility to input directly the raw signal into the machine learning model, but I'm not sure you can do this with CNN.

I think that there is a simple argument in favor of CNN that you should make, which is that supervised machine learning algorithms based on curated features might miss critical information within the signals that CNN will find because it does not need any manual hence subjective definition of the features. CNN use images (most of those models at least), so in order to use them on seismic data you need to transform the signal into something that has the same properties as an image, which are spectrograms. This is largely sufficient to motivate the use of CNN and your study I think.

L249 : You choose to stack the spectrograms, but how is this influencing the global noise of your final spectrograms ? Would it be better to calculate the product of the different spectrograms ? I might be wrong, but by multiplying the spectrograms I think that you will reduce the noise and the influence of propagation effects while bringing out the part of the seismic energy generated by the source ? It might be worth testing in a future work.

L274 : For how many of those 59.608 events the classe has been confirmed manually ? The 2554 you included in the test set or more than this ? If more, what guided your choice of the 2554 events you have in your test set ? Are the events in the training set from those 2554 events ? A better description of your training and testing data sets is needed I think, as suggested in a previous comment.

L303-304 : So you did scan all the 59.608 events manually to remove the electronic

spikes ? See comment above. It should be very clear for the readers if the catalogue you interpret in the following sections is fully automatically made, automatic but fully manually controlled, or automatic but partially manually controlled. Provide numbers.

L400 & 414-415 : Would it be possible to process data with different sizes with other CNN implementations ?

This is a huge advantage of methods based on curated features (RF, SVM), they can work with signals with different durations. However this needs a pre-detection of the event, which can be tricky. This is why we start to see implementations of those approaches on moving windows (Wenner et al., 2021 ; Chmiel et al., 2021). Would it be possible to do the same using a NN such as AlexNet ? If so, it can be interesting to tell the readers in the discussion or in the perspectives how such an implementation could be done and what could be the difficulties.

L435 – 437 : Indeed. This sounds like it could be easily tested. What prevented you from doing so for this study ? Was it too costly in term of computation time ?

Conclusions : I found the last sentence/paragraph a bit vague and underwhelming. I think you have plenty of insights from this first implementation of a CNN that you should share with the readers. They should be highlighted in the conclusion.