Reviewer#5:

Dear reviewer#5, we deeply appreciate your feedback, which has enriched our work. We now focus on your suggestions.

The Authors apply different machine learning techniques to create, from a database of daily seismic registrations (if I understood well) of a certain volcanic area (Deception Island Volcano), the seismic catalogue of another volcanic area (Popocatepetl volcano) labelling the type of event following some criteria (not adequately and quantitatively described). To do this, the Authors use a high-quality database of seismic events, already labelled by human supervision, and collected in another volcanic area. The purpose is to reduce or eliminate the use of human work for labelling the seismic events in catalogue. The purpose is very important and interesting, since the increase of seismic networks in the recent years has the undebatable advantage of having increased the seismic monitoring both in volcanic and in tectonic areas but at the same time it has increased the number of data to be analysed by seismologists. So, an automatic system that can be able to detect and label seismic events in volcanic or tectonic context can be very useful if the system is reliable and it will reduce human supervision to a minimum or even it will eliminate it altogether, working as a human would do.

The Authors conclude that the three ML approaches produce different results and all of them are able to detect a number of events much greater than the existing catalogue (except in one case) and I think that this a very important and intriguing result.

I appreciate the work and the idea, but the manuscript has many problems that I try to list.

Reading the description of the work done, it is not clear how a researcher could verify the results and reply the work with its own database. The description of the method is very confused and only who has already used the same techniques could follow and understand the steps. Moreover, the description both of the method and of the data used is only qualitative and discursive, never detailed and quantitative.

We appreciate the reviewer's constructive comment. In the next version of the manuscript, we will aim to detail the proposed algorithm more thoroughly and in a more didactic manner, including pseudocode for easier understanding. Additionally, we will make the source code available to the scientific community.

The manuscript is full of non-useful repetitions and the Authors should do an effort to re-read the manuscript and be more concise. As an example: Line 281-294. This is a repetition of something already written in the manuscript. Line 170-184. This paragraph should be moved in the Introduction section and rewritten to avoid repetition.

The same acronyms are referred to with other acronyms. As an example, the three techniques employed to achieve the purposes of the manuscript (specified in Line 184-186) are referred to ANN or to ML in different part of the manuscript, generating a

great confusion among all the acronyms. I suggest simplifying and reducing the use of the acronyms to the strictly needed.

Once again, we agree with this constructive comment. We will strive to be more concise and avoid repetitions and redundancies in the new version of the manuscript.

Regarding the used database, the Authors use a qualitative language that does not help to understand. Some examples:

Line 115. Where the original labelled database can be consulted? Is it already released?

Line 134. What do the Authors mean with “subset of data considered the most reliable”? Which criteria did they use for reliability?

Line 139. As above, what do the Authors mean with “ the most representative and of the highest quality”? How do they measure the quality?

Line 149-154. Where the Popocatepetl 2002 catalogue can be found? A citation is missing here

Line 152. Which are the classes of event, adding a table can hep.

Line 229. The phrase “The target domain (denoted as Dt) is the Popo2002 dataset (whose available seismic catalogue will not be considered)” what does it mean? At line 149 the Authors state that Popo2002 consists of 4,883 events, what type of event they are? I can understand that the data used for the target domain is a subset of the Popo2002 collected excluding earthquakes. Is it true? I think that the authors should be more concise and clearer in describing both the technique and the data used.

All of these suggestions will be taken into account and included in the new version of the document to enhance understanding and readability. We are committed to providing a more detailed and straightforward description of these concepts. We will also include references to both catalogs.

In conclusion, I suggest publication after a deep rewriting of the manuscript that does justice to the work done, makes it understandable also to those who have never used the specified ML techniques before and makes the proposed method replicable for other interested scientists.