Reviewer#4:

Dear reviewer#4, many thanks for taking the time to review our work and provide such insightful advice. Below, we respond to your observations.

The authors have taken on an interesting and challenging topic, using machine learning to classify volcanic seismic signals. However, there are several important areas where the paper needs improvement to better explain the methods and show how this research can be useful for volcano monitoring and eruption prediction.

1. The paper focuses mainly on classifying seismic signals but does not clearly explain how this helps us understand volcano activity or predict eruptions. The authors should add more detail about how these results can be used in real-world volcano monitoring systems, including other important data types like geodetic (ground movement) and geochemical data. This would make the study more useful for predicting volcanic hazards.

   The aim of our work is to highlight an issue we have been observing for some time, which we believe needs to be addressed with great depth. There are many automatic seismic signal recognition systems in the literature, and almost all of them perform very well. However, the question is: Are the results of these models genuine, or do they simply reflect how well the systems have learned the information contained in the training catalogs? To explore this, we conducted a series of experiments that support our hypothesis. According to our results, when applying a system trained on a master database to a new database from a different volcanic environment, we find a significant discrepancy between the detected events and those annotated in the original catalog. Based on the approach followed by volcanological observatories, seismic catalogs describe the volcanic behavior or dynamics from a seismic perspective. A biased or incomplete catalog can lead to incorrect conclusions, and when comparing behaviors during future crises, these conclusions could also be biased. Therefore, obtaining more complete catalogs will help in understanding volcanic dynamics. As for the use of other types of data, this falls outside the scope of this study. It will be the task of the observatory to correlate information from different data sources and draw valuable conclusions. Without a doubt, this is an interesting idea for future work, but again we believe it is beyond the scope of this research. Additionally, we do not have access to the data to carry out such a data fusion approach. If the reviewer has access to these data, we sincerely express our openness to initiate the development of this idea through a scientific collaboration.

2. The paper does not provide enough information about the seismic data used. The authors should explain more clearly how they collected the data, what each type of seismic event means, and how the events were labeled. A table or figure

showing how many events of each type were found would help the reader understand the data better. The authors should also show examples of different signal types earlier in the paper to make it clearer how the classification works.

As we have previously mentioned to other reviewers, if this article is ultimately accepted for publication, in the next version, we will improve the description and visualization of the events that make up the different seismic catalogs.

3. The paper lacks detail about how the data was processed. For example, the authors mention using a bandpass filter (1-20 Hz), but they do not explain why. They should also explain which components of the signal were analyzed (e.g., vertical component) and whether the same stations and equipment were used for both volcanoes. Providing these details will make the study more transparent.

   We once again agree with the reviewer's suggestion. In this version, we did not provide a detailed explanation of the data used. If the article is accepted, we will aim to address this suggestion in more detail.

   Regarding the use of the band-pass filter between 1 and 20 Hz, we would like to clarify that this decision was based on expert knowledge of the problem and the parametrization scheme used. According to the source models and characteristics described in Table 1, volcanic seismic signals have discriminatory spectral content between 1 and 20 Hz. Therefore, the different types of events can be characterized based on the information within these bands. To do this, we applied a logarithmic scale filter bank, which increases resolution in the lower frequencies where different events exhibit distinct characteristics, allowing for more detailed analysis. Considering both premises, we believe the band-pass filter between 1 and 20 Hz is justified.

4. The authors suggest that their method can detect more seismic events, but they don't provide enough examples of how this would help in real-time volcano monitoring. It would strengthen the paper if they could show how these improved classifications lead to better volcano hazard assessments or warnings. Additionally, volcanoes can behave differently over time. It would be useful to see if the model works well over different eruption periods or at different volcanoes.

   We thank the reviewer for their comment, as it gives us the opportunity to openly propose collaboration with organizations or researchers who have access to different volcanoes and eruptive periods, allowing us to validate the robustness of our proposal. Regarding the improvement of real-time assistance, in addition to what was previously mentioned in question 1, we commit to including this information in the new version of the manuscript.