https://doi.org/10.5194/egusphere-2026-1201
Preprint. Discussion started: 28 April 2026

Self-Supervised Contrastive Learning in the Context of Volcano-Seismic Datasets

Joe Carthy[1,2], Manuel Titos[1,2], Flavio Cannavó[3], Luciano Zuccarello[4], and Carmen Benítez[1,2]

[1]Department of Signal processing, Telematics and Communications, University of Granada, Granada, 18014, Spain
[2]Research Center on Information and Communication Technologies of the University of Granada (CITIC-UGR)
[3]Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Pisa, Pisa, 56125, Italy
[4]Istituto Nazionale di Geofisica e Vulcanologia, Osservatorio Etneo, Catania, 95123, Italy

Correspondence: Manuel Titos (mmtitos@ugr.es)

Abstract. Volcano-seismic datasets are expensive to label due to the requirement for expertise to understand the signals and the time-intensive nature of extracting and labeling different events that are occurring. This work evaluates whether self supervised methods can enable volcanologists to gain knowledge about the content of volcanic datasets without the use of labels, or reduce the amount of labels required. The aim of this work is to compare several common techniques and illustrate their usefulness for the volcanic community, where labeled data is an even more precious commodity than the wider seismic community. Experiments have been performed on three real-world datasets containing isolated volcano-seismic datasets from Llaima volcano, Colima volcano, and Mount Etna. Time-Series Representation Learning via Temporal and Contextual Contrasting (TS-TCC) shows particularly high performance in this task for finding structures in an self-supervised fashion. This indicates the untapped potential of self-supervised training to aid in different data analysis tasks within the volcano-seismology community.

1 Introduction

There are many volcanic observatories across the world that are all generating vast amounts of seismic data on a daily basis. Understanding the contents of these streams of seismic data is critical for volcanic monitoring. It is possible to extract useful information to determine the evolution of the state of the volcanic system, such as if the volcano is entering a pre-eruptive state, or if there is a change in the activity levels by monitoring these signals (Chouet, 1996; Buurman et al., 2006; Scarfì et al., 2023). Robust machine learning models have been highly successful in aiding the monitoring of volcanic systems. The training of these models requires large amounts of data, and traditionally, a large number of labels for the model to learn about the contents of the signals. Large datasets have been made public thanks to work by IRIS and FDSN smith1987iris, suarez2008international; however, while there is a large amount of data available and a large amount of labelled general seismic data (Woollam et al., 2022), there is a lack of available labels for training supervised machine learning models on volcano-seismic tasks (Malfante et al., 2018; Cortés et al., 2019). This is in part because volcano-seismic datasets are expensive to label due to the requirement for expertise to understand the signals and the time-intensive nature of extracting and labeling different events that are occurring (Bueno Rodriguez et al., 2017); this cost makes it difficult for every volcanic observatory to produce robust seismic catalogs from scratch. There is potential for transferring a trained model to a new volcanic system (Lapins

https://doi.org/10.5194/egusphere-2026-1201
Preprint. Discussion started: 28 April 2026

et al., 2021; Münchmeyer et al., 2021; Münchmeyer et al., 2022; Titos et al., 2019; Bueno et al., 2019), however, there can be issues with model generalization in the context of volcano-seismic event analysis (Mousavi and Beroza, 2023). Studies have shown drops in performance due to changes in the type of volcanic activity, and when transferring to different volcanic systems or different network topologies (Malfante et al., 2018; Cortés et al., 2019). These facts motivate the pursuit of methods that can offer insight into volcano-seismic signals without the constraint of labels, or transferring trained models across different volcanic systems. Approaches using active learning have been previously explored to combat this issue, however there is a lack of studies using self-supervised learning in the context of event identification for volcano-seismic data (Manley et al., 2022).

This study investigates the potential of self-supervised learning (SSL) to build discriminative feature sets when analysing discrete volcano-seismic events without labels. Three different datasets are considered. Two of these datasets were hand-selected events from Colima and Llaima volcano, where the true labels are known and may be used to benchmark the methods. The third dataset is generated by automatically extracting events from a continuous volcano-seismic signal on Mount Etna. The Mount Etna dataset is analysed to highlight the capability of the method for analysing unknown data that contains a high degree of noise. Noise in this context refers both to the background noise expected in active volcanic systems, and noise introduced in the event selection process that may include cut-off events, or false positives introduced by the event extraction system.

Two different self-supervised training strategies are analysed: Sim-CLR (A Simple Framework for Contrastive Learning Visual Representations) (Chen et al., 2020); and TS-TCC (Time Series Representation Learning via Temporal and Contextual Contrasting) (Eldele et al., 2021). These strategies are selected due to their high performance for building representations in different tasks, with TS-TCC being designed specifically for time series signals. The representations created by each method have the potential to increase the clustering capabilities for seismic data. Each approach extracts features from the seismic time series directly. The machine learning models used within the SSL strategy are convolutional models (CNNs). These are selected due to their high performance in various tasks related to seismic and volcano-seismic signals (Lara et al., 2021; Curilem et al., 2018; Salazar et al., 2020). Our results indicate that the CNNs trained in using self-supervised learning have the potential to extract robust, discriminative features.

A clear performance increase is seen for tasks related to volcano-seismic event analysis. In particular, for the hand-crafted datasets, the models that have undergone self-supervised pre-training with TS-TCC exhibit high performance. This builds on previous work done illustrating that TS-TCC has the potential to discriminate well between micro-seismic events and noise (Yang et al., 2023), illustrating its potential on several more complex volcano-seismic datasets. For the Mount Etna dataset, where the event extraction is automated, the additional noise leads and lack of ground truth leads to closer apparent performance between the SimCLR and TS-TCC strategies. Both SSL strategies outperform the benchmark strategy and identify previously unidentifiable groups within the dataset.

The performance of the SSL methods is evaluated on the basis of the increased performance in visualisation and clustering tasks. These tasks are performed using the Uniform Manifold Approximation and Projection (UMAP) method to cluster the events (McInnes et al., 2020). The SSL methods generate a custom feature set that is input to UMAP. The capability to cluster the SSL features is benchmarked against the capability of UMAP to cluster the events using a combination of time-based and frequency-based input feature sets.

https://doi.org/10.5194/egusphere-2026-1201
Preprint. Discussion started: 28 April 2026

The main contributions of this study are as follows:

1. Contrastive learning is introduced and proposed for training convolutional networks to extract powerful representations of volcano-seimsic events. Focus is placed on two contrastive learning strategies: Sim-CLR; and TS-TCC. These strategies are selected due to their high performance for building representations in different tasks, with TS-TCC being designed specifically for time series signals. The representations created by each method have the potential to increase the clustering capabilities for seismic data.

2. The SSL models are shown to be capable of clustering events that are automatically extracted from a previously unseen dataset. This highlights the utility of these methods to the volcano-seismic community for discrete event analysis.

The work is arranged as follows. First, there is an introduction to the self-supervised learning methods used within this work. Second, there is an introduction to the volcano-seismic datasets under investigation in this work. Third, there is an investigation into effective CNN architectures for the encoders, and effective parameters for the self-supervised training. Fourth, the representations are evaluated based on the capability to cluster the data. Fifth, the results are discussed. Finally, the conclusion outlines the key points of this work.

2 Self Supervised Learning Introduction

Contrastive learning methods are a group of self-supervised learning techniques that train models to compress data, while retaining the critical characteristics that differentiate samples from different classes (Le-Khac et al., 2020). The goal of these methods is to build representations of the input data that are well-suited for down stream tasks (Jaiswal et al., 2021). These techniques have seen widespread use in different domains, such as sound and image recognition (Kothinti and Elhilali, 2022; Zeng et al., 2021; Saeed et al., 2021; Wu et al., 2022). They are also extremely popular within the SAR community due to the fact that they can train strong deep learning models even in the case where labels are either limited or unavailable (Bountos et al., 2021, 2022; Agastya et al., 2021). Contrastive learning methods are also gaining traction for feature extraction on seismological datasets (Murshed et al., 2024; Si et al., 2024; Li et al., 2023).

This work compares common self-supervised methods and illustrate their usefulness for the volcano-seismic community where labelled data is an even more precious commodity than the wider seismic community. Two different contrastive methods are considered, SimCLR, and TS-TCC. These methods represent frameworks for training neural network encoders. A 1D convolutional encoder structure is utilised due to this architecture's high performance on volcano-seismic tasks. A general contrastive workflow, the SimCLR framework, and the TS-TCC framework are illustrated in Figure 1, with pseudocode detailing the steps for each training framework available in Appendix A.

2.1 Contrastive Learning Background

The basis for instance-wise contrastive learning is that when an encoder embeds two views or augmentations of the same data sample, both embeddings should be located close to eachother. The embeddings of different samples should be located

https://doi.org/10.5194/egusphere-2026-1201
Preprint. Discussion started: 28 April 2026

further away, particularly in the case where the samples belong to a different class. This method can be well described using the concept of an anchor event alongside a positive and a negative pair. Given the anchor x, we select a positive event, x^+ and a negative event x^-, where the positive sample and the anchor belong to the same class and the negative sample belongs to a different class. A loss function can be constructed to maximize the distance between the anchor, x, and the negative, x^-, and to minimize the distance between the anchor, x, and the positive, x^+, using the triplet loss equation (Schroff et al., 2015):

$$L_{\text{triplet}}\left(x, x^+, x^-\right) = \sum_{x \in X} \max\left(0,\ \left\|f(x) - f(x^+)\right\|_2^2 - \left\|f(x) - f(x^-)\right\|_2^2 + \alpha\right) \tag{1}$$

where α is a margin enforced between positive and negative pairs.

A natural extension of this method is a loss function that accounts for multiple negative pairs for a given positive pair. The Normalized Temperature-scaled Cross Entropy Loss (NT-Xent) allows for this case (Sohn, 2016; Chen et al., 2020). This loss function is defined in Equation 2.

$$\lambda_i^j = -log\left(\frac{\exp(sim(x_i, x_j)/\tau)}{\Sigma_{k=1}^{2N} 1_{[k \neq i]} \exp(sim(x_i, x_k)/\tau)}\right) \tag{2}$$

where $sim(u, v) = u^T v / ||u|| ||v||$ represents the cosine similarity between two vectors, $1_{[k \neq i]}$ evaluates to 1 iff $k \neq i$ and τ is a temperature parameter to be tuned to the problem. This loss function is computed across all positive pairs (i, j) and (j, i) within a batch.

In effect, this loss uses a normalized cross-entropy method that applies a penalty when the similarity of the positive pair is smaller than the sum of similarities across the negative pairs. It is possible to extend to the scenario where there are multiple positives pairs, however this is outside the scope of the current work (Frosst et al., 2019). In this study, SimCLR is used due to its high performance on tasks in computer vision and its related adaptations for time series analysis (Chen et al., 2020; Eldele et al., 2021).

2.2 SimCLR

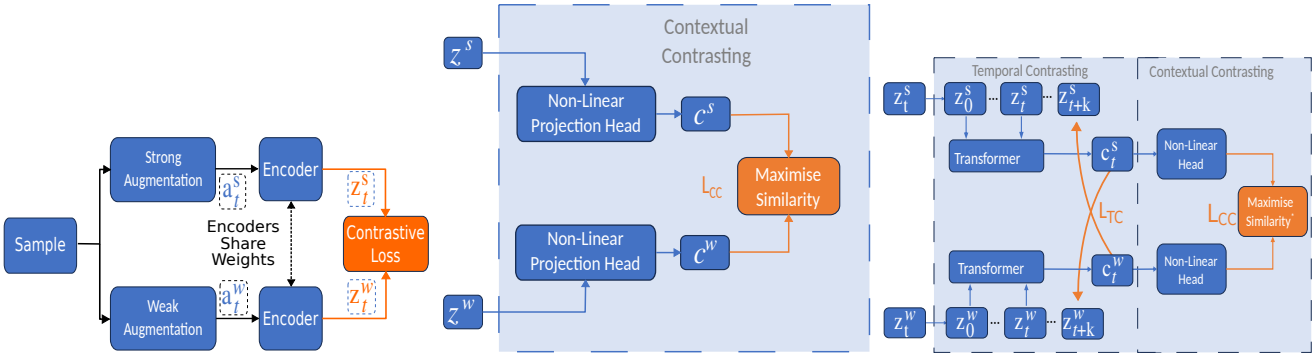
The SimCLR framework offered simplifications to self-supervised training processes. The workflow in Figure 1b illustrates the implementation. The goal is the same as other contrastive methods, to train an encoder to create a useful embedding. The inclusion of a non-linear head after the encoder aids in the creation of the embedding. In our implementation a convolutional backbone is utilised.

This work utilises the NT-Xent loss function described in Equation 2. This method requires a positive pair and multiple negative pairs for each sample. The data is split into batches of N events, with a pair of augmentations for each event, leading to a batch size of 2N. The different pairs of augmentations from the same event represent the positive pairs within the batch. The other 2(N-1) events within the batch are treated as negative pairs. This is the strategy utilised within the work proposing SimCLR, and other contrastive learning schemes (Chen et al., 2020; Eldele et al., 2021; Chen et al., 2017).

https://doi.org/10.5194/egusphere-2026-1201
Preprint. Discussion started: 28 April 2026

Additional negative pairs have been generated from noise data that are available within the Colima dataset. These noise segments are used in addition to the in-batch negatives. The advantage of including in-batch negative pairs is that it should lead to *hard* negatives being included in the learn approach, which is critical for self-supervised learning methods to succeed (Chen et al., 2020).

There may be consequences with respect to the presence of false negatives (Chuang et al., 2020), as there will be different members of the same class within the batch. Alternative schemes have been implemented for negative sample generation within seismology. If an event picker has been utilised to segment a signal, it is possible to sample the un-picked space within the signal (Lee et al., 2023), but this may not be useful for discriminating between different event types, instead leading to the model disciminating activity from inactivity, a potentially simpler task.



(a) Training Workflow (b) SimCLR Loss (c) TS-TCC Loss

Figure 1. (a) General contrastive learning training framework, where the input sample undergoes two different augmentations. These augmentations are encoded with identical encoders that share weights. The encoder outputs z_t^w, z_t^s are the inputs to the contrastive loss functions. The loss value updates the encoder weights. (b) SimCLR loss framework. (c) TS-TCC loss framework.

2.3 Temporal Contrasting

Temporal contrasting adds a new aspect to the loss function based on the temporal nature of the signal. This method aims to exploit the mutual information that is contained between sequential segments to build a useful representation of the signal. The method relies on a tough cross-view prediction task, as illustrated in Figure 1c. A strong and a weak augmentation are applied to the input time series to obtain a^s and a^w. The input time series is then embedded using an encoder network to obtain z_t^s and z_t^w. Both z_t^s and z_t^w remain time series signals, retaining the temporal nature of the augmentations. These time series are then sent to a predictive model, where the goal is to predict the next k time steps of the alternative embedding. The output of the transformer, c_t^s and c_t^w are then sent to a non-linear projection head that aims to project the inputs to a space where the positive pairs are similar and negative pairs are dissimilar.

The loss function is composed of two parts. The first is a temporal contrastive loss that is applied to evaluate the performance of the transformer in the cross-view predictive task. This metric is computed using the predicted value of one view, c (trans-

https://doi.org/10.5194/egusphere-2026-1201
Preprint. Discussion started: 28 April 2026

former output) and the true value of the other view, z. This loss function is defined in Equation 3 and aims to maximise the cosine similarity between c_T and z_T of the same sample, while minimising the similarity with the other $\mathcal{N}_{t,k}$ samples within the batch. Note that the loss is computed for both the strong and weak predictions, leading to λ^s_{TC} and λ^w_{TC}:

$$\lambda^s_{\mathrm{TC}} = -\frac{1}{K}\sum_{k=1}^{K}\log\left(\frac{\exp\left(\mathrm{sim}(W_k(c^s_t), z^w_{t+k})\right)}{\sum_{n\in\mathcal{N}_{t,k}}\exp\left(\mathrm{sim}(W_k(c^s_t), z^w_n)\right)}\right),$$
$$\lambda^w_{\mathrm{TC}} = -\frac{1}{K}\sum_{k=1}^{K}\log\left(\frac{\exp\left(\mathrm{sim}(W_k(c^w_t), z^s_{t+k})\right)}{\sum_{n\in\mathcal{N}_{t,k}}\exp\left(\mathrm{sim}(W_k(c^w_t), z^s_n)\right)}\right) \quad (3)$$

Where W_k is a linear function that is used to map the transfomer output c_t to the same dimensions as z.

The second aspect of the loss function applies the NT-Xent loss described in Equation 2. This loss function can be applied either to the output of the non-linear projection of the transformer outputs or to the output of the CNN embedding block. This aims to futher refine the representation.

The inclusion of the temporal contrastive loss enables the architecture to exploit the temporal relationships that should exist in the embedding to create higher quality features than in the Sim-CLR case where this temporal aspect is not taken into account.

The predictive model utilises a transformer architecture due to its efficiency and speed (Vaswani et al., 2017). The architecture is shown in Figure 2, and matches that used in the original work on TS-TCC (Eldele et al., 2021).

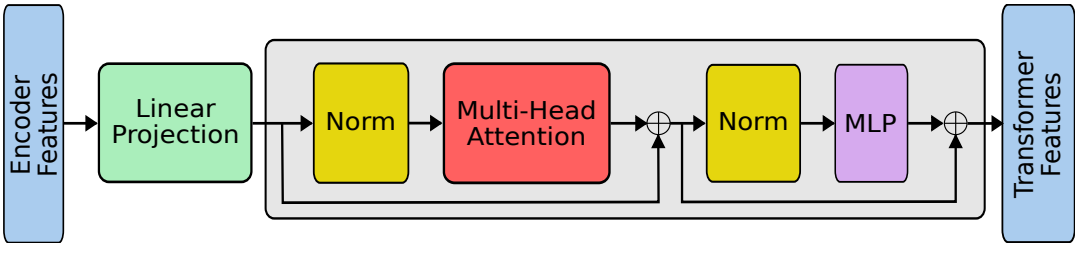


Figure 2. Transformer Architecture used within TS-TCC Method.

3 Dataset utilised

Three datasets from three separate volcanos (Colima volcano, Llaima volcano, Mount Etna) are considered in this study. Both the Colima and the Llaima datasets have been used for the validation exploration of machine learning techniques on volcano-seismic data. The Mount Etna dataset has been created using a reproducable method based on a curated catalog of events selected by the energy content in a short time window relative to a longer time window via an STA/LTA method. For example, a wide search of different supervised methods has been conducted on the Llaima dataset (Canário et al., 2020). Many deep learning techniques have been explored in the context of the Colima volcano (Titos et al., 2018). Each dataset consists of the

https://doi.org/10.5194/egusphere-2026-1201
Preprint. Discussion started: 28 April 2026

seismic signal of isolated time series events are taken from a single station, meaning that no station effects will be examined within this study.

3.1 Llaima Dataset

Llaima volcano is located within the western edge of the Andes region in Chile. The signals within this study were collected at LAV and are related with the Z component. The original sampling rate is 100 Hz and they have been downsampled to 50 Hz for the purposes of this study. The providers of the dataset applied a bandpass filter between 1 and 10 Hz. There are 4 types of events within this dataset; Volcano-Tectonic (VT), Long Period(LP), Tremor (TR), and Tectonic (TC). Examples of each of these types of events can be seen in Figure 3. Further information on the dataset is provided by Canario et al. (Canário et al., 2020).

3.2 Colima Dataset

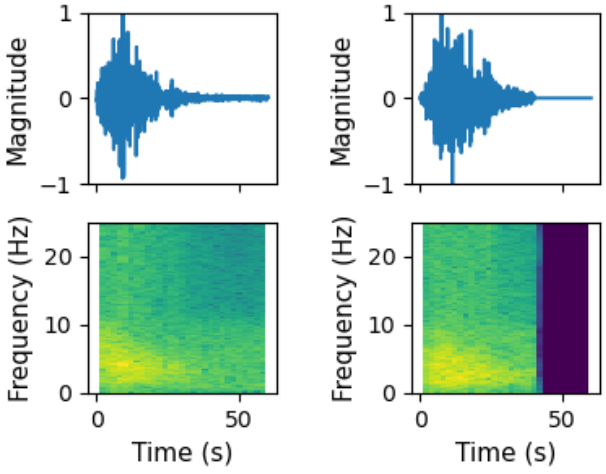
Colima volcano is a highly active stratovolcano in western Mexico (Robin et al., 1991; Luhr and Carmichael, 1980). The signals within this study are obtained from a single station on Colima volcano and are related with the Z component. The original sampling rate is 50 Hz and no filtering or downsampling has been applied. There are 5 categories of events considered of interest in this study; Volcano-Tectonic (VT), Long Period(LP), Tremor (TR), Tectonic (TC), and Explosive (EXP). There are also segments of Noise available for this volcano, that are used within the self-supervised training process. Examples of each of these event types can be seen in Figure 4. Further information on the dataset is provided by Titos et al. (Titos et al., 2018).

3.3 Etna Dataset

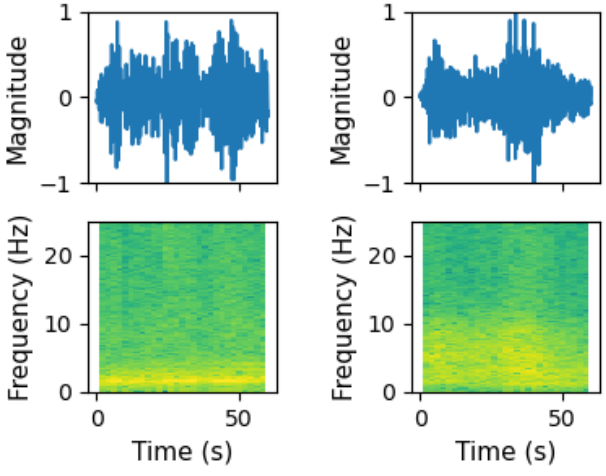
Mount Etna, located in Sicily is one of the most active stratovolcanos in the world (Branca and Del Carlo, 2004). The volcano exhibits both explosive and effusive activity, with high levels of each in the past 30 years (Bisson et al., 2021). The dataset for this experiment has been obtained from the Z component of the ECPN summit station of the permanent INGV network (di Geofisica e Vulcanologia , INGV). The data was originally collected at 100 Hz. The seismic data have been filtered to the band [0.5, 7] Hz, and downsampled to 15 Hz. A range of activity types are expectd to be contained in the dataset, from low frequency to hybrid, to high frequency events. The time period chosen is from a relatively quiet period at the beginning of 2022, with any time periods of activity removed using an activity catalog (Proietti et al., 2024). The exact dates for the training set are January 1st to March 30th, excluding the 8-25 February due to lava fountaining activity. Two test sets have been created, the first a holdout set of 10% of identified events from January to March, and second events identified in April 2022. Isolated events have been extracted using a recursive STA/LTA method with the following settings: $l_{\mathrm{STA}} = 1\,\mathrm{s}$, $l_{\mathrm{LTA}} = 7.5\,\mathrm{s}$, $\tau_{\mathrm{on}} = 2.8$, $\tau_{\mathrm{off}} = 0.5$

Where l_{STA} is the length of the short time window; l_{LTA} is the length of the long time window; τ_{on} is the detection threshold for events, and τ_{off} is the threshold to allow for the next event.

https://doi.org/10.5194/egusphere-2026-1201
Preprint. Discussion started: 28 April 2026



(a) Volcano-Tectonic (VT) (b) Long Period (LP)



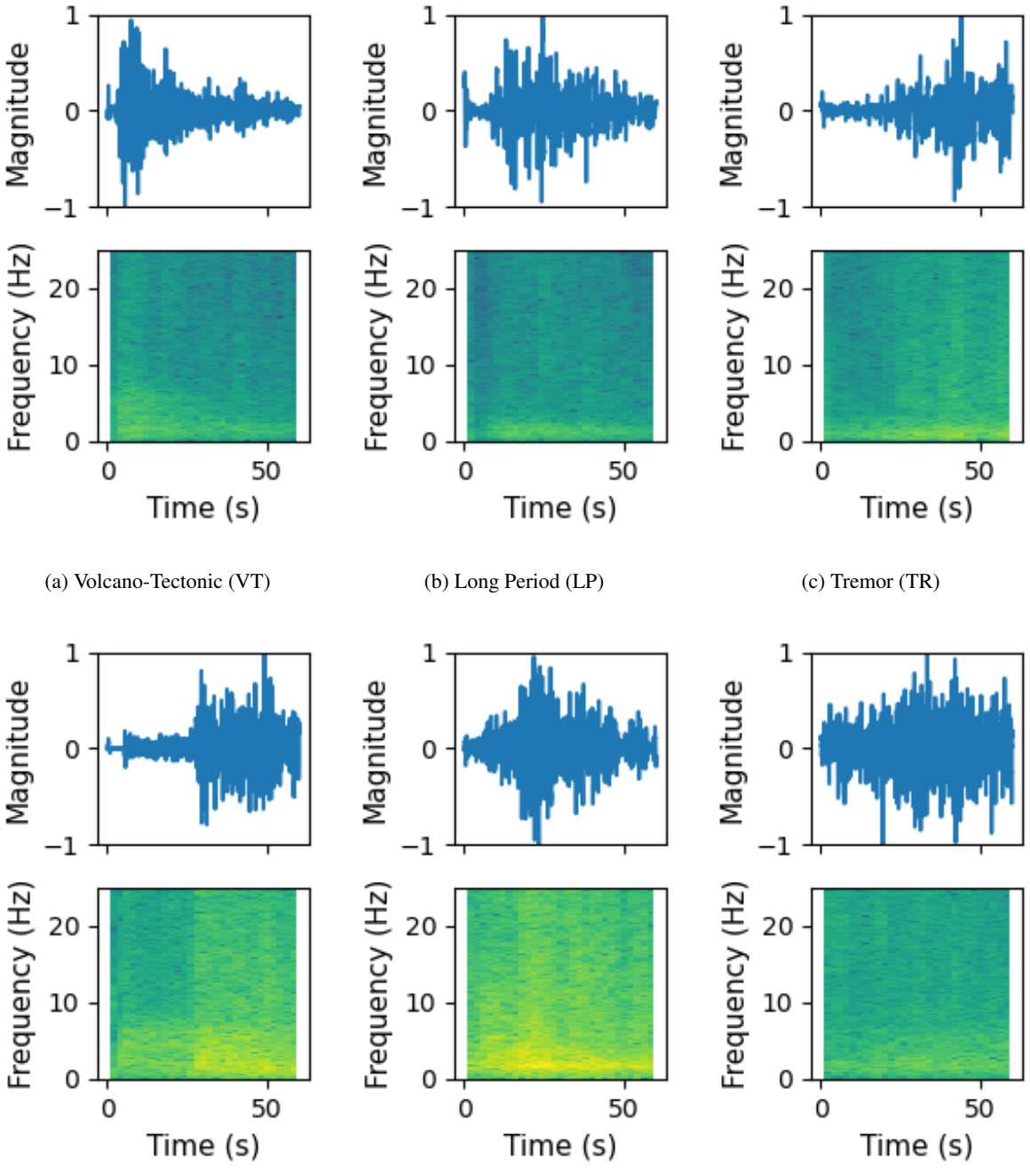
(c) Tremor (TR) (d) Tectonic (TC)

Figure 3. Llaima dataset and different types of events. In total there are 3592 events; VT - 304, LP - 1310, TR - 490, TC - 1488.

3.4 Data Preparation

Preprocessing is a fundamental step for any machine learning problem and can widely impact the result. This is clearly illustrated in the context of machine learning applied to volcano-seismic data (Titos et al., 2018). For this study, a training, validation and hold-out set are used to evaluate performance on the Llaima and the Colima datasets where labels are available. The time series signal for each event is prepared by selecting a cut-off time for each dataset, and padding shorter events with zeros to the cut-off time. The Llaima dataset is provided in this format, with a cut-off pre-determined by the providers of the

https://doi.org/10.5194/egusphere-2026-1201
Preprint. Discussion started: 28 April 2026



(a) Volcano-Tectonic (VT) (b) Long Period (LP) (c) Tremor (TR)

(d) Tectonic (TC) (e) Explosive (EXP) (f) Noise

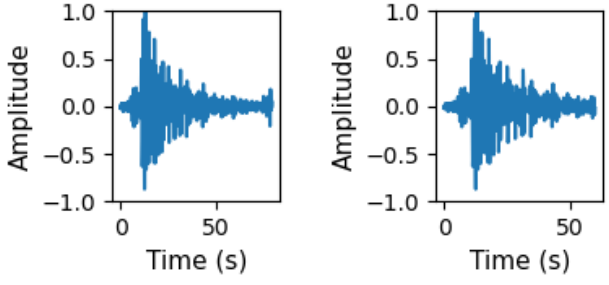
Figure 4. Colima dataset and different types of events. In total there are 5936 events; VT - 1681, LP - 2638, TR - 973, TC - 410, EXP - 233. There are 2784 noise segments.

https://doi.org/10.5194/egusphere-2026-1201
Preprint. Discussion started: 28 April 2026

dataset at $60s$. For the Colima dataset, the cut-off is set at $60s$, with events shorter than $60s$ padded with zeros to the cut-off time. This is based both on the average duration of events, and to maintain consistency with model selection across the Llaima and Colima datasets.

After cutting, the events are scaled to bring them into the same numerical range for the purposes of training the machine learning models. The events are scaled independently in an event-wise manner by the maximum absolute value within each event. This scaling removes the possibility for the model to compare the absolute energy across events, however this is compensated as it brings all events into the same numerical range allowing the extraction of useful features.

The Etna dataset has undergone a different preprocessing pipeline as this dataset contains additional noise due in part to the automated event isolation method. The isolated events are cut or padded to $3s$ before and $6s$ after the onset pick. This is to allow for a degree of variabilty in pick accuracy, and to capture the pre-event behaviour. There are a wider range of outliers for this dataset, as a result, the events can be filtered to remove low quality picks as has been performed in Sciotto et al. (Sciotto et al., 2022). Sciotto et al. study removes picks that lie below the median peak-to-peak amplitude value, and verifies that this retains only picks that are higher than a 30 minute rolling RMS in the relevant frequency band. In our case, a more strict filtering is required to remove events lower than the rolling RMS. Additionally, an emphasis is placed on selecting events that are well-picked, and contained within the 9 second time window. To this end, events where the absolute maximum value lies outside the center of the window are discarded, and additionally, events with lower maximum STA/LTA ratios are also discarded. The exact criteria are that events must have a peak-to-peak amplitude higher than the median, a maximum STA/LTA ratio higher than the median, and have a maximum value before $8.1s$ or 90% of the event has occured in time. These criteria enable the creation of a dataset that meets the criteria of surpassing the 30-minute rolling RMS energy, and based on a visual random sampling, well-picked events.



(a) Original Event (b) Cut-Off or Padded in Time

Figure 5. The events are either padded or cut at $T_{cut-off}$. The cut-off time is selected to match the cut-off for the Llaima dataset.

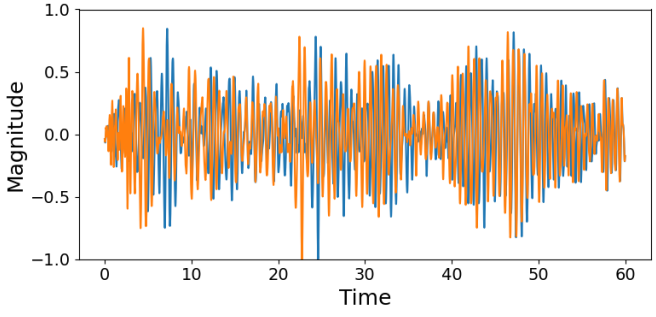
https://doi.org/10.5194/egusphere-2026-1201
Preprint. Discussion started: 28 April 2026

3.5 Augmentations

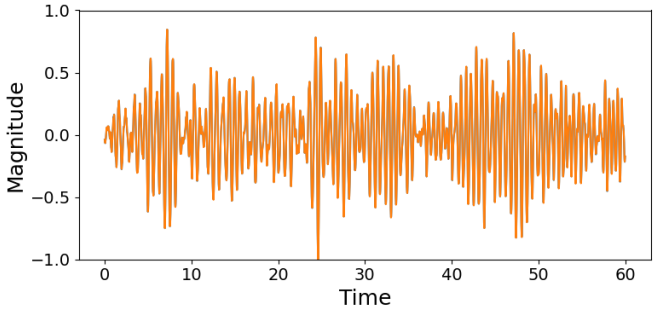
For each preprocessing method, a strong and a weak augmentation have been selected. The reasoning for using two augmentations when using contrastive learning methods is motivated in (Eldele et al., 2021). For each preprocessing method, the weak augmentation selected is “jittering”, this augmentation involves adding random Gaussian noise to the data.

The strong augmentation in the time series case is "window warping" followed by “jittering”. Window warping involves taking a portion of each event (10% in our case) and randomly speeding it up or slowing it down by a factor of 2. It will stretch or slice the entire event to fit the original dimensions.

The augmentations are visualised in Figure 6. Note that these augmentations are only used for the training in the contrastive learning framework where multiple views of the same events are compared to ensure that they are similar in the embedded space. The augmentation step is not used to increase the size of the training set.



(a) Strong Augmentation



(b) Weak Augmentation

Figure 6. Overview of the Data Augmentations. The strong augmentation applies window wawrping, followed by jittering. The weak augmentation applies jittering only.

https://doi.org/10.5194/egusphere-2026-1201
Preprint. Discussion started: 28 April 2026

4 Experimental framework

This section focuses on the encoder architecture selection, the evaluation framework for the different datasets, and the method for selecting the parameters for the contrastive learning approach. The search for the optimal CNN architecture is performed through a supervised learning task on the Llaima dataset, as summarised in this section. The evaluation framework for the unsupervised task involves the calculation of different unsupervised clustering metrics, and qualitative clustering measures. The selection of parameters for the contrastive learning approach is based on a grid search of parameters, evaluating performance based on both on the ability to converge on the contrastive pre-training task, and the performance on the downstream task of unsupervised clustering.

4.1 Selecting the Encoder CNN Architecture

The first step in the contrastive approach is to generate a suitable deep learning model for feature extraction. In this paper, 1D convolutional networks are used to learn patterns in the signal. The structure of these networks is inspired by (Canario et al., 2020) and is illustrated in Figure 7. A search has been conducted to explore different networks that could achieve high performance on a supervised classification task on the Llaima dataset. The results of this supervised search are summarized in Table 1. The difference between performance across the different models is small. For this reason priority was placed on selecting a model that achieved performance on the higher end, while also having a large number of output channels as the width is known to be a beneficial parameter for contrastive learning models (Chen et al., 2020). The performance of these models is similar to the top performance achieved in the reference work on the same dataset (Canario et al., 2020).

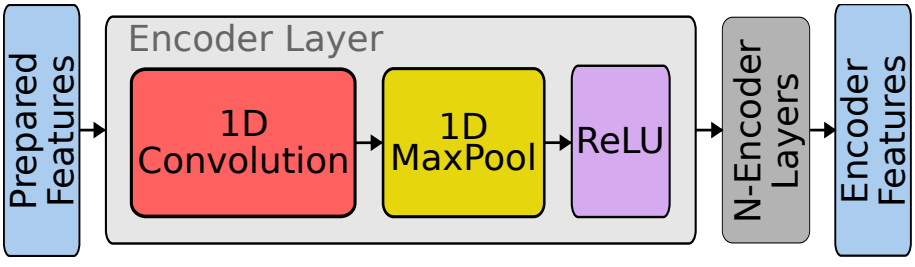


Figure 7. The encoder architecture is made up of convolutional blocks with a dropout of 10% in all cases. The convolutional blocks contain the following sequence: 1D Convolution layer, 1D Maxpool, ReLU, Dropout. The final layer of the encoder does not have a ReLU activation function.

The final architecture selected based on the results from the Llaima dataset contains 4 layers, and has 48 output channels, and 11 features per channel. This architecture is selected as it balances a small number of parameters, while achieving high accuracy in the supervised task. Additionally, the TS-TCC method performs a cross-view prediction in the time domain (along the features per channel). Note that the performance for each convolutional architecture is similar, indicating that the method if not sensitive to changes in model architecture. Based on this result, the same encoder is utilised for the Colima dataset. The events in the Etna dataset have different dimensions, and thus require a different architecture. A simple architecture is selected,

https://doi.org/10.5194/egusphere-2026-1201
Preprint. Discussion started: 28 April 2026

Table 1. High performant models on the Llaima dataset.

Accuracy	# Parameters	# Layers	# Output Channels	# per Channel
97.8	547700	3	48	15
97.8	296380	4	48	11
98.1	83836	4	64	4
97.0	1054180	5	512	1

with 3 layers, 32 output channels, and 6 features per channel. A smaller network is selected as the Etna dataset has smaller dimensions due to the lower sample rate, and the shorter event duration.

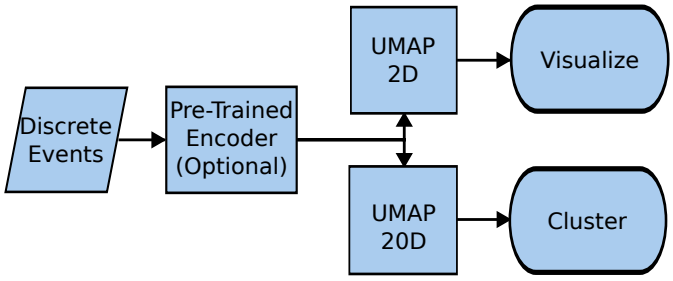


Figure 8. Overview of the Unsupervised Processing. For the original representation, the encoder block is skipped. The visualization pathway utilises a UMAP transform to two dimensions. The clustering pathway utilises a UMAP representation that collapses the input data to 20 dimensions. The increased number of dimensions allows the clustering algorithm to use additional information when generating clusters.

4.2 Clustering Framework

The entire methodology flow is illustrated in Figure 8. Three different common clustering algorithms have been utilised; K means, Spectral, and Agglomerative clustering. In the context of the Llaima and Colima datasets, the approach is evaluated using three supervised metrics: completeness, homogeneity, and V-measure (Rosenberg and Hirschberg, 2007). In the context of the Etna dataset, the evaluation is performed qualitatively due to the lack of public ground truth labels.

The clustering pipeline involves the application of a non-linear projection (UMAP) prior to applying the clustering algorithm. It is common to apply this type of projection to visualise and cluster data across a range of fields. The clustering pipeline that is implemented utilises a non-linear projection both for the output of the contrastive learning models, and for the baseline performance comparison.

4.2.1 Visualisation & Clustering Methods

UMAP is utilised to visualise the events in two dimensions. This algorithm uses a manifold technique to create an embedding of the data (McInnes et al., 2020). The UMAP embedding dimension does not have any physical meaning, and it is possible for

https://doi.org/10.5194/egusphere-2026-1201
Preprint. Discussion started: 28 April 2026

the algorithm to distort the dataset. However, for the purposes of visualising a data set and to aid in clustering, it is a powerful tool and has widespread applications (Allaoui et al., 2020; Dorrity et al., 2020). Alongside the visualisation in 2 dimensions, a UMAP projection to 20 dimensions is utilised in the case of each clustering algorithm. This has two benefits. Firstly, it simplifies the clustering task, and secondly, the clustering algorithms run faster due to the smaller dimensions of each sample.

Three clustering algorithms are applied in this study: K-means, Spectral, and Agglomerative. K-means clustering is an algorithm which attempts to seperate the data into n clusters of equal variance, where n is specified by the user. It scales well to a large number of samples and has widespread applications (Lloyd, 1982; MacQueen et al., 1967). Spectral clustering applies a low dimensional embedding of the affinity matrix between samples, followed by a K-means clustering within the low dimensional space (von Luxburg, 2007). Agglomerative clustering refers to a group of hierarchical clustering methods that use a bottom-up approach by successively merging together observations into clusters. In the context of this study, the clusters are merged based on the sum of squared difference within all clusters (Ward's method) (Murtagh and Legendre, 2014).

Each of these algorithms requires the number of clusters to be specified. The assumption that the number of clusters should equal the number of classes in the dataset has been taken. However, it would also be possible to sweep through a number of values based on the visualisation, or to use an algorithm such as HdbScan, where the number of clusters does not need to be specified (McInnes et al., 2017).

4.3 Clustering Metrics

Three metrics are used to quantitatively analyze the performance of the clustering algorithms; homogeneity, completeness, and V-measure (Rosenberg and Hirschberg, 2007). Homogeneity is a measure of the consistency of each cluster and is calculated using Equation 4. This metric is based on the objective that each cluster should contain only members of a single class. Completeness is a measure of the portion of each class within single clusters and is calculated using Equation 5. This metric is based on the goal that all members of a given class are assigned to the same cluster. Both metrics are bounded between 0 and 1, with a higher score indicating better performance. The V-measure is the harmonic mean of completeness and homogeneity and is calculated using Equation 6.

$$h = 1 - \frac{H(C|K)}{H(C)} \tag{4}$$

$$c = 1 - \frac{H(K|C)}{H(K)} \tag{5}$$

$$v = \frac{(1+\beta)\,h\,c}{(\beta\,h)+c} \tag{6}$$

Where $H(C)$ is the entropy related to the true labels calculated across all classes; $H(K)$ is the entropy related to the assigned labels calculated across all classes; $H(C|K)$ is the conditional entropy of the true labels with respect to the assigned labels calculated across all classes; $H(K|C)$ is the conditional entropy of the assigned labels with respect to the true labels calculated across all classes.

https://doi.org/10.5194/egusphere-2026-1201
Preprint. Discussion started: 28 April 2026

4.4 Qualitative Analysis Aids

For the Etna dataset, the method has been applied in an exploratory manner as there is no ground truth public catalog available. The quality of clustering can be estimated by stacking both the time and frequency representations of the events within each cluster, as done in Figure 11.

4.5 Framework for Selecting Contrastive Learning Parameters

The most important hyper-parameters of interest for the contrastive learning approaches are the number of training epochs, the batch size, and the learning rate. For the TS-TCC method, the number of hidden dimensions in the transformer is also important. These parameters are selected based on performance on the contrastive tasks and the performance for supervised classification. There does appear to be a relationship between the performance on these two tasks, with good performance on the contrastive task indicating high performance on the downstream tasks, however a detailed examination of this is beyond the purview of this work.

Note that contrastive learning methods in general benefit from longer training and larger batch sizes than its supervised counterparts (Chen et al., 2020). In the case of each dataset a grid search of batch size, learning rate, and number of epochs is performed, with batch sizes in the set [32, 48, 64, 96, 128]; number of epochs in the set [30,40,50,60,75]; and learning rates range [1e-4, 1e-2]. For the TS-TCC framework, the number of hidden layers is sweeped from 2 to 5, and the hidden layer dimensions are sweeped in the range [12,48].

4.6 Selection of UMAP Parameters

UMAP is used for dimension reduction in two tasks, clustering and visualisation. There are several UMAP parameters of interest including the number of neighbours, the minimum distance between points in the embedded space, the distance metric, and the number of components to use for clustering the data. These parameters have been hand-tuned for the Llaima dataset to obtain a discriminative representation. For visualisation, the parameters are set to: 2 components; 20 neighbours; 0.1 minimum distance. For clustering, the parameters are set to: 20 components; 20 neighbours; 0 minimum distance.

5 Llaima & Colima Results and Discussion

5.1 Contrastive Learning Parameters

For the Llaima dataset, the optimal parameters identified are a batch size of 32, 60 training epochs, and a learning rate of 2e-3. The size of the hidden layers in the transformer is 24, and the transformer contains 4 layers.

For the Colima dataset, the optimal parameters are a batch size of 48, 40 training epochs, and a learning rate of 2.1e-3. The size of the hidden layers in the transformer is 24, and the transformer contains 4 layers.

https://doi.org/10.5194/egusphere-2026-1201
Preprint. Discussion started: 28 April 2026

5.2 Clustering Results

The quantitative results for the Llaima and Colima datasets are in Table 2 and Table 3. Figures 9, 10 show the UMAP visualisation of the highest performing method based on the V-measure (TS-TCC in each case) and the UMAP visualisation from the raw data. Both the ground truth label and the cluster label are visualised for each event in each representation.

Table 2. Unsupervised Performance for Llaima Dataset

SSL Method	Cluster Method	c	h	v
None	Agglomerative	0.67	0.71	0.69
SimCLR	Spectral	0.75	0.77	0.76
TS-TCC	Spectral	0.89	0.89	0.89

Table 3. Unsupervised Performance for Colima Dataset

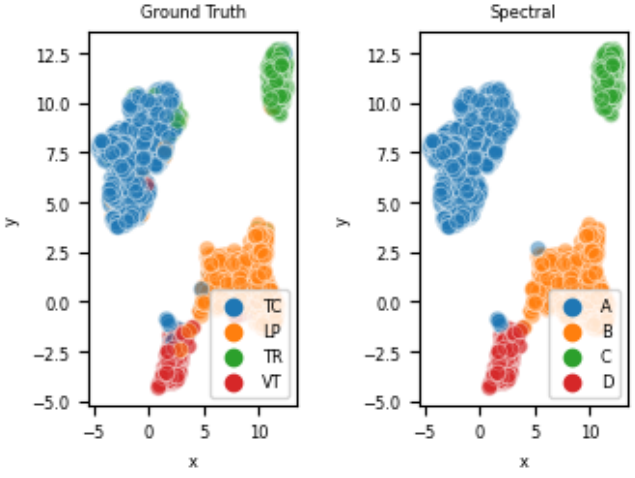
SSL Method	Cluster Method	c	h	v
None	Spectral	0.20	0.24	0.22
SimCLR	Agglomerative	0.48	0.54	0.51
TS-TCC	K means	0.53	0.60	0.56

5.3 Discussion

On the dataset from Llaima volcano, high performance is obtained, and these methods are capable of creating a discriminative representation that outperforms more traditional methods of dimension reduction both quantitatively and qualitatively.

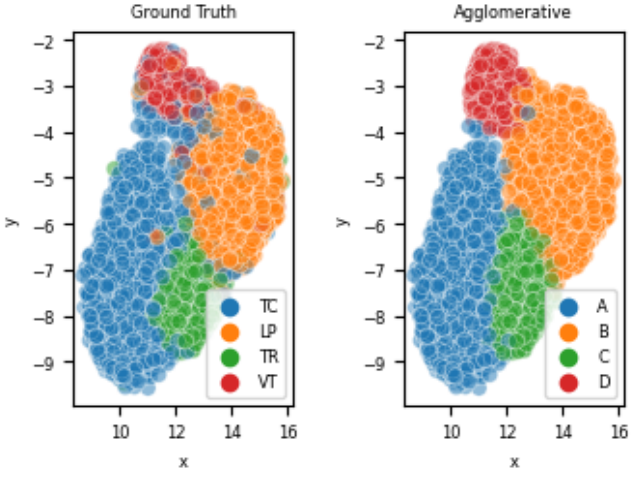
The dataset from Colima volcano contains events with a higher level of noise. The self-supervised methods again show high performance, however in this case the qualitative performance is less clear with a higher level of interclass mixing in the visual representation. This can be attributed to multiple factors. Firstly, the higher level of noise in the dataset. Secondly, this may be attributable to mistakes, or inconsistencies labelling process. The Colima dataset has been labelled by experts, however there is a bias when creating seismic catalogs, and the labels may not be perfectly consistent with the content of individual events. This dataset was created by multiple experts, each labelling different events, and as a result there may be inadvertent class overlaps which are contributing to the poorer performance. Thirdly, the method analyses the first 60s of the events, while the discriminative features may be contained in a different segment of the event that is not being considered by the self-supervised methods.

https://doi.org/10.5194/egusphere-2026-1201
Preprint. Discussion started: 28 April 2026



(a) Llaima UMAP Visualisation of Contrastive Representation - Ground Truth

(b) Llaima UMAP Visualisation of Contrastive Representation - Spectral Labels

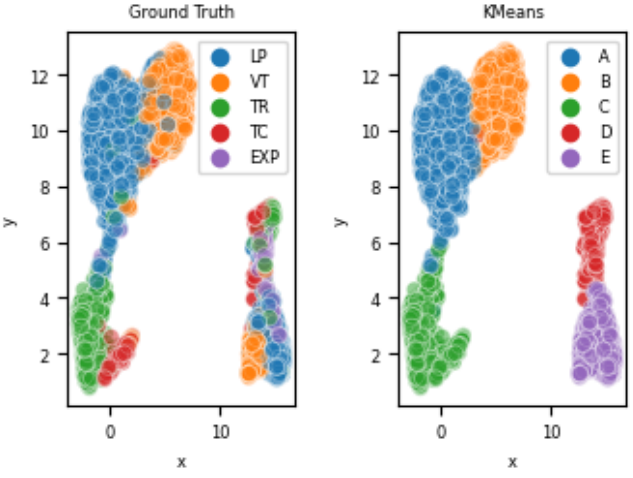


(c) Llaima UMAP Visualisation of Raw Data - Ground Truth

(d) Llaima UMAP Visualisation of Raw Data - Agglomerative Labels

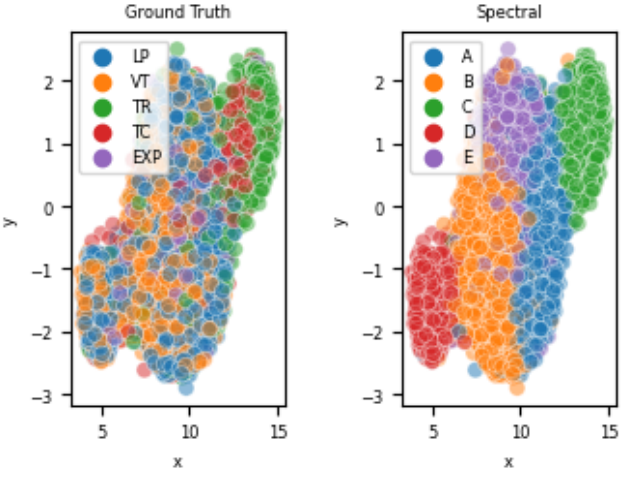
Figure 9. Llaima Representations Visualised

https://doi.org/10.5194/egusphere-2026-1201
Preprint. Discussion started: 28 April 2026



(a) Colima UMAP Visualisation of Contrastive Representation - Ground Truth

(b) Colima UMAP Visualisation of Contrastive Representation - Kmeans Labels



(c) Colima UMAP Visualisation of Raw Data - Ground Truth

(d) Colima UMAP Visualisation of Raw Data - Spectral Labels

Figure 10. Colima Representations Visualised

6 Etna Results and Discussion

For the Etna dataset, the focus is placed on the performance of TS-TCC for clustering the data. The focus is placed on this method, as the lack of a ground truth catalog means that an unbiased performance analysis across different approaches is difficult to achieve.

https://doi.org/10.5194/egusphere-2026-1201
Preprint. Discussion started: 28 April 2026

6.1 Contrastive Learning Parameters

For the Etna dataset, the optimal parameters are a batch size of 96, 75 training epochs, and a learning rate of 2.5e-3. The size of the hidden layers in the transformer is 12, and the transformer contains 3 layers. These are identified based on a qualitative analysis of the clustering capabilities.

6.2 Qualitative Cluster Results

There are several clusters that may be identified based on the visualisation of the UMAP embedding of the Etna dataset in Figure 11a. The structure of the embedding is consistent across the training set, and the two holdout test sets, indicating that the model is capable of generealising well to unseen data.

6.3 Discussion

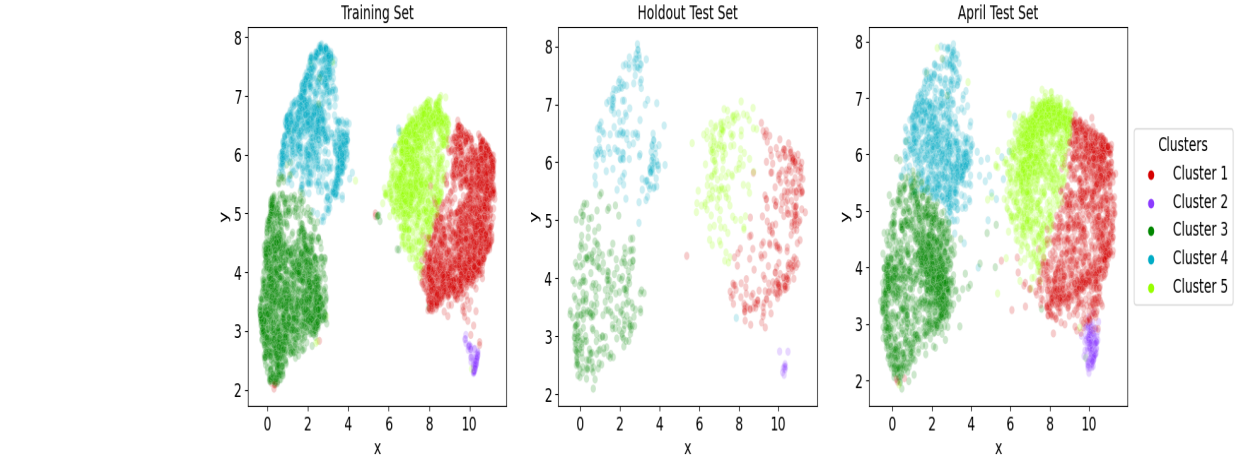
The Mount Etna dataset showcases the self-supervised methods, applied to the dataset that has been automatically picked through the STA/LTA process. This dataset is expected to have a high level of noise, both in the data, and in the picking process. Figure 11 shows that the TS-TCC method is capable of discriminating between events in the context of noisy events. There is a clear relationship between the clusters identified, and the frequency content of the events. Both clusters 3 and 4 have high energy at 1 Hz, with cluster 4 having energy also focused around 2.5 Hz, and cluster 3 having two secondary peaks at 2.5 and 4 Hz. Clusters 1 and 5 each have dominant higher frequency components, at 3.5 and 2.5 Hz respectively. The final cluster, number 2, has energy across the spectrum, with no clear dominant frequency. These clusters are identified consistently across the training set and both test sets. The exploration of strict geophysical differences between the clusters is considered out of scope for this work. However we believe that this method offers a novel first step for investigating seismic datasets, and for exploring the different types of activity that may be taking place.

7 Conclusions

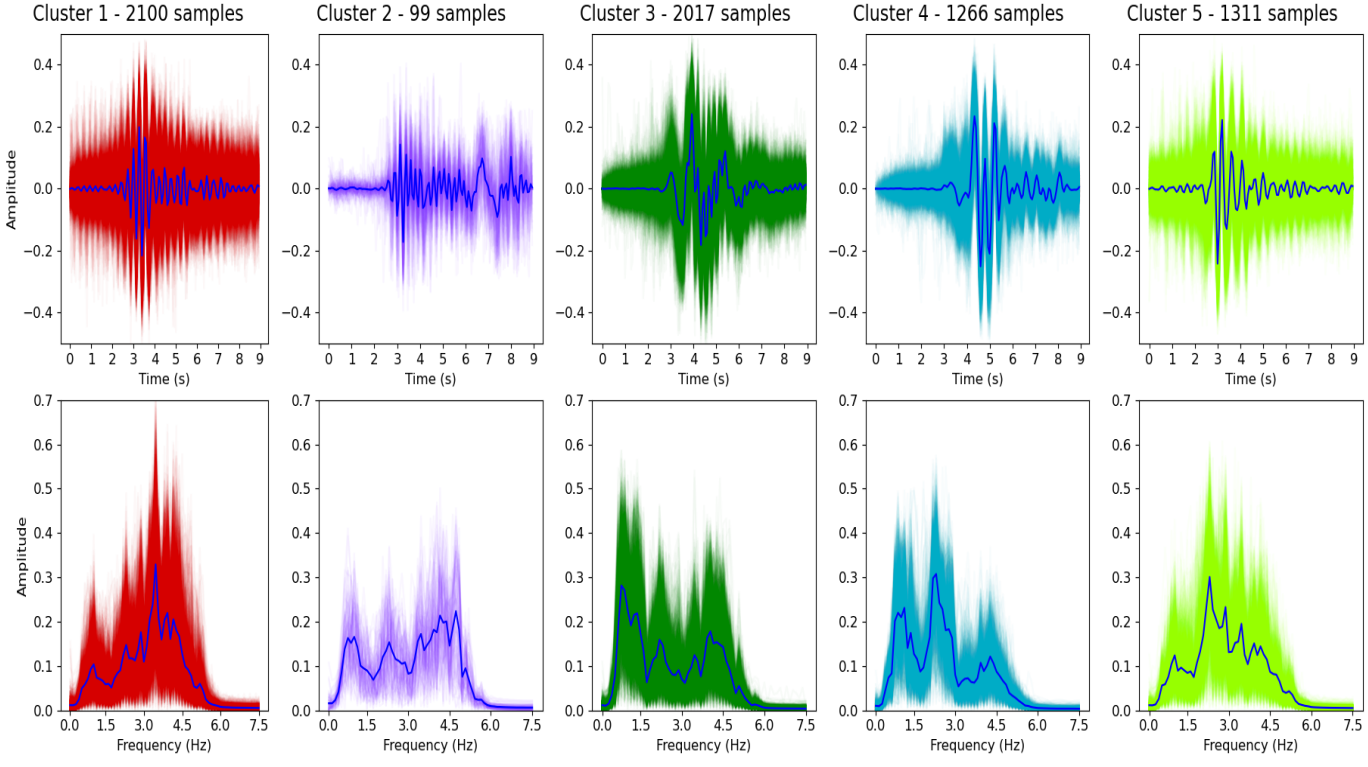
The performance of the self supervised methods has been showcased across a range of datasets, with high performance achieved across a range of volcano-seismic datasets. This result motivates the use of self-supervised learning, and specifically TS-TCC, for feature extraction on downstream tasks in volcano-seismology. It has the potential to separate out different types of volcano-seismic activity based on the temporal data from a single station, as shown in the case of the Llaima volcano. Furthermore, it can be applied in the context of noisy data as shown with the Colima dataset and beyond this in a continuous fashion on the noisy Etna dataset.

There is the potential that a different type of data preparation could allow for more powerful representation to be created. There are studies that extract representations from both the time domain and the frequency domain that have shown good performance in the wider context of time series analysis (Zhang et al., 2022). This exploration of signals with a frequency-based preparation is regarded as a potential direction for future work.

https://doi.org/10.5194/egusphere-2026-1201
Preprint. Discussion started: 28 April 2026



(a) Etna UMAP Visualisation of TS-TCC Representation



(b) Etna Clusters Visualised for TS-TCC Representation of the Training Set.

Figure 11. Etna dataset representations visualised.

370 *Code availability.* Correspondence and requests for materials should be addressed to M.T.

https://doi.org/10.5194/egusphere-2026-1201
Preprint. Discussion started: 28 April 2026

Data availability. Correspondence and requests for materials should be addressed to M.T.

Code and data availability. Correspondence and requests for materials should be addressed to M.T.

Appendix A: Contrastive Framework Pseudocode

A1 SimCLR Algorithm

Input:

- X: Time series data
- N: Number of epochs
- B: Batch size
- Encoder: CNN encoder for feature extraction
- Projection Head: Non-linear head (ReLU + two linear layers)

Output: Model weights trained using the SimCLR framework.

Procedure:

1. Apply weak augmentation to obtain X^w.
2. Apply strong augmentation to obtain X^s.
3. Split X^w and X^s into batches of size B.
4. For $k = 1$ to N:
   (a) For each pair of batches (B^w, B^s):
      i. Encode B^w and B^s to obtain Z^w and Z^s.
      ii. Apply projection head to obtain C^w and C^s.
      iii. Compute NT-Xent loss (positive if $i = j$, negative if $i \neq j$).
      iv. Update encoder and projection head weights.
5. Output trained CNN weights.

https://doi.org/10.5194/egusphere-2026-1201
Preprint. Discussion started: 28 April 2026

A2 TS-TCC Algorithm

Input:

- X: Time series data
- N: Number of epochs
- B: Batch size
- N_p: Number of previous timesteps used for prediction
- N_T: Number of temporal slices
- Encoder: CNN encoder
- Predictor: Transformer predictor
- Projection Head: Non-linear head

Output: Model weights trained using the TS-TCC framework.

Procedure:

1. Apply weak and strong augmentations to obtain X^w and X^s.
2. Split X^w and X^s into batches of size B.
3. For $k = 1$ to N:
   (a) For each pair of batches (B^w, B^s):
      i. Encode to obtain Z^w and Z^s.
      ii. Slice along the time dimension to obtain Z_T^w and Z_T^s.
      iii. For $t = N_p$ to N_T:
         A. Predict C_{t+1}^w from the previous N_p steps of Z_t^s.
         B. Predict C_{t+1}^s from the previous N_p steps of Z_t^w.
         C. Compute TS-TCC loss as cosine distance between (C_{t+1}^w, Z_{t+1}^w) and (C_{t+1}^s, Z_{t+1}^s).
      iv. Compute NT-Xent loss.
      v. Update encoder, transformer, and projection head weights.
4. Output trained CNN weights.

https://doi.org/10.5194/egusphere-2026-1201
Preprint. Discussion started: 28 April 2026

Author contributions. All authors contributed to the conception of the study, conceived and conducted the experiments. All authors analysed and interpreted the results. All authors reviewed the manuscript.

Competing interests. The authors declare no competing interests

Acknowledgements. This study was supported by the following grants. Grant "LEARNING", PID2022-143083NB-I00, "LEARNING", funded by MCIN/AEI /10.13039/501100011033 and by FEDER (EU) "Una manera de hacer Europa". Grant DIGIVOLCAN PLEC2022-009271 funded by 346 MCIN/AEI 10.13039/501100011033 and by "European Union NextGenerationEU/PRTR"; Grant 347 P21_00051 funded by MCIN/AEI/ 10.13039/501100011033 and by "ERDF A way of making Europe"; Grant TED2021-132178B-I00 funded by MCIN/AEI/10.13039/501100011033 and by the "European Union NextGenerationEU/PRTR"; Grant IMPROVE fuded by European Union's Horizon 2020 research and innovation programme under grant agreement No 858092; SINFONIA2, progetto Bando Ricerca Libera 2024 - Delibera 56/2025-INGV.

https://doi.org/10.5194/egusphere-2026-1201
Preprint. Discussion started: 28 April 2026

References

Agastya, C., Ghebremusse, S., Anderson, I., Vahabi, H., and Todeschini, A.: Self-supervised contrastive learning for irrigation detection in satellite imagery, arXiv preprint arXiv:2108.05484, 2021.

Allaoui, M., Kherfi, M. L., and Cheriet, A.: Considerably improving clustering algorithms using UMAP dimensionality reduction technique: a comparative study, in: International conference on image and signal processing, pp. 317–325, Springer, 2020.

Bisson, M., Spinetti, C., Andronico, D., Palaseanu-Lovejoy, M., Fabrizia Buongiorno, M., Alexandrov, O., and Cecere, T.: Ten years of volcanic activity at Mt Etna: High-resolution mapping and accurate quantification of the morphological changes by Pleiades and Lidar data, International Journal of Applied Earth Observation and Geoinformation, 102, 102 369, https://doi.org/https://doi.org/10.1016/j.jag.2021.102369, 2021.

Bountos, N. I., Papoutsis, I., Michail, D., and Anantrasirichai, N.: Self-supervised contrastive learning for volcanic unrest detection, IEEE Geoscience and Remote Sensing Letters, 19, 1–5, 2021.

Bountos, N. I., Michail, D., and Papoutsis, I.: Learning from synthetic InSAR with vision transformers: The case of volcanic unrest detection, IEEE Transactions on Geoscience and Remote Sensing, 60, 1–12, 2022.

Branca, S. and Del Carlo, P.: Eruptions of Mt Etna during the past 3.200 years: a revised compilation integrating the Historical and stratigraphic records., Mt. Etna: volcano laboratory, 2004.

Bueno, A., Benítez, C., De Angelis, S., Moreno, A. D., and Ibáñez, J. M.: Volcano-seismic transfer learning and uncertainty quantification with Bayesian neural networks, IEEE transactions on geoscience and remote sensing, 58, 892–902, 2019.

Bueno Rodriguez, A., Titos Luzón, M., Garcia Martinez, L., Benitez, C., and Ibáñez, J.: Automatic Seismic-Event Classification with Convolutional Neural Networks., in: AGU Fall Meeting Abstracts, vol. 2017, pp. S21E–03, 2017.

Buurman, H., West, M. E., Power, J., and Coombs, M.: Seismic precursors to volcanic explosions during the 2006 eruption of Augustine Volcano, The, pp. 41–57, 2006.

Canario, J. P., Mello, R., Curilem, M., Huenupan, F., and Rios, R.: In-depth comparison of deep artificial neural network architectures on seismic events classification, Journal of Volcanology and Geothermal Research, 401, 106 881, https://doi.org/https://doi.org/10.1016/j.jvolgeores.2020.106881, 2020.

Canário, J. P., de Mello, R. F., Curilem, M., Huenupan, F., and Rios, R. A.: Llaima volcano dataset: In-depth comparison of deep artificial neural network architectures on seismic events classification, Data in Brief, 30, 105 627, https://doi.org/https://doi.org/10.1016/j.dib.2020.105627, 2020.

Chen, T., Sun, Y., Shi, Y., and Hong, L.: On sampling strategies for neural network-based collaborative filtering, in: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 767–776, 2017.

Chen, T., Kornblith, S., Norouzi, M., and Hinton, G.: A simple framework for contrastive learning of visual representations, in: International conference on machine learning, pp. 1597–1607, PMLR, 2020.

Chouet, B. A.: Long-period volcano seismicity: its source and use in eruption forecasting, Nature, 380, 309–316, 1996.

Chuang, C.-Y., Robinson, J., Yen-Chen, L., Torralba, A., and Jegelka, S.: Debiased Contrastive Learning, 2020.

Cortés, G., Carniel, R., Mendoza, M. Á., and Lesage, P.: Standardization of noisy volcanoseismic waveforms as a key step toward station-independent, robust automatic recognition, Seismological Research Letters, 90, 581–590, 2019.

Curilem, M., Canário, J. P., Franco, L., and Rios, R. A.: Using CNN to classify spectrograms of seismic events from Llaima volcano (Chile), in: 2018 International Joint Conference on Neural Networks (IJCNN), pp. 1–8, IEEE, 2018.

https://doi.org/10.5194/egusphere-2026-1201
Preprint. Discussion started: 28 April 2026

di Geofisica e Vulcanologia (INGV), I. N.: https://doi.org/10.13127/SD/X0FXNH7QFY, 2005.

Dorrity, M. W., Saunders, L. M., Queitsch, C., Fields, S., and Trapnell, C.: Dimensionality reduction by UMAP to visualize physical and genetic interactions, Nature Communications, 11, 1537, https://doi.org/10.1038/s41467-020-15351-4, 2020.

Eldele, E., Ragab, M., Chen, Z., Wu, M., Kwoh, C. K., Li, X., and Guan, C.: Time-Series Representation Learning via Temporal and Contextual Contrasting, in: Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21, pp. 2352–2359, 2021.

Frosst, N., Papernot, N., and Hinton, G.: Analyzing and Improving Representations with the Soft Nearest Neighbor Loss, 2019.

Jaiswal, A., Babu, A. R., Zadeh, M. Z., Banerjee, D., and Makedon, F.: A Survey on Contrastive Self-Supervised Learning, Technologies, 9, https://doi.org/10.3390/technologies9010002, 2021.

Kothinti, S. and Elhilali, M.: Temporal contrastive-loss for audio event detection, in: ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 326–330, IEEE, 2022.

Lapins, S., Goitom, B., Kendall, J.-M., Werner, M. J., Cashman, K. V., and Hammond, J. O.: A little data goes a long way: Automating seismic phase arrival picking at Nabro volcano with transfer learning, Journal of Geophysical Research: Solid Earth, 126, e2021JB021 910, 2021.

Lara, F., Lara-Cueva, R., Larco, J. C., Carrera, E. V., and León, R.: A deep learning approach for automatic recognition of seismo-volcanic events at the Cotopaxi volcano, Journal of Volcanology and Geothermal Research, 409, 107 142, 2021.

Le-Khac, P. H., Healy, G., and Smeaton, A. F.: Contrastive representation learning: A framework and review, Ieee Access, 8, 193 907–193 934, 2020.

Lee, D., Aune, E., Langet, N., and Eidsvik, J.: Ensemble and self-supervised learning for improved classification of seismic signals from the Åknes rockslope, Mathematical Geosciences, 55, 377–400, 2023.

Li, K., Liu, W., Dou, Y., Xu, Z., Duan, H., and Jing, R.: CONSS: Contrastive Learning Method for Semisupervised Seismic Facies Classification, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 16, 7838–7849, https://doi.org/10.1109/JSTARS.2023.3308754, 2023.

Lloyd, S.: Least squares quantization in PCM, IEEE transactions on information theory, 28, 129–137, 1982.

Luhr, J. F. and Carmichael, I. S. E.: The Colima Volcanic complex, Mexico, Contributions to Mineralogy and Petrology, 71, 343–372, https://doi.org/10.1007/BF00374707, 1980.

MacQueen, J. et al.: Some methods for classification and analysis of multivariate observations, in: Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, vol. 1, pp. 281–297, Oakland, CA, USA, 1967.

Malfante, M., Dalla Mura, M., Mars, J. I., Métaxian, J.-P., Macedo, O., and Inza, A.: Automatic Classification of Volcano Seismic Signatures, Journal of Geophysical Research: Solid Earth, 123, 10,645–10,658, https://doi.org/https://doi.org/10.1029/2018JB015470, 2018.

Manley, G. F., Mather, T. A., Pyle, D. M., Clifton, D. A., Rodgers, M., Thompson, G., and Londono, J. M.: A deep active learning approach to the automatic classification of volcano-seismic events, Frontiers in Earth Science, 10, 807 926, 2022.

McInnes, L., Healy, J., and Astels, S.: hdbscan: Hierarchical density based clustering, Journal of Open Source Software, 2, 205, https://doi.org/10.21105/joss.00205, 2017.

McInnes, L., Healy, J., and Melville, J.: UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction, 2020.

Mousavi, S. M. and Beroza, G. C.: Machine Learning in Earthquake Seismology, Annual Review of Earth and Planetary Sciences, 51, 105–129, https://doi.org/https://doi.org/10.1146/annurev-earth-071822-100323, 2023.

Münchmeyer, J., Bindi, D., Leser, U., and Tilmann, F.: The transformer earthquake alerting model: A new versatile approach to earthquake early warning, Geophysical Journal International, 225, 646–656, 2021.

https://doi.org/10.5194/egusphere-2026-1201
Preprint. Discussion started: 28 April 2026

Murshed, R. U., Noshin, K., Zakaria, M. A., Uddin, M. F., Amin, A. S., and Ali, M. E.: Real-time seismic intensity prediction using self-supervised contrastive gnn for earthquake early warning, IEEE Transactions on Geoscience and Remote Sensing, 2024.

Murtagh, F. and Legendre, P.: Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion?, Journal of Classification, 31, 274–295, https://doi.org/10.1007/s00357-014-9161-z, 2014.

Münchmeyer, J., Woollam, J., Rietbrock, A., Tilmann, F., Lange, D., Bornstein, T., Diehl, T., Giunchi, C., Haslinger, F., Jozinović, D., Michelini, A., Saul, J., and Soto, H.: Which Picker Fits My Data? A Quantitative Evaluation of Deep Learning Based Seismic Pickers, Journal of Geophysical Research: Solid Earth, 127, e2021JB023 499, https://doi.org/https://doi.org/10.1029/2021JB023499, https://doi.org/10.1029/2021JB023499, 2022.

Proietti, C., De Beni, E., and Cantarero, M.: One hundred lava flows of Mt. Etna, Italy: July 2019–December 2023 update, Journal of Maps, 20, 2380 899, 2024.

Robin, C., Camus, G., and Gourgaud, A.: Eruptive and magmatic cycles at Fuego de Colima volcano (Mexico), Journal of Volcanology and Geothermal Research, 45, 209–225, https://doi.org/https://doi.org/10.1016/0377-0273(91)90060-D, 1991.

Rosenberg, A. and Hirschberg, J.: V-Measure: A Conditional Entropy-Based External Cluster Evaluation Measure, in: Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL), edited by Eisner, J., pp. 410–420, Association for Computational Linguistics, Prague, Czech Republic, https://aclanthology.org/D07-1043, 2007.

Saeed, A., Grangier, D., and Zeghidour, N.: Contrastive learning of general-purpose audio representations, in: ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 3875–3879, IEEE, 2021.

Salazar, A., Arroyo, R., Pérez, N., and Benítez, D.: Deep-learning for volcanic seismic events classification, in: 2020 IEEE Colombian Conference on Applications of Computational Intelligence (IEEE ColCACI 2020), pp. 1–6, IEEE, 2020.

Scarfì, L., Aloisi, M., Barberi, G., and Langer, H.: Observing Etna volcano dynamics through seismic and deformation patterns, Scientific Reports, 13, 12 951, 2023.

Schroff, F., Kalenichenko, D., and Philbin, J.: FaceNet: A unified embedding for face recognition and clustering, in: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, https://doi.org/10.1109/cvpr.2015.7298682, 2015.

Sciotto, M., Cannata, A., Di Grazia, G., and Montalto, P.: Volcanic tremor and long period events at Mt. Etna: Same mechanism at different rates or not?, Physics of the Earth and Planetary Interiors, 324, 106 850, https://doi.org/https://doi.org/10.1016/j.pepi.2022.106850, 2022.

Si, X., Wu, X., Sheng, H., Zhu, J., and Li, Z.: SeisCLIP: A Seismology Foundation Model Pre-Trained by Multimodal Data for Multipurpose Seismic Feature Extraction, IEEE Transactions on Geoscience and Remote Sensing, 62, 1–13, https://doi.org/10.1109/TGRS.2024.3354456, 2024.

Sohn, K.: Improved deep metric learning with multi-class n-pair loss objective, Advances in neural information processing systems, 29, 2016.

Titos, M., Bueno, A., García, L., and Benítez, C.: A Deep Neural Networks Approach to Automatic Recognition Systems for Volcano-Seismic Events, IEEE JSTAR, 11, 1533–1544, https://doi.org/10.1109/JSTARS.2018.2803198, 2018.

Titos, M., Bueno, A., García, L., Benítez, C., and Segura, J. C.: Classification of isolated volcano-seismic events based on inductive transfer learning, IEEE Geoscience and Remote Sensing Letters, 17, 869–873, 2019.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I.: Attention Is All You Need, CoRR, abs/1706.03762, http://arxiv.org/abs/1706.03762, 2017.

von Luxburg, U.: A tutorial on spectral clustering, Statistics and Computing, 17, 395–416, https://doi.org/10.1007/s11222-007-9033-z, 2007.

https://doi.org/10.5194/egusphere-2026-1201
Preprint. Discussion started: 28 April 2026

Woollam, J., Münchmeyer, J., Tilmann, F., Rietbrock, A., Lange, D., Bornstein, T., Diehl, T., Giunchi, C., Haslinger, F., Jozinović, D., Michelini, A., Saul, J., and Soto, H.: SeisBench—A Toolbox for Machine Learning in Seismology, Seismological Research Letters, 93, 1695–1709, https://doi.org/10.1785/0220210324, 2022.

Wu, H.-H., Seetharaman, P., Kumar, K., and Bello, J. P.: Wav2clip: Learning robust audio representations from clip, in: ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 4563–4567, IEEE, 2022.

Yang, Z., Li, H., Tuo, X., Li, L., and Wen, J.: Unsupervised clustering of microseismic signals using a contrastive learning model, IEEE Transactions on Geoscience and Remote Sensing, 61, 1–12, 2023.

Zeng, Z., McDuff, D., Song, Y., et al.: Contrastive learning of global and local video representations, Advances in Neural Information Processing Systems, 34, 7025–7040, 2021.

Zhang, X., Zhao, Z., Tsiligkaridis, T., and Zitnik, M.: Self-Supervised Contrastive Pre-Training For Time Series via Time-Frequency Consistency, 2022.