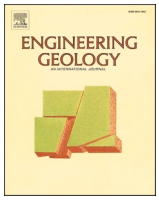
Fast and effective classification of landslide microseismicity: a machine learning perspective

Engineering Geology 367 (2026) 108698

Contents lists available at ScienceDirect

Engineering Geology

journal homepage: www.elsevier.com/locate/enggeo



Fast and effective classification of landslide microseismicity: a machine learning perspective☆

Farbod Khosro Anjom *, Lorena Di Toro, Chiara Colombero

Politecnico di Torino, Department of Environmental, Land and Infrastructure Engineering (DIATI), 10129 Torino, Italy

ARTICLE INFO

Keywords:
Seismic monitoring
Landslide
Geo-hazard
Machine learning
Artificial intelligence

ABSTRACT

The analysis of microseismic events from passive seismic data has proven invaluable for failure prediction in the context of natural hazard and landslide monitoring. Generally, microseismicity associated with fracturing processes exponentially increases prior to failure. However, passive seismic recordings commonly include a broad range of non-fracturing signals, making reliable event classification a prerequisite for near-real-time monitoring. This issue is particularly critical under data-limited monitoring conditions, such as during the early phases of slope monitoring or at minimally instrumented sites, where only a single seismic station or a very sparse sensor configuration is available. This study evaluates and compares the performance of supervised and unsupervised machine-learning approaches for the classification of landslide-related microseismic events using a challenging single-station dataset acquired over a six-month monitoring period at the Bossea Cave (NW Italy). Fracture-related microseismicity at the site exhibits marked temporal variability in spectral features, primarily driven by air-temperature fluctuations and precipitation.

A supervised convolutional neural network (CNN), trained on a limited labeled dataset, is compared with an unsupervised k-means clustering approach applied to time- and frequency-domain features reduced through principal component analysis (PCA). The k-means clustering is strongly influenced by the seasonal variability of spectral features, leading to diffuse and overlapping clusters, whereas the CNN operates directly on signal spectrograms and incorporates Monte Carlo Dropout layers to quantify epistemic uncertainty through stochastic inference. Both approaches achieve satisfactory classification performance; however, the CNN demonstrates slightly higher accuracy and improved robustness under non-stationary spectral conditions. The results emphasize both the potential and the limitations of automated single-station classification workflows for passive seismic monitoring in realistic, data-limited operational scenarios.

1. Introduction

Microseismicity is a valuable monitoring tool across a wide variety of applications, from georesources and climate change mitigation strategies (e.g. reservoir monitoring, geothermal field operations, carbon capture and storage sites; Maxwell et al., 2010; Grigoli et al., 2022; Verliac and Le Calvez, 2021) to the detection of underground stress changes and rock failure for mine safety (e.g., Wilkins et al., 2020; Sedghizadeh et al., 2024). The analysis of microseismicity also plays a key role in understanding and forecasting natural processes related to volcanic activity (Lamb et al., 2022; Pantobe et al., 2023), glacial dynamics (West et al., 2010; Carr et al., 2020) and landslides (Yang et al., 2022; Yfantis et al., 2021; Colombero et al., 2021).

In the context of natural hazards, several studies have reported an increase in fracturing-related microseismic activity prior to failure (Senfaute et al., 2009; Faillettaz et al., 2019; Liu et al., 2024). When recorded by a small network of receivers, the event source can also be located, allowing the identification of potentially unstable volumes as well as active fractures or slip surfaces (Lacroix and Helmstetter, 2011; Tonnellier et al., 2013; Spillmann et al., 2007). Despite this potential, the number, magnitude, and type of recorded microseismic events are generally highly variable from site to site, since the time- and frequency-domain features of the recorded signals strongly depend on the source and network geometry. In addition, many unwanted microseismic signals, not related to the incipient fracturing, might populate the event data set. These events include natural signals, such as earthquakes,

☆ This article is part of a Special issue entitled: 'Landslides Prediction' published in Engineering Geology.

* Corresponding author.
E-mail address: Khosro-anjom.farbod@polito.it (F. Khosro Anjom).

https://doi.org/10.1016/j.enggeo.2026.108698
Received 21 July 2025; Received in revised form 15 February 2026; Accepted 18 March 2026
Available online 21 March 2026

rainstorms, rockfalls or rockslides, as well as anthropogenic signals, such as electronic transients and nearby human activities. While event detection can be effectively calibrated to site-specific condition using adaptive STA/LTA (short time average over long time average) algorithms, event classification remains a challenging task and is critical for tracking microseismic activity over time (Colombero et al., 2018). Early classification strategies relied on the visual inspection of event spectrograms (Helmstetter and Garambois, 2010; Levy et al., 2011). Although accurate, this approach is time-consuming, potentially subjective, and impractical for large datasets or near-real-time monitoring systems. Consequently, substantial research efforts have been devoted to the development of automatic event discrimination methods. Major contributions include clustering approaches based on selected time- and frequency-domain parameters (e.g., Hibert et al., 2014) as well as machine learning (ML) techniques (e.g., Malfante et al., 2018; Falcin et al., 2021; Manley et al., 2022), particularly in volcanic seismology.

For landslide monitoring, ML approaches are growing in importance to predict landslide activation and displacement evolution, through physics-informed neural networks and deep learning methods embedding geotechnical and mechanical constraints or remote sensing time series (e.g., Dahal and Lombardo, 2025; Zhou et al., 2024; Gatto et al., 2025). Applications of ML specifically targeting landslide-related microseismicity are still relatively limited, but recent studies have demonstrated promising results for event detection, denoising, classification, and source characterization (Anikiev et al., 2023).

Regarding classification, Provost et al. (2017) and Wenner et al. (2021) applied Random Forest supervised classifiers to long-term seismic records acquired with multi-station monitoring networks, explicitly addressing operational challenges such as imbalanced datasets, low signal-to-noise ratios, and non-ideal network geometries. By comparison, Langet and Silverberg (2023) employed convolutional neural networks (CNNs) to classify 15 years of continuous recordings from an instrumented rock slope using event spectrograms as input. Random Forest classifiers rely on manually engineered signal features and are generally robust under limited data availability, whereas CNNs learn hierarchical representations directly from the input data but typically require larger labeled datasets and higher computational resources. Hybrid or ensemble approaches combining different classifiers may further improve classification performance by exploiting complementary strengths of individual methods (Xin et al., 2024; Chen et al., 2024). However, such gains must be balanced against increased computational cost and implementation complexity, particularly in operational monitoring scenarios. Alternative lightweight approaches combining feature extraction with shallow classifiers have also been proposed, demonstrating that computationally efficient models can still achieve reliable microseismic waveform classification (Shu et al., 2024). Transfer-learning strategies based on pre-trained computer-vision models offer another promising compromise, although fine-tuning large architectures such as AlexNet or VGG-type networks often require extensive labeled datasets, limiting their applicability to near-real-time early-warning systems (Zhang et al., 2021).

In this context, a key open methodological question concerns the relative performance of supervised and unsupervised learning approaches for microseismic event classification under realistic monitoring constraints. Here, we compare a supervised CNN and an unsupervised k-means clustering approach applied to microseismic events recorded on the slope at the top of a potentially unstable cave. The k-means classification is based on clustering salient, uncorrelated time- and frequency-domain parameters identified through principal component analysis (PCA) on a subset of visually classified and labeled events detected in the first monitored months. Cluster centroids are then used to classify the whole data set. The CNN method is trained on the spectrograms of the same labeled events.

The event dataset selected from the Bossea Cave (Dabove et al., 2023) is characterized by a strong temporal variability of the spectral features of fracture-related microseismic events. In particular, the peak frequency is highly sensitive to air temperature fluctuations and precipitation rates, resulting in progressive shifts over time, while the overall spectral shape remains relatively stable. This pronounced non-stationarity poses significant challenges for conventional feature-based classification approaches. In addition, the microseismic activity was predominantly recorded by a single monitoring station installed on the slope directly above the vertical projection of one of the largest and potentially unstable cave chambers. Conversely, seismic stations deployed along the same slope but outside the cave vertical projection detected only a very limited number of events, likely due to the weak amplitude of the signals and the increased source–receiver distance. As a result, the available dataset does not allow for the application of network-based discrimination techniques, reinforcing the relevance of a single-station analysis framework in this context.

The main scientific objective of this study is therefore to assess and compare the performance of supervised and unsupervised machine-learning approaches for the classification of landslide-related microseismic events under data-limited monitoring conditions. The analysis is intentionally conducted on a short monitoring period (six months) and using data from a single monitoring station, reflecting realistic constraints commonly encountered during the early stages of slope monitoring or in minimally instrumented sites. Specifically, we evaluate the relative strengths and limitations of the CNN approach trained on a limited labeled dataset and the unsupervised k-means clustering in which time- and frequency-domain features exhibit pronounced seasonal variability, leading to broad and overlapping clusters.

2. Event dataset

To test and compare the two classification approaches, we used a microseismic event dataset of 3166 events recorded over six months by a single passive seismic station deployed on the slope at the top of the Bossea Cave (Cuneo, NW Italy, Fig. 1a). The natural cave is the terminal sector of a large karst system and is characterized by a long complex of narrow galleries and few wide underground chambers resulting from the collapse of the karst network (Fig. 1b). Several signs of instability and accumulation of big rock blocks are visible in the main chambers due to historical rockfalls. The limestone roof of the natural cave exhibits indeed modifications in its stability that might be related to precipitation and the water load caused by infiltration and circulation, along with temperature cycles and gradients at the daily and seasonal scales. To monitor the behavior of the cave roof, one 3C seismometer (Trillium Compact 20 s) connected to a Pegasus Digital Recorder (Nanometrics) was deployed on the external slope (S1 in Fig. 1b), above one of the big potentially unstable chambers. The roof thickness at the sensor location is approximately 50 m. The recordings were sampled at 250 Hz and internally stored in daily MiniSEED files. Air temperature and precipitation over the monitoring period are shown in Fig. 1c.

We applied a STA/LTA algorithm for the detection of microseismic events in the continuous noise recordings (STA window length = 0.2 s; LTA window length = 10 s; STA/LTA threshold = 5). After detection, the traces of the events were extracted for a window of variable duration, spanning 2 s before the threshold exceedance and 4 s after returning below the threshold, and were saved as separate segments for further classification and analysis. The cumulative number of events is shown in Fig. 1d, together with their peak frequency in the amplitude spectrum.

The 437 microseismic events detected at S1 in the first three months (December 2021 – February 2022) were visually classified based on their spectrograms (Fig. 2 and Fig. 3). The events potentially related to fracturing showed short duration and a characteristic spectrogram with a sharp onset followed by an exponential decrease of the high frequency energy content (class I, Fig. 2a to 2i). They occurred both as single isolated events and in long sequences of subsequent events with recurrent spectral signatures (Fig. 2c and i). They generally showed peak frequencies higher than 50 Hz, with fluctuations likely correlated with the external air temperature after the winter period (Fig. 1d), as already

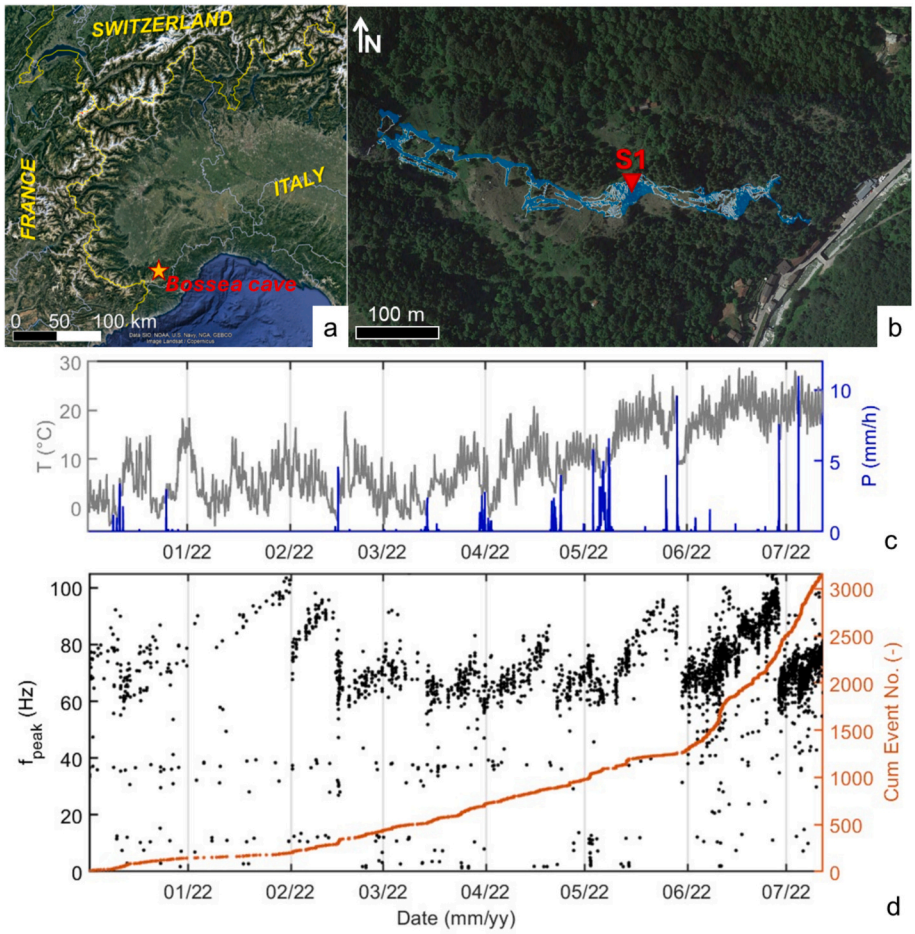


Fig. 1. (a) Geographic location of the monitored site. (b) Location of the passive seismic station S1, deployed on the external slope, with respect to the underground cave network, shown in blue. (c) Hourly air temperature and rainfall data. (d) Cumulative number of detected microseismic events at S1 in the period December 2021–July 2022 and related peak frequencies. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

observed on other monitored potentially unstable rock masses (Colombero et al., 2018; Colombero et al., 2021). However, intense rainfall events seem to strongly affect this temperature driven variation, with concomitant local drops in the peak frequencies (Fig. 1d).

Long-duration events with peak frequencies of around 40 Hz were also frequently recorded and likely related to either rockfalls and other slides of rock materials on the monitored slope or water infiltration and circulation inside the cave roof (class II, Fig. 2j to 2r).

Fewer local and regional earthquakes were also found in the event dataset. They generally showed long duration and peak frequencies below 25 Hz (class III, Fig. 3a to 3i). The source of many of these events was confirmed from seismological catalogues (e.g., https://www.emsc-csem.org/, last access on July 17th, 2025), however not all the events visually belonging to this class were found in there. Due to the similarities in time and frequency domain all the events showing long duration and low peak frequency were associated with this class; however, some might be related to other local seismic sources.

Many high-energy short-duration noise transients were found in the data set (Fig. 3j to 3r), showing variable spectral features. They appeared to be particularly frequent during intense rain events, likely due to electronic transients or raindrops in the vicinity of the stations (e.g., Fig. 3k).

From this preliminary visual analysis and interpretation of the detected events, it appears clear that the only class of events that might be of interest for early warning purposes is class I. The other classes, and especially noise transients, should be therefore removed from the event count to correctly track potential instabilities and failure precursors.

A summary of the labeled events from December 2021 to February 2022 is provided in Table 1. We labeled a total of 355 events, excluding 82 events whose classification was uncertain due to the overlap of different types of events in the same time segment or impossibility to visually assign them to a certain class. These events were consequently not suitable for the clustering and training of the ML model.

3. Methods

3.1. Cluster analysis approach

We tested the effectiveness of a classification approach based on unsupervised k-means clustering of salient event features in discriminating the different classes of Table 1.

Particularly, we considered the labeled events belonging to the first three classes of events (I, II and III) to compute three representative centroids of these clusters, since the variability of the events contained in the noise would not have allowed to describe them as a single cluster defined by a unique centroid. The labeled events belonging to class I were however overrepresented (257 events over a total of 314 events in classes I to III, Table 1). Therefore, we randomly selected 43 events only of class I to get a sub dataset of 100 labeled events for centroid computation.

The aim of this first phase was to extract some key features, capturing both time- and frequency-domain features of the signals of class I to III

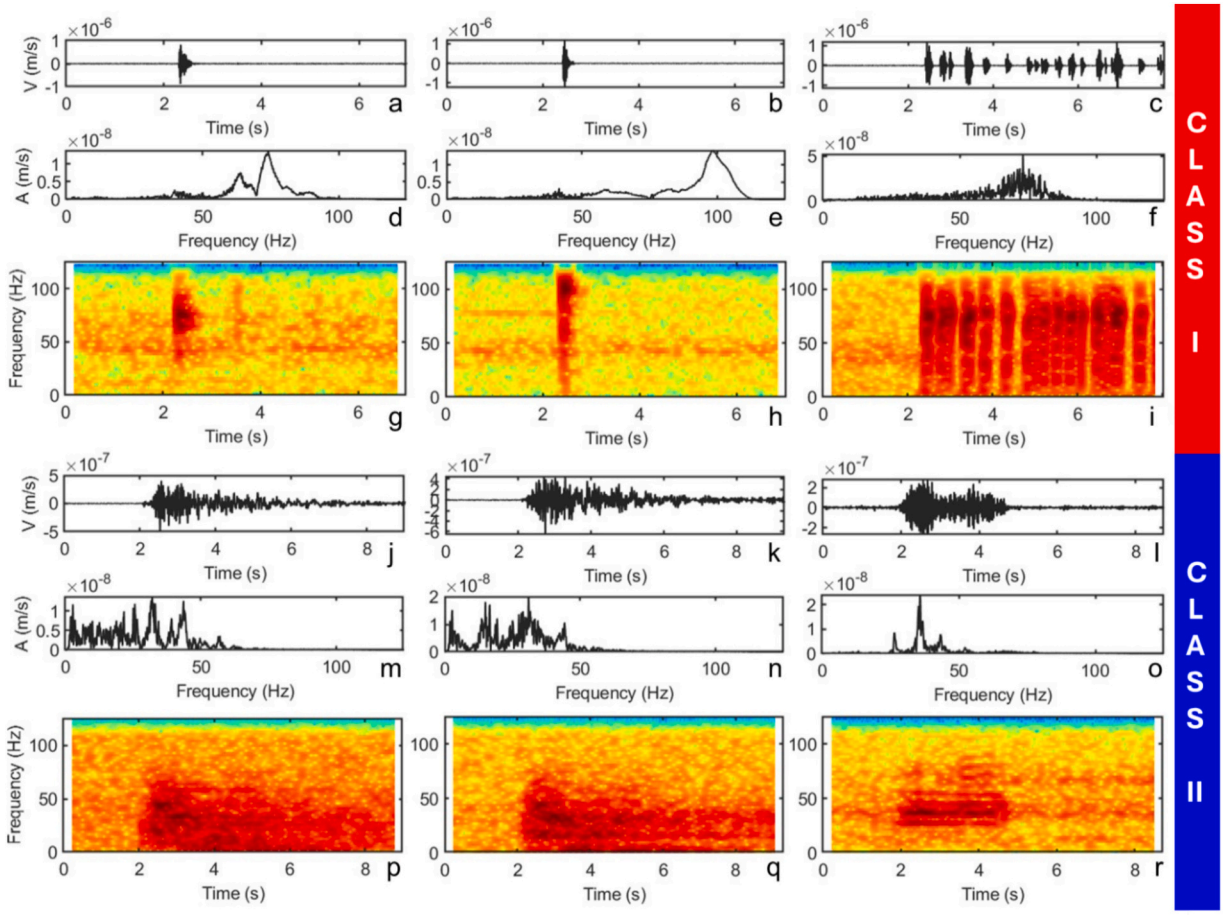


Fig. 2. Examples of events belonging to classes I and II. Each event is shown in time domain, as extracted from the continuous recordings (a to c, j to l), with its related amplitude spectrum (d to f, m to o) and spectrogram (g to i, p to r).

(Hibert et al., 2014; Colombero et al., 2021), to be used as representative input for the clustering. If the clustering of *n* features proves effective in separating the labeled events, each class can then be represented by a centroid in the *n*-dimensional space. The distance between the *n* features of a non labeled event and the reference centroids of each class can then be used to automatically assign all the event to the nearest cluster, thus allowing for an automatic classification of the whole dataset. We assumed that the minimum distance from the centroids in the *n*-dimensional space would exceed a threshold in the case of noise events not represented by a centroid and adopted this strategy to label the events belonging to this additional class.

Among the possible features, we computed five time-domain features, i.e., bracketed duration, uniform duration, maximum-to-average amplitude ratio, maximum amplitude-to-bracketed duration ratio, kurtosis; and three frequency-domain features, namely peak frequency, dominant 5-Hz frequency band and bandwidth.

Given a signal of N samples and amplitude of each i-th sample equal to A_i, the bracketed duration refers to the time interval between the first (t_a) and last (t_b) absolute amplitude values $|A_i|$ exceeding a threshold set to 15% of the maximum amplitude of the signal, following:

$$T_B = t_b - t_a \tag{1}$$

However, to account for the possible occurrence of multiple events in the extracted event window (e.g., Fig. 2c), we also computed the unform duration, as the sum of the time segments (Δt_s, with s = 1,.., S) in which the signal amplitude exceeds the 15% threshold:

$$T_U = \sum_{s=1}^{S} \Delta t_s \tag{2}$$

The difference between bracketed and uniform duration might help in discriminating repeated or overlapping events occurring within the same extracted event file.

To differentiate between short, impulsive events and longer events with less distinct amplitude peaks, the maximum-to-average amplitude ratio (A_r) was computed following:

$$A_r = \frac{max|A_i|}{\mu} \tag{3}$$

where μ is the mean amplitude of the signal:

$$\mu = \frac{1}{N}\sum_{i=1}^{N} A_i \tag{4}$$

Similarly, the maximum amplitude-to-uniform duration ratio (A_d) was calculated for each event:

$$A_d = \frac{max\ |A_i|}{T_U} \tag{5}$$

The kurtosis K of each signal was computed as an additional indicator of the signal's sharpness as the normalized fourth central moment (following the definition implemented in the MATLAB® built-in function):

$$K = \frac{\frac{1}{N}\sum_{i=1}^{N}(A_i - \mu)^4}{\left(\frac{1}{N}\sum_{i=1}^{N}(A_i - \mu)^2\right)^2} \tag{6}$$

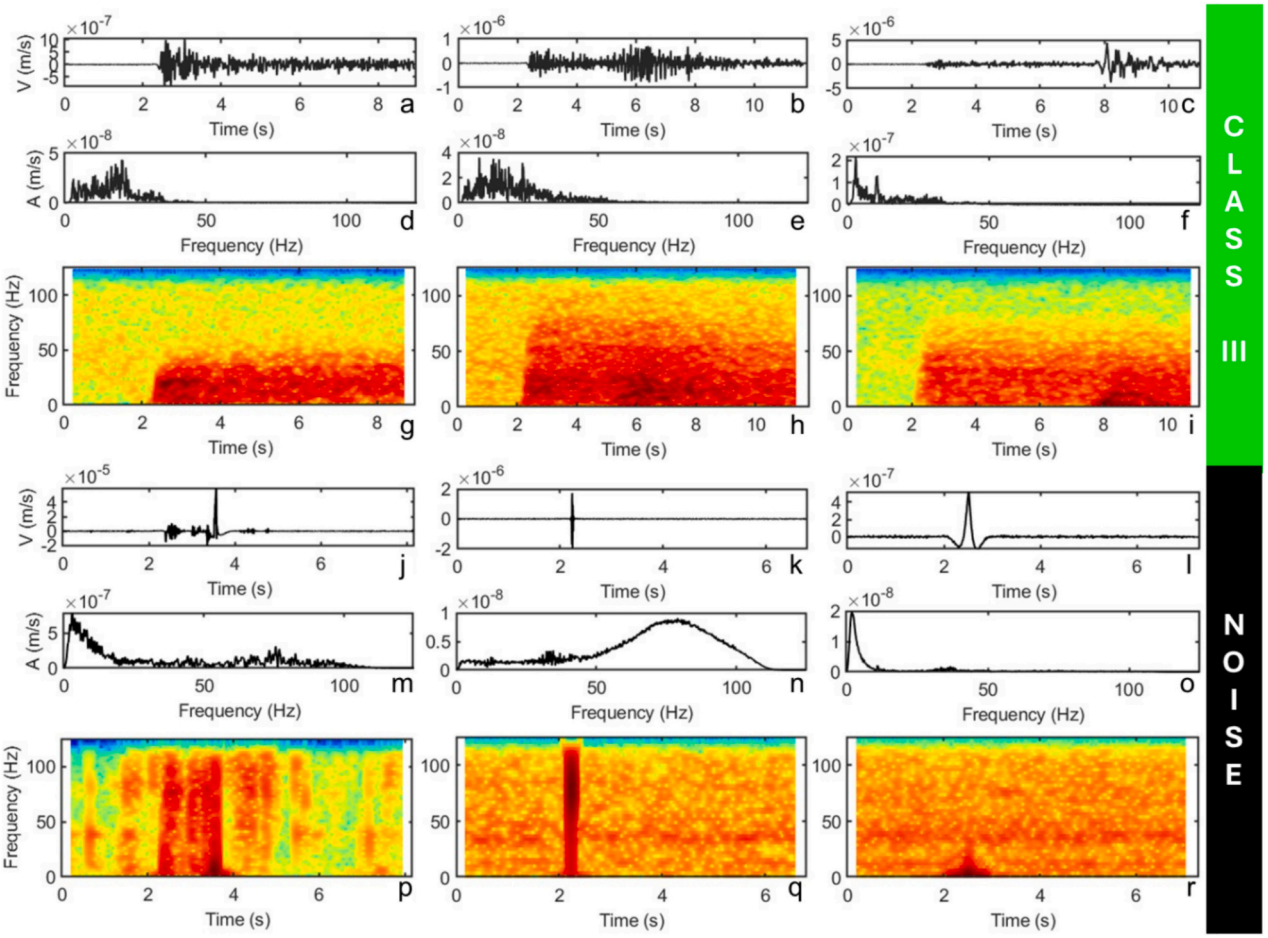


Fig. 3. Examples of events belonging to class III and noise transients. Each event is shown in time domain, as extracted from the continuous recordings (a to c, j to l), with its related amplitude spectrum (d to f, m to o) and spectrogram (g to i, p to r).

Table 1
The occurrence of the various classes of events from the labeled data of the first three monitoring months.

	Dec	Jan	Feb	TOT
CLASS I	84	32	141	257
CLASS II	10	3	8	21
CLASS III	12	11	13	36
NOISE	12	3	26	41
Uncertain – not labeled	*26*	*7*	*49*	*82*
TOT	**144**	**56**	**237**	**437**

The peak frequency was directly extracted from the amplitude spectrum of each event. Let A_j denote the spectral amplitude at the discrete frequency f_j, the peak frequency was consequently defined as:

$$f_{peak} = f_j \tag{7}$$

such that.

$$|A_j| = \max |A_j| = A_{peak} \tag{8}$$

Also in this case, to avoid possible noise spikes in the spectra affecting the estimation of the peak frequency, the spectral content was additionally quantified by the energy contained in subsequent 5-Hz frequency bands, in the 0–100 Hz range. The spectral energy in the k-th 5-Hz frequency band was therefore defined as:

$$E_k = \sum_{j:5(k-1)<f_j\le 5k} |A_j|^2 \Delta f \tag{9}$$

where k = 1, …, 20 and Δf is the frequency resolution of the discrete amplitude spectrum.

We then selected the k-th frequency band (FB_k) having maximum E_k as representative for the event.

To account for the shape and width of the peaks in the amplitude spectrum, the bandwidth exceeding a threshold equal to 80% of the maximum spectral amplitude was finally computed, following:

$$BW = \sum_{j:|A_j|\ge 0.8A_{peak}} \Delta f \tag{10}$$

To address the high dimensionality of these eight classification features and potential correlations among them, we then applied PCA for potential data space reduction. This approach simplifies the dataset by transforming the original variables into a lower-dimensional set of uncorrelated variables, known as principal components. These components are linear combinations of the initial variables where the coefficients are defined as loadings. They are ordered by the amount of variance they capture, with the first component capturing most of the dataset's original variability. We used a 3D biplot of the first two principal components to visually represent the data, projecting variables as vectors and observations as points in a three-dimensional space defined by these components.

The contribution parameter (CTR) was also calculated to quantify the contribution of each variable to the principal components. This is useful to identify the most influential features of the dataset, and was calculated as follows:

$$CTR_{j,k} = \frac{l_j^2{}_{,k}}{\sum_{i=1}^{p} l_i^2{}_{,k}} \bullet 100 \quad (11)$$

where $l_{j,k}$ is the loading of variable j on principal component k and $l_{i,k}$ represents the loadings of all variables on k.

We then applied the k-means clustering algorithm on the five classification parameters that best represent our dataset, as determined by the PCA results. The k-means unsupervised clustering technique partitions a set of n-dimensional observations (with $n = 5$) into k clusters, where k is the only user-defined input parameter. We used $k = 3$ to try to retrieve three representative centroids for the visually identified classes I to III. The results were then compared to the assigned labels to understand the efficiency of this method to cluster the whole dataset according to the minimum distance of each event parameters from the three cluster's centroids.

3.2. Machine learning approach

3.2.1. Data Pre-processing

The only inputs to the ML model were the spectrograms of the extracted events and their corresponding labels. Nevertheless, data preprocessing helped in optimizing the ML model. We consequently applied logarithmic transformation to convert the data into decibel scale, which compresses the dynamic range and enhances the event visibility. We then normalized the decibel data between zero and one by performing range normalization as:

$$S_{norm} = \frac{S - S_{min}}{S_{max} - S_{min}} \quad (12)$$

where S_{min} and S_{max} were the minimum and maximum values in the events data set, respectively. This normalization ensures that all input samples are in a comparable scale, which improves the performance of gradient-based optimization algorithms during training. The normalized data were then resized to 128×128. The resizing was to ensure that the model can be applied to spectrograms with different sizes.

3.2.2. Data balancing

The event dataset exhibited significant class imbalance, with classes II, III and noise being significantly underrepresented (Table 1). To address this issue, we used the synthetic minority over-sampling technique (SMOTE), which generates synthetic samples for minority classes by interpolating between existing samples (Chawla et al., 2002). SMOTE has been shown to improve model performance in imbalanced datasets, especially in cases where underrepresented classes are critical for accurate prediction (Chawla et al., 2002; Han et al., 2005). The method operates by selecting a minority class sample x_i, identifying its k-nearest neighbors, and generating a synthetic sample x_{new} as

$$x_{new} = x_i + \lambda \cdot (x_{neighbor} - x_i) \quad (13)$$

where $\lambda \in [0, 1]$ is a random number, and $x_{neighbor}$ is one of the k-nearest neighbors of x_i. This interpolation between points in the feature space produces new synthetic samples that help balancing the class distribution without merely duplicating data. By applying SMOTE, we ensured that the number of samples in each class are equal.

3.2.3. Model architecture

The classification task was performed using the CNN architecture shown in Fig. 4. The input layer takes in the pre-processed spectrograms, followed by three convolutional layers with 16, 32, and 64 kernels of size 5×5. All convolutional layers used the ReLU (Rectified Linear Unit) activation function, allowing the model to capture complex, non-linear patterns in the spectrograms. The final convolutional layer was connected to two fully connected layers with 256 and 64 nodes. To prevent overfitting, L2 regularization of 0.1 was considered that, penalized large weights, encouraging better generalization of the final model. The Dropout layers with a rate of 0.2 were introduced after the each fully connected layer, randomly deactivating neurons during training to further reduce overfitting (Srivastava et al., 2014). Finally, the output layer uses a SoftMax activation function to provide a probability distribution across the four output classes, including the noise class.

Categorical cross-entropy was used as the loss function, which is suitable for multi-class classification tasks. The loss is given by:

$$Loss = -\sum_{i=1}^{N}\sum_{j=1}^{K} y_{ij} log\left(\widehat{y_{ij}}\right) \quad (14)$$

where N and K are the number of samples and classes, respectively; y_{ij} represents the true label for the ith sample and jth class, whereas $\widehat{y_{ij}}$ is the predicted probability.

4. Results

4.1. Cluster analysis approach

In Fig. 5, we show the results of the PCA applied to the eight computed time- and frequency-domain features of the 100 labeled events belonging to the first three classes of Table 1. The histogram in Fig. 5a shows the variance explained by each principal component, whereas the cumulative line tracks the total explained variance as components increase. The first principal component (PC1) accounts for the largest portion of variance (71.0%), while the second and third principal components (PC2 and PC3) cover 14.9% and 5.8% respectively. The cumulative curve shows that the first two components collectively explain approximately 86% of the variance, indicating that these components effectively capture most of the information in the dataset.

The 2D biplot of Fig. 5b shows the relationships between

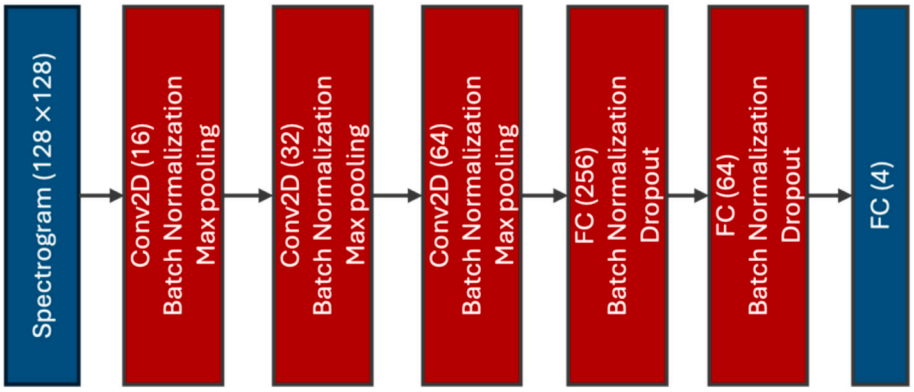


Fig. 4. CNN architecture for event classification.

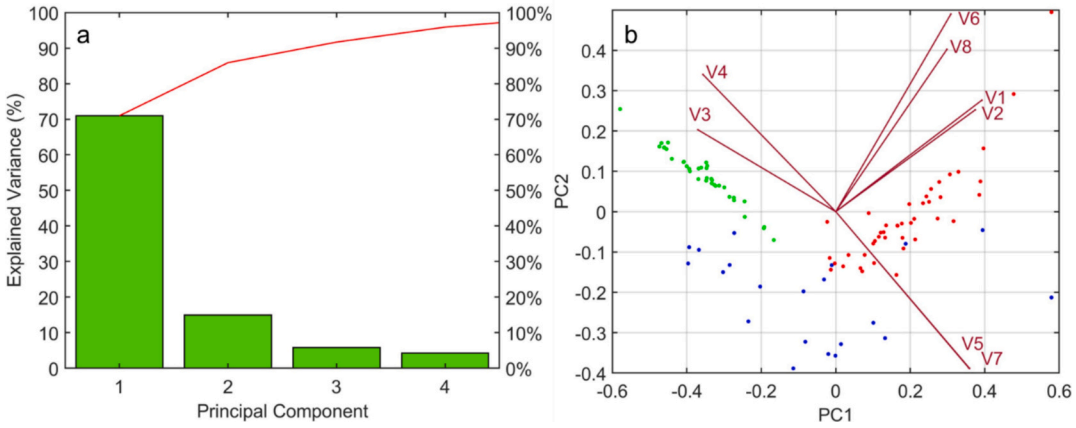


Fig. 5. (a) Histogram of the variance explained by each principal component. (b) 2D biplot showing the relationships among the first two principal components (PC1 and PC2). The dots represent the standardized observations (class I in red, class II in blue, class III in red), while the red lines indicate the feature loadings (V1 to V8, as listed in Table 2). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

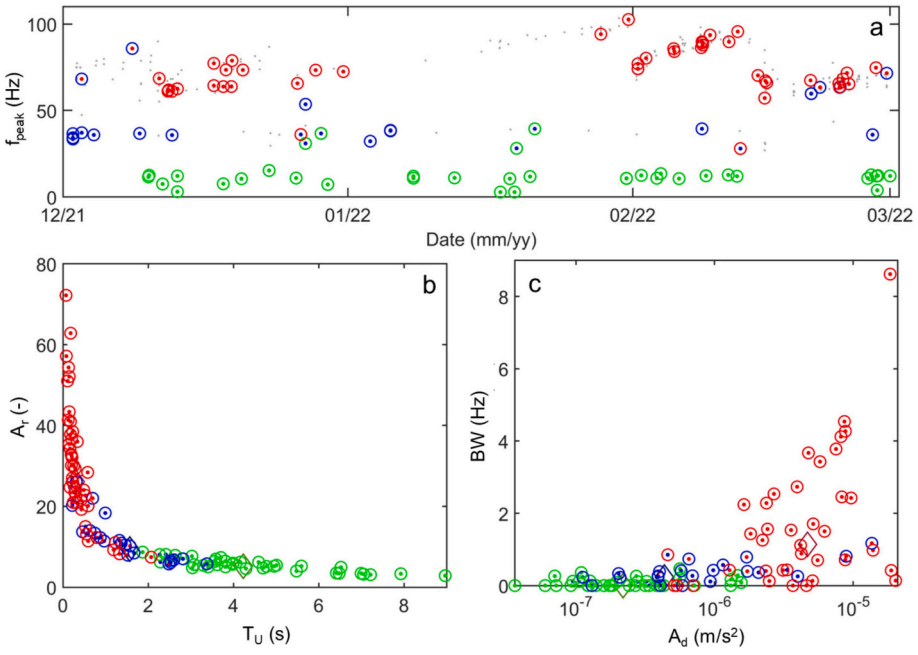


Fig. 6. k-means clustering results shown as function of the adopted classification features: (a) evolution of the peak frequency of the events over the first three monitoring months; (b) maximum-to-average amplitude ratio vs. uniform duration; (c) frequency bandwidth vs. maximum amplitude-to-uniform duration ratio. In each plot the colors refer to the three considered classes (I in red, II in blue, III in green), the circles refer to the original labels, while the dots refer to the clustering results. The cluster centroids are highlighted with colored diamonds in (b) and (c). In (a) the gray dots refer to the events belonging to class I that were excluded from the clustering to balance the number of events of each cluster.

Table 2
CTR parameter for each feature of the dataset in relation to the first two principal components (PC1 and PC2), and global CTR parameter weighted to the variance explained by each principal component (71.0% for PC1, 14.9% for PC2). The features selected for k-means clustering are highlighted in green.

	Feature	PC1	PC2	PC1 + PC2
V1	*maximum-to-average amplitude ratio* (A_r)	15.4	7.6	12.1
V2	*Kurtosis (K)*	14.0	6.4	10.9
V3	*bracketed duration* (T_B)	13.7	4.1	10.4
V4	*uniform duration* (T_U)	12.7	11.6	10.8
V5	*dominant 5-Hz band* (FB_k)	12.9	15.0	11.3
V6	*Frequency bandwidth (BW)*	9.6	24.0	10.4
V7	*peak frequency* (f_{peak})	12.8	15.0	11.3
V8	*maximum amplitude-to-uniform duration ratio* (A_d)	8.9	16.2	8.7

standardized observations and features (listed in Table 2) in the reduced space defined by the first two principal components. Length and direction of the red lines (loadings) indicate how strongly each feature contributes to the components and their correlation.

The maximum-to-average amplitude ratio and kurtosis (V1 and V2 in Fig. 5b) are positively correlated, as their vectors are closely positioned and both pointing in the same direction, suggesting a predominant contribution to PC1 and a significant one to PC2. The bracketed and uniform durations (V3 and V4 in Fig. 5b) primarily align to PC1, but V4 has higher impact on PC2 with respect to V3. The dominant 5-Hz frequency class and peak frequency loadings (V5 and V7 in Fig. 5b) are almost superimposed, as a proof of their physical meaning and the absence of noise spikes in the signal recordings. They contribute almost equally to PC1 and PC2. The frequency bandwidth and the maximum amplitude-to-bracketed duration ratio (V6 and V8 in Fig. 5b) show similarly orientated loadings, and a predominant contribution to PC2.

Computed CTR values, reported in Table 2, confirm this interpretation. Considering that many features showed similar behavior in the orientation of the loadings of Fig. 5b and in CTR values of Table 2, we selected only five features for the k-means clustering (V1, V4, V5, V6 and V8 in Table 2), reducing the dimensionality of the problem.

Fig. 6a shows the k-means clustering results as a function of the peak frequency of the events over the three months considered for the labeling. The same events are displayed in Fig. 6b and 6c as a function of couples of features adopted for the clustering, to simplify the five-dimensional space visualization. In each plot, the dots represent the k-means clustering results, whereas circles indicate the original label assigned to each event. The different colors refer to the different classes and obtained clusters (i.e., class I in red, II in blue, III in green). The k-means clustering appears almost satisfactory in recognizing events visually classified as belonging to the different classes, with some more evident discrepancies in the recognition of class II.

The confusion matrix of Fig. 7 provides an overall view of the k-means classification performance on the sub data set of 100 labeled events. Few misclassifications are present for class I and III, while many events visually classified in class II are assigned to class I by the clustering algorithm.

The accuracy of k-means classification, calculated as the ratio between correctly classified events and total number of considered events is 87%.

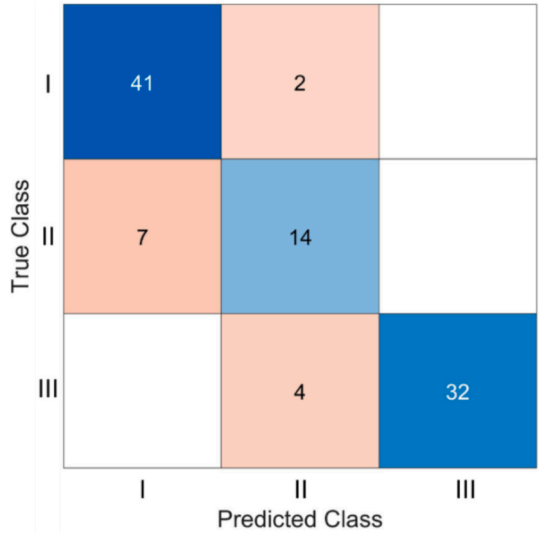


Fig. 7. Confusion matrix comparing the visual classification (true class) with k-means clustering classification (predicted class) of the subset of 100 labeled events.

4.2. Machine learning approach

The training data were limited to 355 samples from December, January and February, with Class I as majority class with 257 samples. The SMOTE method was used to match the number of samples for each group of events, resulting in a total of 1028 samples. Fig. 8 shows the examples of simulated spectrograms from the SMOTE method compared with some real examples of the events.

The labeled dataset was split into 70% for training and 30% for validation and test. The optimal hyperparameters were identified through an exhaustive search within the hyperparameter space, ensuring the model was fine-tuned for the best performance. The model was set to train for up to 1500 epochs with a batch size of 32. The learning rate was initially set to 0.00001 and reduced by 7% every 50 epochs. We used the Adam optimizer for training. Early stopping was applied with a patience of 10 epochs, monitoring the validation loss. Training was stopped before the full 1500 epochs as the early stopping was triggered. The model weights from epoch 766, which achieved the highest validation accuracy, were automatically restored. Fig. 9 shows the loss and the accuracy curves over the course of the training process. In the end, the model achieved an accuracy of 99% on the training set and 96% on the validation set. A slight deviation between the validation and training loss and accuracy is expected due to the limited training samples. While increasing the sample size could reduce this discrepancy, it conflicts with the goal of maintaining a small dataset to support early warning system applications.

The CNN model's performance was evaluated using a confusion matrix on the whole labeled dataset, providing insights into the classification accuracy across different classes, including correct and incorrect predictions for each category (Fig. 10). The confusion matrix shows that the model misclassified only 14 out of 355 events, with 12 of these errors involving Class I and noise. This is expected, as several Class I examples contain significant noise.

We applied the trained model to the entire event dataset recorded by station S1, comprising 3455 events. In Fig. 11, the dots show the peak spectral frequency of each event over time, with the colour of the data points representing the event categories identified by the ML model. The category of the events according to the manual labeling are discriminated with the different colors of the circles. Even though the frequency bands of the classes overlap, the CNN model is able to correctly predict the class in most cases. The prediction accuracy remains consistent over time, showing slightly better performance compared to k-means (Fig. 6a).

5. Discussions

We showed the application of k-means clustering in comparison with a supervised ML method for classification of microseismicity related to possible ground instabilities. The clustering was performed on the salient time- and frequency domain features of the events to describe the physical events. The performance of the supervised ML method, which was designed to classify both physical and non-physical (noise) events, was superior in comparison to the clustering algorithm. The ML model requires a good quality and representative training data set for accurate predictions. Nevertheless, once trained, it can process large datasets and classify events within seconds. The results of the two classification approaches are shown in Fig. 12 in comparison with the meteorological parameters available for the site.

The temporal distribution, total number and cumulative curves of the events classified as class I show high similarity with both classification approaches. These events, potentially related to fracturing, increase as air temperature increases during the last two monitoring months. Their peak frequency is found to be driven by air temperature fluctuations, especially during the warmer months, as further proof of their natural origin as thermally driven fracturing events. Rainfalls cause sudden drops in the peak frequency trend of these events, which

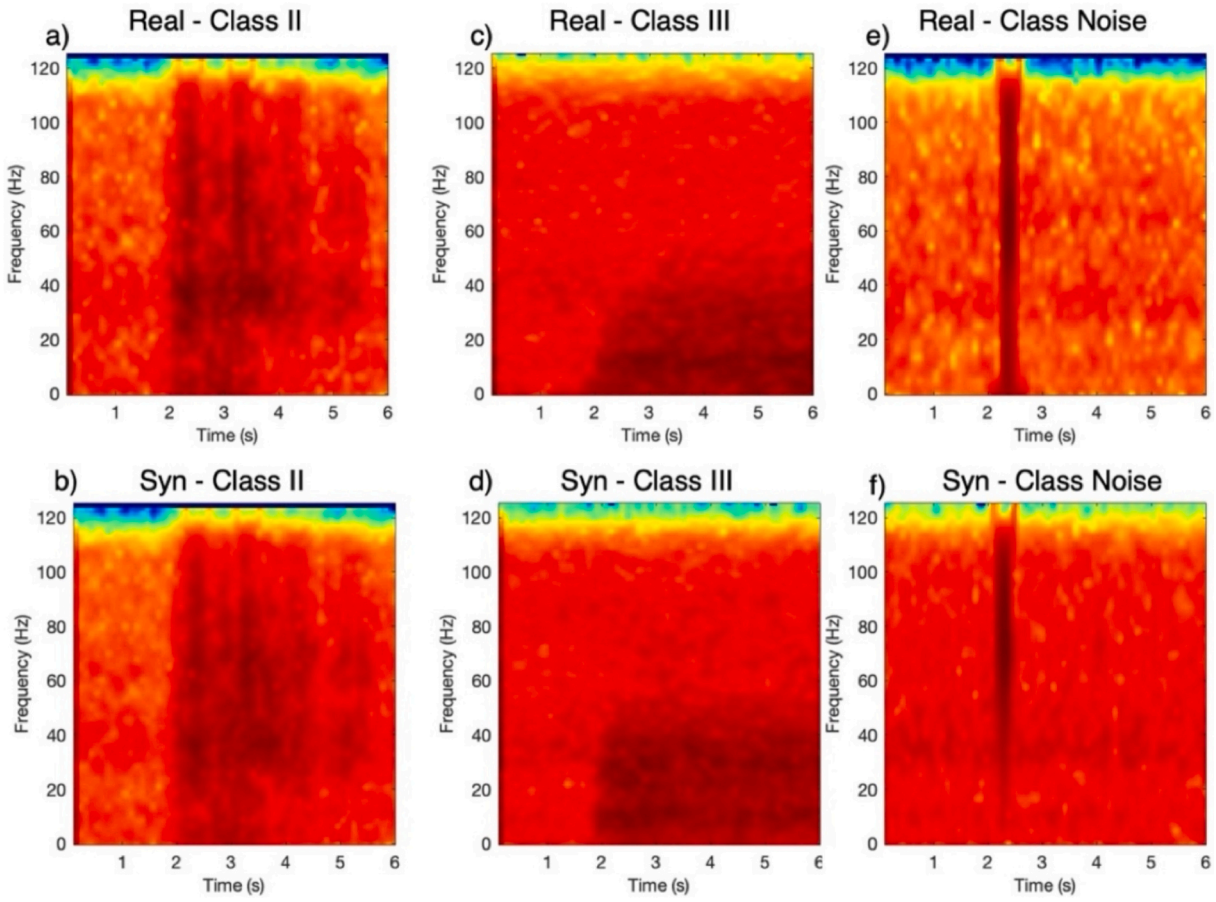


Fig. 8. The real and synthetic spectrogram of the training data for class II, class III and noise events. (a, c, e) real examples. (b, d, f) synthetic examples created by SMOTE.

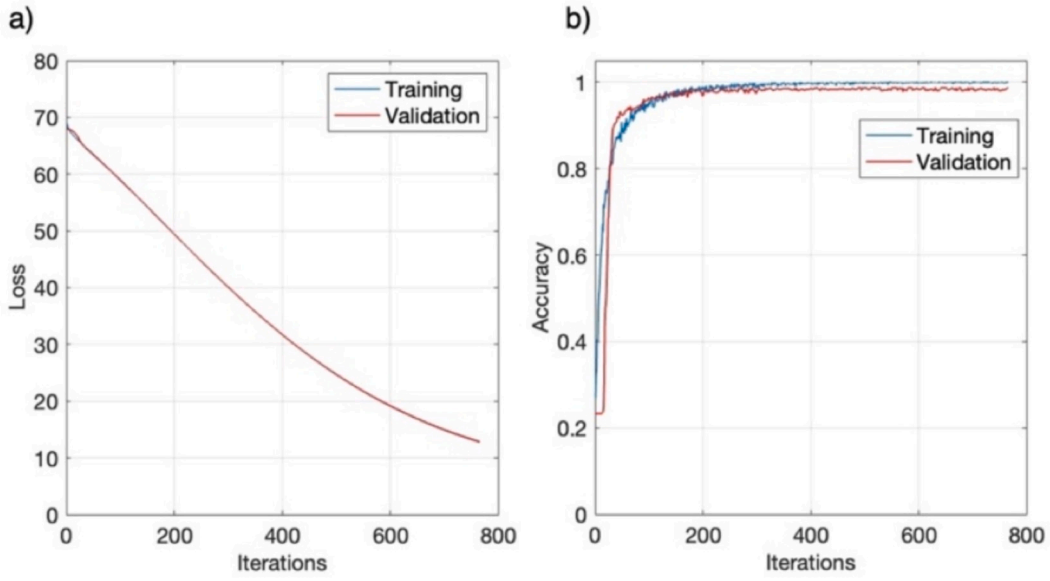


Fig. 9. (a) The loss and (b) accuracy during 766 full iterations of training.

are partially recovered a few weeks after the main precipitation events. In these time frames, the events classified as noise by both methods are found to significantly increase. This increase might be in line with a higher occurrence of electronic transients and raindrops detected in concomitance with intense precipitation. The cumulative curves related to noise show however a major discrepancy between the two classification methods in the last monitored month. This difference might be related to the criterium adopted within the clustering classification approach to deal with these events that were not represented by a cluster centroid but were declared based on a threshold on the distance to the

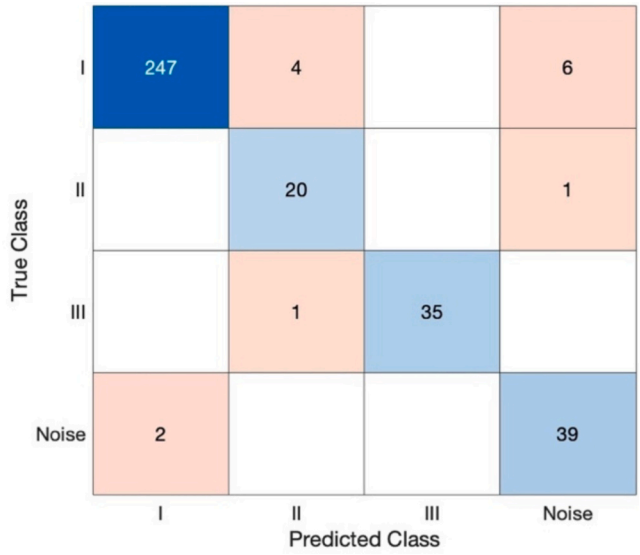


Fig. 10. Confusion matrix, showing the performance of the trained model on the labeled dataset in discriminating class I, II, III and noise events.

centroids of the three available classes. Since the events used for the clustering were selected in the first three monitoring months, it seems also reasonable that the features of the events keep evolving and varying with time. Their distances from the reference centroids are therefore expected to increase and might be the cause of their misclassification as noise events.

The events belonging to class II showed major differences according to the classification methods. It is clear from Fig. 12b and c that the frequency features play a major constraint in defining the class centroid location for the clustering approach, while these events showed more variations in the peak frequency according to the ML classification. According to the confusion matrices of Figs. 7 and 10, the temporal trend highlighted by the ML method appears more reliable. Interestingly, the major increases in the number of these events are found again in concomitance or after rainfall (Fig. 12d), thus supporting their interpretation as events related to sliding of shallow materials or water infiltration inside the rock mass at the top of the natural cave. The events related to class III showed a similar cumulative trend according to the two classification approaches, uncorrelated with the external meteorological parameters. This behavior supports the interpretation of these events as earthquakes and local minor seismic events, not related to the evolution of the cave stability.

We further evaluated the CNN model by comparing its performance to two standard supervised machine learning algorithms that are random forest and XGboost. Table 3 summarizes the class-wise precision, recall, and F1-scores for the CNN and the traditional supervised machine-learning classifiers (Random Forest and XGBoost) applied to the labeled data from December, January and February. For consistency, both Random Forest and XGBoost were trained using the same PCA-based features previously adopted for the k-means clustering, with Random Forest configured with 50 estimators and XGBoost using 50 estimators and a maximum tree depth of 4. These we features as explained before were computed for physically realistic events, excluding the noise events. Overall, the CNN achieved the highest and most stable performance across all classes, with F1-scores exceeding 0.87 for every category, including the noise class. In contrast, Random Forest and XGBoost show more variable behavior, particularly for class II, where both models exhibit a marked reduction in recall and F1-score. While traditional classifiers can reach high precision for some classes, this often comes at the expense of recall, leading to unbalanced performance under class imbalance. The CNN, by directly operating on spectrogram representations and learning task-specific features, maintains a more balanced precision–recall trade-off across classes. These results confirm the superior robustness of the CNN in discriminating seismic event types in a small and highly imbalanced dataset.

Seasonal variability is a known characteristic of microseismic data and represents a potential source of bias when training machine-learning models on limited temporal windows. In this work, model training and hyperparameter tuning were necessarily restricted to winter data due to the six-month availability of labeled events, which constitutes a compromise motivated by the objective of developing a fast and operational early-warning workflow. To evaluate and validate the impact of seasonality, the trained CNN was applied to an unseen data from May, representing warmer seasonal conditions. The confusion matrix in Fig. 13 shows robust performance for the detection of fracturing-related microseismic events (Class I) and perfect identification of earthquake-related signals (Class III), while a reduced classification accuracy is observed for the remaining classes. This decrease is not unexpected given the seasonal shift in signal characteristics and the fact that these data were not used during training. Importantly, the overall estimation of fracturing activity remains sufficiently accurate for the intended early-warning application. These results highlight both the robustness of the proposed approach and the need for future extensions using multi-season and multi-site training data to further mitigate seasonal bias and reduce the risk of over-fitting on seasonally constrained datasets.

Given that the dataset was heavily imbalanced (Table 1), we applied the SMOTE method to balance the training data and reduce classifier bias toward majority classes. To assess the impact of SMOTE on model performance, Table 4 shows the CNN results without synthetic oversampling. When SMOTE is not used, the model shows a noticeable degradation in performance for minority classes, where recall and F1-score decrease substantially. In contrast, the application of SMOTE results in consistently high and well-balanced precision and recall across all classes, with only class II F1-scores below 0.9 (Table 3). This indicates

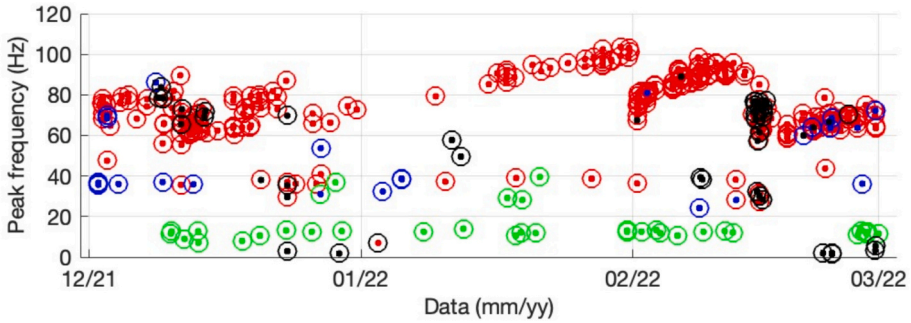


Fig. 11. CNN clustering results on the training data shown as the evolution of the peak frequency of the events over time. In each plot the colors refer to the four considered classes (I in red, II in blue, III in green, noise in black), the circles refer to the original labels, while the dots refer to the clustering results. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

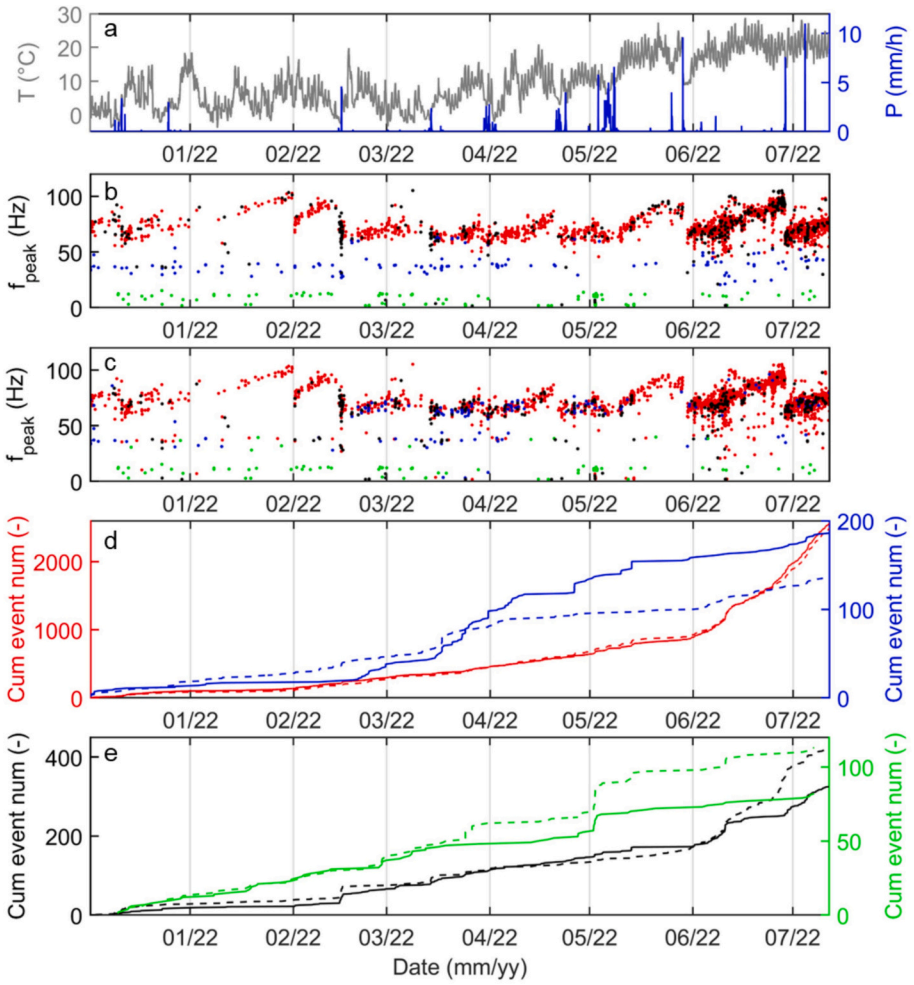


Fig. 12. (a) Hourly air temperature and rainfall data. Peak frequency of the different classes of events (class I in red, II in blue, III in green, noise in black) as obtained from (b) the k-means clustering and (c) machine learning classification approach. Cumulative curves (dashed lines for clustering, solid lines for the ML approach) referred to the number of events classified to (d) class I (red) and II (blue); (e) class III (green) and noise (black). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 3
Class-wise precision, recall, and F1-scores for the proposed CNN model compared with Random Forest and XGBoost classifiers. Random Forest was trained with 50 estimators, while XGBoost used 50 estimators with a maximum tree depth of 4, both relying on the same features from PCA analysis.

Classes	F1	Precision	Recall
Proposed CNN model			
I	0.98	0.99	0.96
II	0.87	0.8	0.95
III	0.99	1	0.97
Noise	0.9	0.85	0.95
Random forest			
Classes	F1 score	Precision	Recall
I	0.9	0.99	0.83
II	0.67	0.6	0.75
III	0.9	0.9	0.9
XGboost			
Classes	F1 score	Precision	Recall
I	0.92	0.86	1
II	0.33	0.5	0.25
III	0.86	0.82	0.9

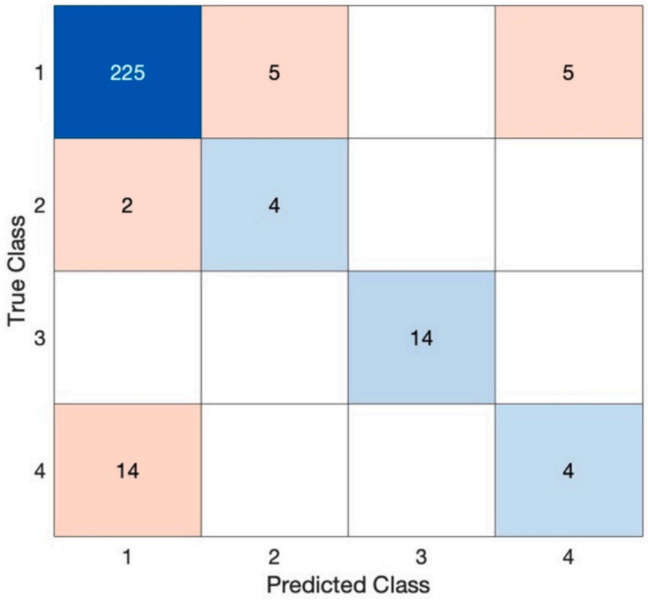


Fig. 13. Confusion matrix, showing the performance of the trained model on the data from May in discriminating class I, II, III and noise events.

that the observed improvements are not limited to synthetic samples but

Table 4
Class-wise precision, recall, and F1-scores obtained with the CNN model trained without SMOTE, highlighting the reduced performance for minority classes under imbalanced training conditions.

DL model without SMOTE			
Classes	F1 score	Precision	Recall
I	0.97	0.93	1
II	0.82	0.89	0.76
III	0.89	0.9	0.89
Noise	0.76	1	0.61

translate into enhanced recognition of real minority-class instances. As SMOTE is applied in the spectrogram feature space rather than on raw waveforms, the generated samples represent interpolations of existing spectral patterns instead of physically reconstructed seismic signals, thereby reducing the risk of introducing unrealistic artifacts. While this approach does not model true physical variability, the results demonstrate that SMOTE provides an effective and practical solution for improving classification robustness in small and highly imbalanced datasets.

Epistemic uncertainty in predicting microseismic events is crucial for understanding the model's confidence in predicting the classes and for evaluating whether fine-tuning or model adjustments are needed over time. To quantify prediction uncertainty, Monte Carlo Dropout is used during inference. Dropout is kept active during prediction, and the model is run multiple times (1000 forward passes). The uncertainty is calculated as the standard deviation of the predicted probabilities across forward passes:

$$\text{Uncertainty} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(p_i - \bar{p})^2} \tag{15}$$

where $\bar{p}$ and p_i are the average and individual probabilities. Fig. 14 shows the histograms of the obtained standard deviations for each prediction across the labeled and unlabeled dataset. In Fig. 14a, the misclassified samples from labeled data set are marked with asterisks, indicating their higher uncertainty levels compared to the distribution of the uncertainty. Notably, most of the misclassified events exhibit significantly elevated uncertainty (above 0.15), suggesting a clear relationship between high uncertainty and incorrect predictions. This demonstrates the model's ability to signal ambiguity in its classifications, offering valuable insights for manual intervention and potential model fine-tuning.

In operational settings, we expect the overall distribution of predictive uncertainty to increase in the presence of changes in data characteristics, thereby acting as an indicator of model degradation and the potential need for fine-tuning with newly labeled data. Similarly, in the context of developing a more generalized model trained on data from multiple sites, uncertainty can serve as an effective metric to assess whether the model is suitable for handling data from unseen locations. When the CNN is applied to the entire dataset, the average standard deviation of the predictive distribution is 0.045 for class I, 0.27 for class II, 0.05 for class III, and 0.057 for noise events. The elevated uncertainty associated with class II is consistent with its higher variability and partial overlap with class I and noise, whereas the lower uncertainty observed for classes I, III, and noise indicates more stable and confident predictions. Overall, uncertainty estimation provides complementary information beyond standard performance metrics and enhances the interpretability and operational relevance of the proposed model.

6. Conclusions

We showed the application of two methods, k-means and CNN, for classifying microseismic events—an essential step toward developing effective, near real-time systems for failure detection in natural hazards such as landslides. The application to challenging data from the top of Bossea caves showed superior accuracy of the CNN method compared to k-means. This is mainly because supervised CNN method can capture complex patterns and relationships in the spectrograms of seismic signals that might not be readily retrievable through clustering approaches, especially when the frequency-based features related to fracturing events overlap with other classes of events. Nevertheless, both k-means and CNN methods provided satisfactory results that showed strong correlation to the meteorological parameters available for the site. The epistemic uncertainty of the CNN model's predictions is estimated using the Monte Carlo Dropout inference approach. This provides valuable insight into the reliability of the results and can help highlight significant changes in the seismic response of the investigated site, as well as the potential need for model fine-tuning. In the future, we intend to expand the training dataset to include data from multiple locations and increase the number of event classes to improve the CNN model's generalizability, with the goal of achieving reliable performance even on data from previously unseen sites.

CRediT authorship contribution statement

Farbod Khosro Anjom: Writing – review & editing, Writing – original draft, Methodology, Data curation. **Lorena Di Toro:**

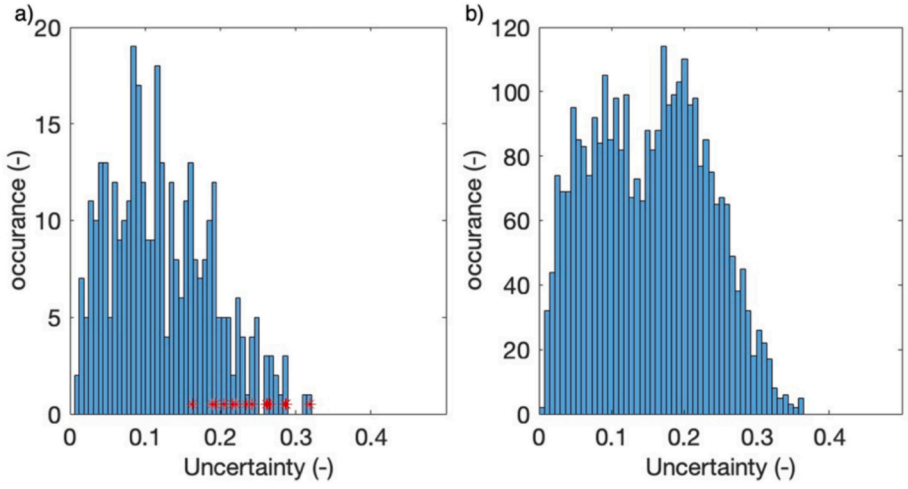


Fig. 14. The computed standard deviation of the predicted results using Monte Carlo Dropout method for (a) labeled data set and (b) unlabeled (whole) data set. In (a) the asterisks show the standard deviation of the misclassified samples.

Methodology. **Chiara Colombero:** Writing – review & editing, Writing – original draft, Supervision, Data curation, Conceptualization.

Declaration of competing interest

None.

Acknowledgement

This study was carried out within the RETURN Extended Partnership and received funding from the European Union Next-GenerationEU (National Recovery and Resilience Plan – NRRP, Mission 4, Component 2, Investment 1.3 – D.D. 1243 2/8/2022, PE0000005) – Vertical Spoke 2.

Data availability

Data will be made available on request.

References

Anikiev, D., Birnie, C., Waheed, U.B., Alkhalifah, T., Gu, C., Verschuur, D.J., Eisner, L., 2023. Machine learning in microseismic monitoring. Earth Sci. Rev. 239, 104371. https://doi.org/10.1016/j.earscirev.2023.104371.

Carr, C.G., Carmichael, J.D., Pettit, E.C., Truffer, M., 2020. The influence of environmental microseismicity on detection and interpretation of small-magnitude events in a polar glacier setting. J. Glaciol. 66 (259), 790–806. https://doi.org/10.1017/jog.2020.48.

Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P., 2002. SMOTE: synthetic minority over-sampling technique. J. Artif. Intell. Res. 16, 321–357. https://doi.org/10.1613/jair.953.

Chen, Z., Cui, Y., Pu, Y., Rui, Y., Chen, J., Mengli, D., Yu, B., 2024. Classification of microseismic signals using machine learning. Processes 12 (6), 1135. https://doi.org/10.3390/pr12061135.

Colombero, C., Comina, C., Vinciguerra, S., Benson, P.M., 2018. Microseismicity of an unstable rock mass: from field monitoring to laboratory testing. J. Geophys. Res. Solid Earth 123 (2), 1673–1693. https://doi.org/10.1002/2017JB014612.

Colombero, C., Godio, A., Jongmans, D., 2021. Ambient seismic noise and microseismicity monitoring of a prone-to-fall quartzite tower (Ormea, NW Italy). Remote Sens. 13 (9), 1664. https://doi.org/10.3390/rs13091664.

Dabove, P., Colombero, C., Salerno Quaroni, A., 2023. Towards the monitoring of underground caves using geomatics and geophysical techniques: 3D analyses and seismic response. ISPRS Int. J. Geo Inf. 12 (5), 191. https://doi.org/10.3390/ijgi12050191.

Dahal, A., Lombardo, L., 2025. Towards physics-informed neural networks for landslide prediction. Eng. Geol. 344, 1–14. https://doi.org/10.1016/j.enggeo.2024.107852.

Faillettaz, J., Funk, M., Beutel, J., Vieli, A., 2019. Towards early warning of gravitational slope failure with co-detection of microseismic activity: the case of an active rock glacier. Nat. Hazards Earth Syst. Sci. 19 (7), 1399–1413. https://doi.org/10.5194/nhess-19-1399-2019.

Falcin, A., Métaxian, J.-P., Mars, J., Stutzmann, É., Komorowski, J.-C., Moretti, R., Malfante, M., Beauducel, F., Saurel, J.-M., Dessert, C., Burtin, A., Ucciani, G., De Chabalier, J.-B., Lemarchand, A., 2021. A machine-learning approach for automatic classification of volcanic seismicity at La Soufrière Volcano, Guadeloupe. J. Volcanol. Geotherm. Res. 411, 107151. https://doi.org/10.1016/j.jvolgeores.2020.107151.

Gatto, M.P.A., Misiano, S., Montrasio, L., 2025. Space-time prediction of rainfall-induced shallow landslides through Artificial Neural Networks in comparison with the SLIP model. Eng. Geol. 344, 107822. https://doi.org/10.1016/j.enggeo.2024.107822.

Grigoli, F., Clinton, J.F., Diehl, T., Kaestli, P., Scarabello, L., Agustsdottir, T., Kristjansdottir, S., Magnusson, R., Bean, C.J., Broccardo, M., Cesca, S., Dahm, T., Hjorleifsdottir, V., Cabrera, B.M., Milkereit, C., Nooshiri, N., Obermann, A., Racine, R., Rinaldi, A.P., et al., 2022. Monitoring microseismicity of the Hengill geothermal field in Iceland. Sci. Data 9 (1), 220. https://doi.org/10.1038/s41597-022-01339-w.

Han, H., Wang, W.-Y., Mao, B.-H., 2005. Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning. Adv. Intel. Comput. 878–887. https://doi.org/10.1007/11538059_91.

Helmstetter, A., Garambois, S., 2010. Seismic monitoring of Séchilienne rockslide (French Alps): Analysis of seismic signals and their correlation with rainfalls. J. Geophys. Res.: Earth Surf. 115 (F3). https://doi.org/10.1029/2009JF001532, 2009JF001532.

Hibert, C., Mangeney, A., Grandjean, G., Baillard, C., Rivet, D., Shapiro, N.M., Satriano, C., Maggi, A., Boissier, P., Ferrazzini, V., Crawford, W., 2014. Automated identification, location, and volume estimation of rockfalls at Piton de la Fournaise volcano. J. Geophys. Res.: Earth Surf. 119 (5), 1082–1105. https://doi.org/10.1002/2013JF002970.

Lacroix, P., Helmstetter, A., 2011. Location of Seismic Signals Associated with Microearthquakes and Rockfalls on the Sechilienne Landslide, French Alps. Bull. Seismol. Soc. Am. 101 (1), 341–353. https://doi.org/10.1785/0120100110.

Lamb, O.D., Lees, J.M., Franco-Marin, L., Lazo, J., Rivera, A., Shore, M.J., Lee, S.J., 2022. Persistent shallow micro-seismicity at Llaima volcano, Chile, with implications for long-term monitoring. J. Volcanol. Geotherm. Res. 426, 107528. https://doi.org/10.1016/j.jvolgeores.2022.107528.

Langet, N., Silverberg, F.M.J., 2023. Automated classification of seismic signals recorded on the Åknes rock slope, Western Norway, using a convolutional neural network. Earth Surf. Dyn. 11, 89–115. https://doi.org/10.5194/esurf-11-89-2023.

Levy, C., Jongmans, D., Baillet, L., 2011. Analysis of seismic signals recorded on a prone-to-fall rock column (Vercors massif, French Alps): Analysis of seismic events preceding a rockfall. Geophys. J. Int. 186 (1), 296–310. https://doi.org/10.1111/j.1365-246X.2011.05046.x.

Liu, J., Yang, S., Li, L., Pan, X., Yin, X., Hu, J., Peng, F., 2024. Precursory characteristics of large-scale rockmass collapse based on multiple microseismic parameters. J. Cent. South Univ. 31 (3), 878–889. https://doi.org/10.1007/s11771-024-5610-y.

Malfante, M., Dalla Mura, M., Mars, J.I., Métaxian, J., Macedo, O., Inza, A., 2018. Automatic classification of volcano seismic signatures. J. Geophys. Res. Solid Earth 123 (12). https://doi.org/10.1029/2018JB015470.

Manley, G.F., Mather, T.A., Pyle, D.M., Clifton, D.A., Rodgers, M., Thompson, G., Londoño, J.M., 2022. A deep active learning approach to the automatic classification of volcano-seismic events. Front. Earth Sci. 10, 807926. https://doi.org/10.3389/feart.2022.807926.

Maxwell, S.C., Rutledge, J., Jones, R., Fehler, M., 2010. Petroleum reservoir characterization using downhole microseismic monitoring. Geophysics 75 (5), 75A129–75A137. https://doi.org/10.1190/1.3477966.

Pantobe, L., Burtin, A., Chanard, K., Komorowski, J.-C., 2023, May 15. Evolution of shallow volcanic seismicity in the hydrothermal system of La Soufrière de Guadeloupe following the April 2018 Mlv 4.1 earthquake. https://doi.org/10.5194/egusphere-egu23-3374.

Provost, F., Hibert, C., Malet, J.-P., 2017. Automatic classification of endogenous landslide seismicity using the Random Forest supervised classifier. Geophys. Res. Lett. 44, 113–120. https://doi.org/10.1002/2016GL070709.

Sedghizadeh, M., van den Berghe, M., Shcherbakov, R., 2024. Leveraging the ETAS model to forecast mining microseismicity. Geophys. J. Int. 238 (3), 1491–1504. https://doi.org/10.1093/gji/ggae236.

Senfaute, G., Duperret, A., Lawrence, J.A., 2009. Micro-seismic precursory cracks prior to rock-fall on coastal chalk cliffs: a case study at Mesnil-Val, Normandie, NW France. Nat. Hazards Earth Syst. Sci. 9 (5), 1625–1641. https://doi.org/10.5194/nhess-9-1625-2009.

Shu, H., Dawod, A.Y., Dong, L., 2024. Recognition and classification of microseismic event waveforms based on histogram of oriented gradients and shallow machine learning approach. J. Appl. Geophys., 105551 https://doi.org/10.1016/j.jappgeo.2024.105551.

Spillmann, T., Maurer, H., Green, A.G., Heincke, B., Willenberg, H., Husen, S., 2007. Microseismic investigation of an unstable mountain slope in the Swiss Alps. J. Geophys. Res. Solid Earth 112 (B7), 2006JB004723. https://doi.org/10.1029/2006JB004723.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 15 (1), 1929–1958.

Tonnellier, A., Helmstetter, A., Malet, J.-P., Schmittbuhl, J., Corsini, A., Joswig, M., 2013. Seismic monitoring of soft-rock landslides: the super-sauze and valoria case studies. Geophys. J. Int. 193 (3), 1515–1536. https://doi.org/10.1093/gji/ggt039.

Verliac, M., Le Calvez, J., 2021. Microseismic monitoring for reliable CO_2 injection and storage—Geophysical modeling challenges and opportunities. Lead. Edge 40 (6), 418–423. https://doi.org/10.1190/tle40060418.1.

Wenner, M., Hibert, C., van Herwijnen, A., Meier, L., Walter, F., 2021. Near-real-time automated classification of seismic signals of slope failures with continuous random forests. Nat. Hazards Earth Syst. Sci. 21, 339–361. https://doi.org/10.5194/nhess-21-339-2021.

West, M.E., Larsen, C.F., Truffer, M., O'Neel, S., LeBlanc, L., 2010. Glacier microseismicity. Geology 38 (4), 319–322. https://doi.org/10.1130/G30606.1.

Wilkins, A.H., Strange, A., Duan, Y., Luo, X., 2020. Identifying microseismic events in a mining scenario using a convolutional neural network. Comput. Geosci. 137, 104418. https://doi.org/10.1016/j.cageo.2020.104418.

Xin, B., Huang, Z., Huang, S., Feng, L., 2024. Ensemble learning improves the efficiency of microseismic signal classification in landslide seismic monitoring. Sensors 24 (15), 4892. https://doi.org/10.3390/s24154892.

Yang, Y., Chen, G., Meng, X., Bian, S., Chong, Y., Shi, W., Jiang, W., Jin, J., Li, C., Mu, X., Yue, D., 2022. Analysis of the microseismicity characteristics in landslide dam failure flume tests: Implications for early warning and dynamics inversion. Landslides 19 (4), 789–808. https://doi.org/10.1007/s10346-021-01783-6.

Yfantis, G., Pytharouli, S., Lunn, R.J., Carvajal, H.E.M., 2021. Microseismic monitoring illuminates phases of slope failure in soft soils. Eng. Geol. 280, 105940. https://doi.org/10.1016/j.enggeo.2020.105940.

Zhang, H., Zeng, J., Ma, C., Li, T., Deng, Y., Song, T., 2021. Multi-Classification of complex Microseismic Waveforms using Convolutional Neural Network: a Case Study in Tunnel Engineering. Sensors 21 (20), 6762. https://doi.org/10.3390/s21206762.

Zhou, C., Cao, Y., Gan, L., Wang, Y., Motagh, M., Roessner, S., Hu, X., Yin, K., 2024. A novel framework for landslide displacement prediction using MT-InSAR and machine learning techniques. Eng. Geol. 334, 107497. https://doi.org/10.1016/j.enggeo.2024.107497.