frontiers | Research Topics

Applications of machine learning in volcanology

Edited by
Bellina Di Lieto, Pierdomenico Romano, Silvia Scarpetta,
Yosuke Aoki and Carmen Benítez

Published in
Frontiers in Earth Science

ISSN 1664-8714
ISBN 978-2-8325-6285-7
DOI 10.3389/978-2-8325-6285-7

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Applications of machine learning in volcanology

Topic editors

Bellina Di Lieto — Vesuvius Observatory, National Institute of Geophysics and Volcanology (INGV), Italy

Pierdomenico Romano — Vesuvius Observatory, National Institute of Geophysics and Volcanology (INGV), Italy

Silvia Scarpetta — University of Salerno, Italy

Yosuke Aoki — The University of Tokyo, Japan

Carmen Benítez — Universidad de Granada, Spain

Citation

Di Lieto, B., Romano, P., Scarpetta, S., Aoki, Y., Benítez, C., eds. (2025). *Applications of machine learning in volcanology*. Lausanne: Frontiers Media SA.

doi: 10.3389/978-2-8325-6285-7

Table of contents

frontiers | Frontiers in Earth Science

TYPE Editorial
PUBLISHED 14 April 2025
DOI 10.3389/feart.2025.1604299

OPEN ACCESS

EDITED AND REVIEWED BY
Valerio Acocella,
Roma Tre University, Italy

*CORRESPONDENCE
Pierdomenico Romano,
pierdomenico.romano@ingv.it

RECEIVED 01 April 2025
ACCEPTED 04 April 2025
PUBLISHED 14 April 2025

CITATION
Romano P, Di Lieto B, Scarpetta S, Benítez C and Aoki Y (2025) Editorial: Applications of machine learning in volcanology.
Front. Earth Sci. 13:1604299.
doi: 10.3389/feart.2025.1604299

Editorial: Applications of machine learning in volcanology

Pierdomenico Romano[1]*, Bellina Di Lieto[1], Silvia Scarpetta[2], Carmen Benítez[3] and Yosuke Aoki[4]

[1]Istituto Nazionale di Geofisica e Vulcanologia, Osservatorio Vesuviano, Napoli, Italy, [2]Dipartimento di Fisica "E. R. Caianiello", Università degli Studi di Salerno, Fisciano, Italy, [3]ETS of Computer and Telecommunications Engineering, University of Granada, Granada, Spain, [4]Earthquake Research Institute, The University of Tokyo, Tokyo, Japan

KEYWORDS

machine learning in volcanology, volcanic eruption forecasting, unsupervised machine learning (UML), supervised machine learning (SML), time series automated analysis

Editorial on the Research Topic
Applications of machine learning in volcanology

Machine learning (ML) techniques, capable of learning and adapting through automated or semi-automated algorithms and statistical models, allow us to classify signals and infer insights from patterns "hidden" in data, hence offering deeper understanding when analyzing and correlating large datasets. In recent decades, geosciences have seen a growth of the data available to the scientific community, as denser monitoring networks surround fault structures, volcanoes, geothermal fields and other regions of geological interest.

ML techniques are applied to volcanology to improve the understanding of volcano dynamics and forecasting volcanic activity (Carniel and Guzmán, 2021). Recent reviews (Karpatne et al., 2019) highlight how ML's adaptability addresses key challenges in geosciences, from noisy data to real-time decision-making under uncertainty. These techniques can be applied to early warning systems, significantly enhancing risk assessment and public safety.

This Research Topic aims to be a compendium of the latest advancements in the use of ML techniques within geosciences. It has attracted 8 contributions, whose introductions follow.

Titos et al. explored advanced techniques for monitoring volcano-tectonic (VT) earthquakes, by leveraging transfer learning, a procedure capable of adapting ML models to specific regions despite limited data available. As traditional methods often struggle with accurately classifying seismic signals in these settings, by using pre-trained models that have already learned to detect earthquake patterns better identifications and categorizations of VT-events have been found by the authors. The proposed approach represents a significant advance for volcano monitoring systems, allowing researchers to identify early warning signs of eruptions more effectively.

Convolutional neural network (CNN) models have been designed to classify first-motion polarities of seismic records in volcanic and tectonic regions, as described in the manuscript by Messuti et al., where the development of the CFM algorithm is reported; this uses CNNs for automated, real-time classification, processing local and regional seismic waveforms and distinguishing between compressional

and dilatational first motions. The network was trained on labeled datasets of seismic events: testing results demonstrate how CFM achieves an accuracy >90%, and adaptability across regions with different noise levels and tectonic structures, highlighting its potential to standardize and streamline polarity classification in seismic studies.

Noguchi and Shoji explored the use of supervised ML to automate the extraction of stratigraphic features from visible images. Traditional methods for analyzing geological stratigraphy are often labor-intensive and subject to human error, making automated techniques highly valuable. Their approach involves training a CNN on a labeled dataset of stratigraphic images, allowing the model to learn to identify various geological formations with high accuracy. Such a CNN model outperformed conventional methods, effectively classifying stratigraphic features across diverse geological environments.

Volcanic eruption forecasting represents an effort valuable for scientists. Rey-Devesa et al. explored this Research Topic by introducing a universal ML approach which uses seismic features. By analyzing seismic data from diverse volcanic settings, the study demonstrates how four different seismic features passed to algorithms like neural networks can effectively predict eruption patterns. Results show the approach's robustness across different volcanoes, highlighting its potential for real-time monitoring and early warning systems, providing a tool for probabilistic short-term volcanic eruption forecasting, easily usable on different volcanic systems.

Mitchinson et al., by using two concurrent unsupervised ML clustering algorithms, found anomalous earthquake swarms at Mt. Ruapehu, New Zealand: the article highlights the challenges of distinguishing swarms from background seismicity, using the chosen algorithms to classify patterns in seismic data. By leveraging historical records and real-time monitoring, the study enhances understanding of swarm characteristics and their links to volcanic processes. The findings support more accurate swarm identification, contributing to improved volcanic hazard assessment and mitigation efforts, and correlating their characteristics to seismic swarms related to fluid migration, more likely to happen near fault systems.

Volcanic thermal emissions could be indicators of impending or ongoing eruptions. Saunders-Shultz et al. describe a specific CNN, the HotLINK system, for automatic identification and quantification of volcanic hotspots in Alaska: the study automates the detection of thermal anomalies from satellite imagery, improving the monitoring of active volcanic regions. The methodology integrates thermal data and predictive modeling to achieve precise identification of hotspots, offering a scalable solution for volcanic hazard management. The results demonstrate its effectiveness in enhancing real-time surveillance and contribute to advancing geoscientific applications.

Giannoulis et al. investigate the drivers of surface temperature changes above geothermal systems by using the deep learning engine DAITAN applied to data recorded at Vulcano Island, Italy. By analyzing a combination of surface and subsurface factors, the study identifies patterns and relationships affecting temperature variations. The approach leverages thermal data and environmental parameters to enhance the understanding of geothermal activity and its external influences. The findings provide valuable insights for monitoring and improving geothermal systems, and predicting potential hazards.

Di Benedetto et al. introduced a grid-search methodology for optimizing Short-Term Average over Long-Term Average (STA/LTA) parameters in seismological applications, focusing on Stromboli's explosion quakes. By testing different parameter combinations by using a quality parameter, the approach improves signal detection and characterization of volcanic seismicity. The study demonstrates how this method refines STA/LTA settings for greater sensitivity to eruption-related seismic signals. The results highlight the methodology's potential not only for enhancing seismic monitoring in active volcanic regions, contributing to more accurate early warning systems, but also for applying it to a wider typology of signals.

Collectively, the contributions proposed in this Research Topic explore how ML techniques can enhance insights into volcanic activities, achieving results that overwhelm classical methods. ML will assume a transformative role in volcanology–from automating labor-intensive tasks to uncovering hidden patterns in complex datasets. As emphasized by Rouet-Leduc et al. (2017), as the field moves toward operational monitoring, future works must balance innovation with interpretability. With increasing availability of data and the growing of population living near active volcanoes, integrating ML techniques into volcanic studies holds the promise of transforming how researchers will forecast eruptions and respond to volcanic hazards, representing a significant step forward in the future of volcanic research, having the potential to redefine volcanic risk mitigation in the decade ahead.

Author contributions

PR: Writing – original draft, Writing – review and editing. BD: Writing – review and editing, Writing – original draft. SS: Writing – review and editing. CB: Writing – review and editing. YA: Writing – review and editing.

Funding

The author(s) declare that no financial support was received for the research and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Generative AI statement

The authors declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

References

Carniel, R., and Raquel Guzmán, S. (2021). *Machine learning in volcanology: a review*. IntechOpen. doi:10.5772/intechopen.94217

Karpatne, A., Ebert-Uphoff, I., Ravela, S., Babaie, H. A., and Kumar, V. (2019). "Machine learning for the geosciences: challenges and opportunities," *IEEE Transactions on Knowledge and Data Engineering* 31 (8), 1544–1554. doi:10.1109/TKDE.2018.2861006

Rouet-Leduc, B., Hulbert, C., Lubbers, N., Barros, K., Humphreys, C. J., and Johnson, P. A. (2017). Machine learning predicts laboratory earthquakes. *Nat. Rev. Earth Environ.* 44, 9276–9282. doi:10.1002/2017GL074677

frontiers | Frontiers in Earth Science

TYPE Original Research
PUBLISHED 20 July 2023
DOI 10.3389/feart.2023.1223686

OPEN ACCESS

EDITED BY
Georg Rümpker,
Goethe University Frankfurt, Germany

REVIEWED BY
Nishtha Srivastava,
Frankfurt Institute for Advanced Studies, Germany
Darren Tan,
University of Alaska Fairbanks, United States

*CORRESPONDENCE
Giovanni Messuti,
g.messuti@studenti.unisa.it
Ortensia Amoroso,
oamoroso@unisa.it

RECEIVED 16 May 2023
ACCEPTED 10 July 2023
PUBLISHED 20 July 2023

CITATION
Messuti G, Scarpetta S, Amoroso O, Napolitano F, Falanga M and Capuano P (2023), CFM: a convolutional neural network for first-motion polarity classification of seismic records in volcanic and tectonic areas.
Front. Earth Sci. 11:1223686.
doi: 10.3389/feart.2023.1223686

CFM: a convolutional neural network for first-motion polarity classification of seismic records in volcanic and tectonic areas

Giovanni Messuti[1,2]*, Silvia Scarpetta[1,2], Ortensia Amoroso[1]*, Ferdinando Napolitano[1], Mariarosaria Falanga[3] and Paolo Capuano[1]

[1]Department of Physics "E.R. Caianiello", University of Salerno, Fisciano, Italy, [2]Section of Naples, National Institute for Nuclear Physics (INFN), Naples, Italy, [3]Department of Information and Electrical Engineering and Applied Mathematics (DIEM), University of Salerno, Fisciano, Italy

First-motion polarity determination is essential for deriving volcanic and tectonic earthquakes' focal mechanisms, which provide crucial information about fault structures and stress fields. Manual procedures for polarity determination are time-consuming and prone to human error, leading to inaccurate results. Automated algorithms can overcome these limitations, but accurately identifying first-motion polarity is challenging. In this study, we present the Convolutional First Motion (CFM) neural network, a label-noise robust strategy based on a Convolutional Neural Network, to automatically identify first-motion polarities of seismic records. CFM is trained on a large dataset of more than 140,000 waveforms and achieves a high accuracy of 97.4% and 96.3% on two independent test sets. We also demonstrate CFM's ability to correct mislabeled waveforms in 92% of cases, even when they belong to the training set. Our findings highlight the effectiveness of deep learning approaches for first-motion polarity determination and suggest the potential for combining CFM with other deep learning techniques in volcano seismology.

KEYWORDS

deep convolutional neural networks, automatic classification, machine learning, self-organizing maps, volcanic and tectonic earthquakes, first-motion polarity

1 Introduction

In the field of Earth sciences, the study of seismic waves generated by earthquakes occupies an important role since it allows us to retrieve the main features of both the propagation medium and the seismic source. As for the seismic source, the attention is mainly devoted to estimating the geometric and kinematic parameters, including the location, magnitude, fault dimension and focal mechanisms. Focal mechanisms are crucial to characterize the seismogenic fault structures and the stress field of a region, from local to nationwide scale, in tectonic (Vavryčuk, 2014; Napolitano et al., 2021a; Uchide et al., 2022), and volcanic areas (Roman et al., 2006; Judson et al., 2018; La Rocca and Galluzzo, 2019; Aoki, 2022; Zhan et al., 2022).

The focal mechanisms can be computed using P-wave first-motion polarity (e.g., FPFIT; Reasenberg, 1985; Snoke et al., 2003; Hardebeck and Shearer, 2002), the waveform information (e.g., Zhao and Helmberger, 1994) or both (Weber, 2018). P-wave polarity

is also used as an additional constraint in the moment-tensor inversion (e.g., in volcanic settings, Dahm and Brandsdottir, 1997; Miller et al., 1998; Pesicek et al., 2012; Alvizuri and Tape, 2016) and full waveform inversion (e.g., for explosion Chiang et al., 2014; Ford et al., 2009). Determining first-motion polarities by manual procedures, mostly done for larger events, is time-consuming, susceptible to human error and can result in different outcomes depending on the expert analyst. In addition, a proper identification of the first-motion polarity can be difficult when dealing with small magnitude earthquakes. This may be due to the unfavorable signal-to-noise ratio. An enhanced method of identifying first-motion polarities will allow us to resolve the focal mechanism of smaller magnitude events, thereby improving our ability to characterize and interpret seismogenic areas. Automated procedures (e.g., Chen and Holland, 2016; Pugh et al., 2016) can avoid drawbacks, such as time consumption and ensure reproducibility. Despite this, identifying first-motion polarity is not a straightforward classification task that can be easily expressed using mathematical procedures. Consequently, the effectiveness of the automated algorithms (not based on machine learning) relies on a limited number of parameters, which require intensive human involvement to fine-tune, and may result in worse performance compared to human analysis (Ross et al., 2018).

Deep learning offers a notable advantage in that prior knowledge of the observed phenomena is not a prerequisite for model development. This is attributed to the capability of Deep Neural Networks (DNNs) to autonomously extract significant features from raw data, eliminating the need for a mathematical representation of the problem. Moreover, when confronted with extensive datasets, deep learning has proved to be a suitable and highly effective methodology to be employed. Hence, the vast amount of seismological data represents an excellent opportunity for the application of DNNs, making deep learning an ideal choice for our purposes. Recent studies demonstrated the possibility of developing effective and competitive applications of DNNs in the study of seismic waves generated by earthquakes, volcanic eruptions, explosions, along with other sources (Mousavi and Beroza, 2022). DNNs have been used for events detection and location (Perol et al., 2018), arrival times picking (Ross et al., 2018; Zhu and Beroza, 2019), data denoising (Richardson and Feller, 2019), classification of volcano-seismic events (López-Pérez et al., 2020), construction of suitable ontologies (Falanga et al., 2022), discrimination of explosive and tectonic sources (Linville et al., 2019; Kong et al., 2022), waveform recognition both focusing on transients and continuous background acquisition (Rincon-Yanez et al., 2022) and for ground motion prediction equations (Prezioso et al., 2022).

Several studies have demonstrated the significant applicability of Convolutional Neural Networks (LeCun et al., 2015) in determining the first-motion polarity. CNNs use convolutional layers to extract spatial patterns from a multi-dimensional input array or matrix-like data. By applying multiple filters with adjustable weights through a process known as convolution, these filters extract relevant features through their scanning process. Stacking multiple convolutional layers allows the network to automatically learn and identify relevant abstract features useful for the task. The ability of CNNs to capture complex spatial relationships has made them particularly effective in a wide range of image and signal processing tasks, including the determination of first-motion polarity. One of the earliest studies in this field, conducted by Ross et al. (2018), involved training a simple CNN on 18.2 million seismograms from the Southern California Seismic Network (SCSN) catalog, achieving a precision in determining polarities of 95%. Hara et al. (2019) established a lower limit on the number of waveforms required for a satisfactory level of performance during training. The same authors explored the possibility of using a CNN to predict waveforms deriving from events located in regions different from those where data used for the training set have been collected. Uchide (2020) derived focal mechanisms and important information about the stress field in Japan exploiting the first-motion polarities determined by using a CNN-based technique. Li et al. (2023) utilized the CNN by Zhao et al. (2023) to develop an automatic workflow for focal mechanism inversion.

In this work, we present the Convolutional First Motion (CFM) neural network, a label-noise robust strategy based on a CNN to automatically identify first-motion polarities of seismic waves. We take advantage of the regularization effects of dropout layers and the implicit regularization properties of Stochastic Gradient Descent (SGD), when used in combination with early stopping, to handle a percentage of mislabelling (often known as noisy labels). CFM is trained on more than 140,000 waveforms derived from INSTANCE dataset (Michelini et al., 2021), and tested both on 8,983 waveforms belonging to different events of the same dataset and on 4,072 waveforms collected from Napolitano et al. (2021b). We found that when CFM is applied to mislabeled waveforms, which we identified through a data visualization procedure, it corrects them in 92% of the cases, even when they belong to the training set. CFM showed high accuracy levels (i.e., 97.4% and 96.3%) when tested on two independent test sets, high reliability and great generalization ability. The approach shown in our study reveals that an appropriate augmentation procedure can make the network able to deal with uncertainty in arrival times, which increases the potential for using CFM in combination with automatic deep learning techniques for phase picking. Such methodology is expected to have a strong impact on any problem related to the source modeling of tectonic and volcanic quakes, whose construction is founded on the best picking and phase recognition.

2 Data

We collected the seismic waveforms included in the INSTANCE dataset (Michelini et al., 2021) and used them to train the neural network and to evaluate its performance. The dataset, specifically compiled to apply machine learning techniques, comprises 1,159,249 waveforms originating from different sources (natural and anthropogenic earthquakes, volcanic eruptions, landslides along with other sources). The waveforms were registered by both velocimeters (HH, EH channels) and accelerometers (HN channel) seismometers belonging to 19 seismic networks operated and managed by several Italian institutions. The dataset includes 54,000 earthquakes that occurred between January 2005 and January 2020 in Italy and surrounding regions, with magnitude ranging from 0.0 to 6.5 (see Michelini et al., 2021 for further details). Each datum consists of a 120 s time window. Each waveform is associated with upward, downward, or undefined polarity. We excluded all those events with undefined polarity. In

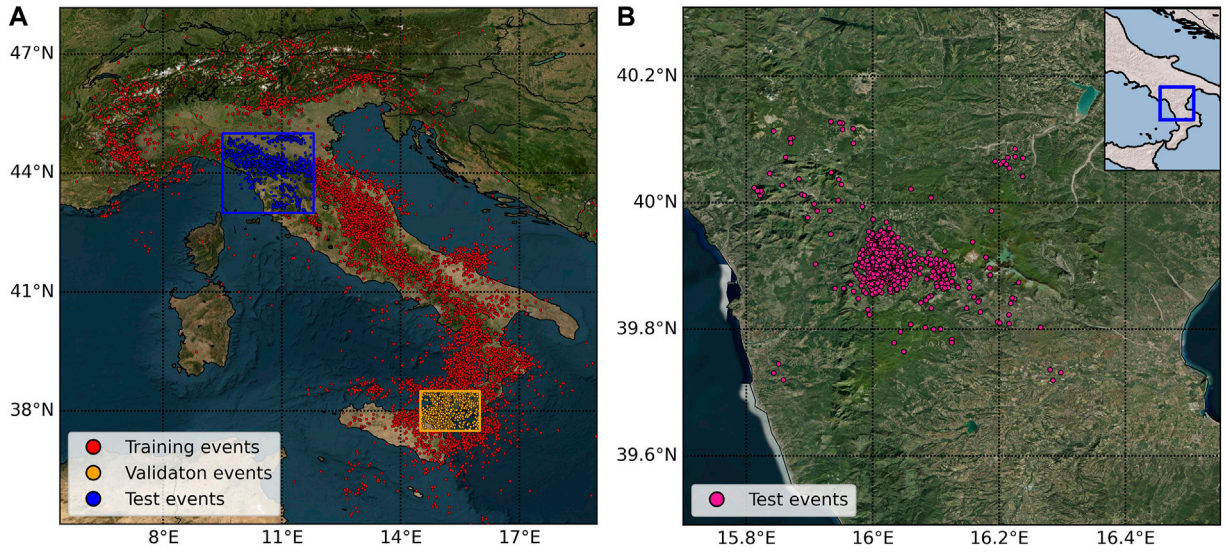


FIGURE 1
Localization of seismic events, shown with circles along the Italian peninsula. **(A)** The 28,551 events considered in *dataset A* (derived from the INSTANCE dataset). Waveforms belonging to events displayed by red circles are used as training data. The orange and blue boxes respectively contain events used to derive validation and test waveforms data. **(B)** The 842 events present in *dataset B (derived from* Napolitano et al., 2021b), located in Southern Italy, whose waveforms are used as a second test set.

addition, selecting only the vertical component of velocimeters data, we achieved 161,198 seismic traces of which 103,530 showed upward polarity and 57,668 downward polarity. We will refer to these waveforms as *dataset A* (Figure 1A). We split this dataset into three subsets, respectively used as.

- Training set: 141,972 waveforms (88.0% of the total data) corresponding to 23,878 events shown as red circles in Figure 1A;
- Validation set: 10,243 waveforms (6.4% of the total data) corresponding to 2,275 events shown as orange circles in Figure 1A;
- Testing set: 8,983 waveforms (5.6% of the total data) corresponding to 2,398 events shown as blue circles in Figure 1A.

The spatial selection was made to avoid correlations between waveforms in the different sets, following the approach proposed by Uchide (2020). It is noteworthy that the validation set comprises earthquakes from the Etna volcano region (orange box in Figure 1A).

Then, we collected the 870 earthquakes (M_L 1.8–5.0), recorded during the 2010–2014 Pollino (Southern Italy) seismic sequence (Figure 1B) by three seismic networks (*Istituto Nazionale di Geofisica e Vulcanologia* (INGV), *Università della Calabria* (UniCal) and *Deutsche GeoForschungsZentrum* (GFZ)) (Passarelli et al., 2012; Margheriti et al., 2013) and located in the new 3D velocity model by Napolitano et al. (2021b). From these events, we selected the vertical components of the waveforms sampled at 100 Hz, registered by velocimeters and with clear P-wave polarity. We refer to this dataset as *dataset B*. It comprises 4,072 manually picked waveforms derived from 824 out of the original 870 events collected. We used *dataset B* as a second test set to evaluate the performance of the neural network on data from a specific Italian tectonic setting. To avoid any possible overlapping between *dataset A* and *dataset B*, we removed the 821 common waveforms in the former dataset.

In addition, we used seismic traces from the Southern California Seismic Network (Ross et al., 2018) and western Japan region (Hara et al., 2019) to evaluate the network's generalization ability on waveforms from completely different regions. For this purpose, we selected the 863,151 waveforms belonging to the 273,882 earthquakes registered at 682 stations from the SCSN dataset. This constitutes the part of the test set with definite polarity used in Ross et al. (2018), whose magnitudes lie in the range [−1.0,7.2]. Similarly, we used 3,930 waveforms (M_L -1.3–6.2) constituting a part of the test set sampled at 100 Hz provided to us by Hara et al. (2019). The waveforms from the western Japan region were registered by stations operated by the National Research Institute for Earth Science and Disaster Prevention (NIED), the National Institute of Advanced Industrial Science and Technology (AIST), the Japan Meteorological Agency (JMA), and Kyoto University (Hara et al., 2019).

3 Methods

3.1 Data visualization with SOM and label noise

Before training the network on part of *dataset A*, the preliminary step of our analysis has been the implementation of a data visualization technique to investigate the waveforms. To this end, we applied the Self Organizing Maps (SOM, Kohonen, T., 2013). This unsupervised machine learning technique is

highly efficient in reducing the dimensionality of large datasets, by leveraging the similarities between the data, to cluster and visualize them in a low-dimensional grid, while preserving their topological structure. In order to focus the SOM on the features of our interest, the map was given a representation of the data in feature space. We normalized the traces to unit variance and we focused our attention on time windows of 0.26 s (26 samples), which include the 0.20 s preceding the P-arrivals and the 0.05 s after. We used 5 samples after the arrival, as they were enough to capture the entire first oscillation of the seismic wave in the case of higher frequency earthquakes, and enough to point out the trend of the oscillation in the case of lower frequency earthquakes. A lower value was not sufficient to capture the trend of oscillations in low-frequency events, whereas with higher values we observed that the analysis also focused on the second oscillations. We employed 20 samples before the arrival as they constituted the minimum number required to capture the essential characteristics of the noise trend in each scenario. Features provided to the SOM were extracted either by the Principal Component Analysis (Bishop and Nasrabadi, 2006), to which the normalized 0.26-seconds-long time windows were provided, and by evaluating averages of 0.16-seconds-long moving temporal windows. The first average was calculated over the time window starting from 0.19 s before the P-arrival, and the subsequent 9 averages were calculated on shifted windows, moving forward by 0.01 s each (1 sample), with the last time window covering the last 0.16 s (from 0.10 s before the arrival to 0.05 s after). In total, we gave the SOM 16 features, namely, the first 6 principal components and 10 moving averages. We chose to consider the principal components up to the sixth because it was a fair trade-off between the number of dimensions taken into account and the explained variance. By using six components, we were able to achieve a 95% explained variance.

In our analysis, the map nodes were organized in a two-dimensional hexagonal 8 × 8 grid (Supplementary Figure S1B gives a representation of the grid). After the SOM training stage, we displayed the waveforms' clusters on the map of nodes. Each single node represents a cluster that contains all those data whose distance in input space is smaller than the distance to all other nodes. Supplementary Figure S1A,S2A,S3A show the mean value of the waveforms contained in each node and one-fifth of the waveforms falling in each of them, respectively using the total, upward, and downward first-motion polarity. The number of waveforms in each cluster is represented by the size of the hexagons in Supplementary Figure S1B,S2B,S3B. We observe that the map places most of the waveform with downward polarity on the left side of the grid (Supplementary Figure S3B), especially in the upper part, while the waveforms with upward polarities are mostly placed on the right side of the grid (Supplementary Figure S2B), with the more populated nodes situated in the lower part. The net separation between the two parts provides a strong indication that, generally, the polarities are resolved in an unambiguous way. Nevertheless, a problem often encountered is that the polarities can be mistakenly labeled. To overcome such difficulty, we investigated the SOM results in more detail.

Figure 2 shows in each cell the weighted percentage of traces with upward polarities contained in it. Since the number of downward polarities is smaller than the upward one in *dataset A*, a weighted percentage is required for a robust analysis. Specifically, the value c_{ij}, showed in the cell relative to the node located in the i-th row and j-th column of the grid, is:

$$c_{ij} = \frac{U_{ij}}{U_{ij} + wD_{ij}}, \tag{1.1}$$

where U_{ij} and D_{ij} are respectively the number of upward and downward waveforms assigned to the node ij, and $w = \sum_{ij} U_{ij} / \sum_{ij} D_{ij}$. We notice the presence of some cells whose percentages of upward polarities are less than 1% or more than 99%. Considering the possibility of labeling errors in the dataset, we hypothesize that the high (low)-populated-upward cells represent nodes where all or most of the waveforms share the same polarity. Consequently, we suppose that the 458 outliers traces falling in those nodes (namely, the waveforms with an assigned polarity different from the majority) are likely to be mislabeled examples.

In fact, we manually checked that at least 100 of the 123 down-labeled traces, which fell in nodes with a weighted percentage of up-labeled data above 99%, had indeed an upward polarity. Analogously, at least 237 of the 336 up-labeled traces, located in nodes with more than 99% down-populated data, were clear waveforms with negative polarity. The remaining traces were mostly unclear waveforms, where extracting polarity information was a challenging task also for a human analyst. We do not exclude the presence of other mislabeled data (respect to the 337 found by the SOM visualization). A visual inspection of 1,000 randomly selected waveforms highlighted that approximately 8% of waveforms are affected by some problems, such as noisy arrival times or not reliable polarity information.

This level of noise is very common in real-world datasets, especially in the case of such large ones, where the ratio of

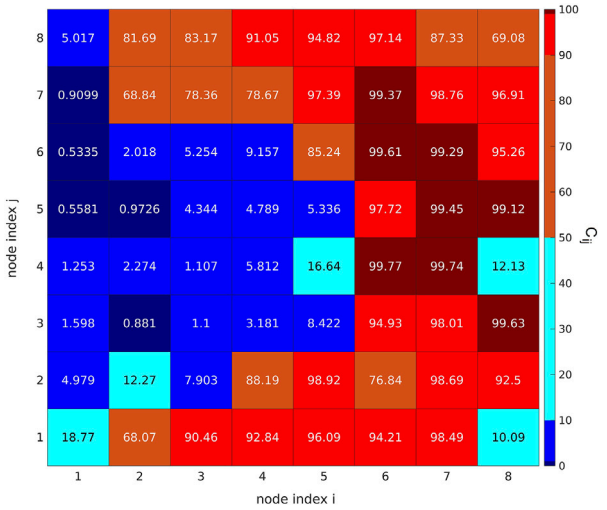


FIGURE 2
Heatmap relative to SOM nodes, showing the weighted percentages c_{ij} of upward waveforms laying in each node. We infer that the waveforms with assigned polarity different from the majority falling in dark blue and dark red cells (percentages less than 1% or more than 99%) are mislabeled data.

corrupted labels can cover, in some cases, up to 40% of the entire dataset (Song et al., 2022). Although it may appear to have drastic consequences to use problematic data to train a classifier, numerous studies have demonstrated that, with appropriate precautions and depending on the nature of the encountered noise, deep learning can exhibit remarkable robustness (Rolnick et al., 2017; Drory et al., 2018). Furthermore, other works highlight that noise can also be useful to better generalize (Damian et al., 2021).

Subsequent investigations revealed that attempting to clean our dataset yielded no significant benefits. Specifically, a second SOM visualization technique, similar to the one previously described, has been applied. This analysis aimed to analyze upward and downward polarity waveforms separately and enabled us to remove from dataset A approximately 10,000 waveforms. We excluded all the waveforms that fell within SOM nodes where we determined the majority of the data to be ambiguous or where extracting polarity information was very challenging. These waveforms comprised elements from the training, validation, and test sets. Supplementary Figure S4 shows some of the excluded nodes. In Supplementary Table S1, we compare the performance of the network trained on the original training set with the network trained on the cleaned training set, presenting the performance on both the cleaned test set and the original test set. Notably, we observed no significant differences in the performance of the two networks, when tested on the same test-set. Therefore, despite the presence of mislabelling in our dataset, we have chosen not to exclude any waveform, but rather, we aimed to design a network that can effectively handle and mitigate the effects of label noise, without the need for a preliminary selection of data points, which can result in information loss.

3.2 CFM architecture and preprocessing stage

The CFM network exclusively utilizes the vertical component of waveforms that have been sampled at a frequency of 100 Hz whose polarity information is available. To ensure consistency of the input data, all waveforms are subjected to a standardized preprocessing stage. Specifically, we subtracted to each waveform the mean value of the noise, from 200 samples (2.0 s) before the corresponding P arrival time to 5 samples before (in order to not include in the value of the mean some unbalanced oscillations due to the seismic phase). Subsequently, the initial wave portion is emphasized by setting a clipping threshold, in order not to neglect any of the smaller oscillations resulting from the signal (Uchide T., 2020). In this work, the threshold is different for each data point. To decide its value, the amplitude of the highest peak among those preceding the arrival time by at least 5 points was considered for each waveform. The threshold is equal to 20 times the value of this amplitude. Each seismogram is normalized to its respective threshold value. The portion of the signal exceeding this threshold is cut off. Previous studies did not highlight a specific filtering standard. Uchide (2020) used a high-pass filter at 1 Hz, while Ross et al. (2018) applied a filter between 1 and 20 Hz. On the other hand, Hara et al. (2019) and Chakraborty et al. (2022) avoided using any filter. CNN (and other deep networks) are known to work well on raw data (Goodfellow et al., 2016), since they learn features during training, in a hierarchical way, where initial layers acquire local features from data and the final layers extract global features representing high-level information. Considering these factors, we decided not to apply any frequency filters to our data.

We chose as our training set the part of dataset A outside the two boxes depicted in Figure 1A. Waveforms were presented in time windows of 160 samples (1.60s, 0.79 preceding the P-arrival and 0.80 after), with the 80th sample corresponding to the declared P-arrival times. During the training stage, we presented to the network both waveforms and their corresponding labels. Specifically, we assigned to a generic waveform x the label $y_x = 1$ if its label in the dataset was "upward" polarity; else $y_x = 0$. As previously stated, dataset A contains 103,530 upward and 57,668 downward polarity waveforms, resulting in an upward/downward ratio of 1.8. Similar level of unbalance is present in the selected data constituting our training set (on 141,972 total waveforms 91,563 showed upward polarity, while 50,409 showed downward polarity). A class imbalance may lead the network to prioritize the majority class, resulting in overlooking the characteristics of the minority class (Wang et al., 2016). For this reason, we balanced the training data applying a data augmentation technique (Uchide T., 2020; Chakraborty et al., 2022; Falanga et al., 2022) that allowed us to use a single data twice: the original trace and the corresponding flipped one, obtained by multiplying −1 and assigning it the opposite polarity. As a result, our augmented training set doubled in size, comprising 283,944 waveforms, with half exhibiting upward polarity and the remaining half exhibiting downward polarity. We did not augment test or validation data.

Figure 3 represents the Convolutional Neural Network architecture used in the present study. The network architecture is divided into two stages, the first of which is represented by the Convolutional layers. They provide a very efficient way to extract relevant features from grid-like data (Goodfellow et al., 2016), such as in the case of 1D time series (Kiranyaz et al., 2021) or 2D grids of pixel, i.e., images (Krizhevsky et al., 2017). The ReLU activation function is employed after each convolutional layer, owing to its well-known benefits in facilitating the training process (Krizhevsky et al., 2017). After three of the five Convolutional layers, a MaxPooling layer is added, which reduces the dimension of the input, preserving the most important features, and helps the network to gain translational invariance (Goodfellow et al., 2016). We also added Dropout layers, which are known to improve performance in case of training with noisy labels (Rusiecki, 2020), and prevent overfitting. In the second part of the network, the classification task is performed. The final layer's sigmoid, or logistic, activation function produces an output in the range [0, 1]. This choice allows the network output to be interpreted as the probability of an input vector to belong to one of the two investigated classes. We have used a threshold value of 0.5, above which we interpret data as having upward polarity and below which we interpret data as having downward polarity.

We set the binary cross-entropy as the loss function to be minimized. To train the network, we used the Stochastic Gradient Descent (Robbins and Monro, 1951). SGD is one of the most simple and effective optimization methods widely used, and it can lead to better generalization performance compared to other more sophisticated methods. SGD is considered to play a central role in the observed generalization abilities of deep learning, since its

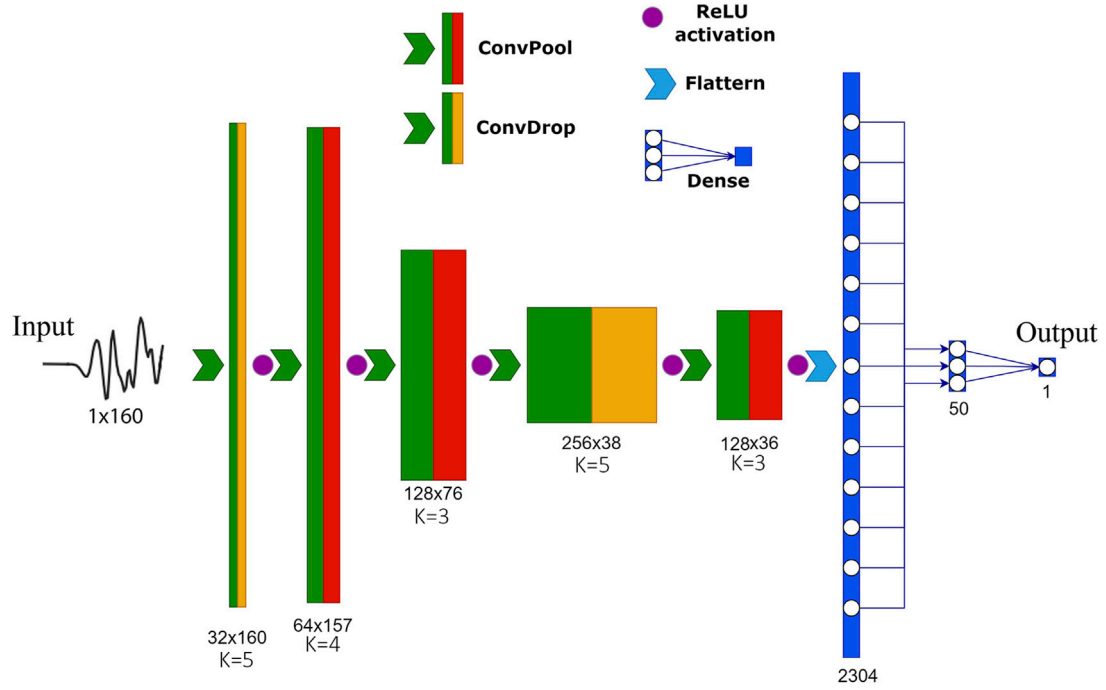


FIGURE 3
Architecture of the CFM, the deep Convolutional neural network for First Motion polarity classification used in this study. Numbers under each layer indicate its shape (i.e., number of channels x number of samples). ConvPool and ConvDrop indicate convolution with maxpooling and convolution with dropout, respectively. The values of K under each convolutional layer indicate the corresponding kernel size. The Flatten procedure (light blue arrow) only reshapes the previous layer in a one-dimensional array, without affecting any value.

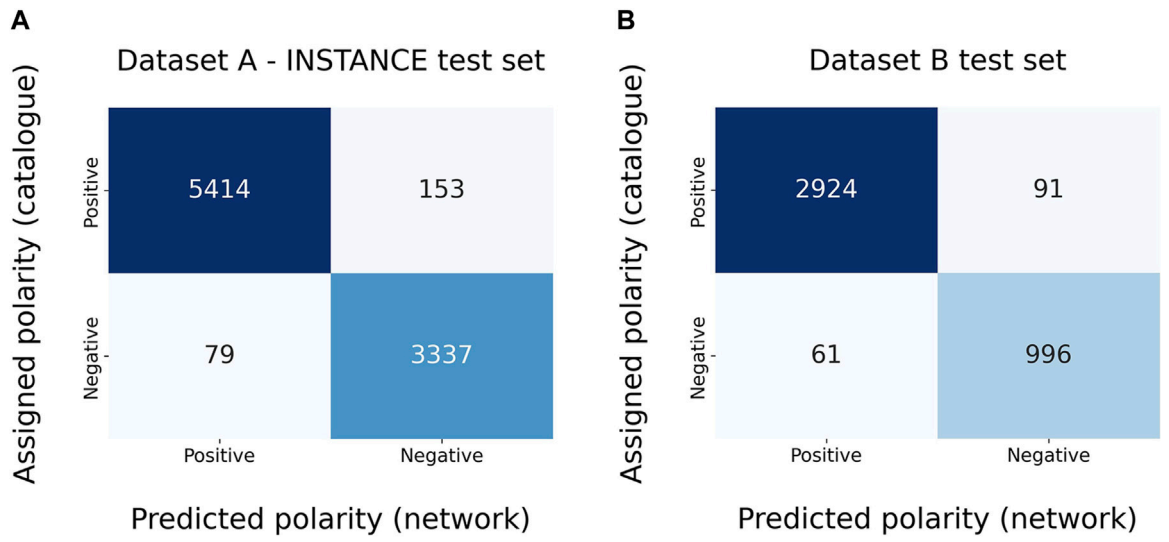


FIGURE 4
Confusion matrices for *dataset A* **(A)** and *dataset B* **(B)** test sets. The *x*-axis shows network prediction, while the *y*-axis reports the labels present in the dataset. The accuracies for *dataset A* and *dataset B* are approximately 97.4% and 96.3%, respectively.

stochasticity, resulting from the mini-batch sampling procedure, can provide a crucial implicit regularization effect (Ali et al., 2020). Moreover, the implicit regularization properties of SGD (Damian et al., 2021) are particularly useful when dealing with noisy data. We exploited the Stochastic Gradient Descend with the addition of Momentum. The default learning rate of 0.01 shows good

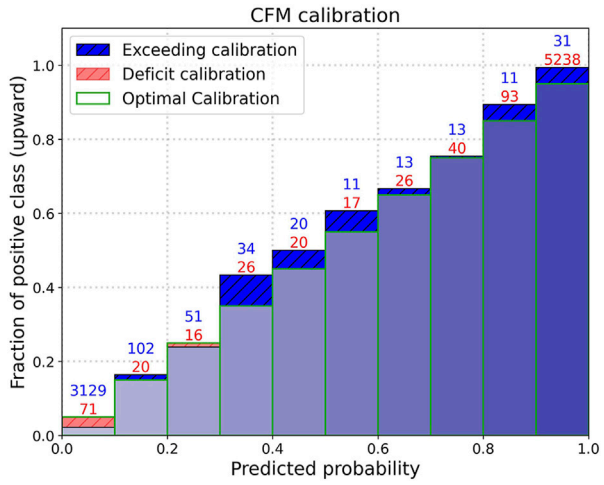


FIGURE 5
Reliability diagram of the network. Predictions made by the model are grouped into bins based on their predicted probabilities. The heights of the bars are the proportion of true positive cases within each bin. Green edges represent the average predicted probability of the bin, i.e., the optimal calibration. Numbers on each bar indicate the upward (red) and downward (blue) polarity waveforms laying in each bin.

performances (multiple training with learning rate in the range [0.007, 0.015] did not highlight substantial differences). We fixed the momentum parameter to be equal to 0.8 and the batch size value equal to 512.

We set the maximum number of epochs to 100 and, to prevent overfitting, we implemented an early stopping technique that interrupts the training if there is no improvement in the validation loss for 7 consecutive epochs. Early stopping is also an effective implicit regularization technique, which has been observed to be surprisingly effective in preventing overfitting to mislabeled data, especially when used in combination with first-order optimization algorithms, such as SGD (Li et al., 2020).

4 Results

CFM was trained on waveforms outside blue and orange boxes in Figure 1A. The early stopping technique stopped the training at epoch number 20. We then evaluated the performance on the test set derived from *dataset A* (Figure 1A, blue box) and on the *dataset B* (Figure 1B), expressing it through confusion matrices (Figures 4A,B, respectively), showing the number of samples labeled consistently with the dataset (top-left and bottom-right) or oppositely (top-right and bottom-left). From them, we computed the accuracies, defined as the number of correct predictions divided by the total ones. The network reached accuracies of 97.4% and 96.3%, respectively.

To provide a measure of the network's reliability, we evaluated its behavior as the output varies on *dataset A* test set. A classifier is said to be 'well-calibrated' when its output probability can be directly interpreted as a confidence level (Dawid, 1982). For instance, a well-calibrated classifier should classify the samples such that among the samples to which it gave a predicted probability close to 0.8, approximately 80% actually belong to the positive class, which in our case is represented by upward polarity. Figure 5 represents a reliability diagram of our network (Niculescu-Mizil and Caruana, 2005), which indicates how often data points assigned a certain forecast output probability interval actually exhibit upward polarity (assigned in the dataset). Mathematically, the value of the height of the rectangle belonging to the bin I_k corresponds to the empirical probability:

$$P\left(y_x = 1 \middle| \widehat{CFM}(x) \in I_k\right) = \frac{\left|\left\{x:\ y_x = 1, \widehat{CFM}(x) \in I_k\right\}\right|}{\left|\left\{x:\ \widehat{CFM}(x) \in I_k\right\}\right|}, \quad (1.2)$$

where x is a generic data point, y_x is its label, $\widehat{CFM}(x)$ the network out probability and $|\cdot|$ represents the cardinality of the ensemble. Although reliability diagrams can be helpful for visualizing calibration, having a scalar summary statistic of calibration is more practical. To this end, we calculated the Expected Calibration Error (Guo et al., 2017):

$$ECE = \sum_{m=1}^{M} \frac{n_m}{n} \left|acc(B_m) - conf(B_m)\right|, \quad (1.3)$$

where m is the number of predictions in bin m, n is the total number of data points, and acc (m) and conf (m) are the accuracy and confidence of bin m, respectively. The ECE values range in the interval [0, 1], and the lower they are, the better the calibration of a model. We obtained an ECE value of 3.7% for our network. In general, ECE values depend on the specific task and dataset involved. For a general comparison, refer to Guo et al., 2017.

4.1 CFM robustness to false annotations

We remember the SOM analysis of Section 3.1 revealed the presence of 337 waveforms with false labels (located within the nodes highlighted in Figure 2). Since the training set covers the majority of *dataset A*, the majority of these outlier waveforms (specifically 311) also belong to it. Despite the fact that the training algorithm forces the network output to match the assigned label, we found that 310 out of the 337 misclassified waveforms are assigned to the correct class by CFM. Figure 6 shows some examples of such waveforms we identified in Section 3.1 and for which the network predicts correct polarities. Given that the network successfully corrected 92% of the false labels, we consider this as evidence of its ability to be robust to overfitting erroneous labels.

4.2 Dealing with uncertain arrival times

In this section, to check the robustness of the network to uncertainty in arrival times, we evaluated the performance of the network including artificial time shifts in arrival times. To this end, we shifted each time-window of *dataset A* test set by a constant value of T samples, with values of T in the range [-20,20]. A value of T = 5, for example, indicates that the time window center is located 5 samples (0.05 s) past the declared

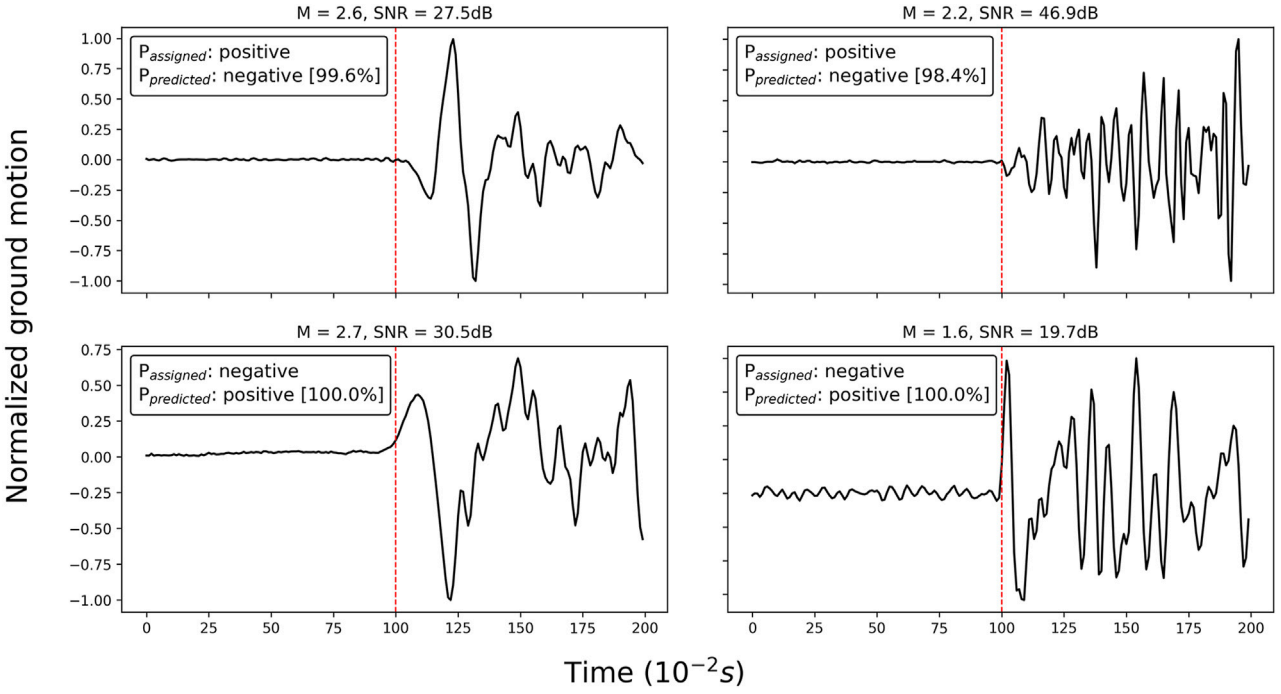


FIGURE 6
Some seismic traces erroneously labeled by the analyst that we identified with the SOM data visualization in Section 3.1. On the top of each subplot, we annotate the magnitude of the event (M) and the signal-to-noise ratio (SNR). $P_{assigned}$ and $P_{predicted}$ refer to the polarity assigned in the dataset and the prediction of the network (with the corresponding probability to belong to the predicted class in square brackets).

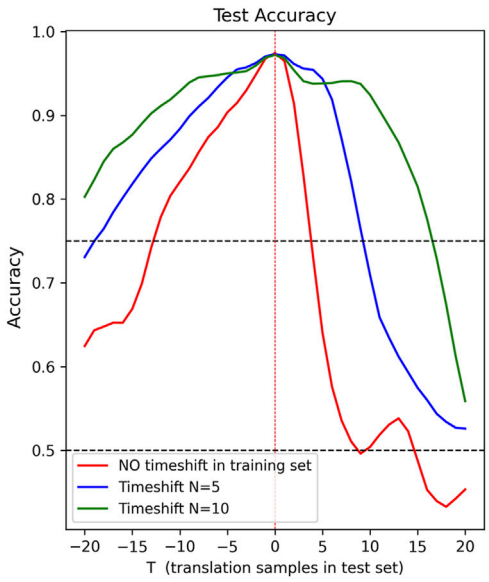


FIGURE 7
The performances of the CFM network on the test set after the two different training strategies. The blue and green lines refer to the trainings with a time-shift in the training set, with a maximum value N of 5 and 10 samples respectively. The red line shows the training without random time-shifts in the training set. Performance is shown as a function of the different shifts T in the test set. Dashed black lines refer to accuracy levels of 0.5 and 0.75.

P-arrival. Red line in Figure 7 shows the behavior of the network (trained on centered time-windows) varying T as test sample-shift. We notice that, as expected, accuracy is highest when there is no shift. Accuracy rapidly declines, dropping to 50% when there is a shift of +10 samples, indicating a significant degradation in performance.

Anyway, uncertainty in fixing the onset of P-wave is a trouble that often affects experimental data becoming much more difficult to manage for different reasons: poor signal-to-noise ratio, magnitude of the events decreases (small-energy/magnitude earthquake), recording stations installed in densely populated areas, complex medium properties, volcanic environment.

For this reason, we explored the possibility of giving the network the ability to deal with uncertainty in P-arrival times. Specifically, we developed an aimed augmentation strategy and performed a second training strategy, including a time-shift in the training set too. We used a time-shift augmentation procedure perturbing the centering of time windows contained in the training set, leaving the validation set unperturbed.

In particular, we selected 50% of the training waveforms and applied two independent uniform random time-shifts to each. The first time-shift was selected from the range [-N, −1], and the second from [1, N]. The original waveform and the two shifted versions were then included in the training set. We conducted two training sessions, on two augmented training sets, with N values of 5 and 10, respectively. Evaluating performance on unperturbed *dataset A* test set (T=0), we observe accuracy levels of 97.2% (in the case of N=10) and of 97.3% (in the case of N=5), which are slightly lower than the

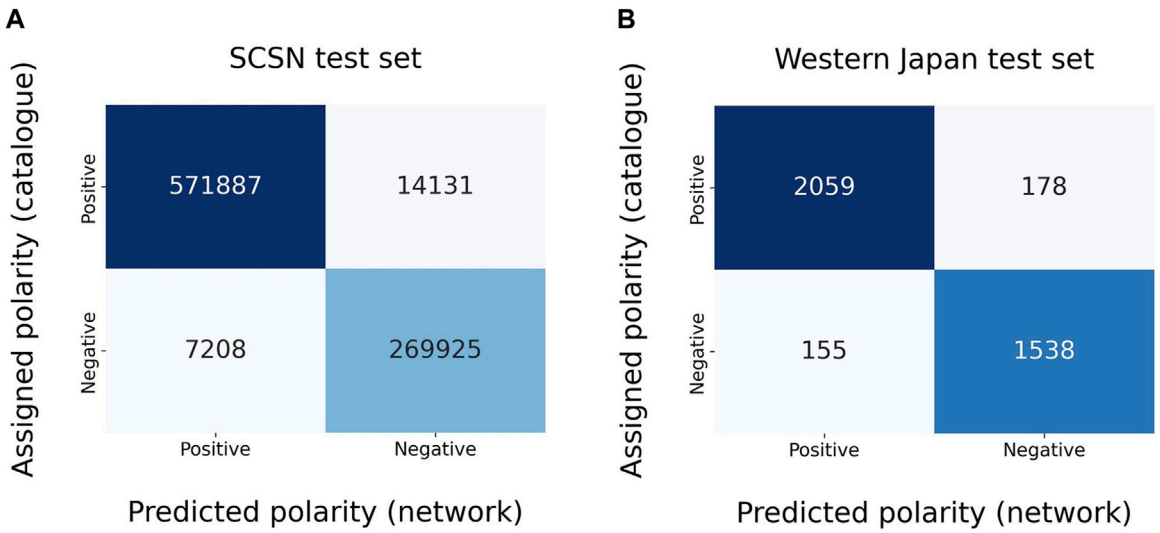


FIGURE 8
Confusion matrices for SCSN **(A)** and western Japan **(B)** test sets. The accuracies are approximately 97.5% and 91.5%, respectively. We recall that the performance on the western Japan test set refers to the different training using 150 as input waveforms.

correspondent obtained by the model trained on unperturbed waveforms. However, as shown by the blue and green lines in Figure 7, adding time-shifts to the training set can lead to an improvement in performance in the presence of uncertain arrival times. In particular, we observed a broader plateau where the accuracy remains above 92.4%, even when dealing with shifts of 10 samples, in the case of N=10 (green line), and it takes 17 translation test samples to reduce the accuracy below 75%.

4.3 Model generalization ability

To evaluate the generalization ability of the CFM network, we checked the capability to generate accurate predictions on new datasets coming from completely different geographic regions (Southern California and western Japan regions), using recordings obtained by different seismic networks, and far from the region (Italy) on which the net was trained on.

We first utilized the SCSN test dataset provided by Ross et al. (2018). We excluded waveforms without assigned polarity, resulting in 863,151 traces suitable for our purposes. The network achieves an accuracy of 98.4% for waveforms with SNR greater than or equal to 10, while the accuracy is 96.3% for waveforms with a SNR less than 10. The overall accuracy is 97.5%, comparable to the model trained by Ross et al. (2018) on the SCSN dataset (i.e. 95%). Figure 8A shows the confusion matrix related to the network prevision on the SCSN dataset.

We furthermore tested the performance on the test set provided by Hara et al. (2019), using only the 3,930 waveforms sampled at 100 Hz. We recall that CFM inputs are 160-sample waveforms, whereas the dataset we received contains 150-sample waveforms. Therefore, we have decided to conduct an additional training while keeping all the settings presented in the previous sections unchanged, except for the input shape, which we have adjusted to 150 samples to ensure compatibility. This additional training resulted in similar performances on both the *dataset A* and *dataset B* test sets when compared to the performance achieved with the 160-sample training. The predictions on the Hara et al. (2019) test set are presented in Figure 8B, from which one can compute an accuracy value of about 91.5%, slightly lower than the 95.4% obtained by the model of Hara et al. (2019).

5 Discussion

First-motion polarity determination can be a challenging task even for expert analysts, mainly when dealing with small-magnitude events, in both tectonic and volcanic environments. Deep learning neural networks have been widely applied in geophysics. Among many other applications, they have been used to detect first-motion polarities (Ross et al., 2018; Hara et al., 2019; Uchide, 2020; Chakraborty et al., 2022).

In this work, we developed the CFM network, a straightforward Convolutional Neural Network that can accurately identify the first-motion waveform polarity. Our results showed that CFM achieved a testing accuracy of 97.4% when applied to previously unseen traces. CFM also shows well generalization abilities, resulting in high accuracies on waveforms recorded from seismic networks located in completely different regions than those utilized to derive the training set (i.e., waveforms derived from the SCSN and western Japan test sets). For the SCSN test set, as noted in the previous works by Ross et al. (2018); Chakraborty et al. (2022), performance is better when dealing with waveforms that have a SNR greater than 10. Even if this is confirmed in our results, our network shows a gap in performance on different SNR of 2.1% when tested on SCSN test set, which is significantly lower than the 7.9% reported by Chakraborty et al. (2022). For the western Japan region, the accuracy achieved by CFM on the Hara et al. (2019) test set, at 91.5%, is slightly lower than

the accuracies obtained on the other test sets and the one reported by Hara et al. (2019) themselves. However, a manual analysis of all 333 misclassified waveforms revealed that the polarity assigned in the test set was correct only in 29 cases, while for 59 waveforms the polarity identified by the model was correct. Other waveforms either presented ambiguous or unextractable polarity (119 waveforms) or had a considerable error in arrival time, up to 35 samples (126 waveforms). Supplementary Figure S5 provides a representation of the various cases. These findings confirm that the instances where the network does not perform well are remarkably limited, and its inferior performance cannot be attributed to shortcomings.

We observed that the employed implicit regularization strategies prevented the network from overfitting mislabeled data, resulting in the network's ability to correct false labeling, even when the mislabeled waveforms are present in the training set. In line with previous studies (Uchide, 2020; Chakraborty et al., 2022), we demonstrated that implementing a time-shift augmentation procedure can lead to a decrease in performance when applied to unperturbed waveforms. However, unlike previous works, our additional training stages uncovered that an accurate augmentation procedure enables the handling of uncertainties in arrival times with only a minimal loss in performance on the unaltered data.

We also observe CFM exhibiting good calibration properties, which is critical for ensuring a high level of reliability in the model's outputs, although we did not carry out any explicit calibration processes (Guo et al., 2017). In addition, we observe (Figure 5) that when the network works on waveforms with defined polarity, as in our case, the vast majority of outputs lie in the ranges [0, 0.1] for downward polarity and [0.9, 1] for upward polarity, resulting in high reliability. Due to its well-calibration properties, CFM is able to produce accurate probability estimates, enabling us to make informed decisions based on the output probability values. For example, a threshold can be introduced to determine when to accept or reject a prediction.

In conclusion, our study introduces the robust and highly adaptable CFM network that holds significant potential for determining the P-wave polarities. The generalization ability of the algorithm in producing accurate prediction on waveforms registered in regions different from those used to derive training data and its ability to rectify previously misclassified polarities are noteworthy contributions of this research. CFM key selling point lies in its capability to efficiently revise or validate large volumes of analyst-derived first-motion polarities in historic catalogs using a consistent method. It is important to note that the algorithm relies on phase arrival times and therefore cannot handle catalogs without this information. Although the application was presented on manually obtained picks, our findings suggest that the CFM network can easily be adapted downstream of the application of an automatic P-phase detection and labeling network, which is currently being worked on as a future development. This integration would enhance its adaptability and streamline the resolution of poorly-determined focal mechanisms in catalogs by quickly and robustly rectifying mislabeled first-motion polarities in databases. Overall, our research lays the foundation for further advancements in accurately characterizing tectonic and volcanic seismic events and improving our understanding of focal mechanisms.

Data availability statement

The datasets used in this study are publicly available for download. The INSTANCE dataset can be accessed at the following link: https://data.ingv.it/en/dataset/471#additional-metadata. The SCSN dataset is accessible at the following link: https://scedc.caltech.edu/data/deeplearning.html. The CFM network and dataset B used in this research can be found in the GitHub repository: https://github.com/Nemenick/CFM.git.

Author contributions

OA, SS, and PC conceived the work. GM performed all the analysis. GM and SS developed the algorithm and implemented the code. FN and OA prepared the seismic catalog. GM, FN, SS, OA, and MF worked on draft manuscript preparation. All authors contributed to the article and approved the submitted version.

Funding

PRIN-2017 MATISSE project, No. 20177EPPN2, funded by the Italian Ministry of Education and Research.

Acknowledgments

We thank Yukitoshi Fukahata for providing us with the dataset of Hara et al. (2019).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/feart.2023.1223686/full#supplementary-material

References

Ali, A., Dobriban, E., and Tibshirani, R. (2020). "The implicit regularization of stochastic gradient flow for least squares," in *International conference on machine learning* (New York: PMLR), 233–244.

Alvizuri, C., and Tape, C. (2016). Full moment tensors for small events (M w< 3) at Uturuncu volcano, Bolivia. *Geophys. J. Int.* 206 (3), 1761–1783. doi:10.1093/gji/ggw247

Aoki, Y. (2022). Earthquake focal mechanisms as a stress meter of active volcanoes. *Geophys. Res. Lett.* 49 (19), e2022GL100482. doi:10.1029/2022GL100482

Bishop, C. M., and Nasrabadi, N. M. (2006). *Pattern recognition and machine learning*. New York: Springer.

Chakraborty, M., Cartaya, C. Q., Li, W., Faber, J., Rümpker, G., Stoecker, H., et al. (2022). PolarCAP–A deep learning approach for first motion polarity classification of earthquake waveforms. *Artif. Intell. Geosciences* 3, 46–52. doi:10.1016/j.aiig.2022.08.001

Chen, C., and Holland, A. A. (2016). PhasePApy: A robust pure Python package for automatic identification of seismic phases. *Seismol. Res. Lett.* 87 (6), 1384–1396. doi:10.1785/0220160019

Chiang, A., Dreger, D. S., Ford, S. R., and Walter, W. R. (2014). Source characterization of underground explosions from combined regional moment tensor and first-motion analysis. *Bull. Seismol. Soc. Am.* 104 (4), 1587–1600. doi:10.1785/0120130228

Dahm, T., and Brandsdóttir, B. (1997). Moment tensors of microearthquakes from the Eyjafjallajökull volcano in South Iceland. *Geophys. J. Int.* 130 (1), 183–192. doi:10.1111/j.1365-246X.1997.tb00997.x

Damian, A., Ma, T., and Lee, J. D. (2021). Label noise sgd provably prefers flat global minimizers. *Adv. Neural Inf. Process. Syst.* 34, 27449–27461. doi:10.48550/arXiv.2106.06530

Dawid, A. P. (1982). The well-calibrated Bayesian. *J. Am. Stat. Assoc.* 77 (379), 605–610. doi:10.1080/01621459.1982.10477856

Drory, A., Avidan, S., and Giryes, R. (2018). *On the resistance of neural nets to label noise*. arXiv preprint arXiv:1803.11410, 2.

Falanga, M., De Lauro, E., Petrosino, S., Rincon-Yanez, D., and Senatore, S. (2022). Semantically enhanced IoT-oriented seismic event detection: An application to Colima and Vesuvius volcanoes. *IEEE Internet Things J.* 9 (12), 9789–9803. doi:10.1109/JIOT.2022.3148786

Ford, S. R., Dreger, D. S., and Walter, W. R. (2009). Identifying isotropic events using a regional moment tensor inversion. *J. Geophys. Res. Solid Earth* 114 (B01306), 1–12. doi:10.1029/2008JB005743

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep learning*. Cambridge: MIT press.

Guo, C., Pleiss, G., Sun, Y., and Weinberger, K. Q. (2017). "July) on calibration of modern neural networks," in *International conference on machine learning* (New York: PMLR), 1321–1330.

Hara, S., Fukahata, Y., and Iio, Y. (2019). P-wave first-motion polarity determination of waveform data in Western Japan using deep learning. *Earth Planets Space* 71 (1), 127. doi:10.1186/s40623-019-1111-x

Hardebeck, J. L., and Shearer, P. M. (2002). A new method for determining first-motion focal mechanisms. *Bull. Seismol. Soc. Am.* 92 (6), 2264–2276. doi:10.1785/0120010200

Judson, J., Thelen, W. A., Greenfield, T., and White, R. S. (2018). Focused seismicity triggered by flank instability on Kīlauea's Southwest Rift Zone. *J. Volcanol. Geotherm. Res.* 353, 95–101. doi:10.1016/j.jvolgeores.2018.01.016

Kiranyaz, S., Avci, O., Abdeljaber, O., Ince, T., Gabbouj, M., and Inman, D. J. (2021). 1D convolutional neural networks and applications: A survey. *Mech. Syst. Signal Process.* 151, 107398. doi:10.1016/j.ymssp.2020.107398

Kohonen, T. (2013). Essentials of the self-organizing map. *Neural Netw.* 37, 52–65. doi:10.1016/j.neunet.2012.09.018

Kong, Q., Wang, R., Walter, W. R., Pyle, M., Koper, K., and Schmandt, B. (2022). Combining deep learning with physics based features in explosion-earthquake discrimination. *Geophys. Res. Lett.* 49 (13), e2022GL098645. doi:10.1029/2022GL098645

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. *Commun. ACM* 60 (6), 84–90. doi:10.1145/3065386

La Rocca, M., and Galluzzo, D. (2019). Focal mechanisms of recent seismicity at Campi Flegrei, Italy. *J. Volcanol. Geotherm. Res.* 388, 106687. doi:10.1016/j.jvolgeores.2019.106687

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521 (7553), 436–444. doi:10.1038/nature14539

Li, M., Soltanolkotabi, M., and Oymak, S. (2020). "June). Gradient descent with early stopping is provably robust to label noise for overparameterized neural networks," in *International conference on artificial intelligence and statistics* (New York: PMLR), 4313–4324.

Li, S., Fang, L., Xiao, Z., Zhou, Y., Liao, S., and Fan, L. (2023). FocMech-flow: Automatic determination of P-wave first-motion polarity and focal mechanism inversion and application to the 2021 yangbi earthquake sequence. *Appl. Sci.* 13 (4), 2233. doi:10.3390/app13042233

Linville, L., Pankow, K., and Draelos, T. (2019). Deep learning models augment analyst decisions for event discrimination. *Geophys. Res. Lett.* 46 (7), 3643–3651. doi:10.1029/2018GL081119

López-Pérez, M., García, L., Benítez, C., and Molina, R. (2020). A contribution to deep learning approaches for automatic classification of volcanovolcano-seismic events: Deep Gaussian processes. *IEEE Trans. Geosci. Remote Sens.* 59 (5), 3875–3890. doi:10.1109/TGRS.2020.3022995

Margheriti, L., Amato, A., Braun, T., Cecere, G., D'Ambrosio, C., De Gori, P., et al. (2013). Emergenza nell'area del Pollino: Le attività della Rete sismica mobile. *Rapp. Tec. INGV* 252, 1–40.

Michelini, A., Cianetti, S., Gaviano, S., Giunchi, C., Jozinović, D., and Lauciani, V. (2021). INSTANCE–the Italian seismic dataset for machine learning. *Earth Syst. Sci. Data* 13 (12), 5509–5544. doi:10.5194/essd-13-5509-2021

Miller, A. D., Julian, B. R., and Foulger, G. R. (1998). Three-dimensional seismic structure and moment tensors of non-double-couple earthquakes at the Hengill–Grensdalur volcanic complex, Iceland. *Geophys. J. Int.* 133 (2), 309–325. doi:10.1046/j.1365-246X.1998.00492.x

Mousavi, S. M., and Beroza, G. C. (2022). Deep-learning seismology. *Science* 377 (6607), eabm4470. doi:10.1126/science.abm4470

Napolitano, F., Amoroso, O., La Rocca, M., Gervasi, A., Gabrielli, S., and Capuano, P. (2021b). Crustal structure of the seismogenic volume of the 2010–2014 Pollino (Italy) seismic sequence from 3D P-and S-wave tomographic images. *Front. Earth Sci.* 9, 735340. doi:10.3389/feart.2021.735340

Napolitano, F., Galluzzo, D., Gervasi, A., Scarpa, R., and La Rocca, M. (2021a). Fault imaging at Mt Pollino (Italy) from relative location of microearthquakes. *Geophys. J. Int.* 224 (1), 637–648. doi:10.1093/gji/ggaa407

Niculescu-Mizil, A., and Caruana, R. (2005). "Predicting good probabilities with supervised learning," in Proceedings of the 22nd international conference on Machine learning, New York, August 2005, 625–632. doi:10.1145/1102351.1102430

Passarelli, L., Roessler, D., Aladino, G., Maccaferri, F., Moretti, M., Lucente, F., et al. (2012). *Pollino seismic experiment (2012-2014)*.

Perol, T., Gharbi, M., and Denolle, M. (2018). Convolutional neural network for earthquake detection and location. *Sci. Adv.* 4 (2), e1700578. doi:10.1126/sciadv.1700578

Pesicek, J. D., Sileny, J., Prejean, S. G., and Thurber, C. H. (2012). Determination and uncertainty of moment tensors for microearthquakes at Okmok Volcano, Alaska. *Geophys. J. Int.* 190 (3), 1689–1709. doi:10.1111/j.1365-246X.2012.05574.x

Prezioso, E., Sharma, N., Piccialli, F., and Convertito, V. (2022). A data-driven artificial neural network model for the prediction of ground motion from induced seismicity: The case of the Geysers geothermal field. *Front. Earth Sci.* 10, 2193. doi:10.3389/feart.2022.917608

Pugh, D. J., White, R. S., and Christie, P. A. F. (2016). Automatic Bayesian polarity determination. *Geophys. J. Int.* 206 (1), 275–291. doi:10.1093/gji/ggw146

Reasenberg, P. A. (1985). FPFIT, FPPLOT, and FPPAGE: Fortran computer programs for calculating and displaying earthquake fault-plane solutions. *U. S. Geol. Surv. Open-File Rep.*, 85–739.

Richardson, A., and Feller, C. (2019). *Seismic data denoising and deblending using deep learning*. arXiv preprint arXiv:1907.01497. doi:10.48550/arXiv.1907.01497

Rincon-Yanez, D., De Lauro, E., Petrosino, S., Senatore, S., and Falanga, M. (2022). Identifying the fingerprint of a volcano in the background seismic noise from machine learning-based approach. *Appl. Sci.* 12 (14), 6835. doi:10.3390/app12146835

Robbins, H., and Monro, S. (1951). A stochastic approximation method. *Ann. Math. Stat.* 22, 400–407. doi:10.1214/aoms/1177729586

Rolnick, D., Veit, A., Belongie, S., and Shavit, N. (2017). *Deep learning is robust to massive label noise*. arXiv preprint arXiv:1705.10694. doi:10.48550/arXiv.1705.10694

Roman, D. C., Neuberg, J., and Luckett, R. R. (2006). Assessing the likelihood of volcanic eruption through analysis of volcanotectonic earthquake fault–plane solutions. *Earth Planet. Sci. Lett.* 248 (1-2), 244–252. doi:10.1016/j.epsl.2006.05.029

Ross, Z. E., Meier, M. A., and Hauksson, E. (2018). P wave arrival picking and first-motion polarity determination with deep learning. J. Geophys. Res. *Solid Earth* 123 (6), 5120–5129. doi:10.1029/2017JB015251

Rusiecki, A. (2020). "Standard dropout as remedy for training deep neural networks with label noise," in *Theory and applications of dependable computer systems: Proceedings of the fifteenth international conference on dependability of computer systems DepCoS-RELCOMEX, june 29–july 3, 2020, brunów, Poland* (Germany: Springer International Publishing), 534–542. doi:10.1007/978-3-030-48256-5_52

Snoke, J. A., Lee, W. H. K., Kanamori, H., Jennings, P. C., and Kisslinger, C. (2003). Focmec: Focal mechanism determinations. *Int. Handb. Earthq. Eng. Seismol.* 85, 1629–1630. doi:10.1016/S0074-6142(03)80291-7

Song, H., Kim, M., Park, D., Shin, Y., and Lee, J. G. (2022). Learning from noisy labels with deep neural networks: A survey. *IEEE Trans. Neural Netw. Learn. Syst.*, 1–19. doi:10.1109/TNNLS.2022.3152527

Uchide, T. (2020). Focal mechanisms of small earthquakes beneath the Japanese islands based on first-motion polarities picked using deep learning. *Geophys. J. Int.* 223, 1658–1671. doi:10.1093/gji/ggaa401

Uchide, T., Shiina, T., and Imanishi, K. (2022). Stress map of Japan: Detailed nationwide crustal stress field inferred from focal mechanism solutions of numerous microearthquakes. *J. Geophys. Res.* 127 (6), e2022JB024036. doi:10.1029/2022JB024036

Vavryčuk, V. (2014). Iterative joint inversion for stress and fault orientations from focal mechanisms. *Geophys. J. Int.* 199 (1), 69–77. doi:10.1093/gji/ggu224

Wang, S., Liu, W., Wu, J., Cao, L., Meng, Q., and Kennedy, P. J. (2016). "Training deep neural networks on imbalanced data sets," in 2016 international joint conference on neural networks (IJCNN), Vancouver, 24-29 July 2016 (IEEE), 4368–4374. doi:10.1109/IJCNN.2016.7727770

Wéber, Z. (2018). Probabilistic joint inversion of waveforms and polarity data for double-couple focal mechanisms of local earthquakes. *Geophys. J. Int.* 213 (3), 1586–1598. doi:10.1093/gji/ggy096

Zhan, Y., Roman, D. C., Le Mével, H., and Power, J. A. (2022). Earthquakes indicated stress field change during the 2006 unrest of Augustine Volcano, Alaska. *Geophys. Res. Lett.* 49, e2022GL097958. doi:10.1029/2022gl097958

Zhao, L. S., and Helmberger, D. V. (1994). Source estimation from broadband regional seismograms. *Bull. Seismol. Soc. Am.* 84 (1), 91–104.

Zhao, M., Xiao, Z., Zhang, M., Yang, Y., Tang, L., and Chen, S. (2023). DiTingMotion: A deep-learning first-motion-polarity classifier and its application to focal mechanism inversion. *Front. Earth Sci.* 11, 335. doi:10.3389/feart.2023.1103914

Zhu, W., and Beroza, G. C. (2019). PhaseNet: A deep-neural-network-based seismic arrival-time picking method. *Geophys. J. Int.* 216 (1), 261–273. doi:10.1093/gji/ggy423

frontiers | Frontiers in Earth Science

TYPE Original Research
PUBLISHED 03 August 2023
DOI 10.3389/feart.2023.1204832

OPEN ACCESS

EDITED BY
Georg Rümpker,
Goethe University Frankfurt, Germany

REVIEWED BY
Matthew Haney,
Alaska Volcano Observatory (AVO), United States
Mario La Rocca,
University of Calabria, Italy

*CORRESPONDENCE
Manuel Titos,
mmtitos@ugr.es

RECEIVED 12 April 2023
ACCEPTED 20 July 2023
PUBLISHED 03 August 2023

CITATION
Titos M, Gutiérrez L, Benítez C, Rey Devesa P, Koulakov I and Ibáñez JM (2023), Multi-station volcano tectonic earthquake monitoring based on transfer learning.
Front. Earth Sci. 11:1204832.
doi: 10.3389/feart.2023.1204832

Multi-station volcano tectonic earthquake monitoring based on transfer learning

Manuel Titos[1]*, Ligdamis Gutiérrez[2], Carmen Benítez[1], Pablo Rey Devesa[2], Ivan Koulakov[3] and Jesús M. Ibáñez[2]

[1]Information Technology and Telecommunications Research Center, Department of Signal Processing, Telematic and Communications, University of Granada, Granada, Spain, [2]Faculty of Sciences, Department of Theoretical Physics and the Cosmos, University of Granada, Granada, Spain, [3]Laboratory for Seismic Forward and Inverse Problems, Institute of Petroleum Geology and Geophysics, Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia

Introduction: Developing reliable seismic catalogs for volcanoes is essential for investigating underlying volcanic structures. However, owing to the complexity and heterogeneity of volcanic environments, seismic signals are strongly affected by seismic attenuation, which modifies the seismic waveforms and their spectral content observed at different seismic stations. As a consequence, the ability to properly discriminate incoming information is compromised. To address this issue, multi-station operational frameworks that allow unequivocal real-time management of large volumes of volcano seismic data are needed.

Methods: In this study, we developed a multi-station volcano tectonic earthquake monitoring approach based on transfer learning techniques. We applied two machine learning systems—a recurrent neural network based on long short-term memory cells (RNN–LSTM) and a temporal convolutional network (TCN)—both trained with a master dataset and catalogue belonging to Deception Island volcano (Antarctica), as blind-recognizers to a new volcanic environment (Mount Bezymianny, Kamchatka; 6 months of data collected from June to December 2017, including periods of quiescence and eruption).

Results and discussion: When the systems were re-trained under a multi correlation-based approach (i.e., only seismic traces detected at the same time at different seismic stations were selected), the performances of the systems improved substantially. We found that the RNN-based system offered the most reliable recognition by excluding low confidence detections for seismic traces (i.e., those that were only partially similar to those of the baseline). In contrast, the TCN-based network was capable of detecting a greater number of events; however, many of those events were only partially similar to the master events of the baseline. Together, these two approaches offer complementary tools for volcano monitoring. Moreover, we found that our approach had a number of advantages over the classical short time average over long time-average (STA/LTA) algorithm. In particular, the systems automatically detect VTs in a seismic trace without searching for optimal parameter settings, which makes it a portable, scalable, and economical tool with relatively low computational cost. Moreover, besides obtaining a preliminary seismic catalog, it offers information on the confidence of the detected events. Finally, our approach provides a useful tentative label for subsequent analysis carried out by a human operator. Ultimately, this study contributes a new framework for rapid and easy volcano monitoring based on temporal changes in monitored seismic signals.

KEYWORDS

automatic volcanic monitoring, real-time monitoring, artificial intelligence, transfer learning, recurrent neural networks, temporal convolutional networks

1 Introduction

Active volcanoes are often monitored by different ground and space-based instruments, which provide essential data for understanding the volcanic system, quantifying impacts, mitigating risk, and contributing to the preparedness of governments and society as a whole (Barsotti et al., 2020; Barsotti et al., 2023). However, identifying transitions in volcanic state is complex and involves the study of various physics processes. Given the large volumes of data now available from permanent monitoring seismic networks, volcanic seismology plays a critical role in volcano monitoring.

Volcanic dynamics generate an exchange of energy with the surrounding medium that propagates in the form of elastic or seismic waves. Owing to the complexity of volcanic processes, these seismic waves can have varying characteristics in both the time and frequency domains (Ibáñez et al., 2000). Identifying and characterizing these signals with the aim of associating them with internal dynamic processes is a key scientific challenge. Accurate recognition (identification and classification) is the basis for developing eruption forecasting based on precursors (Sparks et al., 2012; McNutt et al., 2015; Machacca et al., 2023), and is critical for improving knowledge of volcanic dynamics. Signals are generally classified based on the source model built to explain them. Low frequency signals (LF), such as so-called long period (LP) events and some types of volcanic tremor (TR), are associated with fluid dynamics. However, the most common type of seismic signal recorded in many volcanic environments is volcano tectonic (VT) earthquakes (Chouet, 2003). VT earthquakes are the consequence of stress-induced fluid dynamics inside the volcano (Roman and Cashman, 2006). In general, the source mechanisms of VT events can be described using classical approaches in seismology (Aki and Richards, 2002). However, as indicated by (Chouet and Matoza, 2013), owing to the involvement of fluids, this task is very complex in many volcanic environments. VTs are commonly considered to be potential precursors (McNutt and Roman, 2015), and so new methodologies and advances, including the use of artificial intelligence (AI), are increasingly being used to improve their recognition.

A key aspect of VT seismicity is that it contains much more information than that presented in each waveform. Recent studies have performed source modeling analysis (Sigmundsson et al., 2018; Sigmundsson et al., 2022; Cubuk-Sabuncu et al., 2021), focal mechanisms analysis, and 4D tomography showing the temporal evolution of volcanic structures in (Abacha et al., 2023). (Díaz-Moreno et al., 2015) used spatial and temporal analyses of VT foci evolution; for example, in their study, VTs generated during magma injection were assumed to reflect the effect of hydraulic fracturing, highlighting areas of the crust where stress was propagating as a consequence of magma migration. Seismic tomography allows us to reconstruct the internal structure of a volcano and infer the physical and dynamic characteristics of the volcanic system by studying the travel times of the first arrivals of VT waves (i.e., tomography of velocity (D'Auria et al., 2022), or by studying their loss of energy (i.e., attenuation tomography (Prudencio et al., 2013; Castro-Melgar et al., 2021) showed that volcanic structures are highly attenuating, which causes the waveform of the recorded signals to undergo strong changes, including loss of a large part of their spectral component, especially in the high frequency range. Similarly, (Titos et al., 2018), showed that VT earthquakes can be confused as LP-type events at a certain distance, which has consequences for the interpretation of internal dynamics of the volcanic system. However, these approaches all require data from large numbers of reliable earthquakes. Therefore, developing effective approaches that allow real-time management of large volumes of seismic data has become an important challenge.

Recent advances in machine learning (ML) have encouraged the development of advanced automatic data processing and analysis pipelines. Typically, new automatic approaches are built by learning from large seismic catalogues. These data-driven systems have proven to be very efficient tools in an ever-changing and streaming data environment; however, they have remarkably poor learning and adaptability outcomes owing to the incompleteness of many seismic catalogues. Nonetheless, building complete and reliable catalogues is technically challenging owing to the high cost of data-labelling. This issue has grown in importance in light of recent work, since catalogue-based learning can introduce bias when constructing predictive monitoring tools.

In this study, we developed a new automatic multi-station system for exclusively recognizing and labelling VT earthquakes. As discussed, owing to attenuation, many LP events annotated in seismic catalogues could actually be highly attenuated VT earthquakes. Therefore, we employed a multi-station process to improve the identification of VTs. To control for bias derived

TABLE 1 Classification accuracy (acc. %), number of parameters tuned, and training times for optimal configurations of the recurrent neural network based on long short-term memory cells (RNN–LSTM) and temporal convolutional network (TCN) architectures using the master catalogue (Deception Island volcano, Antarctica).

	RNN-LSTM	TCN
Test 1 acc. (%)	86.61	82.82
Test 2 acc. (%)	80.13	80.10
Test 3 acc. (%)	82.05	84.99
Test 4 acc. (%)	89.09	78.74
Avg. acc. (%)	**84.47**	81.66
No. of parameters	219,455	**63,105**
Training time (s)	16,834	2,073

Notes: A cross-validation procedure with four dataset partitions was used. Classification accuracy corresponds to frames or windows properly classified. Training times are reported in seconds and correspond with those of one partition. Bold values represents best architecture in terms of accuracy and parameters.

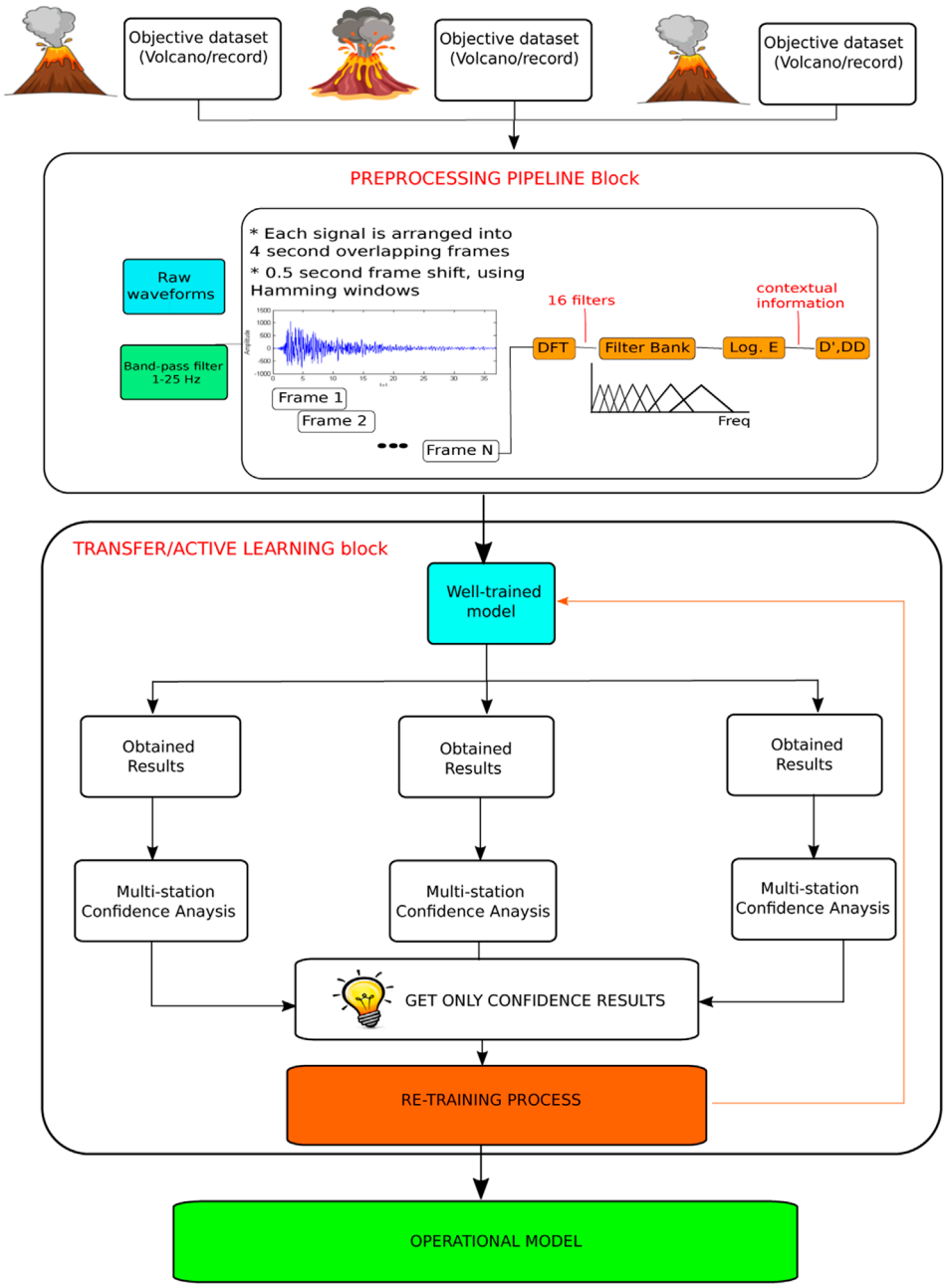


FIGURE 1
Transfer learning methodology used to develop the new volcano seismic monitoring system.

from seismic catalogue incompleteness, we employed transfer learning techniques (Weiss et al., 2016), which are helpful in domain-adaption problems, where the objective is to develop a monitoring system focused on available domain-specific data (Anantrasirichai et al., 2018; Titos et al., 2018; Bueno et al., 2019; Titos et al., 2019; Lapins et al., 2021; Jozinović et al., 2022). In contrast, our new monitoring system does not require prior domain-specific knowledge. Assuming a scenario in which there is no previous information related to the seismic dynamics of the volcano, instead of building a system from scratch (which would require an expensive data-labelling process), we used a recurrent neural network based on long short-term memory cells (RNN–LSTM) and a temporal convolutional network (TCN) (Titos et al., 2018; Titos et al., 2022) trained with a master catalogue belonging to Deception Island volcano (Antarctica) as a baseline. These models were then used as blind-recognizers for a different volcanic environment, that of Mount Bezymianny (Kamchatka). When these systems were re-trained under a multi correlation-based approach, where only reliable seismic traces identified at the same time at different seismic stations were selected and manually labeled, the performance of the systems improved substantially, resulting in a remarkable capability of confidently recognizing seismic traces. In

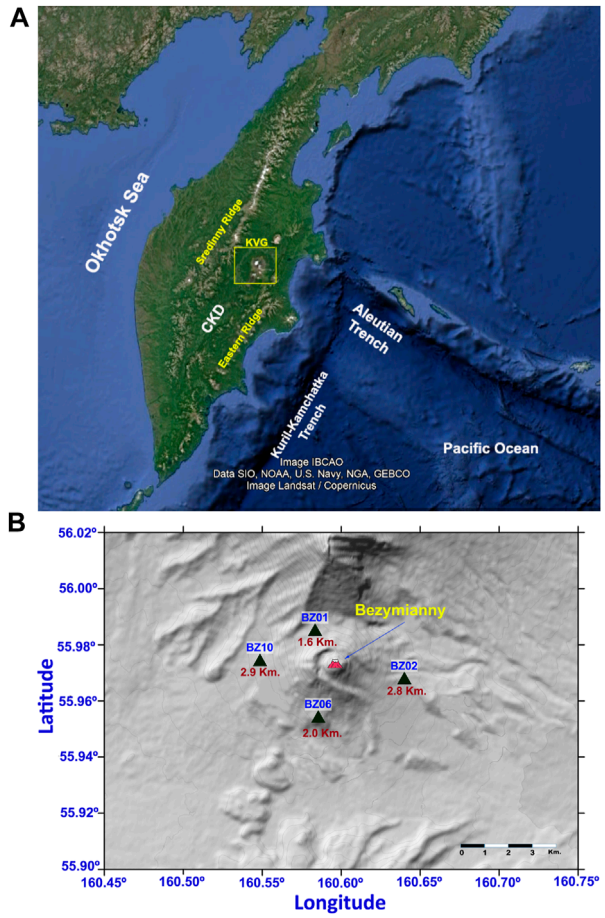


FIGURE 2
(A) Geological framework of Bezymianny volcano and **(B)** seismic station locations used in this study. Figure obtained and modifies from Google Earth resources.

summary, our approach provides a rapid and easy-to-use framework for real-time monitoring of temporal changes in seismic signals at any volcano.

2 Experimental framework and methodology

2.1 Methodology and experimental settings

In this study, we developed a new real-time multi-station seismic monitoring system for volcanoes without any prior knowledge within a transfer learning framework. Although some classical ML techniques such as Markov models have been used in sequence modeling tasks, neural networks (NN), including both RNN and TCN architectures (LeCun et al., 2015; Lea et al., 2016), have optimal temporal modeling capabilities. By generating a spatio-temporal sequence of hierarchical features, both architectures have been applied in complex and emerging geosciences research fields, including seismo-volcanic monitoring (Titos et al., 2018; Bueno et al., 2021), climate change (Yan et al., 2020), remote sensing (Račič et al., 2020), and human activities recognition (Nair et al., 2018). Accordingly, in this work, an RNN based on long short-term memory cells (RNN–LSTM) and a TCN (Titos et al., 2018; Titos et al., 2022) trained with a master catalogue belonging to Deception Island volcano (Antarctica) were proposed as a baseline. These were then used as a blind-recognizer for the data from a different volcanic environment, Mount Bezymianny (Kamchatka). The master database belonging to Deception Island volcano is unbalanced; however, it has been thoroughly reviewed by experts on the volcano. According to (Titos et al., 2018), the Deception Island dataset is composed of five seismic categories: background noise (BGN), tremor (TR), hybrid (HYB), VTs, and LPs; Table 1 summarizes the performances of the two approaches (RNN–LSTM and TCN) using the master catalog, based on the percentage of events correctly recognized.

Then, assuming a scenario in which the monitoring agency does not have any previous information related to the seismic dynamics of a volcano, a new monitoring tool was obtained as follow (see Figure 1):

1. Data parameterization: Raw streaming data belonging to each seismic station within the new volcanic environment were parameterized following the parameterization scheme of (Titos et al., 2018) to obtain the baseline systems.
2. Preliminary seismic catalog: By utilizing parameterized streaming traces as inputs, the pre-trained system generates a preliminary seismic catalog that consists of identified events along with their respective timing and probabilities assigned to each event class. It is important to note that when applying transfer learning without any domain-adaptation process, the seismic categories detected in a new volcanic environment will correspond to the seismic categories used in the master catalogue. Therefore, since the parameterization scheme adopted here was based on the spectral content of the seismic traces, events completely different from those described in the master catalog were categorized into these classes, based on their spectral similarity.
3. Probabilistic event detection: Using the preliminary seismic catalog, a probabilistic event selection process was used to obtain a new dataset from which to re-train or adapt the pre-trained system (RNN-LSTM or TCN) for the new volcanic environment. This process involved five steps:

    - The seismic station detecting the largest number of events was selected as the reference station (RS).
    - For each detected event at the RS, the confidence of the detection was analysed using a probabilistic event detection matrix with per-class probabilities output by the softmax layer (this layer is useful in multiclass classification problems as it converts the output values of the neural network into probabilities to each possible class). We assumed that low per-class probabilities reflect a change in the description of the analysed information. Therefore, only reliable events (those whose per-class probabilities were greater than a given threshold) were selected.
    - For each previously selected event, a multi correlation-based approach was applied to identify if they could be detected at the same time at different seismic stations. If the same event was reliably detected (per-class probabilities

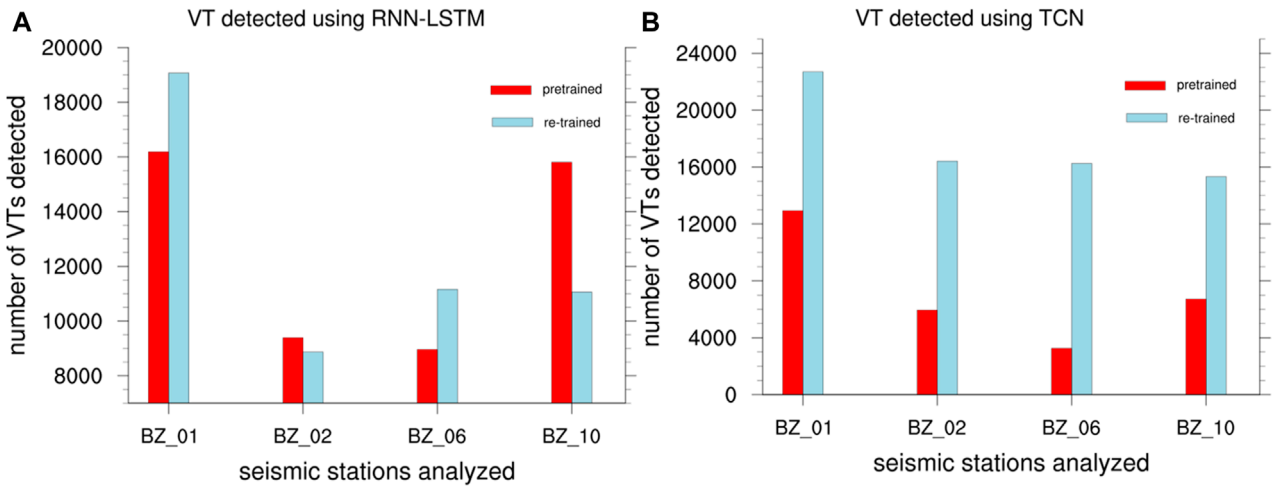


FIGURE 3
Total number of volcano tectonic (VT) earthquakes detected by **(A)** a recurrent neural network based on long short-term memory cells (RNN–LSTM) and **(B)** a temporal convolutional network (TCN) before and after the re-training process.

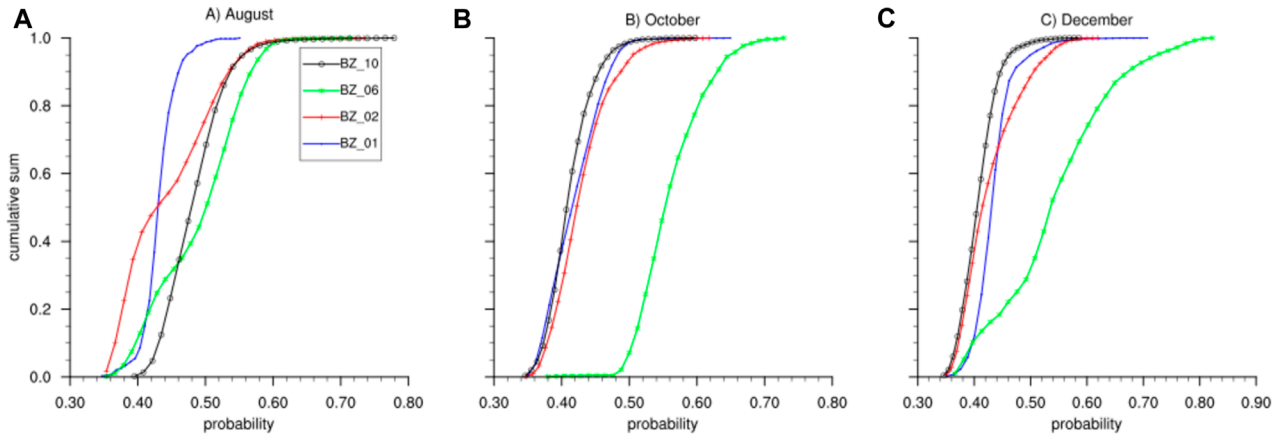


FIGURE 4
Monthly per-station cumulative distribution function (CDF) for the recurrent neural network based on long short-term memory cells (RNN–LSTM) before re-training. The x-axis represents the probabilities assigned by the models to those events detected as volcano tectonic (VT) earthquakes; the y-axis represents the normalized cumulative sum of events predicted within that class. **(A)** August 2017. **(B)** October 2017. **(C)** December 2017.

greater than a given threshold) at the same time at least two seismic stations, it was included as a training instance.
- Once the new training set was created, all instances were manually analyzed and newly labeled by experts in order to refine the bounding of the events.
- Finally, the pre-trained systems were re-trained using the new dataset and labels.

4. The final stage comprised further iterations of the probabilistic event detection (see point 3 above) in order to reach an optimal level of performance.

The pipeline used for this study is suitable for application to other baseline systems and parameterization schemes.

2.2 Geological framework: Bezymianny volcano

Bezymianny volcano (55.6°N, 160.3°E) is an explosive basaltic–andesitic stratovolcano belonging to the Klyuchevskaya (KVG) volcanic group on the Kamchatka Peninsula, Russia. It is located in the central depression of Kamchatka (CKD), which covers > 4,000 km^2 between the Sredinny and Eastern ridges. This region marks the northeastern corner of the Pacific subduction plate, which is formed by the Kuril–Kamchatka and Aleutian trenches (Figure 2A). According to its eruptive history, the volcano was considered inactive for more than 1,000 years (Braitseva and Kiryanov, 1982), until the lateral eruption in 1956. Bezymianny has experienced an active period since 2000, with more than 15 eruptive

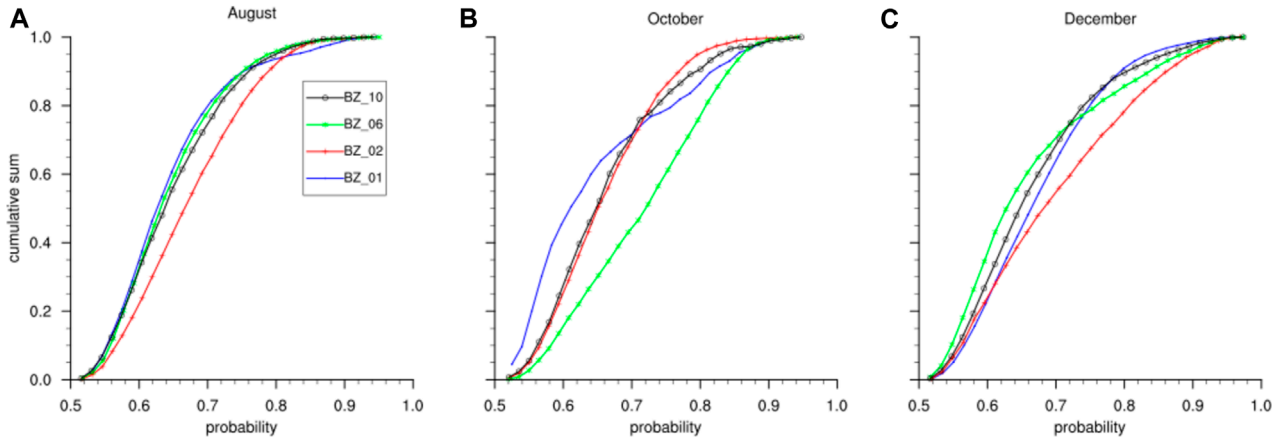


FIGURE 5
Monthly per-station cumulative distribution function (CDF) for the recurrent neural network based on long short-term memory cells (RNN–LSTM) after re-training. The x-axis represents the probabilities assigned by the models to those events detected as volcano tectonic (VT) earthquakes; the y-axis represents the normalized cumulative sum of events predicted within that class. **(A)** August 2017. **(B)** October 2017. **(C)** December 2017.

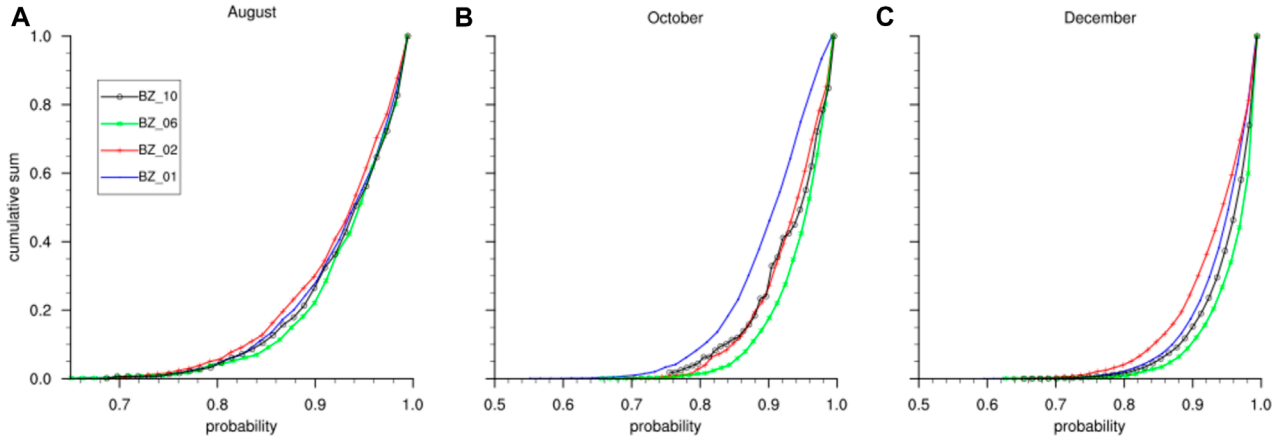


FIGURE 6
Monthly per-station cumulative distribution function (CDF) for the temporal convolutional network (TCN) before re-training. The x-axis represents the probabilities assigned by the models to those events detected as volcano tectonic (VT) earthquakes; the y-axis represents the normalized cumulative sum of events predicted within that class. **(A)** August 2017. **(B)** October 2017. **(C)** December 2017.

episodes (Van Manen et al., 2010). Among its recent eruptive episodes, that of 20 December 2017 (Girina et al., 2018) produced an eruption column that exceeded 15 km in height, representing a potential hazard to air traffic (Neal et al., 2009; McGimsey et al., 2014). The seismic database associated with this eruption is reliable and complete; therefore, it was selected for testing the approach developed in this study.

The seismic data used in this study were collected by a temporary network composed of 10 seismic stations, installed during the 2017–2018 period (Koulakov et al., 2021). However, only data corresponding to four stations (Figure 2B) were selected. Criteria for selecting the seismic stations were motivated by both the availability and quality (signal-to-noise ratio) of the data. In addition, to further determine the reliability of the monitoring system proposed, two additional eruptive phases (a pre-eruptive stage characterized by little activity and a syn-eruptive stage with tens of thousands of events) containing 6 months of seismic data from June to December 2017 were also selected.

3 Results

In this study, we analyzed results for four seismic stations over a 6 month period. However, to facilitate discussion of the results, here, we focus on VTs detected during 3 months of data, —August, October, and December 2017—which correspond to quiescent, pre-eruptive, and syn-eruptive phases, respectively.

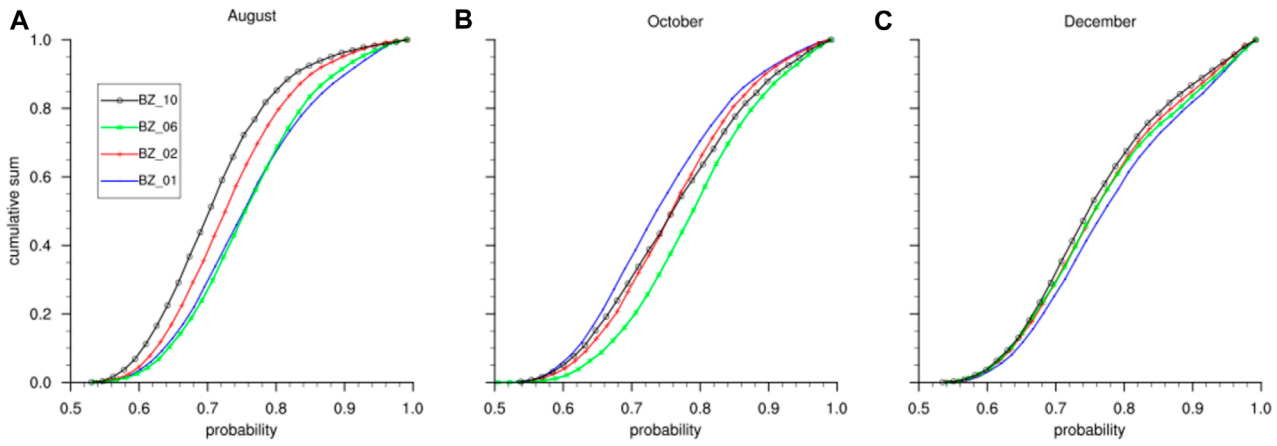


FIGURE 7
Monthly per-station cumulative distribution function (CDF) for the temporal convolutional network (TCN) after pre-training. The x-axis represents the probabilities assigned by the models to those events detected as volcano tectonic (VT) earthquakes; the y-axis represents the normalized cumulative sum of events predicted within that class. **(A)** August 2017. **(B)** October 2017. **(C)** December 2017.

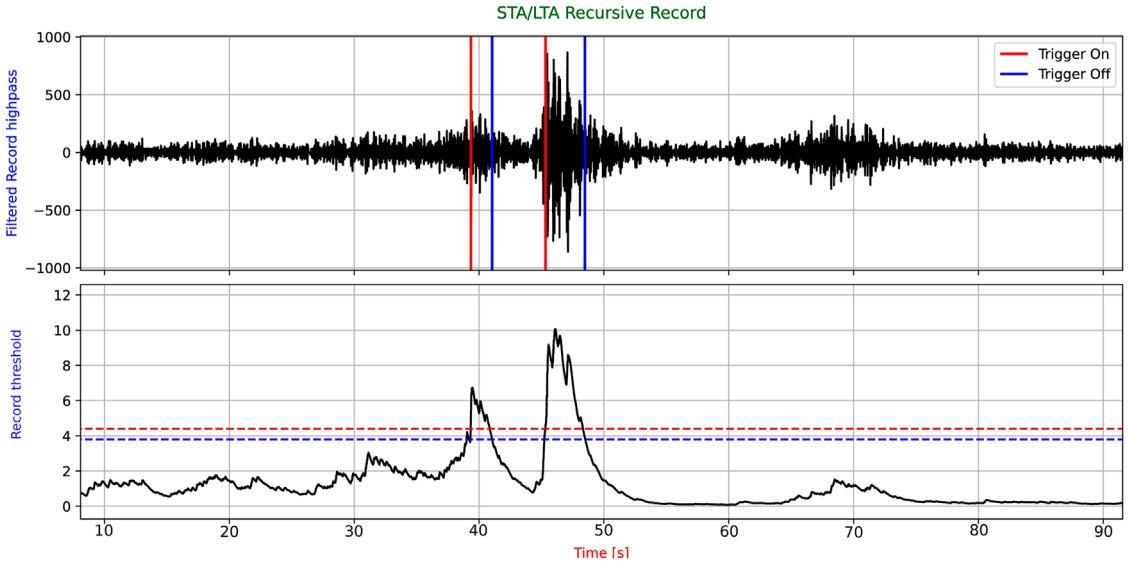


FIGURE 8
Overview of short time average over long time-average (STA/LTA) triggering thresholds in this work.

3.1 RNN-LSTM outcomes

Figure 3A summarizes the VTs detected by the pre-trained RNN–LSTM system before and after re-training using the new (Bezymianny volcano) dataset. Contrary to expectations, the number of VTs detected at some stations using the RNN–LSTM remained constant or decreased after being re-trained. Figures 4, 5 show comparisons of the monthly per-station cumulative distribution function (CDF) before and after the re-training process, representing the probabilities and normalized cumulative sums of events predicted as VTs. Before re-training, while a high number of events were detected, the confidence of such detections was low. More specifically, in August 2017 (Figure 4A), almost 70% of the events detected had probabilities of between 35% and 55%; in October and December 2017 (Figures 4B, C), except at station BZ06 (where 50% of the events detected had probabilities of < 55%), no recognized event exceeded 55%. After re-training, there was a clear change in the trend, with fewer recognized VT earthquakes depending on the station (Figure 3) but much higher confidences of the detections (Figure 5).

3.2 TCN outcomes

Figure 3B summarizes the VTs detected by the pre-trained TCN system before and after re-training using the new (Bezymianny

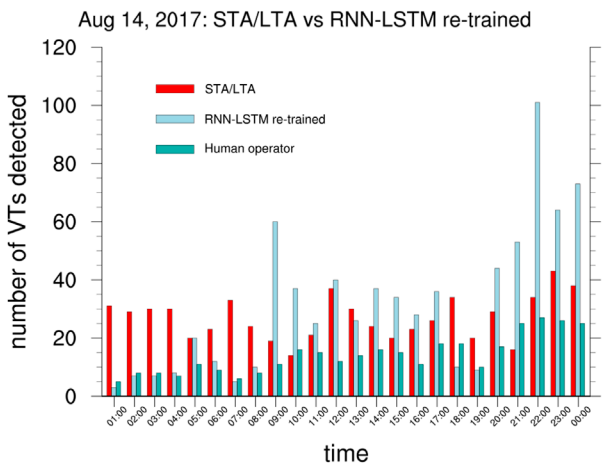


FIGURE 9
Number of volcano tectonic (VT) earthquakes detected hourly on 14 August 2017 using the short time average over long time-average (STA/LTA) trigger algorithm and the re-trained recurrent neural network based on long short-term memory cells (RNN–LSTM). Results obtained by both architectures have been compared with human operator criteria

volcano) dataset. Figures 6, 7 show comparisons of the monthly per-station CDF before and after the re-training process. In contrast to the RNN–LSTM, the TCN architecture saw an increase in the total number of earthquakes detected after being re-trained but a significant decrease in the confidence of the recognitions. Before re-training, 90% of events were detected with probabilities of > 80%; after re-training, depending on the station, only 40%–60% of recognized events had probabilities of > 80%.

3.3 STA/LTA comparison

To determine the robustness of our system, we compared our results before and after re-training to those of a classical approach, the short time average over long time-average (STA/LTA) trigger algorithm (Trnkoczy, 2009). We selected a single day on which several hundred earthquakes occurred and analyzed the results on an hourly timescale. Given that the TCN always detected a greater number of events than the RNN–LSTM, we assumed that the VTs detected by the RNN–LSTM were a subset of those detected by the TCN and selected only those events for analysis. On the chosen

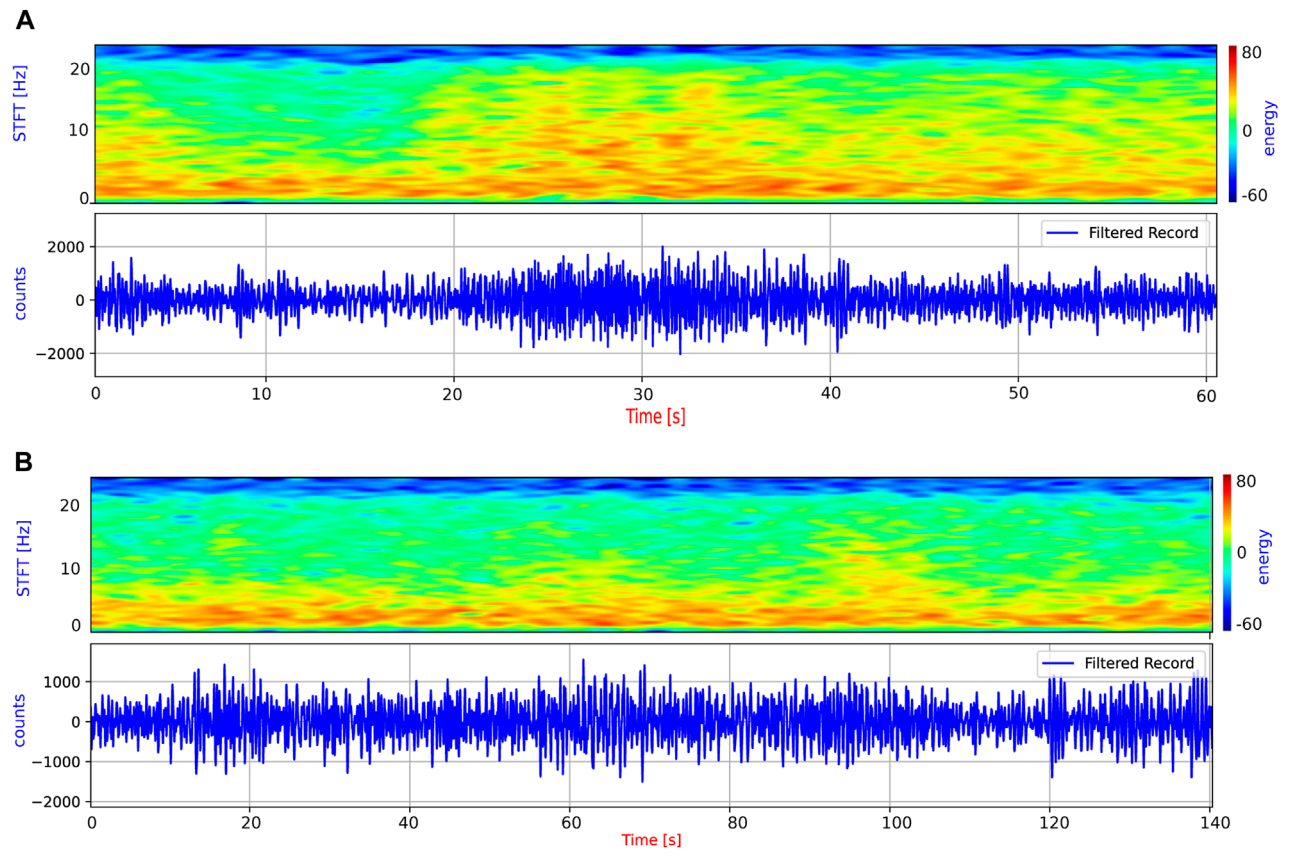


FIGURE 10
Recognition analysis before and after re-training of the recurrent neural network based on long short-term memory cells (RNN–LSTM). Before re-training, the system labeled seismic traces partially matching with volcano tectonic (VT) earthquakes as VTs with low probability. After re-training, only high confidence VTs were detected and labeled; low probability events were categorized as undefined events. **(A)** Example of a low probability (52%) VT earthquake detected before re-training. **(B)** Example of a low probability VT earthquake (60%) labeled as 'undefined' after re-training. With the purpose of enhancing visualization, the raw seismic signals were subjected to a filtering process, limiting the frequency range between 1 and 20 Hz.

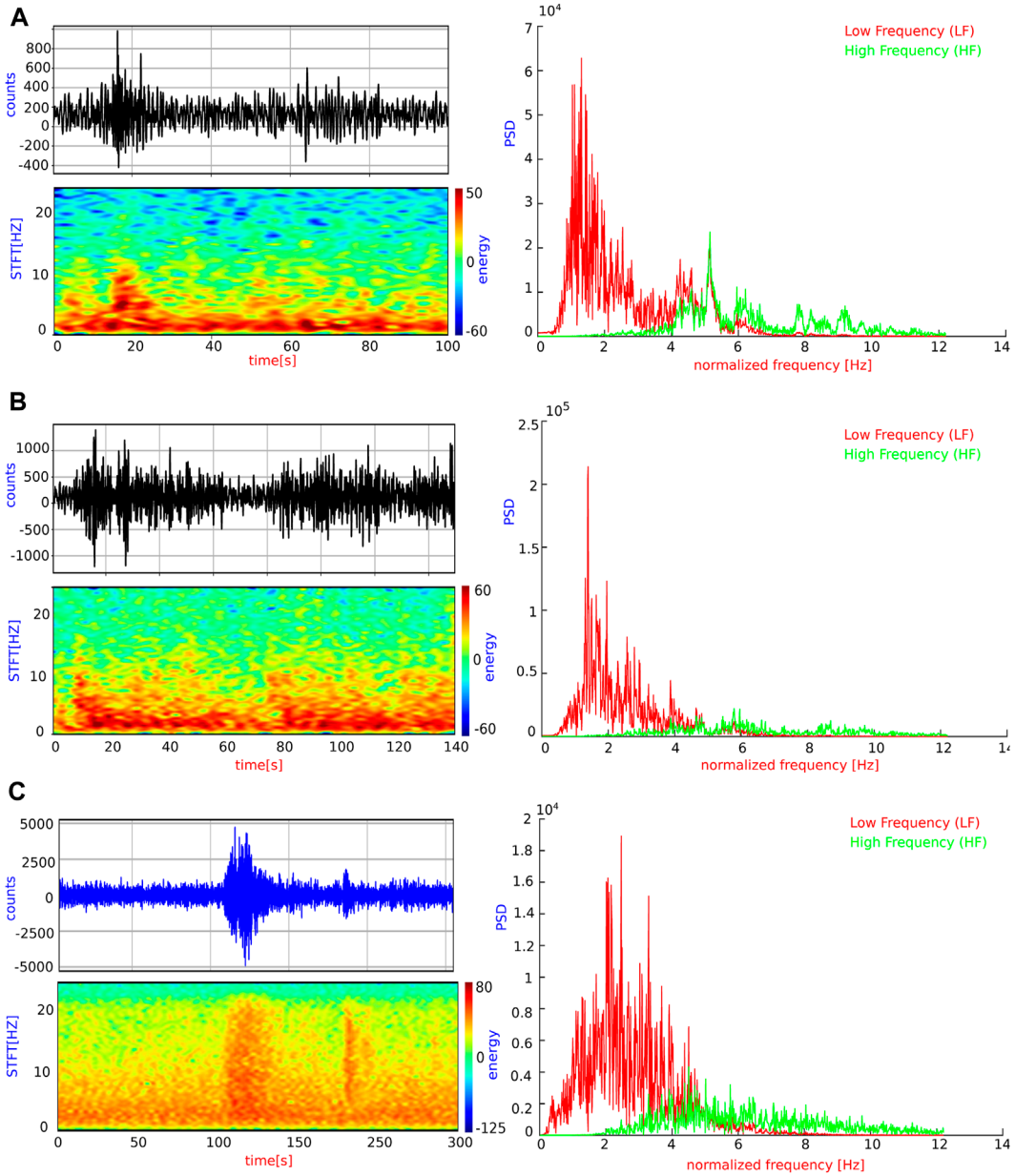


FIGURE 11
Spectrograms and power spectral densities (PSD) of different events belonging to Bezymianny volcano and Deception Island. **(A)** Volcano tectonic (VT) earthquake from Deception Island. **(B)** Volcanic tremor (TR) event from Deception Island. **(C)** VT from Bezymianny volcano.

day, the RNN–LSTM did not recognize any VTs before it was re-trained; after re-training, all of the VTs recognized had previously been categorized as TR events.

Figure 8 presents an overview of the STA/LTA triggering thresholds. For proper operation of the STA/LTA algorithm, four parameters should be tuned: the short window length (STA), long window length (LTA), activation threshold level, and deactivation threshold level. The STA/LTA trigger parameter settings are always a tradeoff among sensitivity and specificity. While sensitivity may also include a tolerable number of false triggers, specificity correctly detects only particular instances, therefore decreasing the number of detections. Considering that the algorithm computes the average absolute amplitude of a seismic signal in two consecutive moving-time windows, only events exceeding pre-set values describing the triggering thresholds of both STA and LTA were identified. Figure 9 compares the number of VTs detected hourly during 14 August 2017, by the STA/LTA trigger algorithm and re-trained RNN-LSTM architecture; overall, the results show that the RNN–LSTM recognized a higher number of VTs than the STA/LTA algorithm (782 vs. 648).

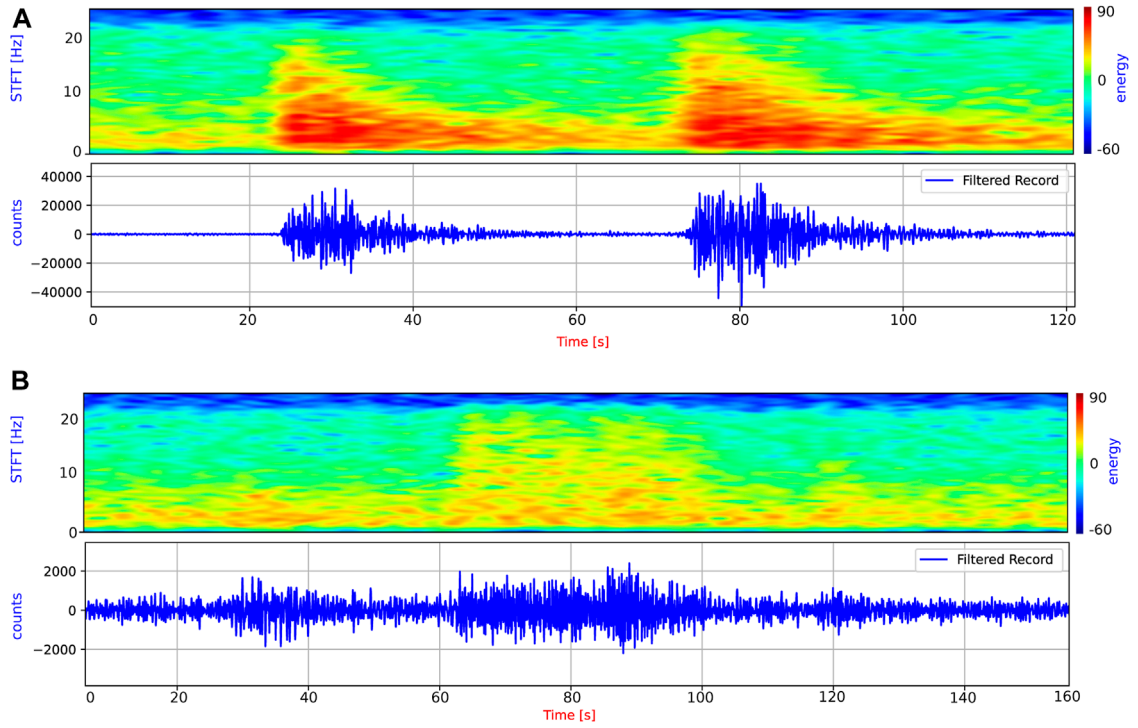


FIGURE 12
Recognition analysis before and after re-training of the temporal convolutional network (TCN). Before re-training, the system labeled seismic traces partially matching with volcano tectonic (VT) earthquakes as VTs with high probability; low probability events were marked as 'undefined'. After re-training, all seismic traces having high frequencies and variable length were detected as VTs; however, only clear VT earthquakes were detected with high probability. **(A)** Example of a high probability VT earthquake (99%) detected before retraining, in which several earthquakes occurring closely together in a short timeframe have been classified as a single event following the volcanic dynamics to the master catalog. Since the system does not detect background noise windows between the two earthquakes due to the multi-resolution dilated skip connections between layers, it labels the two earthquakes as one. **(B)** Example of a low probability (70%) VT earthquake detected after re-training. With the purpose of enhancing visualization, the raw seismic signals were subjected to a filtering process, limiting the frequency range between 1 and 20 Hz.

4 Discussion

4.1 RNN-LSTM considerations

The noticeable difference in the performance of the system after re-training can be explained from a geophysical perspective. First, owing to the noisy content registered in the new volcanic environment, it is possible that many of the events detected before the re-training were low probability VTs (or mis-recognized VT) corresponding to seismic traces characterized by high frequencies and variable length. Since the system fitting approach is a density estimation problem and such seismic traces partially match observed VT features in the master catalog (Deception Island), the estimated probability density function and its parameters cannot explain the underlying distribution of the new input data; as such, it assigns a low probability. Figure 10A provides a clear overview of this issue, in which a seismic trace partially matches with source earthquakes with a low probability (52%).

Second, many of the VTs detected after re-training were originally recognized as TR, which can be explained by the differences between the learned representation at source and target underlying distributions. Figure 11 shows the spectrograms and power spectral densities (PSD) of VTs from Bezymianny volcano and Deception Island, and a TR from Deception Island. The figure shows very similar spectral energy distributions. The beginning of the Deception Island TR (Figure 11B) has a short and overlapped package of high frequency waves (up to 20 Hz). These high frequency signals are associated with the explosive step of pressure in the source region when LP events are generated near the seismic station (no visible exponential decay in frequency is observed) and with small earthquakes. At Bezymianny volcano, many VTs have a higher energy component at low frequencies (Figure 11C); therefore, as our parameterization scheme performs energy analysis by frequency bands that are more sensitive in lower frequencies, the pre-trained system failed to recognizing these energetic low frequency VT events. After re-training, the global number of recognized VTs was similar, but confidence of the detections was much higher.

In summary, the system (i.e., the probability distribution and associated parameters) is fixed to maximize a likelihood function that best explains the joint probability distribution of the new volcanic dynamics (in this case, Bezymianny volcano). As a result, following re-training, only confident VTs were detected and labeled. Those previously mis-recognized as TR were now confidently detected, while those events with low probabilities (< 65%) were labeled as undefined events (e.g., Figure 10B). Such events require

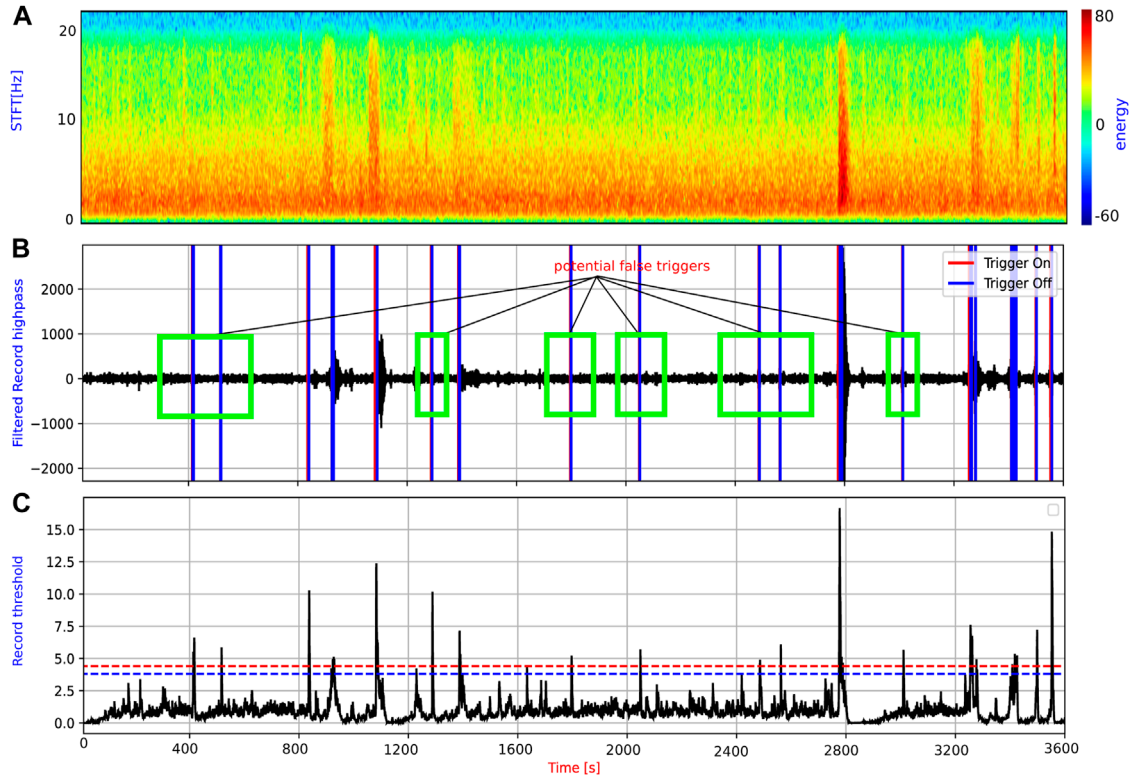


FIGURE 13
Potential false triggers corresponding to short duration energy changes recognized by the short time average over long time-average (STA/LTA) model. **(A)** Spectrogram of the seismic signal selected. **(B)** High pass filtered seismogram. **(C)** short time average over long time-average (STA/LTA) triggering results. Owing to lack of prior knowledge, the recurrent neural network based on long short-term memory cells (RNN–LSTM) model discounted these events as volcano tectonic (VT) earthquakes, since the duration was shorter than the average duration of VT earthquakes in the Deception Island catalog. With the purpose of enhancing visualization, the raw seismic signals were subjected to a filtering process, limiting the frequency range between 1 and 20 Hz.

careful review by experts. As conclusion, before re-training, only VTs with probabilities of > 80% could be included in the new catalog; after re-training, all VTs detected with probabilities of > 65% could be easily identified.

4.2 TCN considerations

The noticeable difference in the performance of the system after re-training can be explained by the greater specialization ability of TCN compared with RNN–LSTM owing to the multi-resolution dilated skip connections between layers and deeper hierarchical features. The new Bezymianny volcanic dynamics often exhibit consecutive seismic events that bear partial resemblance to earthquakes occurring closely together in a short timeframe. Before retraining, the system could avoid recognizing such new volcanic dynamics based on high frequencies and short length as isolated events. Therefore, when such concatenated events were detected, focusing only on those volcanic dynamics that were similar to the master catalog, the system considered them all as a whole, not as isolated events (Figure 12A).

In summary, before re-training, the system labeled seismic traces partially matching with earthquakes as high probability VTs, while low probability events were labeled as 'undefined'. After re-training, all seismic traces with high frequencies and variable length were detected as VTs, decreasing the number of undefined events. However, only clear VTs were detected with high probability. Figure 12B shows an example of a detected low probability (70%) VT. Before re-training, this seismic trace was labeled as 'undefined' with a high probability of assignment to VT (> 90%). After re-training, the system decreased the probability of assignment to VT. In this way, before re-training, many VTs were mis-recognized; after re-training, all VTs detected with probabilities higher than of > 85% could be included in the new catalog.

4.3 STA/LTA considerations

However, in some time slots, STA/LTA detected a greater number of events. These results may be explained by the nature of the STA/LTA algorithm, its trigger parameter settings, and the grammar imposed on the proposed models, which was responsible for improving the interpretability of the models based on geophysical knowledge of the volcano (Titos et al., 2018).

Since there was no previous information related to the seismic catalog, the STA/LTA triggering thresholds were fixed so that the

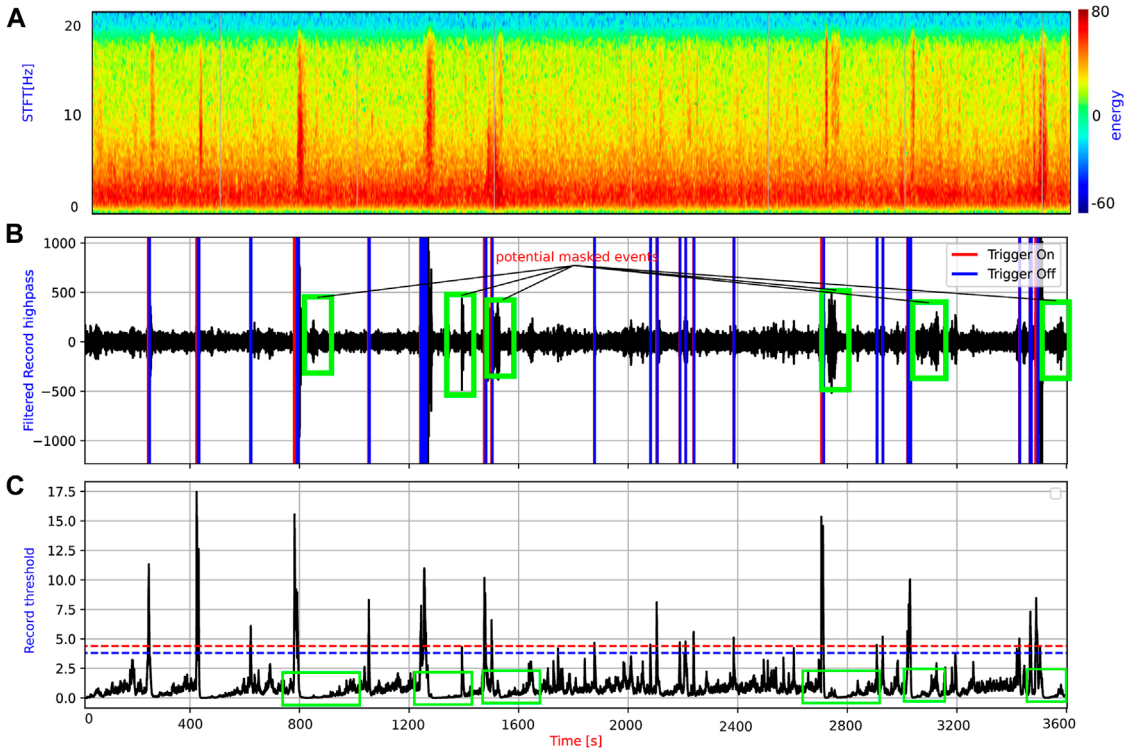


FIGURE 14
Potential masked events, for which the short time average over long time-average (STA/LTA) algorithm computed the average absolute amplitude of a seismic signal from two consecutive moving-time windows containing a low energy event immediately following a very high energy event. The average energy masks the occurrence of the low energy event, decreasing the number of recognized events. **(A)** Spectrogram of the seismic signal selected. **(B)** High pass filtered seismogram. **(C)** short time average over long time-average (STA/LTA) triggering results. With the purpose of enhancing visualization, the raw seismic signals were subjected to a filtering process, limiting the frequency range between 1 and 20 Hz.

system was more sensitive than specific. The goal was to obtain as much information as possible, and all energy changes, even small ones, were detected. This scenario resulted in a tolerable number of false triggers. In contrast, the RNN–LSTM (and TCN) system imposed the use of grammar (a set of rules) based on geophysical knowledge of Deception Island volcano to improve interpretability. The average duration of seismic events belonging to the master dataset (Deception Island) in combination with the per-class probabilities output by the models in the new volcanic environment allowed us to check that the predictions were consistent with the expected lengths of events. Since no information was provided on the average duration of the seismic-volcanic events of the new volcanic environment, the grammar only recognized those events that, on average, had durations that were greater than or similar to those described in the master dataset. Events whose durations were less than the average duration of events in the master database, even if recognized with high per-class probabilities, were labeled as background noise or unknown events.

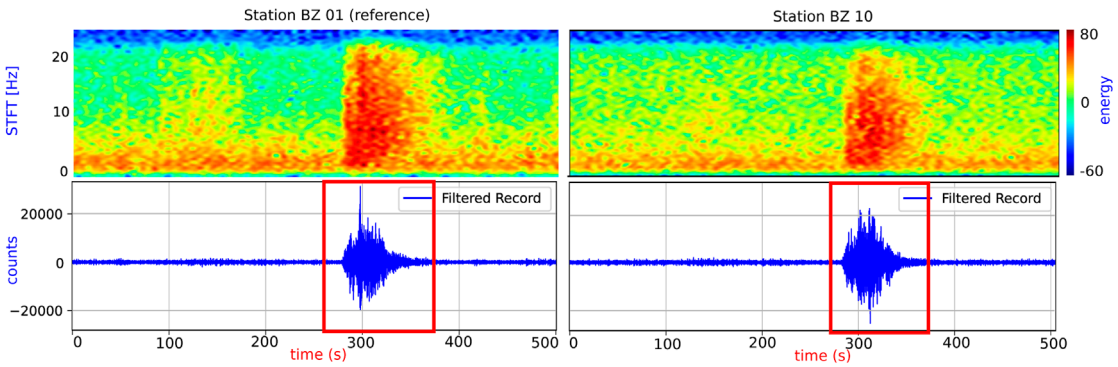
For the time slots in which STA/LTA detected a greater number of VT events compared with RNN–LSTM, many of the events recognized by the STA/LTA model corresponded to short duration energy changes (Figure 13). In contrast, the RNN–LSTM model discarded these events (i.e., labeled them as background noise when the output VT per-class probabilities were low and as unknown events when the output per-class probabilities were high) since the durations were shorter than the average duration of VT earthquakes at Deception Island.

For the time slots in which STA/LTA detected a lower number of VT events compared with the re-trained RNN–LSTM, a possible explanation is the behavior of the STA/LTA algorithm in a seismic swarm state. Seismic swarms, which are a common volcanic phenomenon, involve a sequence of seismic events that occur within a relatively short period of time within a very local area. Given that the STA/LTA algorithm computes an average absolute amplitude of the seismic signal in two consecutive moving-time windows, when a low energy event occurs immediately after a high energy event, the averaging process masks the occurrence of the least energetic one, decreasing the number of recognized events (Figure 14). In contrast, as the RNN–LSTM analyzes signals based on spectral features, it has the ability to analyze a concatenated occurrence of events, such as that observed during a seismic swarm.

Based on our results, once it has been re-trained and the average duration of the seismic-volcanic events has been fixed, our RNN–LSTM has a number of advantages over STA/LTA. In particular, the system will automatically detect VTs present in the seismic trace without searching for optimal parameter settings, which makes it a portable, scalable, and economical tool with relatively low computational cost. Another important advantage is

A

Station	Ref. Station	BGN	TR	HYB	VT	LP	Selected
BZ 01	1	14	37	1	**47**	1	VT
BZ 10	0	17	20	0	**63**	0	VT



B

STATION	REF. Station	BGN	TR	HYB	VT	LP	Selected
BZ 01	1	30	31	1	**37**	0	VT
BZ 10	0	**63**	10	0	26	1	BGN

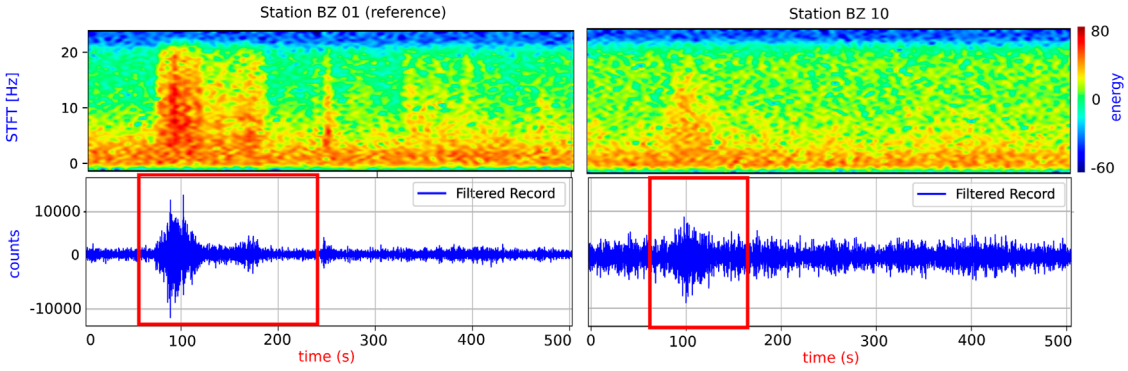


FIGURE 15
Robustness of the obtained seismic catalogs based on per-class probabilities output by the models. **(A)** Example of an earthquake recognized as a volcano tectonic (VT) earthquake with relatively high probability from at least two different seismic stations. **(B)** Example of an attenuated earthquake recognized as a VT, noise, or undefined event depending on the seismic station. With the purpose of enhancing visualization, the raw seismic signals were subjected to a filtering process, limiting the frequency range between 1 and 20 Hz.

that, besides obtaining a preliminary seismic catalog (composed of several types of events), it offers information on the confidence of the recognition. Importantly, for multi-station seismic networks, these probabilities will serve to obtain more reliable seismic catalogs. The recognition of an event characterized by high frequencies at one station provides an indisputable condition to obtain a reliable label at another station where the event has been attenuated or is virtually unrecognizable. An example of this scenario is shown in Figure 15, in which one earthquake is included in the new seismic catalog by both techniques (STA/LTA and RNN-LSTM), while an attenuated one, in addition to presenting difficulty during detection using classical techniques owing to threshold adjustment, could only be considered as an earthquake using our multi-station analysis. Finally, our approach provides a very useful tentative label for subsequent analysis carried out by a human operator.

5 Conclusion

This study provides a comprehensive analysis of how to build a multi-station seismo-volcanic monitoring system based on transfer learning techniques. We evaluated the ability of several operational systems trained using a master seismic catalogue (from Deception Island volcano) to adapt to a new volcanic environment (Bezymianny volcano), without prior domain-specific knowledge.

Our results are significant in at least two major respects. First, transfer learning is shown to offer a robust, effective, and rapid alternative when developing volcano-seismic event monitoring systems in volcanic environments without any previous knowledge or seismic catalogue. Second, depending on the architecture used as a baseline, the final behavior of the system (and consequently the results obtained) can be different. We found that RNN-based systems offer the most reliable recognition by excluding

low confidence detections for seismic traces that are only partially similar to those of the baseline. In contrast, TCN-based networks are capable of detecting a greater number of events; however, many of those events are only partially similar to the master events of the baseline (i.e., the confidence of detections is low). Considering these findings and drawing upon our experience as a guiding factor, we can firmly conclude that among the overall count of events identified as earthquakes, those exhibiting a membership probability surpassing 80% after retraining, can be considered accurately classified. Together, these two approaches offer complementary tools for volcano monitoring, and volcanological observatories should choose the approach that best meets their needs; that is, RNN–LSTM for fine-grained seismic catalogs and TCN for coarse-grained seismic catalogs.

Finally, our study provides a basis for more sophisticated weakly supervised models that could be useful in developing universal monitoring tools able to work accurately across different volcanic systems, even when faced with scenarios without prior domain-specific knowledge.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: The datasets analyzed for this study can be found online at ZENODO (https://doi.org/10.5281/zenodo.7755506). The open source code developed in this work can be downloaded at Github (https://github.com/mmtitos/Multi-station-volcano-tectonic-earthquakes-monitoring-based-on-transfer-learning.git).

Author contributions

Conceptualization: MT, CB, JI. Methodology: MT, CB, LG, PR, JI. Software: MT, LG, CB. Writing—original draft: MT, CB, JI. Writing—review and editing: MT, CB. Funding: CB, JI. All authors contributed to the article and approved the submitted version.

Funding

This study was partially supported by the Spanish FEMALE (PID 2019-106260GB-I00) and PROOF-FOREVER (EUR2022.134044) projects. This work has been partially supported by the project EUR 2022-134044 founded by MCIN/AEI/10.13039/501100011033 in the framework PROYECTOS "EUROPA EXCELENCIA" 2022, CORRESPONDIENTES AL PROGRAMA ESTATAL PARA AFRONTAR LAS PRIORIDADES DE NUESTRO ENTORNO, SUBPROGRAMA ESTATAL DE INTERNACIONALIZACIÓN, DEL PLAN ESTATAL DE INVESTIGACIÓN CIENTÍFICA, TÉCNICA Y DE INNOVACIÓN PARA EL PERIODO 2021-2023, EN EL MARCO DEL PLAN DE RECUPERACIÓN TRANSFORMACIÓN Y RESILIENCIAthe. English language editing was performed by Tornillo Scientific.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author CB declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

References

Abacha, I., Bendjama, H., Boulahia, O., Yelles-Chaouche, A., Roubeche, K., Rahmani, S. T.-E., et al. (2023). Fluid-driven processes triggering the 2010 beni-ilmane earthquake sequence (Algeria): Evidence from local earthquake tomography and 4d vp/vs models. *J. Seismol.* 27, 77–94. doi:10.1007/s10950-022-10130-8

Aki, K., and Richards, P. G. (2002). *Quantitative seismology*.

Anantrasirichai, N., Biggs, J., Albino, F., Hill, P., and Bull, D. (2018). Application of machine learning to classification of volcanic deformation in routinely generated insar data. *J. Geophys. Res. Solid Earth* 123, 6592–6606. doi:10.1029/2018jb015911

Barsotti, S., Oddsson, B., Gudmundsson, M., Pfeffer, M., Parks, M., Ófeigsson, B., et al. (2020). Operational response and hazards assessment during the 2014–2015 volcanic crisis at bárðarbunga volcano and associated eruption at holuhraun, Iceland. *J. Volcanol. Geotherm. Res.* 390, 106753. doi:10.1016/j.jvolgeores.2019.106753

Barsotti, S., Parks, M. M., Pfeffer, M. A., Óladóttir, B. A., Barnie, T., Titos, M. M., et al. (2023). The eruption in fagradalsfjall (2021, Iceland): How the operational monitoring and the volcanic hazard assessment contributed to its safe access. *Nat. Hazards* 116, 3063–3092. doi:10.1007/s11069-022-05798-7

Bueno, A., Benitez, C., De Angelis, S., Moreno, A. D., and Ibanez, J. M. (2019). Volcano-seismic transfer learning and uncertainty quantification with bayesian neural networks. *IEEE Trans. Geoscience Remote Sens.* 58, 892–902. doi:10.1109/tgrs.2019.2941494

Bueno, A., Titos, M., Benítez, C., and Ibáñez, J. M. (2021). Continuous active learning for seismo-volcanic monitoring. *IEEE Geoscience Remote Sens. Lett.* 19, 1–5. doi:10.1109/lgrs.2021.3121611

Castro-Melgar, I., Prudencio, J., Del Pezzo, E., Giampiccolo, E., and Ibanez, J. M. (2021). Shallow magma storage beneath mt. etna: Evidence from new attenuation tomography and existing velocity models. *J. Geophys. Res. Solid Earth* 126, e2021JB022094. doi:10.1029/2021JB022094

Chouet, B. A., and Matoza, R. S. (2013). A multi-decadal view of seismic methods for detecting precursors of magma movement and eruption. *J. Volcanol. Geotherm. Res.* 252, 108–175. doi:10.1016/j.jvolgeores.2012.11.013

Chouet, B. (2003). Volcano seismology. *Pure Appl. Geophys.* 160, 739–788. doi:10.1007/pl00012556

Cubuk-Sabuncu, Y., Jónsdóttir, K., Caudron, C., Lecocq, T., Parks, M. M., Geirsson, H., et al. (2021). Temporal seismic velocity changes during the 2020 rapid inflation at mt. Þorbjörn-svartsengi, Iceland, using seismic ambient noise. *Geophys. Res. Lett.* 48, e2020GL092265. doi:10.1029/2020gl092265

D'Auria, L., Koulakov, I., Prudencio, J., Cabrera-Pérez, I., Ibáñez, J. M., Barrancos, J., et al. (2022). Rapid magma ascent beneath la palma revealed by seismic tomography. *Sci. Rep.* 12, 17654. doi:10.1038/s41598-022-21818-9

Díaz-Moreno, A., Ibáñez, J., De Angelis, S., García-Yeguas, A., Prudencio, J., Morales, J., et al. (2015). Seismic hydraulic fracture migration originated by successive deep

magma pulses: The 2011–2013 seismic series associated to the volcanic activity of el hierro island. *J. Geophys. Res. Solid Earth* 120, 7749–7770. doi:10.1002/2015jb012249

Girina, O., Loupian, E., Melnikov, D., Manevich, A., Sorokin, A., Kramareva, L., et al. (2018). Bezymianny volcano eruption on december 20, 2017. *Mod. Probl. Remote Sens. Earth Space* 15, 88–99. doi:10.21046/2070-7401-2018-15-3-88-99

Ibáñez, J. M., Pezzo, E. D., Almendros, J., La Rocca, M., Alguacil, G., Ortiz, R., et al. (2000). Seismovolcanic signals at deception island volcano, Antarctica: Wave field analysis and source modeling. *J. Geophys. Res. Solid Earth* 105, 13905–13931. doi:10.1029/2000jb900013

Jozinović, D., Lomax, A., Štajduhar, I., and Michelini, A. (2022). Transfer learning: Improving neural network based prediction of earthquake ground shaking for an area with insufficient training data. *Geophys. J. Int.* 229, 704–718. doi:10.1093/gji/ggab488

Koulakov, I., Plechov, P., Mania, R., Walter, T. R., Smirnov, S. Z., Abkadyrov, I., et al. (2021). Anatomy of the bezymianny volcano merely before an explosive eruption on 20 12 2017. *Sci. Rep.* 11, 1–12. doi:10.1038/s41598-021-81498-9

Lapins, S., Goitom, B., Kendall, J.-M., Werner, M. J., Cashman, K. V., and Hammond, J. O. (2021). A little data goes a long way: Automating seismic phase arrival picking at nabro volcano with transfer learning. *J. Geophys. Res. Solid Earth* 126, e2021JB021910. doi:10.1029/2021jb021910

Lea, C., Vidal, R., Reiter, A., and Hager, G. D. (2016). "Temporal convolutional networks: A unified approach to action segmentation," in *Computer vision–ECCV 2016 workshops* (Amsterdam, Netherlands: Springer), 47–54. October 8-10 and 15-16, 2016, Proceedings, Part III 14.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *nature* 521, 436–444. doi:10.1038/nature14539

Machacca, R., Lesage, P., Tavera, H., Pesicek, J., Caudron, C., Torres, J., et al. (2023). The 2013–2020 seismic activity at sabancaya volcano (Peru): Long lasting unrest and eruption. *J. Volcanol. Geotherm. Res.* 435, 107767. doi:10.1016/j.jvolgeores.2023.107767

McGimsey, R. G., Neal, C. A., Girina, O. A., Chibisova, M., and Rybin, A. V. (2014). *2009 volcanic activity in Alaska, kamchatka, and the kurile islands—summary of events and response of the Alaska volcano observatory*.

McNutt, S. R., and Roman, D. C. (2015). "Volcanic seismicity," in *The encyclopedia of volcanoes* (Elsevier), 1011–1034.

McNutt, S. R., Thompson, G., Johnson, J., De Angelis, S., and Fee, D. (2015). "Seismic and infrasonic monitoring," in *The encyclopedia of volcanoes* (Elsevier), 1071–1099.

Nair, N., Thomas, C., and Jayagopi, D. B. (2018). "Human activity recognition using temporal convolutional network," in *Proceedings of the 5th international workshop on sensor-based activity recognition and interaction* (New York, NY, USA: Association for Computing Machinery). iWOAR '18. doi:10.1145/3266157.3266221

Neal, C., Girina, O., Senyukov, S., Rybin, A., Osiensky, J., Izbekov, P., et al. (2009). Russian eruption warning systems for aviation. *Nat. hazards* 51, 245–262. doi:10.1007/s11069-009-9347-6

Prudencio, J., Ibánez, J. M., García-Yeguas, A., Del Pezzo, E., and Posadas, A. M. (2013). Spatial distribution of intrinsic and scattering seismic attenuation in active volcanic islands–ii: Deception island images. *Geophys. J. Int.* 195, 1957–1969. doi:10.1093/gji/ggt360

Račič, M., Oštir, K., Peressutti, D., Zupanc, A., and Čehovin Zajc, L. (2020). "Application of temporal convolutional neural network for the classification of crops on sentinel-2 time series," in *Isprs - international Archives of the photogrammetry, remote Sensing and spatial information sciences XLIII-B2-2020*, 1337–1342. doi:10.5194/isprs-archives-XLIII-B2-2020-1337-2020

Roman, D. C., and Cashman, K. V. (2006). The origin of volcano-tectonic earthquake swarms. *Geology* 34, 457–460. doi:10.1130/g22269.1

Sigmundsson, F., Parks, M., Pedersen, R., Jónsdóttir, K., Ófeigsson, B. G., Grapenthin, R., et al. (2018). "Magma movements in volcanic plumbing systems and their associated ground deformation and seismic patterns," in *Volcanic and igneous plumbing systems* (Elsevier), 285–322.

Sigmundsson, F., Parks, M., Hooper, A., Geirsson, H., Vogfjörd, K. S., Drouin, V., et al. (2022). Deformation and seismicity decline before the 2021 fagradalsfjall eruption. *Nature* 609, 523–528. doi:10.1038/s41586-022-05083-4

Sparks, R., Biggs, J., and Neuberg, J. (2012). Monitoring volcanoes. *Science* 335, 1310–1311. doi:10.1126/science.1219485

Titos, M., Bueno, A., García, L., Benítez, M. C., and Ibañez, J. (2018). Detection and classification of continuous volcano-seismic signals with recurrent neural networks. *IEEE Trans. Geoscience Remote Sens.* 57, 1936–1948. doi:10.1109/tgrs.2018.2870202

Titos, M., Bueno, A., García, L., Benitez, C., and Segura, J. C. (2019). Classification of isolated volcano-seismic events based on inductive transfer learning. *IEEE Geoscience Remote Sens. Lett.* 17, 869–873. doi:10.1109/lgrs.2019.2931063

Titos, M., García, L., Kowsari, M., and Benítez, C. (2022). Toward knowledge extraction in classification of volcano-seismic events: Visualizing hidden states in recurrent neural networks. *IEEE J. Sel. Top. Appl. Earth Observations Remote Sens.* 15, 2311–2325. doi:10.1109/jstars.2022.3155967

Trnkoczy, A. (2009). "Understanding and parameter setting of sta/lta trigger algorithm," in *New manual of seismological observatory practice (NMSOP) (Deutsches GeoForschungsZentrum GFZ)*, 1–20.

Van Manen, S., Dehn, J., and Blake, S. (2010). Satellite thermal observations of the bezymianny lava dome 1993–2008: Precursory activity, large explosions, and dome growth. *J. Geophys. Res. Solid Earth* 115, B08205. doi:10.1029/2009jb006966

Weiss, K., Khoshgoftaar, T. M., and Wang, D. (2016). A survey of transfer learning. *J. Big data* 3, 9–40. doi:10.1186/s40537-016-0043-6

Yan, J., Mu, L., Wang, L., Ranjan, R., and Zomaya, A. Y. (2020). Temporal Convolutional Networks for the Advance Prediction of ENSO. *Sci. Rep.* 10, 8055. doi:10.1038/s41598-020-65070-5

frontiers | Frontiers in Earth Science

TYPE Original Research
PUBLISHED 02 November 2023
DOI 10.3389/feart.2023.1264701

OPEN ACCESS

EDITED BY
Carmen Benítez,
Universidad de Granada, Spain

REVIEWED BY
Manuel Titos,
Icelandic Meteorological Office, Iceland
Guillermo Cortés,
University of Granada, Spain

*CORRESPONDENCE
Rina Noguchi,
r-noguchi@env.sc.niigata-u.ac.jp

RECEIVED 21 July 2023
ACCEPTED 12 October 2023
PUBLISHED 02 November 2023

CITATION
Noguchi R and Shoji D (2023), Extraction of stratigraphic exposures on visible images using a supervised machine learning technique.
Front. Earth Sci. 11:1264701.
doi: 10.3389/feart.2023.1264701

Extraction of stratigraphic exposures on visible images using a supervised machine learning technique

Rina Noguchi[1]* and Daigo Shoji[2]

[1]Faculty of Science, Niigata University, Niigata, Japan, [2]Institute of Space and Astronautical Science, Japan Aerospace Exploration Agency, Sagamihara, Kanagawa, Japan

As volcanic stratigraphy provides important information about volcanic activities, such as the eruption style, duration, magnitude, and their time sequences, its observation and description are fundamental tasks for volcanologists. Since outcrops are often obscured in nature, the first task would be identifying stratigraphic exposures in many cases. This identification/selection process has depended on humans and has difficulties in terms of time and effort consumption and in biases resulting from expertise levels. To address this issue, we present an approach that utilizes supervised machine learning with fine-tuning and forms the backbone to automatically extract the areas of stratigraphic exposures in visible images of volcanic outcrops. This study aimed to develop an automated method for identifying exposed stratigraphy. This method will aid in planning subsequent field investigations and quickly outputting results. We used U-Net and LinkNet, convolutional neural network architectures developed for image segmentation. Our dataset comprised 75 terrestrial outcrop images and their corresponding images with manually masked stratigraphic exposure areas. Aiming to recognize stratigraphic exposures in various situations, the original images include unnecessary objects such as sky and vegetation. Then, we compared 27 models with varying network architectures, hyperparameters, and training techniques. The highest validation accuracy was obtained by the model trained using the U-Net, fine-tuning, and ResNet50 backbone. Some of our trained U-Net and LinkNet models successfully excluded the sky and had difficulties in excluding vegetation, artifacts, and talus. Further surveys of reasonable training settings and network structures for obtaining higher prediction fidelities in lower time and effort costs are necessary. In this study, we demonstrated the usability of image segmentation algorithms in the observation and description of geological outcrops, which are often challenging for non-experts. Such approaches can contribute to passing accumulated knowledge on to future generations. The autonomous detection of stratigraphic exposures could enhance the output from the vast collection of remote sensing images obtained not only on Earth but also on other planetary bodies, such as Mars.

KEYWORDS

machine learning, stratigraphic exposures, volcanic stratigraphies, autonomous extraction, U-Net, LinkNet

Introduction

Volcanic stratigraphy provides important information about volcanic activities. For example, from the thickness of the tephra, we can obtain information about the duration and magnitude of the eruption. Its spatial distribution/variation has been used to estimate the wind direction at the time of eruption and the erupted volume, and to evaluate the eruption magnitude (e.g., Bonadonna et al., 2016 and references therein). The analyses of the grain size contributed to evaluating explosivity during an eruption (e.g., Kueppers et al., 2006a). Applying fractal theory, we can quantify and compare explosivity among several eruptions (e.g., Kueppers et al., 2006b; Perugini et al., 2011). The grain shape and texture analyses of tephra contributed to inferring with the eruption style such as external water participation (e.g., Wohletz and Heiken, 1992; Miwa et al., 2015; Dürig et al., 2021). Statistical analysis and machine learning techniques provide new insights into the characterizations and classifications of the eruption (Leibrandt and Le Pennec, 2015; Liu et al., 2015; Shoji et al., 2018). Thus, the observation and description of volcanic stratigraphy are fundamental tasks for volcanologists to obtain details and histories of eruptions. Such approaches are common to other geological surveys on different targets, fields, and other solid bodies.

Depending on each research subject, the essential requirement in selecting which areas/parts of the outcrop are worth examining and describing is the clear appearance of the layering structure. Since talus and vegetation often obscure outcrops in nature, identifying the areas/parts of stratigraphic exposures is the first task in the field survey. Modifications of outcrops to expose clear stratigraphy are performed occasionally, although those contributions are limited and prohibited in protected areas. Therefore, we often must find naturally clear stratigraphic exposures on outcrops without any modifications. The identification process of stratigraphic exposures has traditionally been dependent on visual observation by humans. Human observation contains difficulties in terms of time and effort consumption, as well as biases resulting from the expertise levels of each person. In the field, time is often limited due to weather, imminent danger, and accessibility. Huge outcrops and large research areas make this problem more serious.

As a solution, recently the unmanned aerial vehicle (UAV) has been actively used for geological surveys. The programmed flights of UAVs can search vast and challenging areas compared with humans' survey (Smith and Maxwell, 2021). Combining the technologies of image analysis, UAVs also show potential to select appropriate locations automatically. On the other hand, UAVs have several difficulties compared with humans. One of these difficulties is to identify stratigraphic exposures. The identification and discrimination of stratigraphic exposures can be challenging for non-experts. Automating the identification of stratigraphic exposures on outcrops significantly contributes to solving issues related to time, effort, and expertise levels.

Difficulties setting the threshold for unparameterizable issues in landforms have been solved by applying machine learning techniques and could also be applied for scouting exposed stratigraphies. One of the typical target landforms for the application of machine learning is impact craters on terrestrial bodies. Using the Mars Orbiter Laser Altimeter digital elevation model, Stepinski et al. (2009) presented the automated cataloging of impact craters and found a regional decrease in the crater depth/diameter ratio, which may relate to subsurface ice. For finding new impact craters which previously depended on manual detection strongly biased by thermal inertia, Wagstaff et al. (2022) trained and evaluated the classifier with visible images obtained using the Context Camera onboard the Mars Reconnaissance Orbiter. Another application example is the detection of volcanic rootless cones (Palafox et al., 2017). Compared with support vector machine approaches, Shozaki et al. (2022) demonstrated the recognition and classification of Martian chaotic terrains using convolutional neural network models. Their models showed over 90% accuracy for the classification and contributed to revealing the global distribution of chaos features on Mars. Thus, automation using machine learning techniques can contribute to identification and discrimination of landforms, which depends on geological expertise so far and could be suitable for finding stratigraphic exposures.

In this study, we present an approach that utilizes supervised machine learning to automatically extract areas of stratigraphic exposures in any type (e.g., ground view, from UAVs) of visible images of volcanic outcrops. Semantic segmentation, a deep learning algorithm, used for the extraction of obstacles and anomalies is adapted for the extraction of stratigraphic exposures among outcrop images in a short time. For example, Silburt et al., 2019 built a convolutional neural network (CNN) architecture to extract lunar impact craters from the Moon digital elevation model. Their trained CNN showed a high precision for crater detection as human-generated. Inspired by Silburt et al., 2019, Latorre et al., 2023 implemented several transfer learning approaches including fine-tuning and presented its capability for the autonomous detection of impact craters across the Moon and Ceres, which have different geological features. Thus, semantic segmentation shows potential to extract specific geological features autonomously. Furthermore, implementing transfer learning approaches for CNN contributes to wider use across target bodies. Here, we demonstrate semantic segmentation for the autonomous extraction of stratigraphic exposures from outcrop images, implementing fine-tuning and pre-trained weights. The main purpose of this study is to provide an automatic identification method of exposed stratigraphy which will be helpful in the planning/preparation of subsequent investigations and in the prompt outputting of investigated results.

Methods

In this work, we demonstrated the automated identification of stratigraphic exposure on outcrop images using a machine learning algorithm. The procedure of this study involves the 1) preparation of original images, 2) generation of masked images, 3) augmentation of original and masked images, and 4) training and evaluation of the U-Net and LinkNet networks to detect stratigraphic exposures (Figure 1).

Preparation of outcrop images

We prepared outcrop images that contained stratigraphic exposures, sky, vegetation, artifacts, and talus to train the network (Figure 1). The total number of prepared images (hereafter, original images) is 75, of which 30 images were our holdings and 45 images were obtained using *google-image-download*, a Python script published on

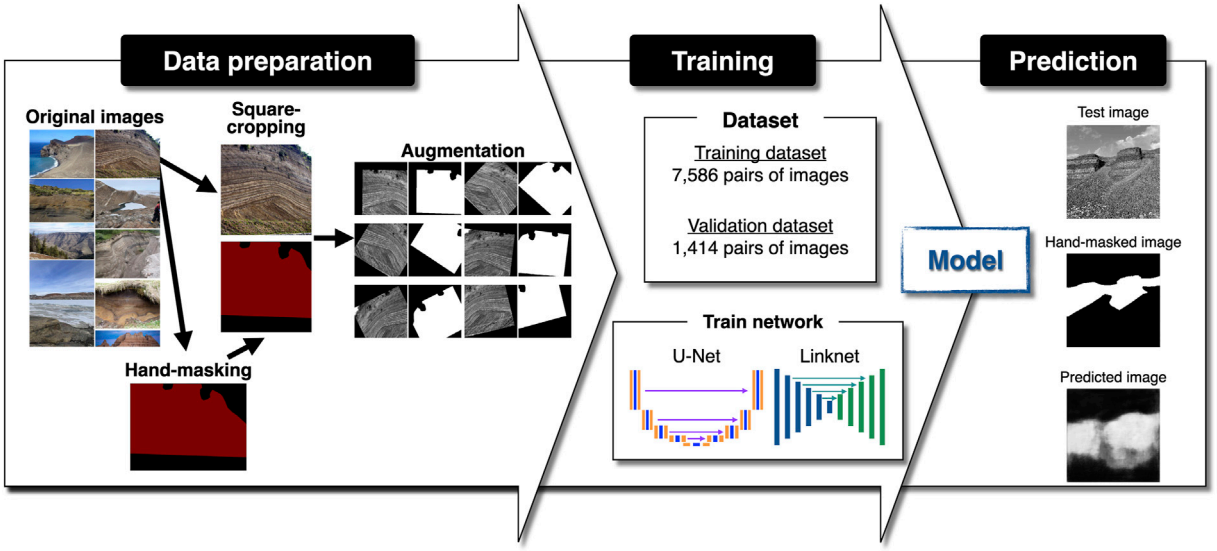


FIGURE 1
Procedure of dataset preparation, training, and evaluation in this study.

GitHub (https://github.com/Joeclinton1/google-images-download/tree/patch-1, Copyright © 2015–2019 Hardik Vasa). This script was developed for searching Google Images on keywords/key phrases and downloading images to locals. Using this script, we obtained outcrop images with "geologic layer," "tephra layer," "volcanic tephra layer," and "volcano geologic layer" as keywords/key phrases. Therefore, our dataset includes both volcanic and non-volcanic stratigraphies. This is acceptable because we focus on the extraction of stratigraphic structures itself in this study. The whole images we obtained by *google-image-download* were labeled as non-commercial reuse with modification. The original images were taken from various distances (meters to kilometers), which correspond to a scale ranging from *in situ* observations to aerial surveys. Aiming to recognize stratigraphic exposures in various situations, the original images include talus, gullies, vegetation, snowy regions, artifacts, and other unnecessary objects (Figure 1). A total of 60, 10, and 5 of those images were used as training, validation, and testing images, respectively (i.e., 80%, 13.3%, and 6.6% splitting). The original images are available in a public repository at https://doi.org/10.5281/zenodo.8396332.

Hand-masked image generation

The masking of stratigraphic exposure areas was carefully performed manually, and unnecessary objects were excluded (Figure 1). Images that are masked at the stratigraphic exposure's region for each original image were generated using *labelme* (https://github.com/wkentaro/labelme, Copyright © 2016–2018 Kentaro Wada), a tool that allows graphical annotation on images. We annotated the region of stratigraphic exposures as polygons and then saved it as a binarized image (hereafter, hand-masked images). Those hand-masked images are available in a public repository at https://doi.org/10.5281/zenodo.8396332.

Image augmentation

To increase the generalization ability of the network, we augmented the original and the hand-masked training/validation images by rotation, horizontal and vertical shifts, and horizontal flip, and converted them into 256 x 256 grayscale images (Figure 1). Data augmentation is a common technique in the training of neural networks to overcome the small amount of data. Before augmentation, we cropped each image as a square because the original images were not square and its aspect ratio will be modified by resizing in augmentation. To augment both original and hand-masked images, Keras ImageDataGenerator (Chollet, 2015) was used. The rotation range was 45°. The maximum width and height shifts were 20% against the width and height. In the ImageDataGenerator, we did not use zoom and shear functions during augmentation because the unfixable aspect ratio could generate pseudo-layering structures in zoom and shear. The fill mode was "constant" because the default "nearest" generates pseudo-layering structures. The angle and width of the rotation and shift were randomly determined within the range, and horizontal flipping occurred randomly. The size of augmented images was adjusted to 256 x 256 pixels to input the following algorithm. Due to the limitation to our computing system, we augmented images to be less than 10,000 in total. As a result, we obtained 7,586 and 1,414 pairs of augmented original/hand-masked images for training and validation datasets, respectively.

Training

To extract certain areas from images by our system, we applied image segmentation algorithms. In this study, we compared two architectures: U-Net (Ronneberger et al., 2015) and LinkNet (Chaurasia and Culurciello, 2017). U-Net is a fully convolutional network originally developed for biomedical image segmentation. It

TABLE 1 Loss, dice, and binary accuracy at the end of training in each model.

ID	Model	Fine-tuning	Backbone	Step	Training stopped epoch	Training			Validation			Duration (min)	
						Loss	Dice loss	Accuracy	Loss	Dice loss	Accuracy	Total	Per epoch
U-1	U-Net	No	No	50	58	0.582	0.422	0.785	0.573	0.569	0.646	512	8.83
U-2				100	52	0.568	0.416	0.802	0.582	0.573	0.638	298	5.73
U-3				150	48	0.563	0.411	0.810	0.577	0.573	0.663	282	5.88
U-4		No	Yes (ResNet18)	50	74	0.468	0.290	0.785	0.632	0.509	0.648	90	1.22
U-5				100	62	0.446	0.263	0.791	0.631	0.484	0.659	70	1.13
U-6				150	62	0.428	0.255	0.799	0.644	0.485	0.670	76	1.23
U-7		No	Yes (ResNet50)	50	73	0.450	0.289	0.800	0.591	0.506	0.672	121	1.66
U-8				100	49	0.451	0.273	0.790	0.616	0.493	0.665	78	1.59
U-9				150	44	0.465	0.300	0.781	0.587	0.506	0.671	76	1.73
U-10		Yes	Yes (ResNet18)	50	63	0.322	0.214	0.882	0.584	0.505	0.686	76	1.21
U-11				100	55	0.315	0.220	0.911	0.567	0.496	0.697	61	1.11
U-12				150	43	0.258	0.158	0.906	0.643	0.469	0.693	53	1.23
U-13		Yes	Yes (ResNet50)	50	138	0.254	0.195	0.954	0.507	0.471	0.733	229	1.66
U-14				100	62	0.206	0.139	0.938	0.567	0.442	0.728	99	1.60
U-15				150	63	0.187	0.128	0.947	0.529	0.428	0.748	108	1.71
L-1	LinkNet	No	Yes (ResNet18)	50	28	0.731	0.455	0.531	0.738	0.607	0.509	17	0.61
L-2				100	21	0.772	0.510	0.467	0.680	0.629	0.599	14	0.67
L-3				150	20	0.852	0.567	0.443	0.692	0.667	0.620	15	0.75
L-4		No	Yes (ResNet50)	50	20	0.666	0.412	0.586	0.712	0.583	0.493	30	1.50
L-5				100	108	0.519	0.321	0.742	0.615	0.506	0.646	146	1.35
L-6				150	134	0.508	0.293	0.751	0.645	0.486	0.647	190	1.42
L-7		Yes	Yes (ResNet18)	50	23	0.777	0.351	0.582	1.090	0.535	0.421	14	0.61
L-8				100	20	0.677	0.417	0.593	0.723	0.572	0.514	13	0.65
L-9				150	138	0.393	0.230	0.824	0.587	0.478	0.705	99	0.72
L-10		Yes	Yes (ResNet50)	50	88	0.328	0.206	0.873	0.586	0.483	0.689	126	1.43
L-11				100	104	0.314	0.219	0.898	0.562	0.494	0.697	141	1.36
L-12				150	88	0.303	0.183	0.896	0.678	0.465	0.706	128	1.45

is designed to work with a small number of training images and produce precise segmentation results. This network classifies each pixel and then outputs segmentation maps, and has been applied to segmentation in terrestrial and planetary remote sensing images (e.g., Silburt et al., 2019; Wieland et al., 2019; Zhu et al., 2021; Latorre et al., 2023). LinkNet is a fully convolutional neural network for fast image semantic segmentation and has been developed to recover the spatial information on images more efficiently through the decoder which was diminished during the encoding procedure [employs ResNet18 (He et al., 2016) as a backbone in the original].

Training, validation, and testing were performed using the Keras package (https://keras.io), which is free and written in Python (Chollet, 2015). We used the original U-Net architecture (Ronneberger et al., 2015) as "simple U-Net" using the *unet* script (https://github.com/zhixuhao/unet, Copyright © 2019 xhizuhao), which does not implement fine-tuning. The segmentation model package (Iakubovskii, 2019), a Python library with neural networks for image segmentation based on Keras and TensorFlow (Abadi et al., 2016), was used for U-Net and LinkNet model buildings with backbones and with/without encoder weights. The path of U-Net comprises unpadded 2 x 2 convolutions [followed by rectified linear units (ReLUs)], 2 x 2 max pooling operations with stride 2 for downsampling, and 2 x 2 upconvolutions. The lowest resolution of images in training in our network is 16 x 16 pixels. The total number of convolutional layers in this network is 23. The LinkNet used comprises three decoder blocks that use the UpSampling Keras layer. Supplementary Figure S1,S2

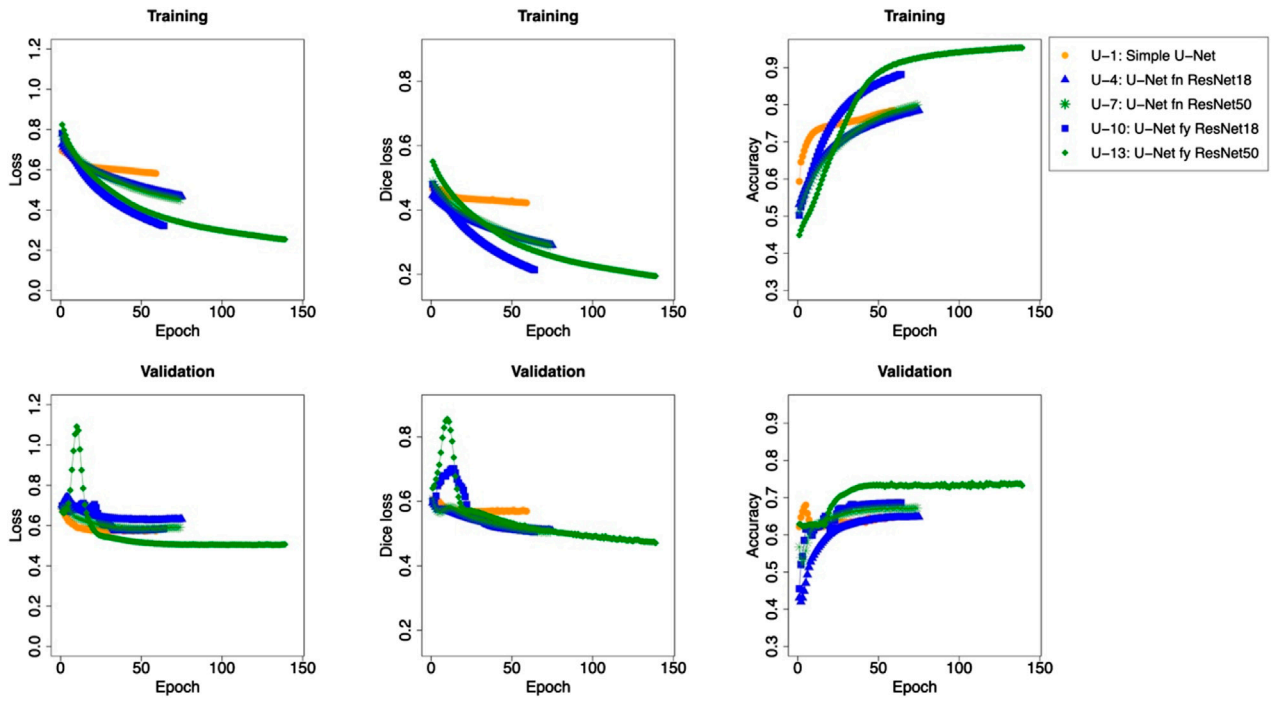


FIGURE 2
Loss, dice loss, and accuracy plots and curves during training for U-Net models with a setting of 50 steps per epoch.

shows the model architectures trained in this study. The training was executed in Mac Studio, Apple M1 Max, 64 GB memory. In Keras, an epoch is an arbitrary cutoff, generally defined as "one pass over the entire dataset," used to separate training into distinct phases, which is useful for logging and periodic evaluation, and "steps per epoch" is a total number of steps (batches of samples) before declaring one epoch finished and starting the next epoch (Chollet, 2015). In our training, the "steps per epoch" ranged from 50 to 150. Epochs for training were fixed to 300, considering an accuracy/time-cost trade. To avoid overfit and further efficient training, we adopt an early stopping callback, which stops training when a monitoring metric has stopped improving (Chollet, 2015). In this study, the monitored metric was validation loss during training. Training was stopped when it had not been improved within the last 20 epochs and the weight from the best epoch was restored. We used the Adam optimization algorithm (Kingma and Ba, 2014) for the training of our network with the learning rate of 1e−6.

We evaluated two training techniques: fine-tuning and the use of a backbone network in the encoder (Table 1). Fine-tuning is an approach that trains with a pre-trained model's weights on new data (Hinton and Salakhutdinov, 2006). The backbone is the recognized architecture or network used for feature extraction (Elharrouss et al., 2022). Both techniques have been used to obtain higher classification accuracy. In this study, we verified two backbones: ResNet18 and ResNet50 (He et al., 2016). As a result, we compared 27 models with varying training steps, model architectures, and training techniques (Table 1).

The performance of the trained network was evaluated by the loss function and two metric functions: the dice loss (Milletari et al., 2016) and binary accuracy. The loss function is a function that calculates gaps between facts and predictions. In this study, we use binary cross entropy in the Keras library for the loss function. This loss function was used in training. The dice loss is a common metric that optimizes networks based on the dice overlap coefficient between the predicted segmentation result (i.e., predicted regions) and the ground truth annotation (i.e., hand-masked), which can solve the data imbalance problem (Milletari et al., 2016). The binary accuracy is the fraction of correctly classified pixels in the image.

The quantitative and qualitative evaluation of our models was performed using the validation data and five test images. The validation data were used to evaluate the network after each epoch of the training. Test images were our holdings which were not included in the image dataset used for training (i.e., not used for training and validation datasets). The evaluation of test image predictions was made from the four points of view: exclusions of sky, vegetation, artifacts, and talus.

Result

Summary of the training

The loss function, dice loss, and accuracy of the epoch in our training are shown in Figures 2–7. The steps, the training stopped epoch, losses, dice losses, and accuracies of training and validation at the end of the training, as well as the duration, are shown in Table 1. We trained networks several times and confirmed that the corresponding changes in binary accuracy were negligible. In all training processes, the early stopping function interrupted the training before the 300th epoch due to a lack of improvement for

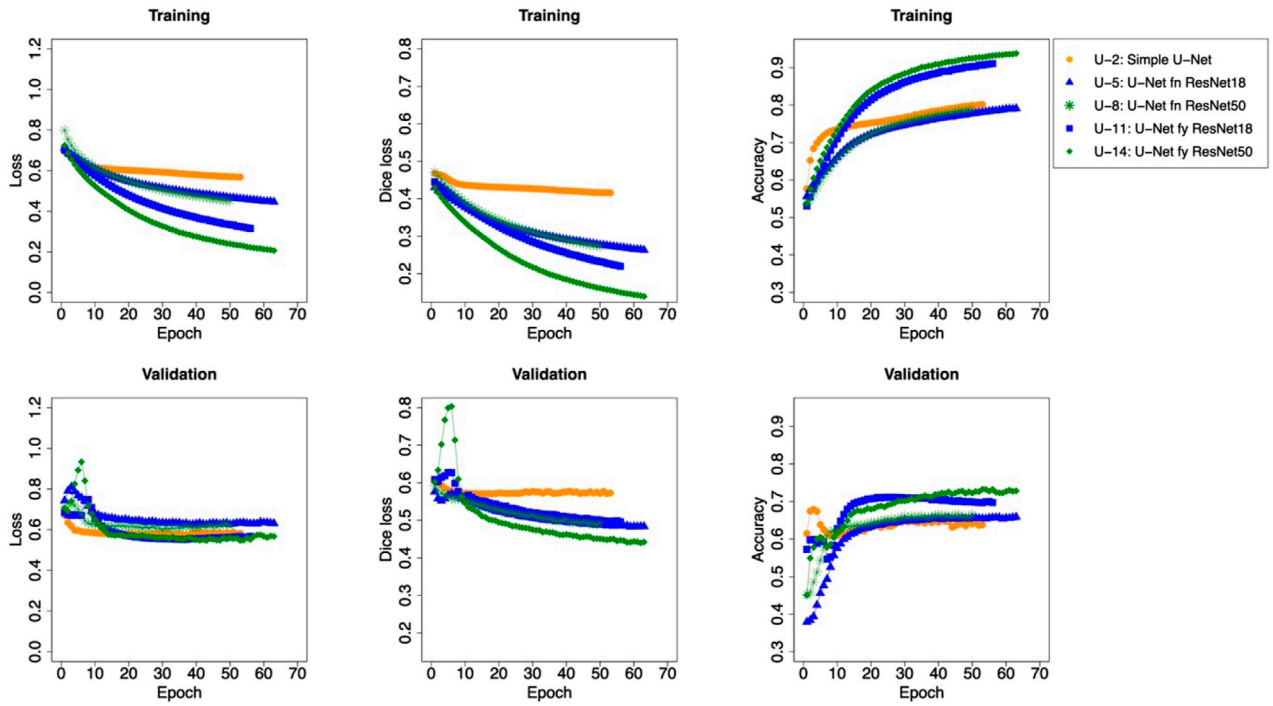


FIGURE 3
Loss, dice loss, and accuracy plots and curves during training for U-Net models with a setting of 100 steps per epoch.

the validation loss within the last 20 epochs. Training of some LinkNet models (IDs: L-1, L-2, L-3, L-4, L-7, and L-8) failed to minimize the validation loss and stopped before the 30th epoch. These indicate training failures. The largest epoch was 138 in the 50-step U-Net training with fine-tuning and a ResNet50 backbone (ID: U-13) and in the 150-step LinkNet training with fine-tuning and a ResNet18 backbone (ID: L-9).

In our computer system, the duration per epoch in the case of the simple U-Net training took ~6 times longer than that of U-Net and LinkNet training with fine-tuning and/or backbones (8.83 min per epoch, Table 1).

The three best validation accuracies were obtained in U-Net training with fine-tuning and a ResNet50 backbone (0.748, 0.733, and 0.728 for model ID U-15, U-13, and U-14, respectively, Table 1). Training failed models (IDs: L-1, L-2, L-3, L-4, L-7, and L-8) showed low validation accuracies (<0.62). For the successfully trained models, the validation accuracy has a 10%–20% gap to the training validation. In many cases, the models with higher "steps per epoch" show higher validation accuracies. Models ResNet50 backbone implemented have higher validation accuracies relative to those ResNet18 backbone implemented. Fine-tuned models show higher validation accuracies than those of non-fine-tuned models.

Predicted stratigraphic exposure regions by trained networks

We verified the fidelity of prediction (masking regions of stratigraphic exposures) using test images that were not used in both training and validation (Figures 8, 9). Training-failed models (IDs: L-1, L-2, L-3, L-4, L-7, and L-8) showed poor predictions. The models U-Net trained show higher fidelities of prediction, especially the exclusion of the sky than those that are LinkNet-trained. This fidelity is higher in models trained with higher steps. For vegetation, artifacts, and talus, those exclusions by our model were incomplete; predicted regions as stratigraphic exposures include them. For vegetation, incomplete extraction often occurred in denser regions. It is common to both U-Net- and LinkNet-trained models than models trained with fine-tuning, and a ResNet50 backbone showed better fidelities of the prediction.

Discussion

Increment of the prediction fidelity

Since our trained network extracts stratigraphic exposures incompletely, the training procedure should be reconsidered. Considering the stability of validation accuracy and the fidelity of predicted images, approximately 100-step training is appropriate for the dataset and the training networks used in this study. The higher fidelities of exclusions of the sky relative to vegetation, artifacts, and talus are probably due to their significantly different textures. Our success implies that color is not necessary for those exclusions because the training dataset was prepared as grayscale images. However, to increase the exclusion fidelity of vegetation, artifacts, and talus, the training procedure should be reconsidered and improved.

The distinguishing of vegetation from stratigraphic exposures is not a difficult task for humans. In general, humans identify

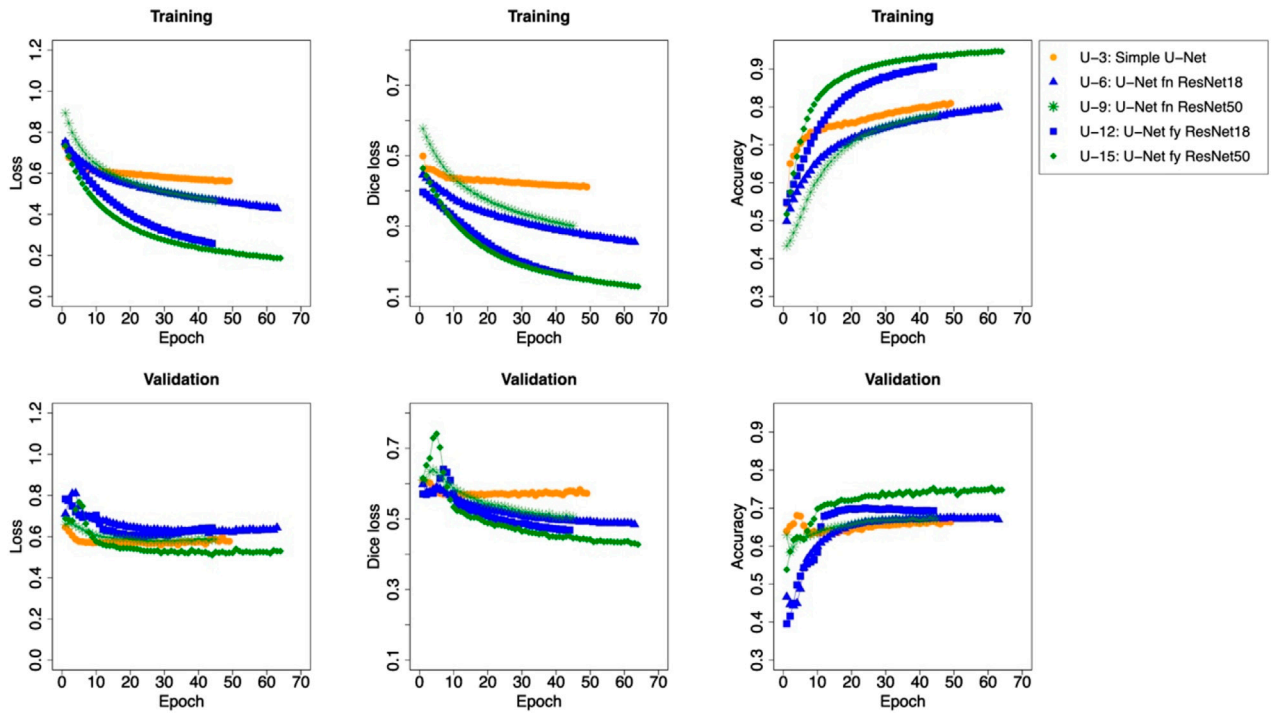


FIGURE 4
Loss, dice loss, and accuracy plots and curves during training for U-Net models with a setting of 150 steps per epoch.

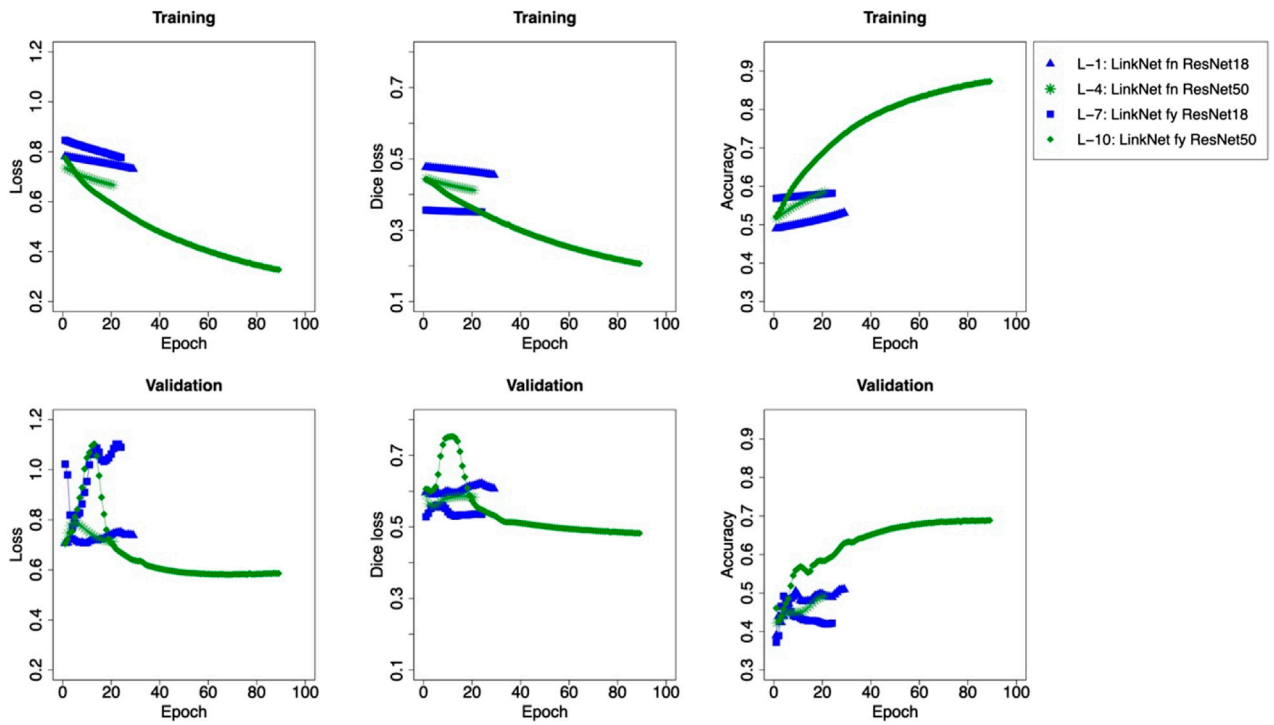


FIGURE 5
Loss, dice loss, and accuracy plots and curves during training for LinkNet models with a setting of 50 steps per epoch.

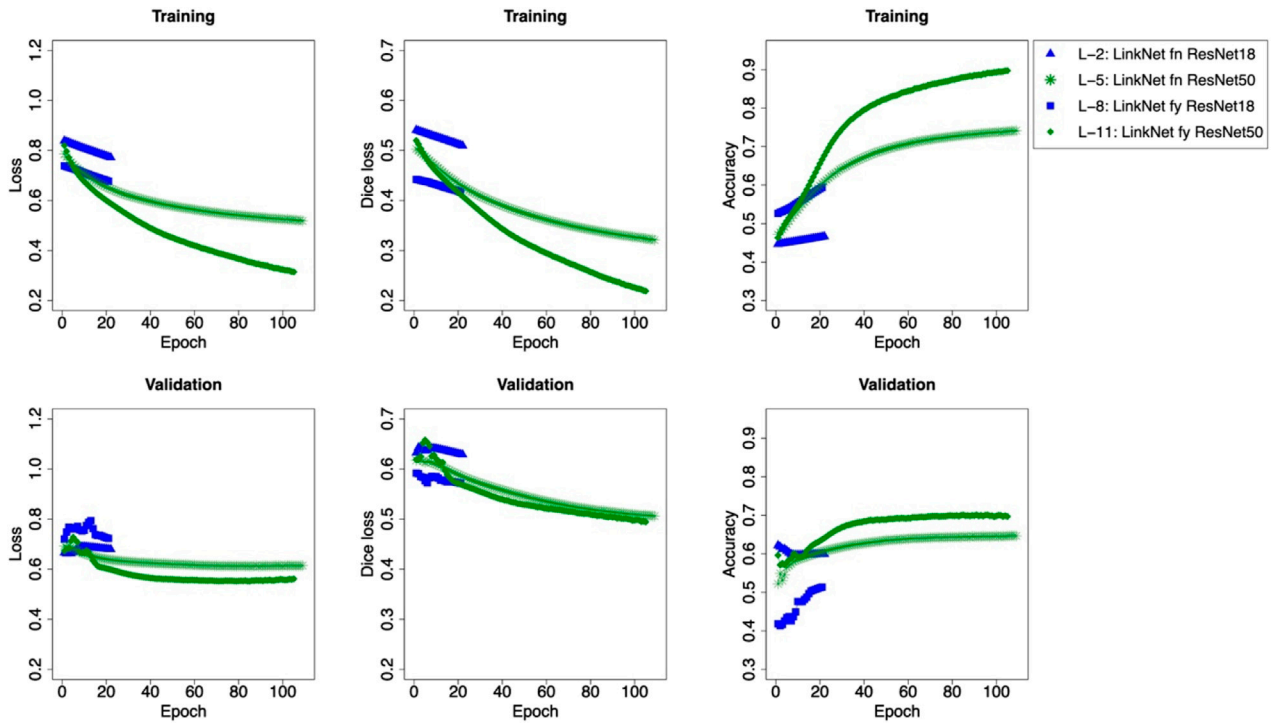


FIGURE 6
Loss, dice loss, and accuracy plots and curves during training for LinkNet models with a setting of 100 steps per epoch.

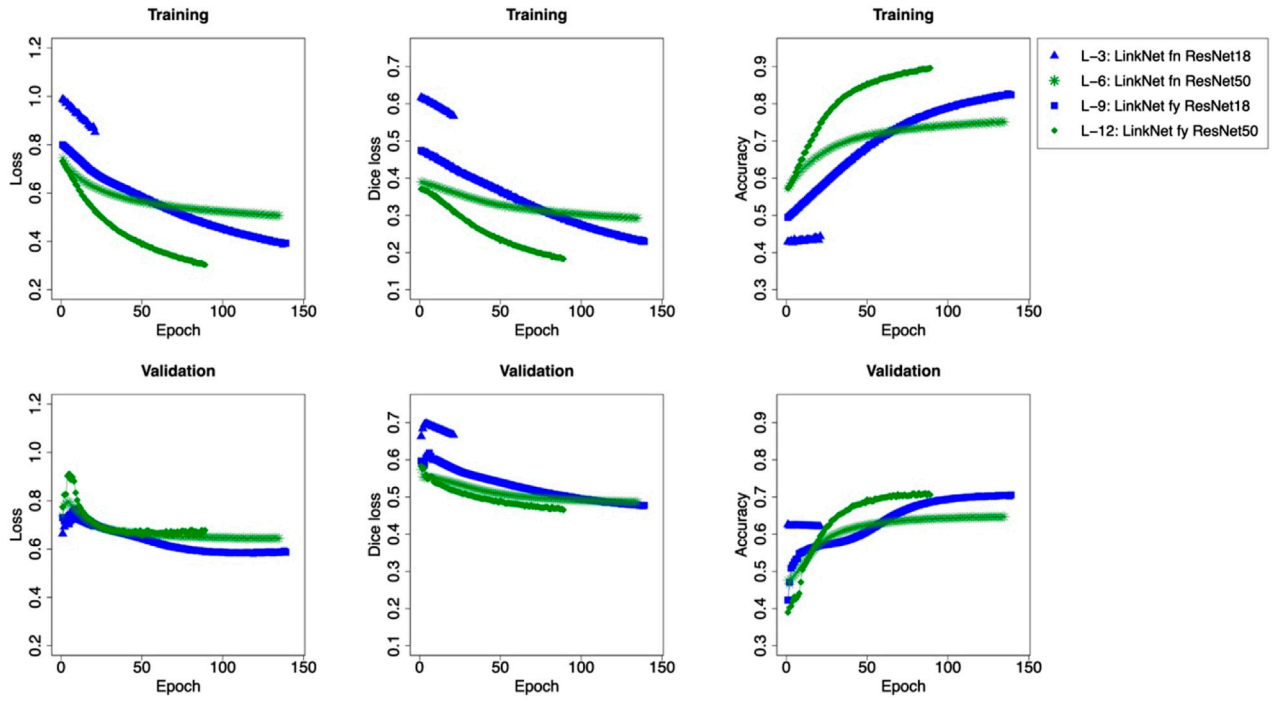


FIGURE 7
Loss, dice loss, and accuracy plots and curves during training for LinkNet models with a setting of 150 steps per epoch.

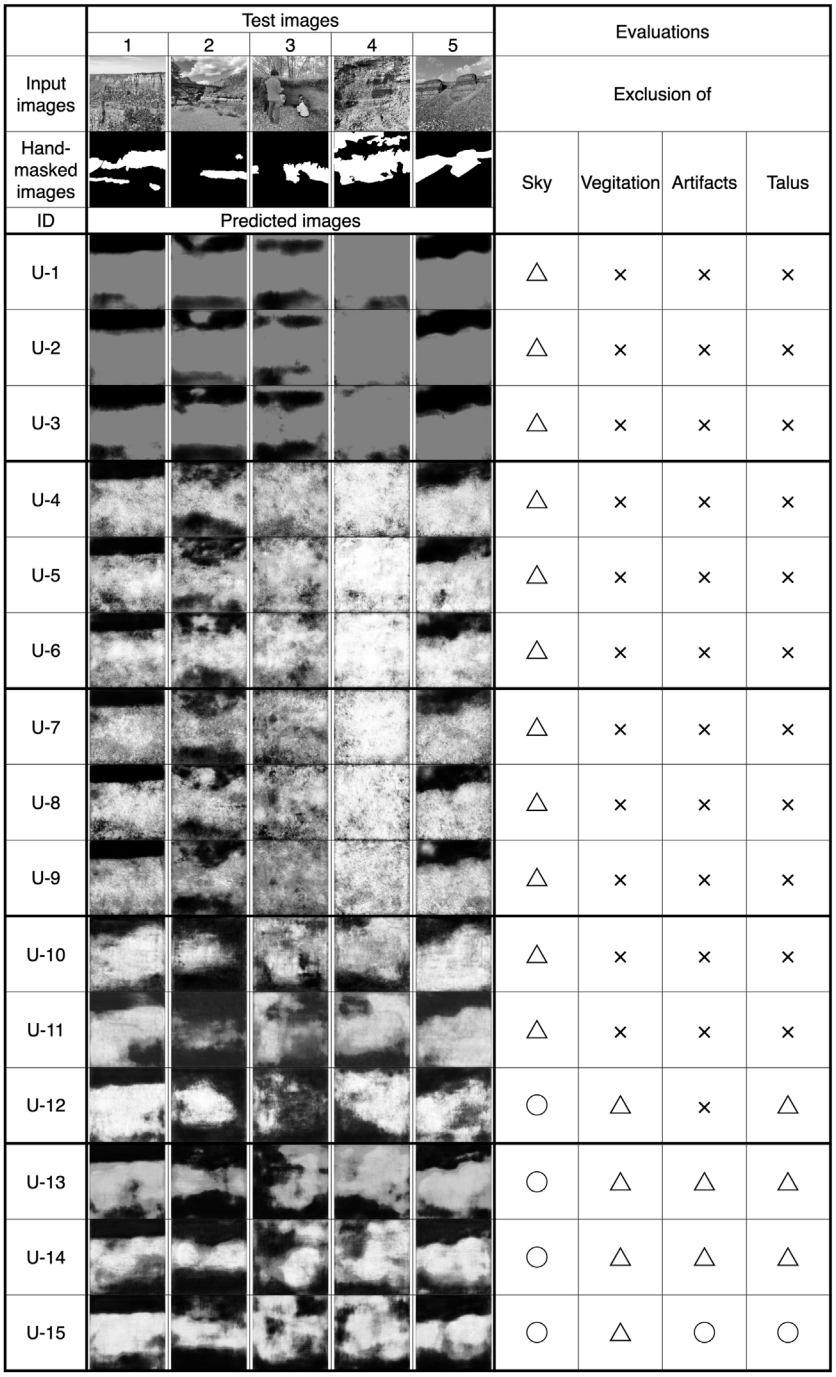


FIGURE 8
Predicted results for test images using U-Net trained models.

vegetation as the accumulation of elongated/oval greenish/brownish objects like leaves, stems, branches, trunks, and roots. Our training dataset was prepared as grayscale images, and the network should learn the exclusion of vegetation by the difference of texture, not by the color difference. The 256 x 256 pixels of augmented images were considered to have a poor resolution for this texture-only-guided distinguishment, although satisfying in the exclusion of the sky. Training with higher resolution and color images will contribute to identifying vegetation that has several types of texture and color (Sodjinou et al., 2022), although time and computing costs are concerned. Since denser vegetation has higher exclusion difficulty (Figures 8, 9), another idea to increase the prediction fidelity is to include images with dense vegetation in the training.

The distinguishing of talus regions with stratigraphic exposures is often difficult for non-experts. This is because its constituent materials are supplied from upper stratigraphies and are indistinguishable from stratigraphic exposure at the same height/

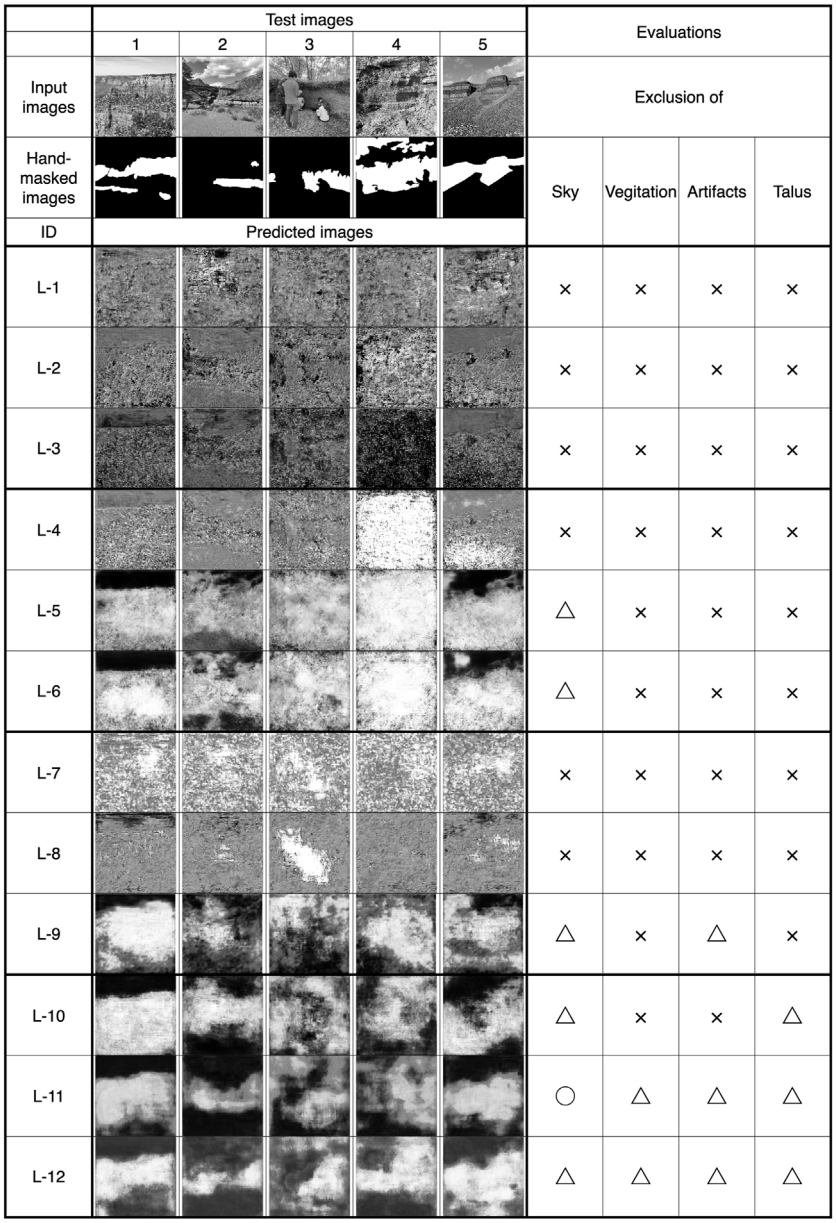


FIGURE 9
Predicted results for test images using LinkNet trained models.

elevation occasionally. It means color/texture analyses will have less contribution to those discriminations, and our strategies for talus exclusion using U-Net and LinkNet architectures, identifying materials as a region, could be reasonable. To increase the talus exclusion fidelity, training with higher-resolution augmented images in which stratigraphic layering can be identified could help contribute toward fidelity.

A stricter masking of stratigraphic exposures will contribute to increasing prediction fidelity, although its time and effort consumptions also increase. Further investigation and verification of reasonable training settings and network structures for obtaining higher prediction fidelities in lower time and effort costs are required.

Strategies for subsequent processes

After the suggestion of stratigraphic exposure regions, the further contribution of computing will be the identification/discrimination of each layer. The interface of each layer is a drastic change in constituent materials. Those changes appear as differences in texture and color in visible images. Since color contains unexpected changes such as wetness and shadow, layer discrimination should also use texture information. Evidently, deep convolutional neural networks are one of the most powerful solutions, as displayed in control and trajectory planning for automated vehicles (e.g., Notle et al., 2018; Dewangan and Sahu, 2021). As a non-deep learning approach, the gray-level co-occurrence matrix (Haralick et al., 1973) and other methods/combinations

(e.g., Armi and Fekri-Ershad, 2019) will contribute to texture-based layer discrimination. Using layer-discriminated (i.e., boundary-drawn) images, we can calculate the thickness of each layer, although the actual scale input (and strictly, strike and dip) is necessary. Furthermore, this kind of texture analysis will provide brief information for constituent material (e.g., lava or pyroclast, lapilli, or ash).

The autonomous measurement of each layer thickness greatly contributes toward decreasing time and effort costs. Although strike and dip should be considered especially on deformed outcrops, the shortest distance between two layer boundaries on scaled front-side images corresponds to layer thickness. The autonomous calculation of those distances can be used for the automatic drawing of stratigraphic columns generally produced in geological surveys. Since the measurement of each layer by hand takes time and has difficulty in unreachable heights, automation helps the researcher in both saving time and effort.

Application to satellite/aerial terrestrial/extraterrestrial images

The automatic identification of stratigraphic exposures would have proven its worth in satellite and aerial images since they often comprise huge datasets. Combination with geological information system tools will contribute to suggesting the locations of outcrops with coordinate values. However, the network of stratigraphic exposure identification should be trained with images that have the same scaling (resolution) as that of target datasets. In this study, our network was trained by outcrop images taken from the ground view; its use may not work for satellite images and aerial photographs. For the use of satellite/aerial images, the training dataset should also have consisted of those images.

The difficulty of the autonomous identification of stratigraphic exposure on extraterrestrial outcrops will be lower than that on terrestrial outcrops because of the lack of vegetation on those bodies. Similar to the terrestrial case, tuned training using images taken on each target body is necessary. Since the data volume obtained on extraterrestrials reaches challenging amounts for remote sensing analysis as mentioned for Mars by Palafox et al. (2017), our improved scheme will be a powerful tool for geological surveys on other bodies.

Conclusion

The automatic extraction of stratigraphic exposure in visible images using a trained network will play an important role in the lower time/effort costs during geological surveys. In this work, we trained U-Net and LinkNet, with fine-tuning and backbones, and demonstrated the successful exclusion of the sky and clouds and the difficulties for those of vegetation, artifacts, and talus. Considering the stability of validation accuracy, the fidelity of predicted images, and time/computing costs, approximately 100-step training is appropriate for the dataset and architectures used in this study. Further surveys of reasonable training settings, network architectures, and techniques for obtaining higher prediction fidelities in lower time and effort costs are necessary. In this study, we presented the usability of image segmentation algorithms in the observation and description of geological outcrops. Such approaches could contribute to passing accumulated knowledge on to further generations. Our improved model will enhance the output from the vast collection of remote sensing images obtained not only on Earth but also on other planetary bodies, such as Mars.

Data availability statement

The datasets presented in this study can be found in online repositories at: https://doi.org/10.5281/zenodo.8396332.

Author contributions

RN: conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, resources, software, supervision, validation, visualization, writing–original draft, and writing–review and editing. DS: methodology, software, validation, writing–original draft, and writing–review and editing.

Funding

The author(s) declare that the financial support was received for the research, authorship, and/or publication of this article. RN was supported by Kakenhi grant no. 22K14082.

Acknowledgments

The authors would like to thank the developers of *google-image-download* and *labelme*. Discussions with Nobuo Geshi, Motomaro Shirao, Keiichiro Fujimoto, and Junichi Haruyama were greatly helpful.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/feart.2023.1264701/full#supplementary-material

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., et al. (2016). *Tensorflow: large-scale machine learning on heterogeneous distributed systems*. arXiv preprint arXiv:1603.04467.

Armi, L., and Fekri-Ershad, S. (2019). *Texture image analysis and texture classification methods-A review*. arXiv preprint arXiv:1904.06554.

Bonadonna, C., Cioni, R., Costa, A., Druitt, T., Phillips, J., Pioli, L., et al. (2016). MeMoVolc report on classification and dynamics of volcanic explosive eruptions. *Bull. Volcanol.* 78, 84. doi:10.1007/s00445-016-1071-y

Chaurasia, A., and Culurciello, E. (2017). "Linknet: exploiting encoder representations for efficient semantic segmentation," in 2017 IEEE visual communications and image processing (VCIP) (IEEE), 1–4.

Chollet, F. (2015). Keras. Available at: https://keras.io.

Dewangan, D. K., and Sahu, S. P. (2021). RCNet: road classification convolutional neural networks for intelligent vehicle system. *Intell. Serv. Robot.* 14 (2), 199–214. doi:10.1007/s11370-020-00343-6

Durig, T., Ross, P. S., Dellino, P., White, J. D. L., Mele, D., and Comida, P. P. (2021). A review of statistical tools for morphometric analysis of juvenile pyroclasts. *Bull. Volcanol.* 83, 79. doi:10.1007/s00445-021-01500-0

Elharrouss, O., Akbari, Y., Almaadeed, N., and Al-Maadeed, S. (2022). *Backbones-review: feature extraction networks for deep learning and deep reinforcement learning approaches. arXiv preprint arXiv:2206.08016*.

Haralick, R. M., Shanmugam, K., and Dinstein, I. H. (1973). Textural features for image classification. *IEEE Trans. Syst. man, Cybern.* (6), 610–621. doi:10.1109/tsmc.1973.4309314

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 770–778.

Hinton, G. E., and Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *science* 313 (5786), 504–507. doi:10.1126/science.1127647

Iakubovskii, P. (2019). Segmentation models GitHub repository. Available at: https://github.com/qubvel/segmentation_models.

Kingma, D. P., and Ba, J. (2014). *Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980*.

Kueppers, U., Perugini, D., and Dingwell, D. B. (2006a). Explosive energy during volcanic eruptions from fractal analysis of pyroclasts. *Earth Planet. Sci. Lett.* 248 (3-4), 800–807. doi:10.1016/j.epsl.2006.06.033

Kueppers, U., Scheu, B., Spieler, O., and Dingwell, D. B. (2006b). Fragmentation efficiency of explosive volcanic eruptions: A study of experimentally generated pyroclasts. *J. Volcanol. Geotherm. Res.* 153 (1-2), 125–135. doi:10.1016/j.jvolgeores.2005.08.006

Latorre, F., Spiller, D., Sasidharan, S. T., Basheer, S., and Curti, F. (2023). Transfer learning for real-time crater detection on asteroids using a Fully Convolutional Neural Network. *Icarus* 394, 115434. doi:10.1016/j.icarus.2023.115434

Leibrandt, S., and Le Pennec, J. L. (2015). Towards fast and routine analyses of volcanic ash morphometry for eruption surveillance applications. *J. Volcanol. Geotherm. Res.* 297, 11–27. doi:10.1016/j.jvolgeores.2015.03.014

Liu, E. J., Cashman, K. V., and Rust, A. C. (2015). Optimising shape analysis to quantify volcanic ash morphology. *GeoResJ* 8, 14–30. doi:10.1016/j.grj.2015.09.001

Miwa, T., Shimano, T., and Nishimura, T. (2015). Characterization of the luminance and shape of ash particles at Sakurajima volcano, Japan, using CCD camera images. *Bull. Volcanol.* 77, 5–13. doi:10.1007/s00445-014-0886-7

Nolte, M., Kister, N., and Maurer, M. (2018). "Assessment of deep convolutional neural networks for road surface classification," in 2018 21st International Conference on Intelligent Transportation Systems (ITSC) (IEEE), 381–386.

Palafox, L. F., Hamilton, C. W., Scheidt, S. P., and Alvarez, A. M. (2017). Automated detection of geological landforms on Mars using convolutional neural networks. *Comput. geosciences* 101, 48–56. doi:10.1016/j.cageo.2016.12.015

Perugini, D., Speziali, A., Caricchi, L., and Kueppers, U. (2011). Application of fractal fragmentation theory to natural pyroclastic deposits: insights into volcanic explosivity of the Valentano scoria cone (Italy). *J. Volcanol. Geotherm. Res.* 202 (3-4), 200–210. doi:10.1016/j.jvolgeores.2011.02.008

Ronneberger, O., Fischer, P., and Brox, T. (2015). "U-net: convolutional networks for biomedical image segmentation," in International Conference on Medical image computing and computer-assisted intervention, Cham (Springer), 234–241.

Shoji, D., Noguchi, R., Otsuki, S., and Hino, H. (2018). Classification of volcanic ash particles using a convolutional neural network and probability. *Sci. Rep.* 8 (1), 8111. doi:10.1038/s41598-018-26200-2

Shozaki, H., Sekine, Y., Guttenberg, N., and Komatsu, G. (2022). Recognition and classification of Martian chaos terrains using imagery machine learning: A global distribution of chaos linked to groundwater circulation, catastrophic flooding, and magmatism on Mars. *Remote Sens.* 14 (16), 3883. doi:10.3390/rs14163883

Silburt, A., Ali-Dib, M., Zhu, C., Jackson, A., Valencia, D., Kissin, Y., et al. (2019). Lunar crater identification via deep learning. *Icarus* 317, 27–38. doi:10.1016/j.icarus.2018.06.022

Smith, Z. D., and Maxwell, D. J. (2021). Constructing vertical measurement logs using UAV-based photogrammetry: applications for multiscale high-resolution analysis of coarse-grained volcaniclastic stratigraphy. *J. Volcanol. Geotherm. Res.* 409, 107122. doi:10.1016/j.jvolgeores.2020.107122

Sodjinou, S. G., Mohammadi, V., Mahama, A. T. S., and Gouton, P. (2022). A deep semantic segmentation-based algorithm to segment crops and weeds in agronomic color images. *Inf. Process. Agric.* 9 (3), 355–364. doi:10.1016/j.inpa.2021.08.003

Stepinski, T. F., Mendenhall, M. P., and Bue, B. D. (2009). Machine cataloging of impact craters on Mars. *Icarus* 203 (1), 77–87. doi:10.1016/j.icarus.2009.04.026

Wagstaff, K. L., Daubar, I. J., Doran, G., Munje, M. J., Bickel, V., Gao, A., et al. (2022). Using machine learning to reduce observational biases when detecting new impacts on Mars. *Icarus* 386, 115146. doi:10.1016/j.icarus.2022.115146

Wieland, M., Li, Y., and Martinis, S. (2019). Multi-sensor cloud and cloud shadow segmentation with a convolutional neural network. *Remote Sens. Environ.* 230, 111203. doi:10.1016/j.rse.2019.05.022

Wohletz, K., and Heiken, G. (1992). *Volcanology and geothermal energy*, 432. Berkeley: University of California Press.

Zhu, Q., Zhang, Y., Wang, L., Zhong, Y., Guan, Q., Lu, X., et al. (2021). A global context-aware and batch-independent network for road extraction from VHR satellite imagery. *ISPRS J. Photogrammetry Remote Sens.* 175, 353–365. doi:10.1016/j.isprsjprs.2021.03.016

frontiers | Frontiers in Earth Science

TYPE Correction
PUBLISHED 31 January 2024
DOI 10.3389/feart.2023.1339426

OPEN ACCESS

APPROVED BY
Frontiers Editorial Office,
Frontiers Media SA, Switzerland

*CORRESPONDENCE
Rina Noguchi,
r-noguchi@env.sc.niigata-u.ac.jp

RECEIVED 16 November 2023
ACCEPTED 27 November 2023
PUBLISHED 31 January 2024

CITATION
Noguchi R and Shoji D (2024), Corrigendum: Extraction of stratigraphic exposures on visible images using a supervised machine learning technique.
Front. Earth Sci. 11:1339426.
doi: 10.3389/feart.2023.1339426

Corrigendum: Extraction of stratigraphic exposures on visible images using a supervised machine learning technique

Rina Noguchi[1]* and Daigo Shoji[2]

[1]Faculty of Science, Niigata University, Niigata, Japan, [2]Institute of Space and Astronautical Science, Japan Aerospace Exploration Agency, Sagamihara, Kanagawa, Japan

KEYWORDS

machine learning, stratigraphic exposures, volcanic stratigraphies, autonomous extraction, U-Net, LinkNet

A Corrigendum on
Extraction of stratigraphic exposures on visible images using a supervised machine learning technique

by Noguchi R and Shoji D (2023). Front. Earth Sci. 11:1264701. doi: 10.3389/feart.2023.1264701

In the published article, there was an error in the legend for Figure 2 as published. The symbols were attributed to the incorrect IDs. The blue triangle—U-Net fn ResNet18—was described as U-7 but should have been U-4. The green star—U-Net fn ResNet50 - was described as U-13 but should have been U-7. The blue square—U-Net fy ResNet18 was described as U-4 but should have been U-10. The green diamond—U-Net fy ResNet50—was described as U-10 but should have been U-13. The order has been rearranged into the numerical order of the IDs. The correct figure and legend appears below.

In the published article, there was an error in the legend for Figure 3 as published. The symbols were attributed to the incorrect IDs. The blue triangle—U-Net fn ResNet18—was described as U-8 but should have been U-5. The green star—U-Net fn ResNet50 - was described as U-14 but should have been U-8. The blue square—U-Net fy ResNet18 was described as U-5 but should have been U-11. The green diamond—U-Net fy ResNet50—was described as U-11 but should have been U-14. The order has been rearranged into the numerical order of the IDs. The correct figure and legend appears below.

In the published article, there was an error in the legend for Figure 4 as published. The symbols were attributed to the incorrect IDs. The blue triangle—U-Net fn ResNet18—was described as U-9 but should have been U-6. The green star—U-Net fn ResNet50—was described as U-15 but should have been U-9. The blue square—U-Net fy ResNet18 was described as U-6 but should have been U-12. The green diamond—U-Net fy ResNet50—was described as U-12 but should have been U-15. The order has been rearranged into the numerical order of the IDs. The correct figure and legend appears below.

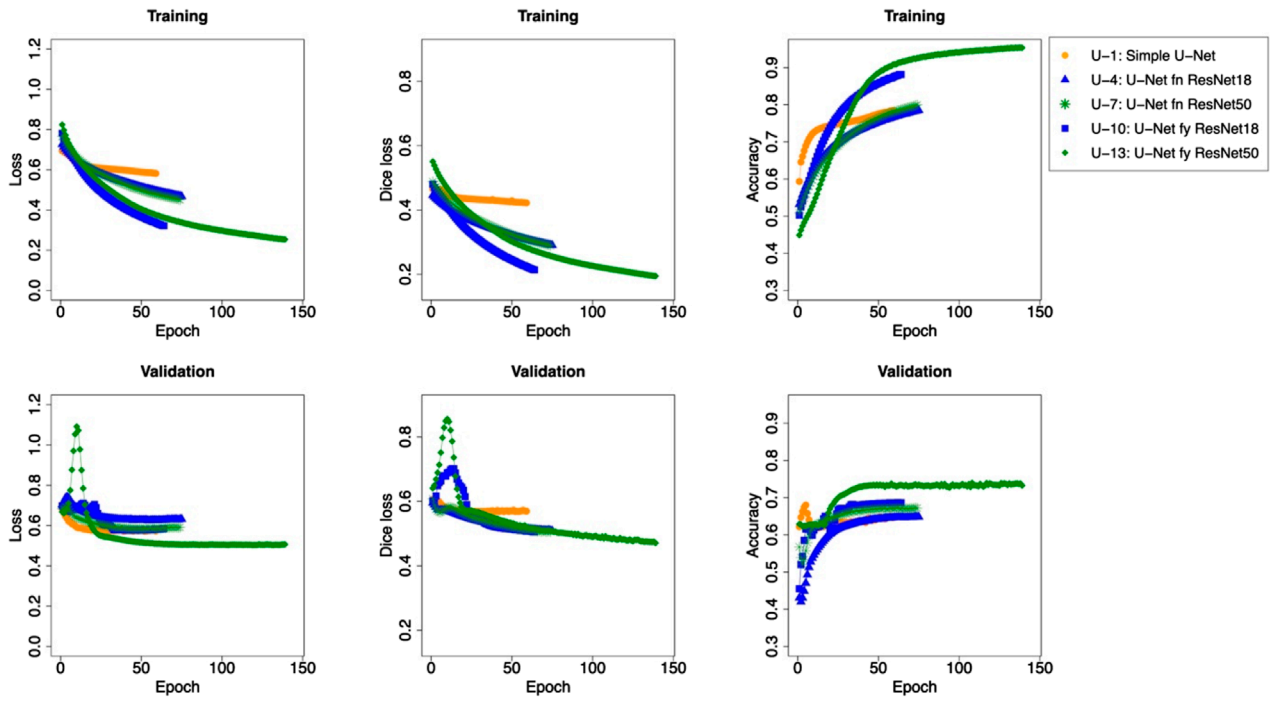


FIGURE 2
Loss, dice loss, and accuracy plots and curves during training for U-Net models with a setting of 50 steps per epoch.

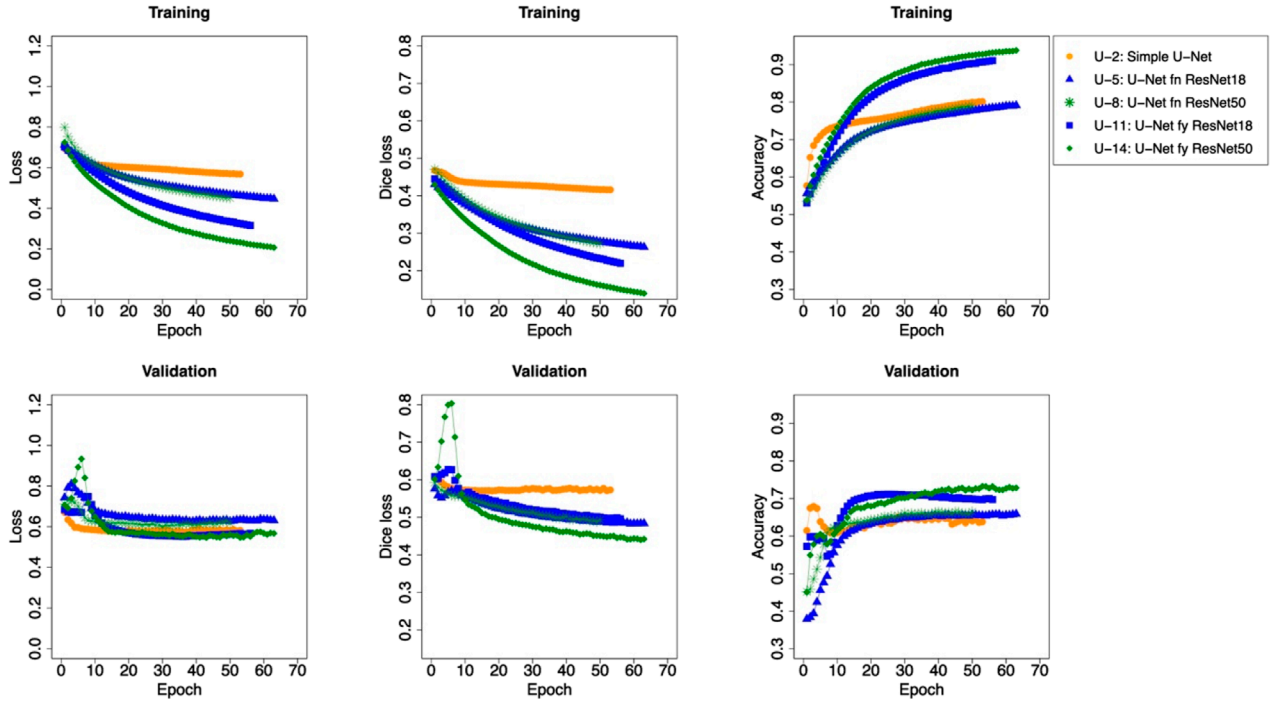


FIGURE 3
Loss, dice loss, and accuracy plots and curves during training for U-Net models with a setting of 100 steps per epoch.

In the published article, there was an error in the legend for Figure 5 as published. The symbols were attributed to the incorrect IDs. The blue triangle–LinkNet fn ResNet18 - was described as L-4 but should have been L-1. The green star—LinkNet fn ResNet50—was described as L-10 but should have been L-4. The blue square—LinkNet fy ResNet18 was described as L-1 but should have been L-7. The green diamond—LinkNet fy ResNet50—was described as L-7 but should have been L-10. The order has been rearranged into the numerical order of the IDs. The correct figure and legend appears below.

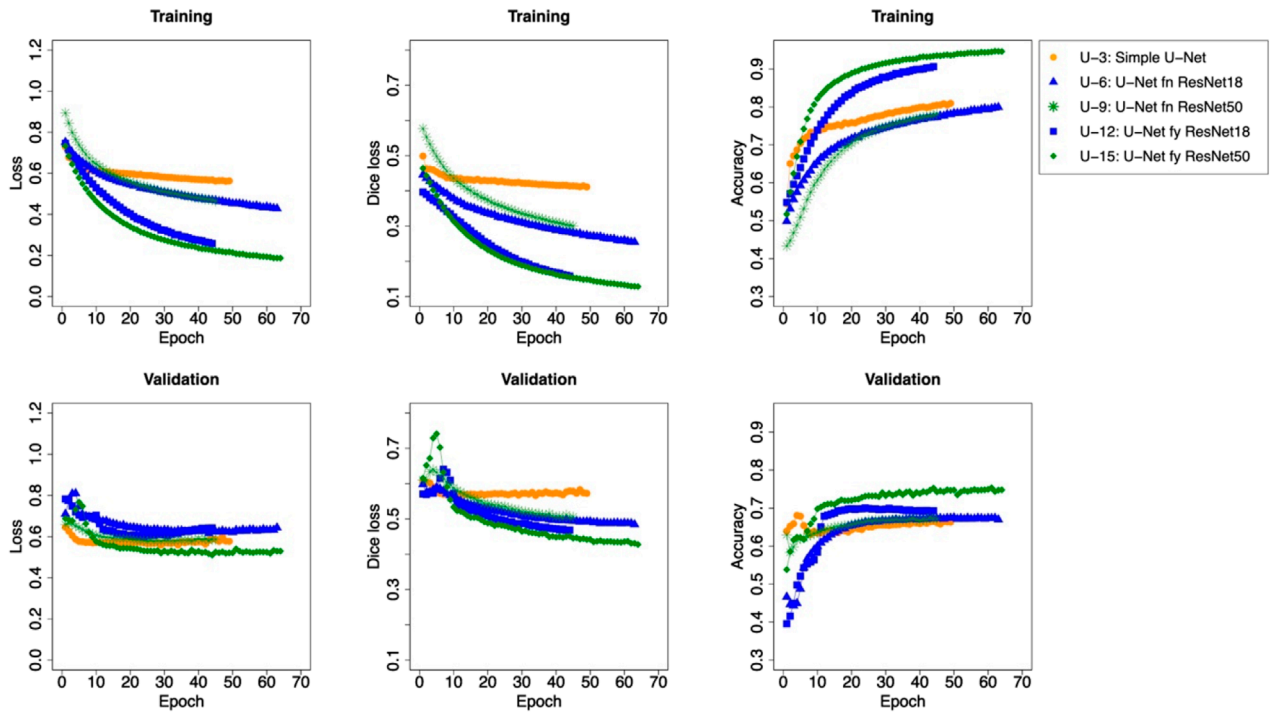


FIGURE 4
Loss, dice loss, and accuracy plots and curves during training for U-Net models with a setting of 150 steps per epoch.

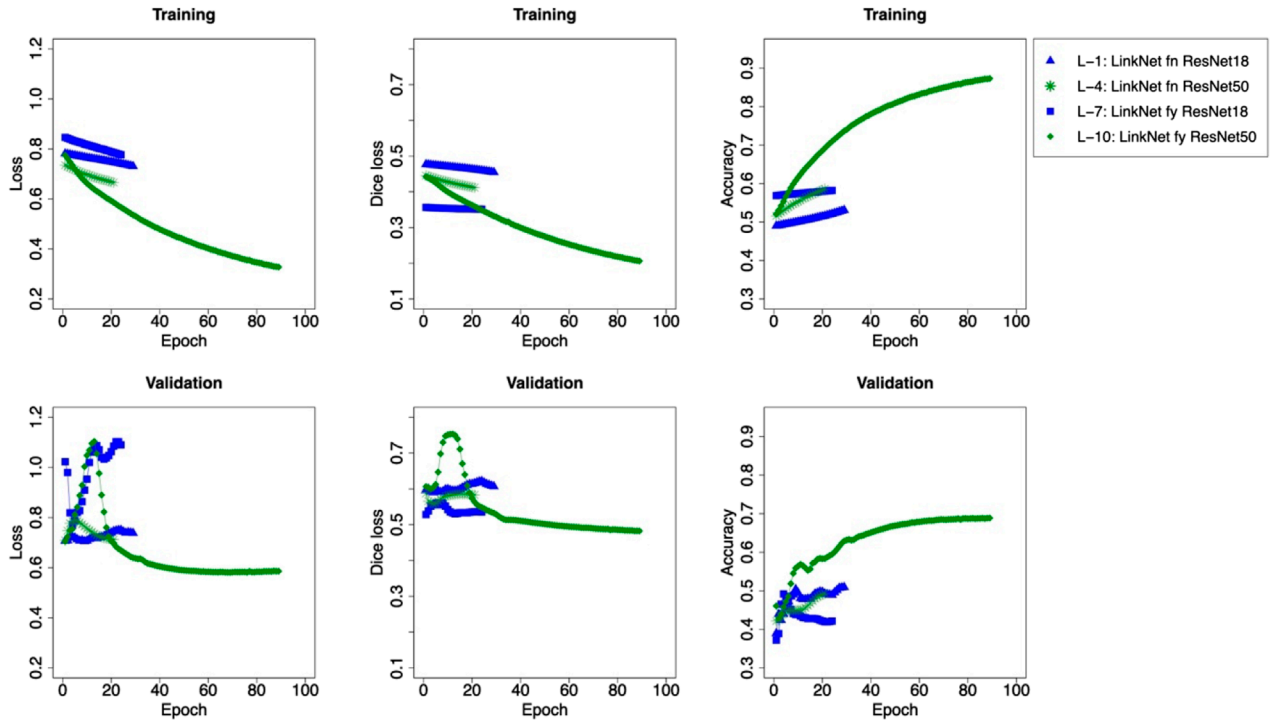


FIGURE 5
Loss, dice loss, and accuracy plots and curves during training for LinkNet models with a setting of 50 steps per epoch.

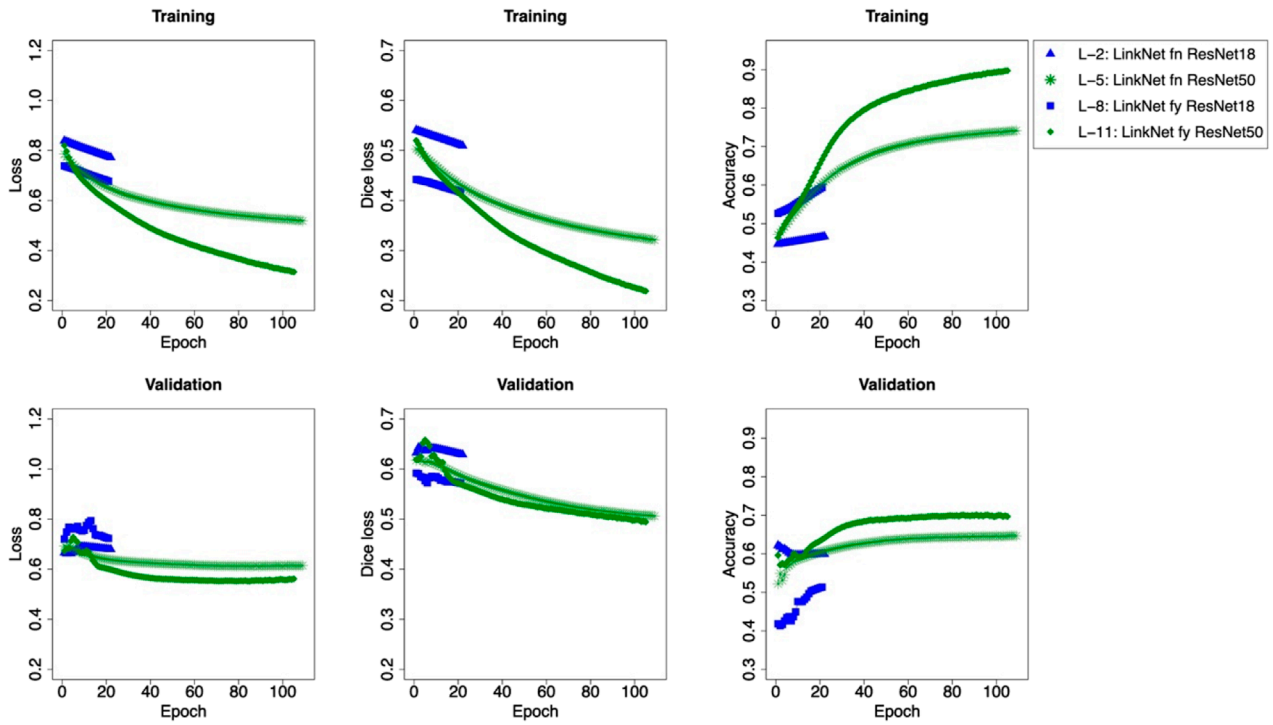


FIGURE 6
Loss, dice loss, and accuracy plots and curves during training for LinkNet models with a setting of 100 steps per epoch.

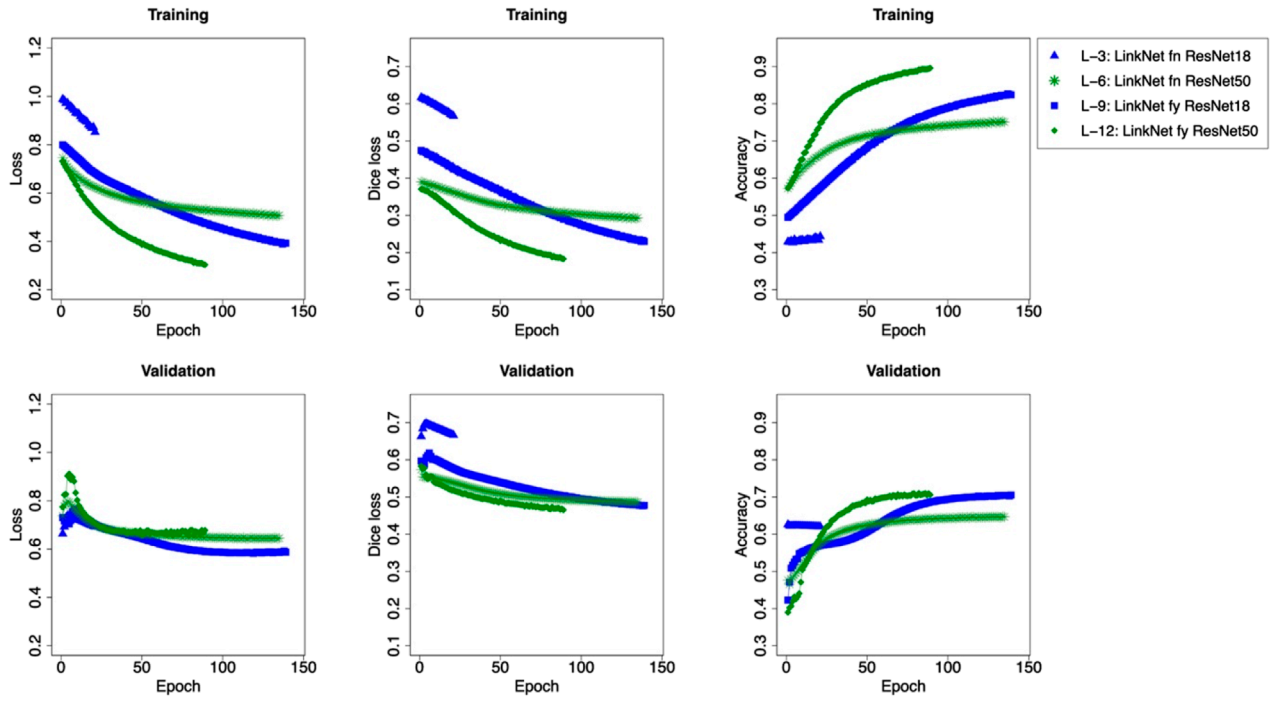


FIGURE 7
Loss, dice loss, and accuracy plots and curves during training for LinkNet models with a setting of 150 steps per epoch.

TABLE 1 The loss, dice, and binary accuracy at the end of training in each model.

ID	Model	Fine-tuning	Backbone	Step	Training stopped epoch	Training			Validation			Duration (min)	
						Loss	Dice loss	Accuracy	Loss	Dice loss	Accuracy	Total	Per epoch
U-1	U-Net	No	No	50	58	0.582	0.422	0.785	0.573	0.569	0.646	512	8.83
U-2				100	52	0.568	0.416	0.802	0.582	0.573	0.638	298	5.73
U-3				150	48	0.563	0.411	0.810	0.577	0.573	0.663	282	5.88
U-4		No	Yes (ResNet18)	50	74	0.468	0.290	0.785	0.632	0.509	0.648	90	1.22
U-5				100	62	0.446	0.263	0.791	0.631	0.484	0.659	70	1.13
U-6				150	62	0.428	0.255	0.799	0.644	0.485	0.670	76	1.23
U-7		No	Yes (ResNet50)	50	73	0.450	0.289	0.800	0.591	0.506	0.672	121	1.66
U-8				100	49	0.451	0.273	0.790	0.616	0.493	0.665	78	1.59
U-9				150	44	0.465	0.300	0.781	0.587	0.506	0.671	76	1.73
U-10		Yes	Yes (ResNet18)	50	63	0.322	0.214	0.882	0.584	0.505	0.686	76	1.21
U-11				100	55	0.315	0.220	0.911	0.567	0.496	0.697	61	1.11
U-12				150	43	0.258	0.158	0.906	0.643	0.469	0.693	53	1.23
U-13		Yes	Yes (ResNet50)	50	138	0.254	0.195	0.954	0.507	0.471	0.733	229	1.66
U-14				100	62	0.206	0.139	0.938	0.567	0.442	0.728	99	1.60
U-15				150	63	0.187	0.128	0.947	0.529	0.428	0.748	108	1.71
L-1	LinkNet	No	Yes (ResNet18)	50	28	0.731	0.455	0.531	0.738	0.607	0.509	17	0.61
L-2				100	21	0.772	0.510	0.467	0.680	0.629	0.599	14	0.67
L-3				150	20	0.852	0.567	0.443	0.692	0.667	0.620	15	0.75
L-4		No	Yes (ResNet50)	50	20	0.666	0.412	0.586	0.712	0.583	0.493	30	1.50
L-5				100	108	0.519	0.321	0.742	0.615	0.506	0.646	146	1.35
L-6				150	134	0.508	0.293	0.751	0.645	0.486	0.647	190	1.42
L-7		Yes	Yes (ResNet18)	50	23	0.777	0.351	0.582	1.090	0.535	0.421	14	0.61
L-8				100	20	0.677	0.417	0.593	0.723	0.572	0.514	13	0.65
L-9				150	138	0.393	0.230	0.824	0.587	0.478	0.705	99	0.72
L-10		Yes	Yes (ResNet50)	50	88	0.328	0.206	0.873	0.586	0.483	0.689	126	1.43
L-11				100	104	0.314	0.219	0.898	0.562	0.494	0.697	141	1.36
L-12				150	88	0.303	0.183	0.896	0.678	0.465	0.706	128	1.45

In the published article, there was an error in the legend for Figure 6 as published. The symbols were attributed to the incorrect IDs. The blue triangle–LinkNet fn ResNet18—was described as L-5 but should have been L-2. The green star—LinkNet fn ResNet50 - was described as L-11 but should have been L-5. The blue square—LinkNet fy ResNet18 was described as L-2 but should have been L-8. The green diamond—LinkNet fy ResNet50—was described as L-8 but should have been L-11. The order has been rearranged into the numerical order of the IDs. The correct figure and legend appears below.

In the published article, there was an error in the legend for Figure 7 as published. The symbols were attributed to the incorrect IDs. The blue triangle—LinkNet fn ResNet18—was described as L-6 but should have been L-3. The blue square—LinkNet fy ResNet18 was described as L-3 but should have been L-9. The green star—LinkNet fn ResNet50—was described as L-12 but should have been L-6. The green diamond—LinkNet fy ResNet50—was described as L-9 but should have been L-12. The order has been rearranged into the numerical order of the IDs. The correct figure and legend appears below.

In the published article, there was an error in Table 1 as published.

Values for the model ID L-1 and L-7 were incorrect. The corrected Table 1 and its caption appear below.

The authors apologize for these errors and state that this does not change the scientific conclusions of the article in any way. The original article has been updated.

Publisher's note

frontiers | Frontiers in Earth Science

TYPE Original Research
PUBLISHED 04 March 2024
DOI 10.3389/feart.2024.1343874

OPEN ACCESS

EDITED BY
Silvia Scarpetta,
University of Salerno, Italy

REVIEWED BY
Angelo De Santis,
National Institute of Geophysics and Volcanology (INGV), Italy
Melody Whitehead,
Massey University, New Zealand

*CORRESPONDENCE
Sam Mitchinson,
s.mitchinson@uea.ac.uk

RECEIVED 24 November 2023
ACCEPTED 16 February 2024
PUBLISHED 04 March 2024

CITATION
Mitchinson S, Johnson JH, Milner B and Lines J (2024), Identifying earthquake swarms at Mt. Ruapehu, New Zealand: a machine learning approach.
Front. Earth Sci. 12:1343874.
doi: 10.3389/feart.2024.1343874

Identifying earthquake swarms at Mt. Ruapehu, New Zealand: a machine learning approach

Sam Mitchinson[1]*, Jessica H. Johnson[1], Ben Milner[2] and Jason Lines[2]

[1]School of Environmental Sciences, University of East Anglia, Norwich, United Kingdom, [2]School of Computing Sciences, University of East Anglia, Norwich, United Kingdom

Mt. Ruapehu is an active andesitic stratovolcano, consisting of several peaks with the summit plateau at 2,797 m, making it the tallest active volcano in New Zealand. The extent of the volcano spreads 40 km across with a series of complex faults encompassing almost the entire base of the volcano. A series of earthquakes occurring 20 km west of the summit of Mt. Ruapehu, near the small town of Erua, which preceded the 1995/1996 major volcanic eruption sequence has been proposed as a medium-term precursor for eruptions at Mt. Ruapehu. We use unsupervised machine learning clustering algorithms HDBSCAN and DBSCAN to define anomalous earthquake swarms in the region and determine whether the Erua swarm was unique by identifying key characteristics in space, time and magnitude distribution. HDBSCAN found six spatial cluster zones to the west of Mt. Ruapehu, which have temporal seismic bursts of activity between 1994 and 2023. DBSCAN identified the seismic swarm that preceded the 1995/1996 major eruption, along with one other similar cluster in the same region, which did not coincide with any documented magmatic unrest, suggesting distal seismic swarms at Mt. Ruapehu may not serve as a reliable eruption precursor when observed in isolation. We instead found that earthquake swarms are relatively common at Mt. Ruapehu and the temporal evolution of the earthquake clusters west of Mt. Ruapehu share similar characteristics to seismic swarms identified in other settings related to fluid migration, typical of fault-valve models.

KEYWORDS

HDBSCAN, DBSCAN, Ruapehu volcano, unsupervised learning, machine learning, time series, earthquake sequence

1 Introduction

The ability to accurately forecast when volcanic eruptions will occur is of great importance in the field of volcanology. Ground deformation and volcanic seismicity are some of the most reliable precursory signals for forecasting volcanic eruptions, particularly after long periods of repose (Kilburn, 2018). Even subtle changes in the stress experienced by active volcanoes can lead to volcanic earthquakes and, sometimes, eruptions (De La Cruz-Reyna et al., 2010). In a period between 1994 and 1995, a series of earthquakes occurred near the small town of Erua some 20 km from the summit of the active stratovolcano Mt. Ruapehu (Hurst and McGinty, 1999), which forms part of the Tongariro Centre, at the most southwestern point of the The Taupo Volcanic Zone (TVZ) in New Zealand, North Island (Figure 1). Months later, the 1995/1996 major volcanic eruption sequence began

(Bryan and Sherburn, 1999). Petrological studies of the erupted materials have suggested a magma mixing event, which coincides temporally with increased seismicity prior to the eruption (Hurst et al., 2018). The timing and depth of the earthquakes suggest a process of fresh magma being fed into the magma reservoir from a deeper source, 5 months prior to the eruption (Kilgour et al., 2014). It was hypothesised that the influx of magma created a change in stress loading in the neighbouring rock, resulting in the observed earthquake swarm (Hurst et al., 2018), meaning the earthquake swarm may have been a precursor to the volcanic eruption. This led to the suggestion that earthquake swarms at distal faults could serve as useful mid-term forecasts at Mt. Ruapehu (Kilgour et al., 2014; Hurst et al., 2018). Earthquake swarms at distal faults and their relationship with volcanic eruptions may indeed be of scientific relevance (White and McCausland, 2016). For example, in Indonesia, a seismic swarm detected >10 km distance from the Mt. Agung summit, has been deemed significant regarding the lateral migration of magma to the central magma reservoir prior to the eruption (Albino et al., 2019). The swarm occurred 2 months prior to the 2017 eruption, which began with a phreatomagmatic phase before a sustained magmatic eruption (Albino et al., 2019), similar to the 1995 eruption at Mt. Ruapehu (Bryan and Sherburn, 1999). We propose using unsupervised machine learning clustering to answer the question of whether the Erua earthquake swarm could have been a reliable forecasting tool for the major magmatic eruption of Mt. Ruapehu in 1995. To do this, we will statistically, and without bias or *a priori* knowledge, define an anomalous earthquake swarm in the region and determine whether the Erua swarm was unique by identifying key characteristics in space, time and magnitude distribution.

The timely identification of seismic signals have lead to the formation of many successful eruption forecasts in history (McNutt, 2002). Earthquakes in volcanic settings can be the result of different processes. As magma moves below the surface it may exert strain on the brittle surrounding rock causing it to fracture, which is detected as earthquakes (Hill et al., 2002). Similarly, an injection of magma into a crack will cause stress on the neighbouring rock and result in a seismic signal (Chouet and Matoza, 2013). Conversely, tectonic earthquakes causing stress changes may promote the movement of fluids towards the surface (Seropian et al., 2021). The frequency of earthquake occurrences has been shown to increase exponentially when rock is deformed at a constant strain rate (Kilburn, 2012, 2018). These episodes of sudden increases in the number of earthquakes in a region are referred to as earthquake *swarms* and are often associated with volcanic activity (Chouet and Matoza, 2013; White and McCausland, 2016; Pesicek et al., 2018). "Swarm-like" earthquake sequences have been described in the literature as having the following characteristics: 1) earthquakes occur in close proximal space to each other, compared to the more uniform background seismicity, 2) earthquake events occur relatively close together in time at a more frequent rate than the background seismicity, and 3) the largest and second-largest earthquakes usually have a similar magnitude with a maximum difference of around ΔM 0.5, and tend to not follow a typical mainshock-aftershock sequence (Gudmundsson, 2020). Swarm-like sequences also tend to have the largest event occurring during the middle of the sequence (Vidale and Shearer, 2006). This somewhat subjective description of how earthquake swarms are defined can be detrimental for how they can be used for forecasting eruptions. For example, studies have shown that the duration of swarms that result in volcanic activity can vary significantly (McNutt, 2005), making eruption forecasting difficult. Moreover, in many cases, a sudden increase in earthquake activity, even for highly active volcanoes, has not resulted in a volcanic eruption (Ramis et al., 2018; Gudmundsson, 2020). To enhance the precision and consistency of earthquake swarms in volcanic eruption forecasts, it would be advantageous to develop a more objective and statistically-driven methodology for identifying various types of earthquake sequences. Utilising unsupervised clustering algorithms enables the impartial identification of spatio-temporal anomalies in the earthquake catalogue, and their relevance to documented historical eruptions, ultimately allowing us to confirm whether the Eura swarm (e.g., Hurst and McGinty, 1999; Hurst et al., 2018) represents a unique and distinguishable signal.

2 Mount Ruapehu

Mt. Ruapehu is an active andestic stratovolcano, consisting of several peaks with the summit plateau at 2,797 m, making it the tallest active volcano in New Zealand. The extent of the volcano spreads 40 km across from the Raurimo fault to the west and the Rangipo fault to the east (Conway et al., 2016) (Figure 2). The oldest dated eruptive material dates to c.200 ka, although there are parts of the lava formation and clasts which may date back even further (Gamble et al., 2003). Historical eruptive composition is mainly made up of andesites stored from depths of 5–10 km (Leonard et al., 2021), though there is evidence of composition ranging from basaltic andesite to dacite (53–66 wt.% silica) throughout the eruptive history of Mt. Ruapehu (Gamble et al., 2003). The active magma reservoir is coupled with a hydrothermal system (Christenson and Wood, 1993), which feeds radiation and volatile-rich gases into a crater lake at the summit containing c.9x10^6 m^3 of acid-rich water (Manville et al., 2007). The hydrothermal system frequently produces phreatic eruptions, which often lack any precursory warning (Houghton et al., 1987), and can cause the water in the lake to burst its banks causing lahars (e.g., Manville et al., 2007; Schaefer et al., 2018). Other frequently erupted material includes tephra, with sizes ranging from ashfall to clasts and blocks, and block lava flows that have historically flowed from the summit and three flank vents (Houghton et al., 1987).

2.1 Eruption history

The current period of volcanic activity has been ongoing for the last 2 ka., described as periodic low-volume (<0.05 km^3) phreatomagmatic eruptions which occur every 25–30 years (Kilgour et al., 2013; Conway et al., 2016). Since the first historically recorded eruption in 1830, there have been 35 eruptions with a recorded Volcanic Explosivity Index (VEI) >1 (Hurst et al., 2018). Major eruptions have a repose period of around 50 years with the last three occurring in 1895, 1945 and finally 1995/1996. Notably, the 1945 major eruption resulted in the creation of a new lava dome and debris barrier confining the crater lake (Johnston et al., 2000). Crater lake breakouts at Mt. Ruapehu are relatively common and have occurred repeatedly throughout the 20th and 21st century

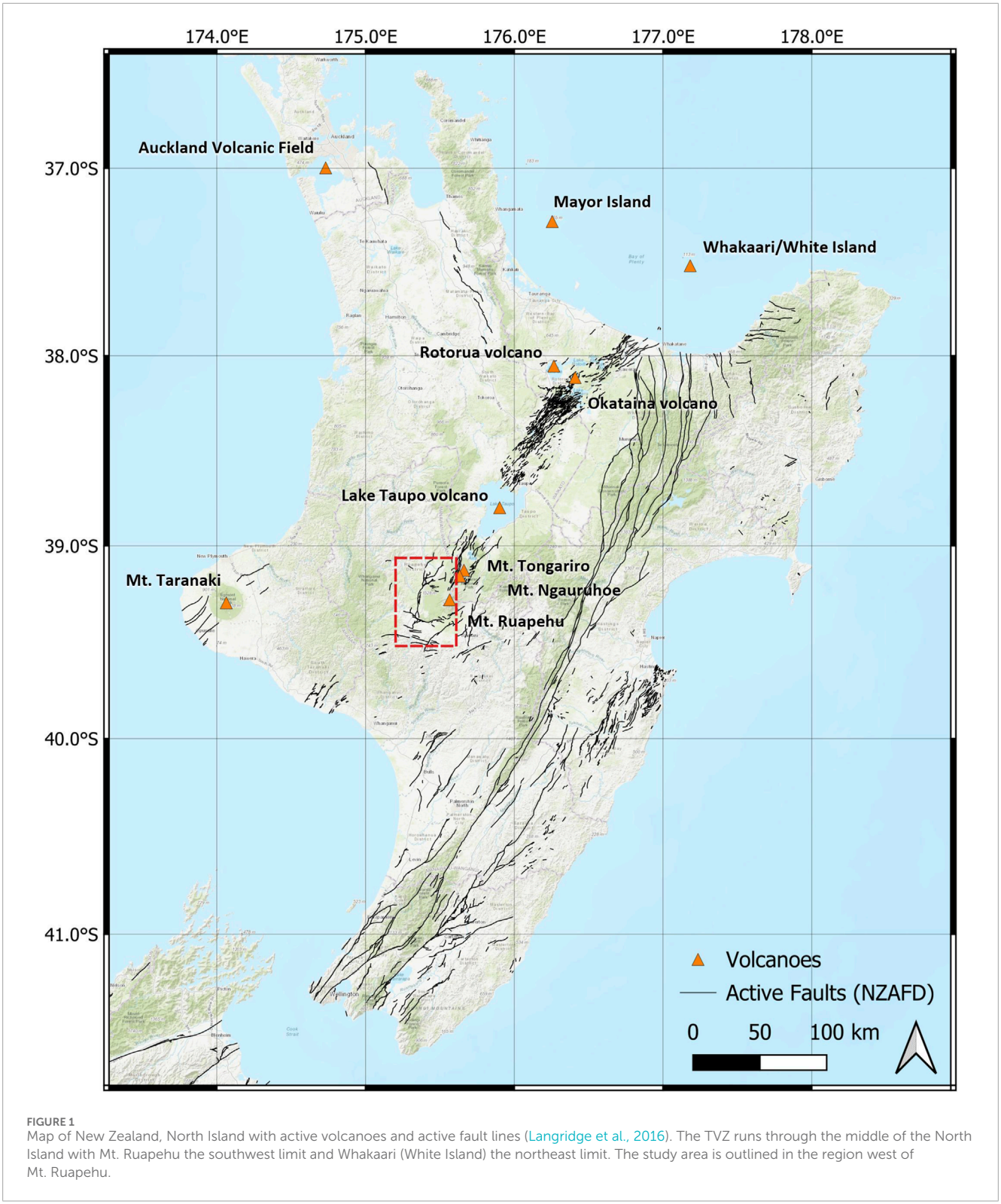


FIGURE 1
Map of New Zealand, North Island with active volcanoes and active fault lines (Langridge et al., 2016). The TVZ runs through the middle of the North Island with Mt. Ruapehu the southwest limit and Whakaari (White Island) the northeast limit. The study area is outlined in the region west of Mt. Ruapehu.

(Houghton et al., 1987; Schaefer et al., 2018). In 1953, the sudden collapse of the crater debris wall created a crater lake breakout and the formation of a lahar which destroyed the Tangiwai railway bridge causing the death of 151 people (Houghton et al., 1987; Johnston et al., 2000). The 1995/1996 eruption sequence was the largest in the 50 years prior (Sherburn et al., 1999). The sequence began with a localised phreatomagmatic eruption on 17 September 1995 (Hurst et al., 2018) and developed into a major eruption undergoing 10 eruptive phases, before ending in December 1996 (Bryan and Sherburn, 1999). The most recent recorded eruptions

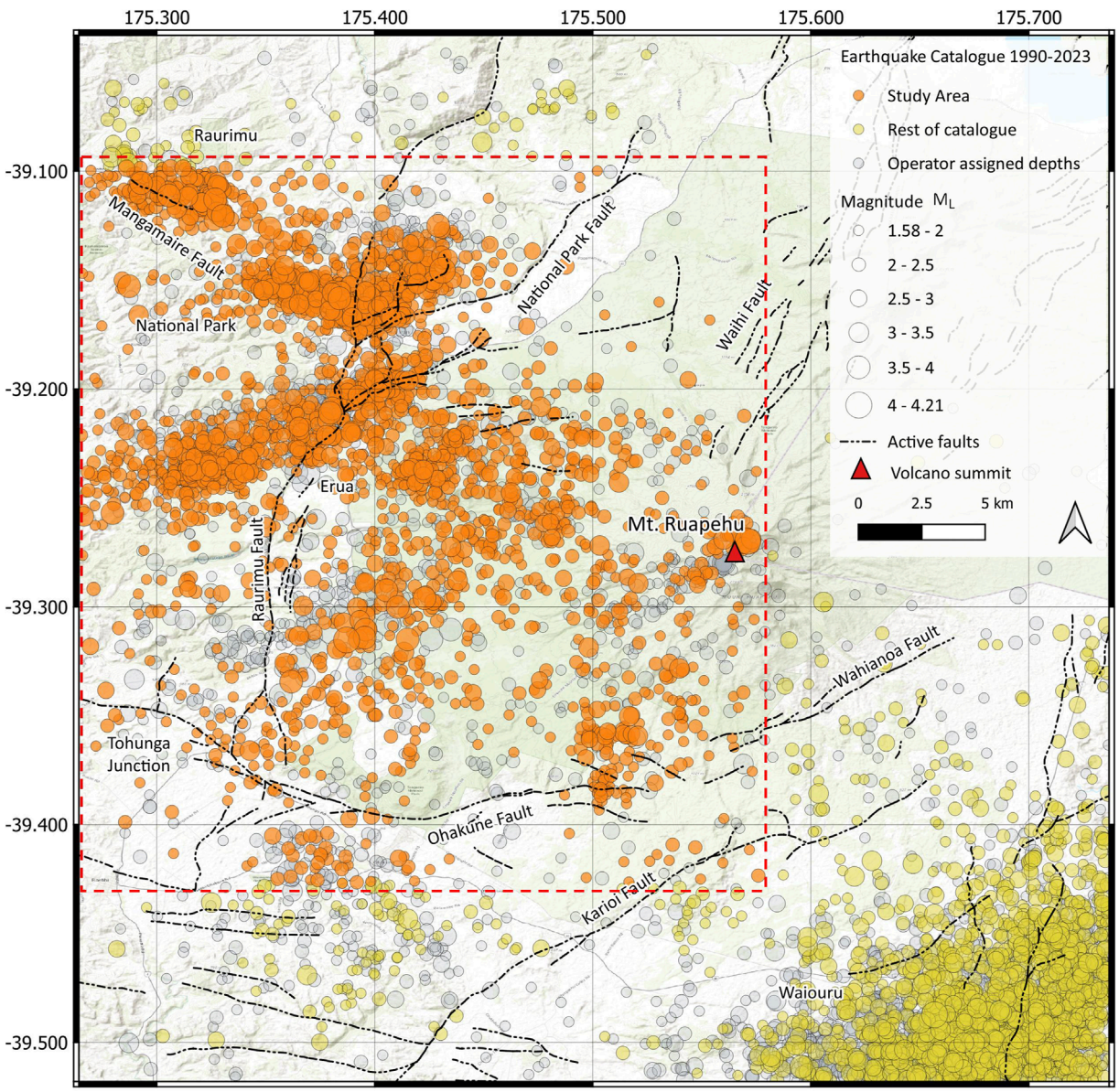


FIGURE 2
Map of the Ruapehu seismic region, including the study region outlined in a red box, showing the distribution of earthquake locations and magnitude between 1990 and 2023 at Mt. Ruapehu, with non-operator and operator assigned earthquakes. It is argued that the distribution of earthquakes are similar for both data sets, and that the removal of operator assigned earthquakes does not reduce the information of the catalogue. The boundary of the study area was selected to contain the seismically active region to the northwest of the summit, which has been analysed in previous studies at Mt. Ruapehu and associated with the 1995/1996 eruption sequence (e.g., Hurst and McGinty, 1999; Sherburn et al., 1999; Hurst et al., 2018).

were two steam-driven eruptions occurring on 4 October 2006 and 25 September 2007 (Jolly et al., 2010; Keats et al., 2011; Carniel et al., 2013). The sub-aqueous 4 October 2006 eruption may not have produced an eruption column, but did significantly raise the crater lake water level (Jolly et al., 2010), which has been linked to the 18 March 2007 tephra dam collapse and subsequent lahar on 17 March 2007 (Carrivick et al., 2008). Another, larger phreatic eruption occurred on 25 September 2007 (Jolly et al., 2010). Although there were insufficient precursory data suggesting an eruption was imminent at the time, minor volcano-tectonic earthquakes and tremor were found in the seismic data to have preceded the eruption by around 10 min (Jolly et al., 2010). The main eruption lasted no longer than 1 minute and produced an eruption column which ejected ballistics 2.5 km to the north, with lahars entering two local catchments and disrupting the Whakapapa Ski field (Kilgour et al., 2010).

2.2 Regional seismic activity

Mt. Ruapehu hosts its own tectonic setting, with faults circumscribing almost the entire base of the volcano some 20 km from the summit (Leonard et al., 2021). Shallow seismicity

(<20 km depth) over the last 30 years occurs in distinctive high-density clusters at depths along the Raurimu Fault, which runs north to south along the west flank of Mt. Ruapehu (Figure 2). Analysis of earthquake epicentre distributions from the GeoNet earthquake catalogue, suggests there are two distinctive regions, to the northwest and southeast of the Mt. Ruapehu summit, where earthquakes tend to occur more frequently than others in characteristic clusters (Figure 2). Seismicity originating from deeper sources reflects the slip geometry of the Pacific Plate subducting beneath the Australian Plate (e.g., Yarce et al., 2019) from the southwest at depths 50–100 km, to the northeast at depths >150 km.

To the southeast, approximately 20 km from the summit, is a band of near-constant seismicity along the Rangipo (Desert Road) Fault, near the town of Waiouru (Hayes, 2004) (Figure 2). Average b-values have been calculated at 1.06 at mid-crust depths ranging from 11 to 21 km implying a tectonic origin (Hayes, 2004). The b-value describes the relationship between large and small earthquakes in a catalogue, commonly used to describe the nature of the seismic activity of a region (Gudmundsson, 2020). Increased seismic activity was detected prior to the 1995 major eruption and once again in 2000/2001, with b-values sustaining high levels through the 1995/1996 eruption sequence reaching 1.71 in 2000 (Hayes, 2004). Although the earthquake cluster seems connected to volcanic episodes at Mt. Ruapehu, it is unlikely to be of volcanic origin itself and instead may be associated with regional stress changes following the 1995 major eruption (Hayes, 2004).

The seismicity to the west of Mt. Ruapehu is mainly concentrated northwest of the summit at the north extent of the Raurimu fault near National Park. There is also another distinct band of seismicity, which can be observed near Erua (Figure 2). Earthquake locations then become more uniformly distributed to the west and southwest towards Horopito and the Tohunga Junction. The cluster of earthquakes near Erua has been defined as precursory signal to the 1995/1996 major eruption sequence (Hurst and McGinty, 1999; Hurst et al., 2018). Studies by White and McCausland (2016) and Meyer et al. (2021) have developed methods, originating from McGarr (1976), for estimating the intrusive volume of magma intrusions from tectonic earthquake sequences at distal faults to suggest they might be suitable mid-term eruption forecasts. Hurst et al. (2018) argues the petrological timescale of erupted materials coincides with the seismicity near Erua, before the 1995 eruptions, suggesting that the seismic signal was an indicator of magma movement. Other studies have also suggested that the 1995 Erua swarm was not of tectonic origin, but a signal of stress changes associated with fluid movement within the volcanic system (e.g., Hayes, 2004). This may not be limited temporally to the 1995/1996 eruption, and may instead be an ongoing signal typical of pore-fluid pressure triggered by fluid movement in the mid-crust (e.g., Keats et al., 2011). However, a lack of observed seismic rate increase prior to the 2006 and 2007 eruptions makes eruption forecasts using this seismic signal uncertain. For the swarm near Erua to be considered a reliable forecasting signal (e.g., Hurst et al., 2018), it must be unique to the months preceding volcanic eruptions at Mt. Ruapehu, and should be absent during periods of volcanic inactivity (e.g., Ardid et al., 2022).

3 Data and methods

In this research, we classified spatio-temporal earthquake sequences in the GeoNet earthquake catalogue using a combination of density-based clustering algorithms. The magnitude distributions were analysed to see if the earthquake sequences exhibit typical mainshock-aftershock (MS-AS) or swarm-like characteristics. MS-AS sequences typically have a dominant initial event, called the mainshock, followed by a subsequent Omori-type (Omori, 1895) aftershock decay (Petersen and Pankow, 2023). It has been shown that analysis of the temporal evolution of earthquake sequences coupled with their associated cumulative seismic moment can be useful for differentiating between different types of seismic activity (Vidale et al., 2006; Lanza et al., 2022).

3.1 Earthquake catalogue

The earthquake catalogue for the Mt. Ruapehu region was downloaded for a period from 1990 to 2023, containing 28,522 earthquakes at shallow depths (<20 km) containing information on earthquake event type, time, location, depth and magnitude (Figure 2). Information including hypocentre location error is included in the Supplementary Material. Event types not relating to naturally occurring seismic events (e.g., quarry blasts and explosions) were removed, as were earthquakes located at the Mt. Ngauruhoe and Tongariro summits, which neighbours Mt. Ruapehu to the northeast, along with the near-persistent earthquake hot spot the southeast near Waiouru (e.g., Hayes, 2004) (Figure 2). The final study area was refined to a rectangular area to the west of Ruapehu [−39.09, 175.27, −39.43, 175.58] spanning approximately 1,800 km^2 to the west of the summit, to focus on activity exclusively to the west of Mt. Ruapehu, where the precursory signals at distal faults were identified by Hurst et al. (2018) (Figure 2).

Approximately half of the earthquakes in the catalogue (14,587) were assigned depths of 0, 5, or 12 km by an internal operator. These approximations for the hypocentre of the earthquake may be necessary when there are fewer seismic stations available to make a measurement, or when there is uncertainty in the velocity model used. Keeping earthquakes with locations biased to the assigned depth would be detrimental to the unsupervised machine learning approach we adopt and were therefore removed from the catalogue. Although this is a large proportion of the earthquakes in the catalogue, we argue that these earthquakes generally fit the underlying geographical trend of the seismicity of the region, and do not add any significant information to the overall catalogue (Figure 2). Therefore, the operator assigned events were removed from the catalogue for the purpose of this research.

The final data processing step was to assess the magnitude of completeness (M_c) of the catalogue with *operator assigned* earthquakes removed. This was done by estimating the point of maximum curvature for the cumulative and non-cumulative frequency magnitude distribution (FMD) (Pavlenko and Zavyalov, 2022). Using the Maximum Curvature (MAXC) method (Wiemer, 2000), bin sizes of ΔM_L 0.1 were evaluated to find the best fit M_c value, which was calculated as 1.6 with a b-value of 1.04 ± 0.02. Finally, the earthquake catalogue was filtered to contain only earthquake events ≥M_c, totalling 2,795 individual events.

3.2 Detecting spatial anomalies

Density-based clustering algorithms are effective at analysing noisy data sets and can handle irregular cluster shapes, making them a good method for analysing the clustering and zoning of earthquakes (Scitovski, 2018). Clustering algorithms, such as DBSCAN (density-based clustering algorithms with noise) (Ester et al., 1996) have been effectively used to identify and classify seismic swarms in earthquake catalogues (e.g., Petersen and Pankow, 2023). Where the methodology described in Petersen and Pankow (2023) begins with identifying temporal clusters in the catalogue before spatial analysis, this study aims to identify spatial clusters prior to the temporal cluster analysis using density-based clustering algorithms. The seismic activity in the months prior to the 1995 eruption was confined to a distinctive cluster Hurst and McGinty (1999), therefore we attempt to observe whether this kind of seismic signal is unique in specific regions (e.g., Hurst et al., 2018) prior to volcanic eruptions or whether it repeats during periods of volcanic quiescence.

Hierarchical density-based clustering algorithms with noise (HDBSCAN) (Campello et al., 2013) is an unsupervised machine learning clustering algorithm that is particularly useful at identifying clusters of varying shape and densities within a noisy data set. HDBSCAN begins by computing the density of data points based on their proximity to neighbouring points using a similar methodology as the popular density-based algorithm DBSCAN (Ester et al., 1996). DBSCAN begins by selecting a random point in the data set ($\mathbf{x}$) and, using Euclidean distance, calculates whether there are enough points in its neighbourhood to begin building a cluster, this is labelled a core point. Once a core point has been satisfied, the cluster is expanded to detect all core points which are directly-reachable in the latitude, longitude and depth dimensions. Once all of the core points for a cluster have been assigned, the algorithm adds the non-core border points to the cluster. The border points cannot be used to connect any other points to extend the cluster, even if another non-core point is within the distance threshold of the border point. Data points which fall outside the threshold defined by the input parameters are labelled as outliers and given the value −1. These steps are processed sequentially, meaning a non-core point which is within the distance threshold (ε) of two or more separate clusters will be assigned to the cluster which was calculated first (Figure 3).

HDBSCAN builds upon the DBSCAN algorithm by allowing the computation of clusters with varying ε distance, which is effective for grouping clusters with a range of densities. This is achieved by using the concept of mutual reachability distance (MRD), where the influence of multiple points are considered in relation to their local neighbourhood density, defined as:

$$d_{\text{mreach}-k}(a,b) = \max\{\text{core}_k(a), \text{core}_k(b), d(a,b)\}$$

Where $d(a, b)$ is the original distance metric between points a and b. HDBSCAN then constructs a minimum spanning tree (MST) from the MRD values, which is designed to highlight the strongest connections between points. The *hierarchical* part of the algorithm begins by using a single-linkage approach to repeatedly merge clusters close in density, ensuring that the merged cluster is associated with the cluster having the highest minimum density of the two. As more clusters are merged, the algorithm constructs a condensed tree that captures the hierarchy of cluster merges based on density. Each level of the tree represents a different level of minimum density required for a cluster to exist. This allows for a range of cluster sizes and shapes to be detected. The stability of a cluster is measured by how often it appears across different density levels in the condensed tree. Clusters that are more stable (i.e., they persist across multiple levels) are considered more significant. The final clustering result is obtained by selecting the clusters with the highest stability, while noise points and outliers are also accounted for. This allows HDBSCAN to automatically determine the total optimal number of clusters and handle clusters of varying densities.

Where HDBSCAN inherently selects the optimal number of clusters and corresponding distance between points, the algorithm requires a domain knowledge for the minimum number of samples per cluster, i.e., the minimum number of earthquake events for a cluster, (*min_cluster_size*). The *min_cluster_size* parameter determines the size of clusters calculated in the data set and impacts the total number of noise values (Starczewski et al., 2020) and has a default value of 5 data points per cluster. Although there are no constraints as to the number of earthquake events that make up a swarm, we can assume the earthquake signals we are attempting to identify contain greater than 5 events per cluster. In order to select the optimal number of points per cluster, one can run a range of *min_cluster_size* values through the clustering algorithm and test each result against a fitness function in a method known as *Tournament Selection* (TS) (e.g., Karami and Johansson, 2014). The best-fit *min_cluster_size* candidate based on the score of the fitness function is then selected for the clustering algorithm. Although there is no silver bullet fitness function for evaluating clustering algorithms, the Silhouette Coefficient θ (Rousseeuw, 1987) is a simple analytical method that can be used to find the optimal number of clusters for a variety of different clustering algorithms (e.g., Wang et al., 2017; Angmo et al., 2021) and does not require any prior training to produce results (Shutaywi and Kachouie, 2021). The Silhouette Coefficient θ is available in Scikit-learn and begins by comparing the distances of each sample to its own cluster and the neighbouring cluster, with scores ranging from −1 to 1:

$$\theta(i) = \frac{b-a}{max(a,b)}$$

Where a is the average distance of sample point i within a to other point within cluster a and b is the distance of sample point i to its nearest cluster. The function then calculates the mean Silhouette Coefficient over all cluster to return a single score for all clusters:

$$\theta = \frac{1}{n}\sum_{i=1}^{n}\theta(i)$$

where n is the number of samples and $\theta(i)$ is the Silhouette Coefficient for the i-th sample. A score of 1 would be achieved for two perfectly distinguished clusters at infinite distance from each other. A score of zero means that clusters are insignificant and the distance between them is not significant. The Silhouette Coefficient evaluated the goodness of fit for *min_cluster_size* 5–40 where the

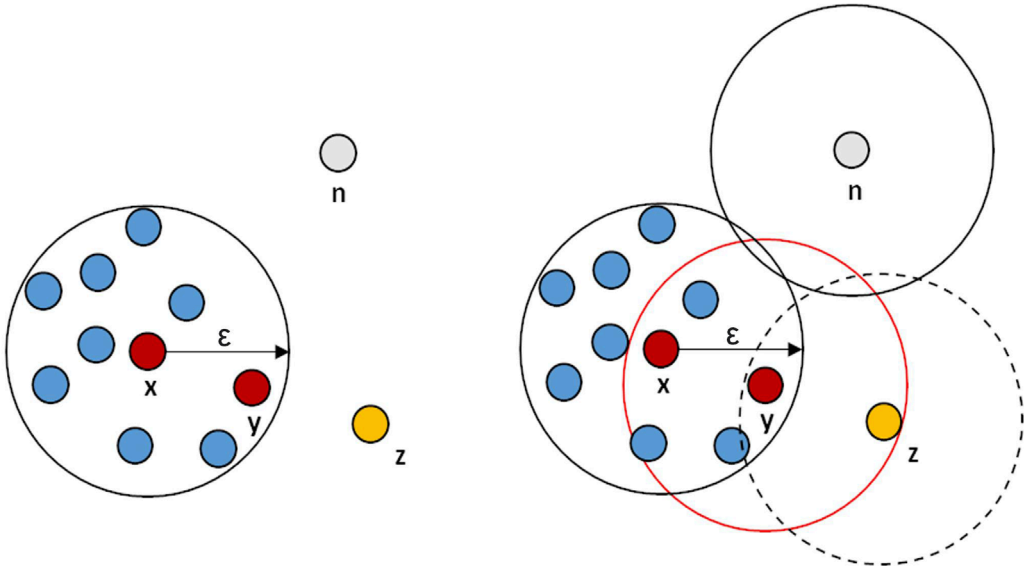


FIGURE 3
Illustration of a density-based clustering algorithm. Point **x** is a randomly selected starting point within the data set. The minimum number of points which define a cluster is 10, within a distance ε. Point **y** is directly density-reachable and is also a core point. Point **z** is density connected to point **y**, however is not reachable for any other points. Therefore, **z** becomes a border point and signifies the end of the cluster. Data point **n** is unreachable from any other data point and is therefore not part of the cluster and labelled as an outlier or noise.

optimal number of samples in a cluster was 31 with a score of 0.19. We found that the cluster size did not change much for *min_cluster_size* 30–40. However, the signal near Erua was most prominent for *min_cluster_size* 30–32 and the Silhouette Coefficient reduces as sample size >33. Therefore, we selected a *min_cluster_size* of 31 for the HDBSCAN algorithm. We noticed an increase in volatility for the total number of clusters, given a sample size <30 samples per cluster, where clusters tend be more unstable and form smaller groups. A summary of HDBSCAN results with *min_cluster_size* 5–40 can be seen in the Supplementary Material.

3.3 Detecting temporal anomalies

DBSCAN was used to identify temporal clusters in the time series for each of the regions classified using HDBSCAN. Temporal anomalies represent periods when the number of events within the HDBSCAN cluster region is unusually high, which is a proxy for earthquake sequences such as mainshock-aftershock (MS-AS), foreshock-mainshock aftershock (FS-MS-AS) or swarm sequences. The date and time of each event was converted to numeric using matplotlib's *date2num()* with the epoch set to "1990-01-01T00:00:00," where "1990-01-02T00:00:00" = 1. The DBSCAN method requires two input parameters, the minimum distance between points (ε), and the minimum number of points (*MinPts*) accepted to assign a cluster. The distance metric ε selected is then based on the number of days, converted to numeric time. To identify temporal anomalies, we employ a rolling average window for a 30 days time period and calculate the number of events in each window, which is stepped forward in daily increments. The z-score is then calculated for each rolling average and the inverse cumulative distribution function (PPF) of the standard normal distribution is calculated to define the anomaly threshold. This threshold corresponds to a 99% confidence interval and can be interpreted as the number of events above which rolling averages are considered anomalous, or the earthquake rate above usual background levels, which is considered anomalous. We found that the number of events considered anomalous within a 30-day window was 25 events. Using a 30-day window allows for a sufficient number of data points within each window, whilst facilitating the identification of longer-term patterns, which represents the nature of earthquake swarms. 25 events within a 30-day window also generally corresponds to the earthquake counts to the west of Mt. Ruapehu preceding the 1995/1996 eruption sequence (Hurst and McGinty, 1999; Sherburn et al., 1999; Hurst et al., 2018). The selected DBSCAN parameters are then applied to each of the geospatial clusters identified using HDBSCAN.

Using an unsupervised machine learning clustering approach, such as DBSCAN, is advantageous as it has the ability to identify moments in the time series where there may be anomalously high numbers of earthquakes over a period of time, without prior domain knowledge of earthquake sequences, and it also removes the need to set absolute boundaries on the length of time for each sequence. Meaning one can automatically capture earthquake sequences over a range of time-frames without bias, providing the minimum threshold set by the z-score is achieved.

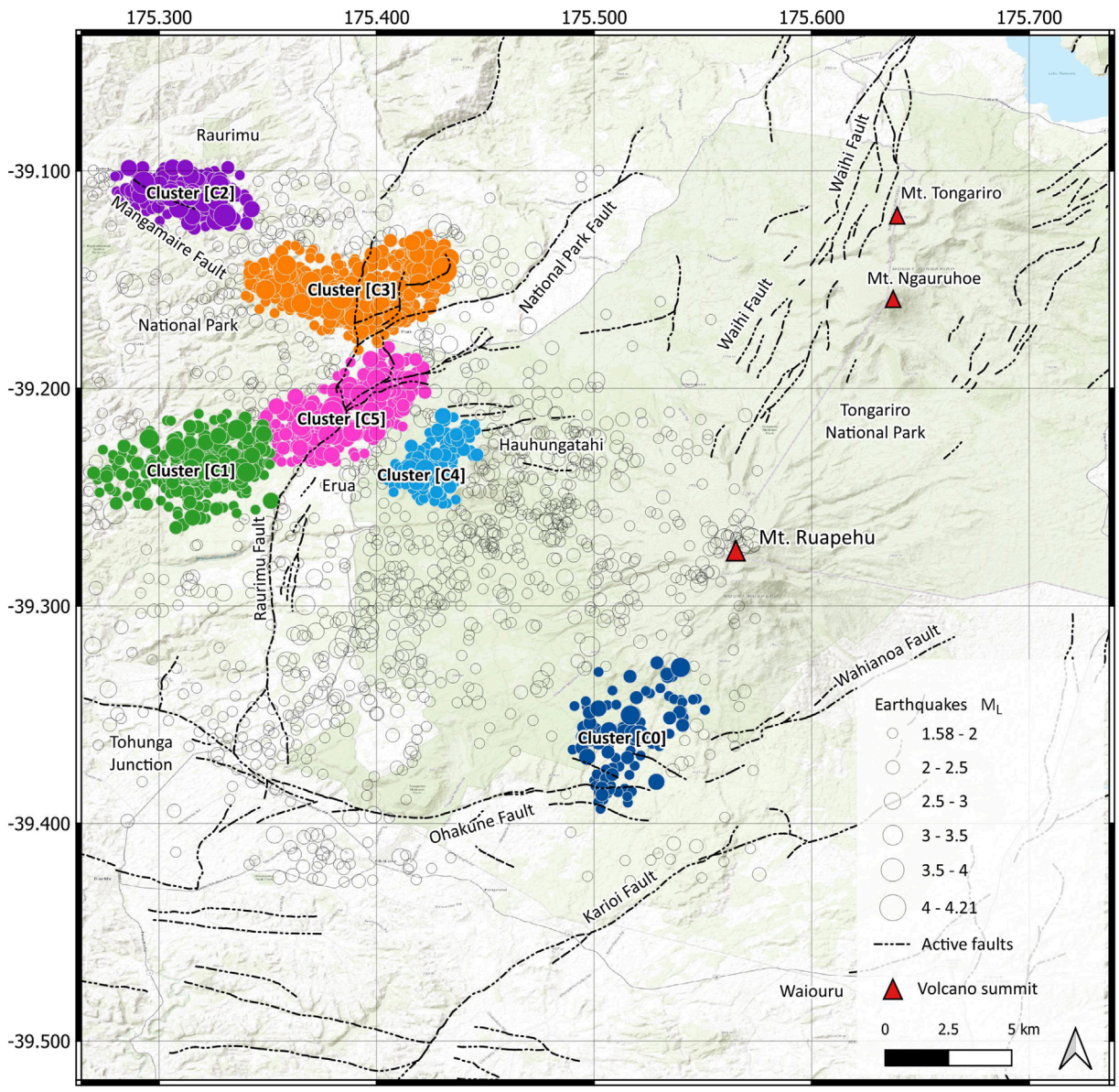


FIGURE 4
A map of Ruapehu volcano and Tonagariro National Park to the northeast, with the study region earthquake epicenter events. Six clusters were detected using HDBSCAN clustering and the noise data points achieving no cluster are labelled as the background seismicity. More information on the earthquake characteristics for each cluster can be seen in Table 1.

3.4 Cumulative moment

Evaluating the seismic energy associated with geospatial and temporal anomalies is an effective method for gaining a deeper understanding of the underlying characteristics and mechanisms at play. It is noteworthy that the seismic energy release is intrinsically linked to the magnitude of the earthquake event (McGarr, 1976). Methods for determining the earthquake local magnitude M_L can vary depending on the geology of the region and seismic network administration (Bormann and Di Giacomo, 2011). When performing seismic hazard assessments, it is convenient to convert local magnitude estimates to be consistent with moment magnitude M_w calculations (Rhoades et al., 2021), which is widely considered the best estimation tool for calculating earthquake magnitude (Ristau, 2009). In New Zealand, the relationship between M_L and M_w for shallow earthquakes (i.e., <33 km depth) (Ristau, 2009), can be described as:

$$M_L = (0.88 \pm 0.03)\,Mw + (0.73 \pm 0.20)$$

Seismic moment M_0 gives information regarding the total energy release of an earthquake and is proportional to the moment magnitude M_w defined as:

$$M_w = \frac{3}{2}\log_{10}(M_0) - 9.0$$

Where M_0 is in N m (Nm) and 9.0 is an empirical calculation of the shear stress of the crust. The cumulative seismic moment of earthquake sequences has been identified as a tool for volcanic eruption forecasting (e.g., Thelen et al., 2010; White and McCausland, 2016; Meyer et al., 2021). These kinds of approaches represent a significant step towards statistically quantifying key volcanic processes, including swarm-like earthquake sequences, departing from conventional descriptive measures and providing a more data-driven foundation for volcanic eruption forecasting.

4 Results

We can observe that the region to the west of Mt. Ruapehu has maintained seismic activity through the 33 years catalogue with low magnitude earthquakes typically in the range of 1.58–4.21 M_L. Using density-based clustering algorithms for the spatial and temporal analysis of the earthquake catalogue, we were able to identify six earthquake regions (clusters) where a number of earthquakes all occur in close proximity to each other, from 1990 to 2023. DBSCAN discovered seven earthquake sequences within the HDBSCAN clusters, where there were periods of seismicity which exceeded the threshold of 25 events for a 30-day rolling window. Two of the clusters have repeating periods of increased earthquake rate discovered using DBSCAN. Some cluster regions have clear temporal "bursts" of activity lasting from a few days to many weeks, yet remain relatively inactive for the majority of the 33-year period, and other regions have almost constant seismic activity with occasional small bursts. This section begins by presenting the geospatial anomalies detected using HDBSCAN, followed by the temporal and cumulative magnitude time series results.

4.1 Geospatial anomalies - HDBSCAN

The HDBSCAN algorithm took 0.05 s to run for the earthquake catalogue containing 2,795 events, identifying six distinct earthquake regions (clusters), for a cluster size minimum of 32 samples per cluster, amounting 63% of earthquakes in the catalogue. The clusters [C0, C1, C2, C3, C4, C5] exhibited various characteristics, including geographical distribution, density, and shape (Figure 4). The algorithm performed well at grouping a variety of cluster densities and shapes, and was also effective at discerning between boundaries of high levels of seismic activity, particularly between C5 and C1 (Figure 4). The location of earthquakes is spread non-uniformly across the study area west of Mt. Ruapehu, with occasional regions of geospatial density, particularly >5 km northwest of the summit. Earthquake activity decreases north of the summit towards Tongariro National Park. There is no clear correlation between earthquake activity and the active fault lines, with many earthquakes occurring away from documented fault lines, particularly in the region on the west flank of Mt. Ruapehu within 10 km of the summit, although, it is possible that not all active fault lines are documented. There are also few earthquakes located in close proximity to the summit.

Cluster C0 (blue) in Figure 4 is a comparatively low-density cluster, located 5–10 km south of the summit on the eastern limit of the Ohakune Fault, which runs west towards the Tohunga Junction. It contained the fewest number of earthquakes (113 events), with typical magnitudes ranging from 1.75 to 2.16, at depths typically ranging from 14 to 17 km. C1 (green) and C5 (pink) sit in a trajectory orientated approximately 40° northeast, with C1 earthquakes occurring away from the Raurimu Fault at a greater depth, on average between 14.7 and 18.1 km, compared to C5, which typically occur within the range of 11.2–14.6 km below the surface. There is a distinct difference in the magnitude of earthquakes, with significantly more M_L ≥3 occurring in C5 (Table 1). There is a subtle difference in the distribution density of events between C1 and C5, with C5 having a higher density of events. The cluster density tends to decrease as earthquakes are detected away from the Raurimu Fault to the west. C2 (purple), situated > 20 km from the Mt. Ruapehu summit, is in proximity to the Mangamaire Fault. Given the cluster has relatively few events, event magnitudes are characterised by some relatively large, M_L ≥3, events, coinciding with a range of smaller 1.73–2.28 M_L events. Although some of the larger events occur along the fault line, there is evidence of a group of events spreading northeast of the fault. C3 (orange) is mainly situated on the junction between the north extent of the Raurimu Fault and the southwest limit of the National Park Fault, but also extents northwest, away from the Raurimu Fault. C3 has the most number of events of all clusters, with 666 earthquakes at an average depth of 12.8 km and 16 events with a M_L >3.0. C4 (cyan) is located 12 km northwest of Mt. Ruapehu at the eroded extinct volcano Hauhungatahi (Cameron et al., 2010) (Figure 4).

TABLE 1 Clusters identified in the earthquake catalogue using HDBSCAN (minimum samples=31).

Cluster	Events	Max M_L	M_L >3	Avg. Depth (km)
C0	113	3.1	3	14.9
C1	369	3.4	4	16.3
C2	188	3.8	12	12.7
C3	666	4.2	16	12.8
C4	154	3.4	1	13.3
C5	283	3.8	13	13.0

4.2 Temporal anomalies - DBSCAN

DBSCAN took <10 s to iterate over the seven clusters identified by HDBSCAN (Figure 5), including the noise cluster containing background seismicity (Figure 6). Using the z-score threshold outlined in the methodology, and a 99% confidence interval, the number of events per cluster above the threshold was calculated as 25 events. This means, for all core points within a 30 day period, DBSCAN checks for any other events within 30 days. These are border events and are added to the cluster, but not considered density reachable for any other events. If there are at least 25 core and border events in total, a cluster is formed representing a period of earthquakes above the background level. For an ε distance of 30 days and a minimum number of 25 events per cluster, the DBSCAN

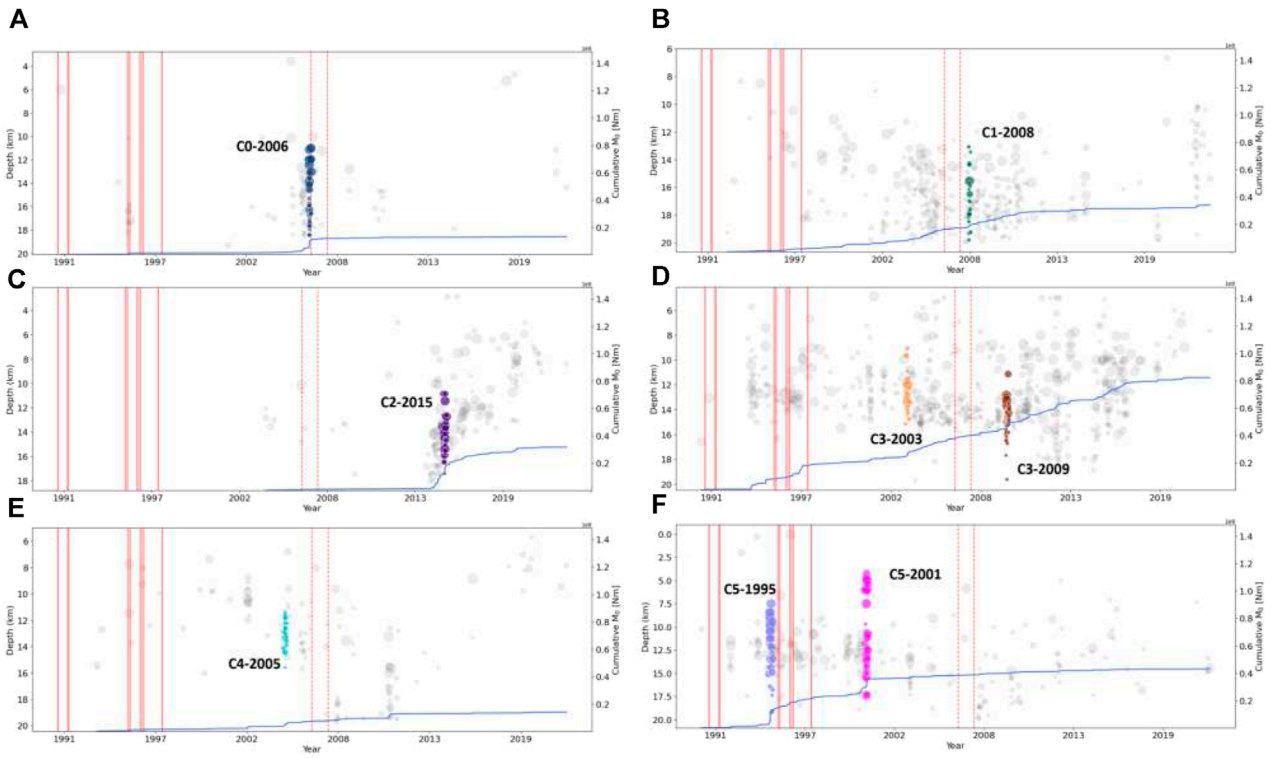


FIGURE 5
Time series scatter plot of earthquake epicenter events with depths shown on the y-axis and magnitudes with data point size for each HBDSCAN cluster. Eight clusters detected using DBSCAN, were determined by ε distance of 30 days for a minimum of 25 events, and represent moments in time where the earthquake rate is abnormally high. Activity in C0 **(A)** increased above the threshold in 2006, C1 **(B)** in 2008, C2 **(C)** in 2015, C3 **(D)** had a burst in 2003 and 2009, C4 **(E)** in 2005, and finally two sequences occurred in 1995 and 2001 C5 **(F)**. Magmatic eruptions are labelled as vertical red lines. Two small (VEI~1) eruptions were recorded in 1991 and 1992 before the major 1995/1996 eruption sequence (VEI~3), along with another small eruption in 1997. Two phreatic eruptions in 2006 and 2007 are labelled with a dotted red line. The cumulative moment magnitude represented by the blue line is an estimation for the amount of energy released in the study region through time.

algorithm identified eight temporal clusters within the time series derived from the HDBSCAN earthquake regions. The clusters identified by HDBSCAN are labelled by adding the year of when the sequence began to the original cluster name. The occurrence of earthquake sequences identified by DBSCAN are unique to each HDBSCAN cluster. Each cluster "lights up" with activity at different times throughout the 33 years and do not seem to have any correlation with any other earthquake sequence. Some clusters have near constant background activity with occasional bursts in earthquake rate (e.g., C3-2003 and C3-2009), whereas others have a single moment of activity before returning to quiescence (e.g., C0-2006). Earthquake sequences detected above the anomaly threshold vary in duration, but tend to last no longer than 100 days. The temporal clusters shown in the time series graphs (Figure 5), including the background seismicity (Figure 6), were plotted to show the geographical extent of each earthquake sequence (Figure 7).

C0-2006 is situated in a typically inactive region and had a significant increase in earthquake rate in 2006, where 71 events occurred. Activity peaked between 23 August–2 September 2006, with 40 events. On 28 August, a total of 16 events occurred in a single day. The earthquake rate in this period is correlated with a sudden increase in cumulative seismic moment. DBSCAN automatically highlighted a period between 23 August–8 October 2006 as an anomalous earthquake sequence, containing a total of 45 events, with an average magnitude of 2.04 M_L and two events M_L >3. Overall, DBSCAN attributed 39.8% of earthquakes within C0 to a single earthquake sequence over the 33 year period, with an average earthquake rate of 1.1 events per day.

The time series for the C1 region shows near constant earthquake activity, with an increase of activity between 2004 and 2012, also corresponding to an increase in cumulative seismic moment during the same time frame, before flattening out again. Within the period 2004–2012, when the cumulative seismic rate increases, DBSCAN identified one period (C1-2008) where the earthquake rate surpassed the background seismicity threshold. The sequence consisted of 27 small magnitude earthquakes ranging from 1.59 to 2.53 M_L, lasting 48 days from 6 April 2008–24 May 2008 at a rate of 0.56 earthquakes per day.

There was no seismicity recorded in the C2 region before 2004, perhaps due to a limitation of seismic apparatus available. Since 2014, the seismicity has increased in the region significantly, with 92.6% of total events in the cluster region occurring after July 2014 with a relatively deep source and getting shallower through time. C2-2015 had an average earthquake rate of 0.44 per day, which was deemed higher than the threshold by DBSCAN from 2 July 2015–8 October 2015. During this period, three earthquakes M_L >3

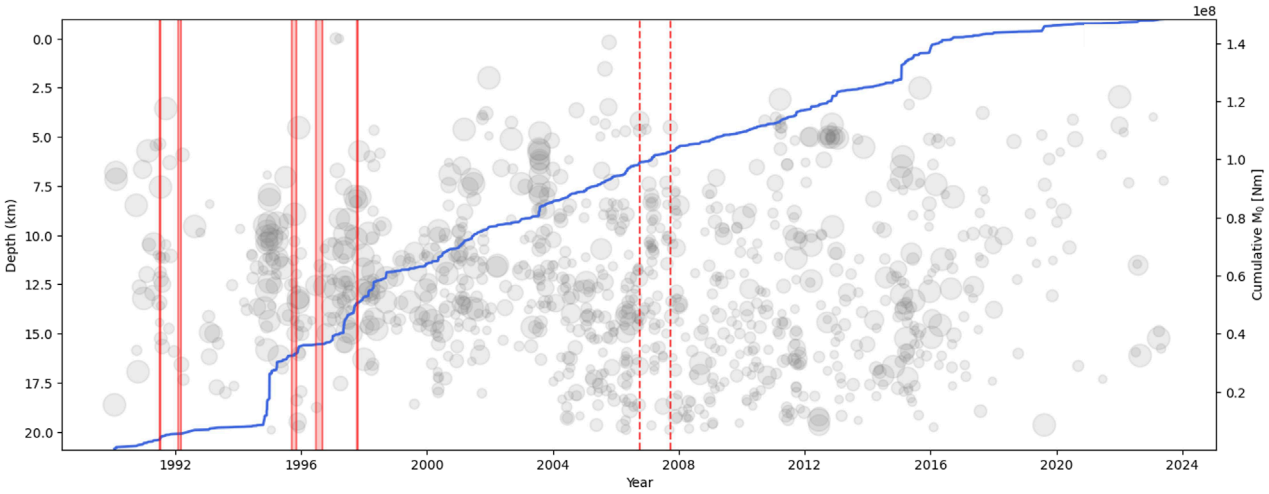


FIGURE 6
Time series scatter plot of background seismicity of the full study area excluding the HDBSCAN clusters, with the cumulative earthquake moment of the region. Earthquake epicenter events with depths shown on the y-axis and magnitudes shown as the size of the data point. Documented eruption periods are labelled using vertical red lines.

were recorded including a 3.54 M_L, which is reflected in the sudden increase in cumulative seismic moment.

The C3 time series is characterised as near-constant earthquake activity at relatively shallow depths of <16 km, with occasional temporal bursts in activity. There are two occasions where DBSCAN detected an increase in seismic rate greater than the background threshold. The first sequence (C3-2003) consists of 34 events at a rate of 0.4 earthquakes per day, located laterally across the north extent of the Raurimu fault, with the largest event occurring long the fault line and smaller events moving east and west. The second sequence (C3-2009) was made up of 60 events, mainly located in the region west of the fault with an approximate 20° orientation. This sequence had a higher earthquake rate than the previous sequence at approximately 0.7 events per day.

The small, dense cluster located in proximity to Hauhungatahi (C4), had three time periods where there was a sequence of earthquakes and small increases in cumulative seismic moment. The first sequence occurred in 2002 at shallow depths from 8 to 12 km and was not identified by DBSCAN. The second sequence (C4-2005), which was identified by DBSCAN due to an increase in earthquake rate of 0.45 events per day over 76 days, began in January 2005 and consisted of 34 relatively small events at depths of 13 km on average. There was a third sequence, similar to the events in 2002, originating from deeper sources (15–19 km), which began with a 3.42 M_L earthquake before following a typical mainshock-aftershock pattern.

C5 is contained to the Raurimu fault in close proximity to the town of Erua. Although earthquake activity can be observed throughout the time series, the majority of events have occurred before 2004 within C5-1995 and C5-2001, with epicentres located in a narrow band between 10 and 15 km depth. The two earthquake sequences highlighted by DBSCAN had anomalous earthquake rates, consistent with sudden increases in cumulative seismic moment. C5-1995 and C5-2001 are similar in geographical orientation, average depths, earthquake rate, and maximum magnitude. However, there were more events M_L >3 in C5-1995 and a slightly lower average rate of 0.4 earthquakes per day. C5-1995 sequence occurred in close temporal proximity to the onset of the 1995 major eruption. Whereas the C5-2001 earthquake sequence began on 30 December 2000, lasting 88 days with an average daily rate of 0.45 events, and has no temporal proximity to any volcanic activity at Mt. Ruapehu. The number of earthquakes during the peak of seismic activity was different for C5-1995 and C5-2001, with the peak of the C5-1995 sequence totalling 16 events over 15 days, and 26 events over 25 days in total for C5-2001. However, although the largest event in each sequence had a similar magnitude (3.4 M_L), the mean magnitude for C5-1995 was higher (2.6 M_L) than C5-2001 (2.2 M_L).

The slow increase of cumulative seismic moment across the entire region is coupled with occasional rapid increases (Figure 6), particularly leading up to the 1995 eruption. We have labelled this earthquake activity as *background seismicity*, due to the fact that earthquakes occurred in regions with no geospatial relationship. The largest increase in cumulative seismic moment began in 1995, driven by the first earthquake sequence C5-1995, before the onset of the 1995 major eruption and waned during the main eruption sequence before increasing again before and during the small eruption in 1997. There were also smaller increases in 2003/2004, which occur due to C3-2003, and slightly larger increase in 2015/2016 which coincides C2-2015.

5 Discussion

Using the temporal anomaly results obtained using DBSCAN, and the seismic moments of the earthquake sequences, we have assessed the statistical similarities between sequences and their implications for the 1995 Erua swarm (Hurst and McGinty, 1999; Hurst et al., 2018). An unsupervised algorithm with no *a priori* domain knowledge of the system has identified six regions to

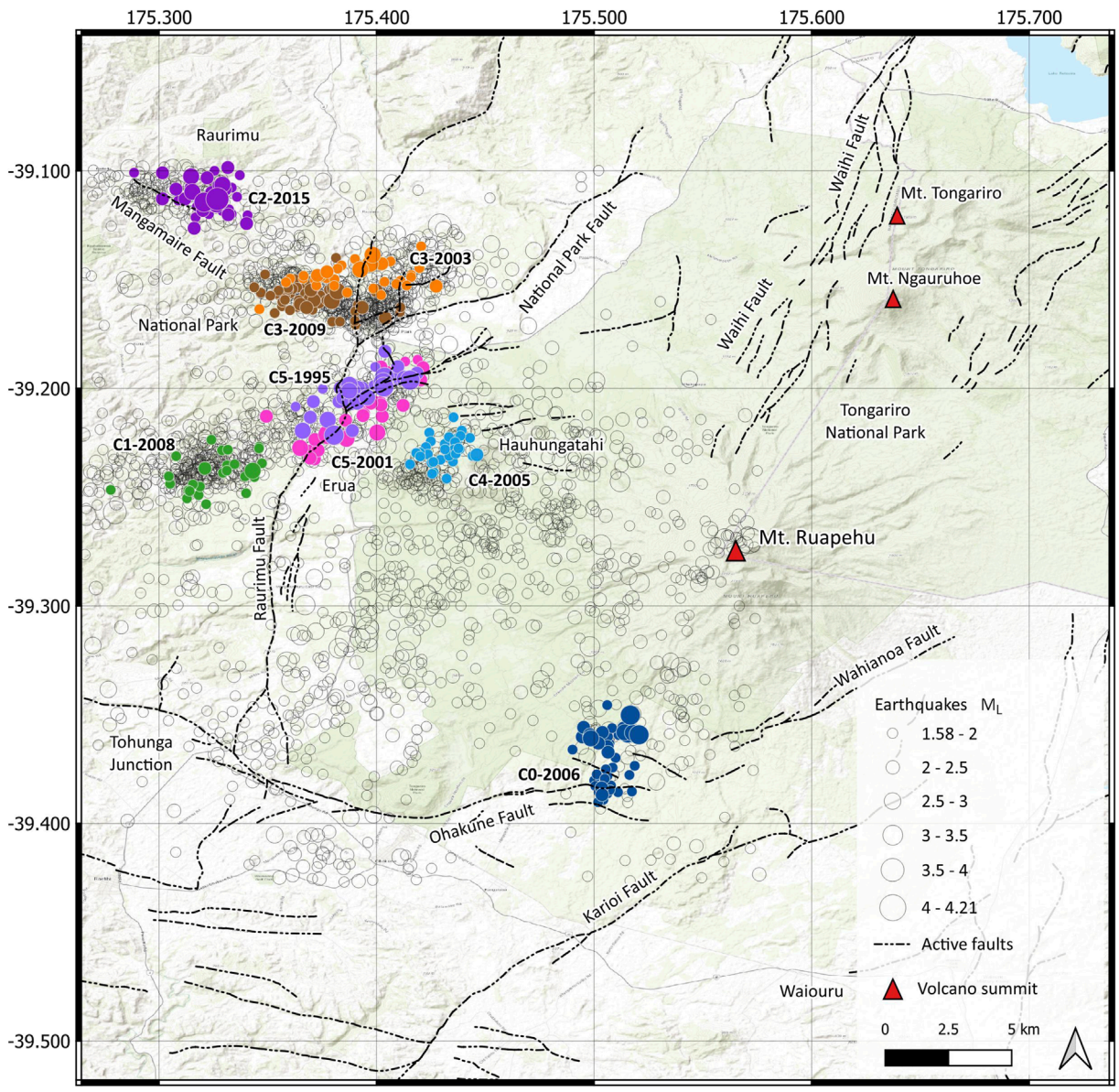


FIGURE 7
Geospatial representation of eight earthquake sequences, with background seismicity, identified at Mt. Ruapehu the 1990–2023 GeoNet earthquake catalogue using DBSCAN (Figures 5, 6).

the west of Mt. Ruapehu where earthquake activity tends to be concentrated more than in other areas. Earthquake sequences occurring in distinct clusters are regularly observed at volcanoes and are often indicative of where stress changes are concentrated during the natural dynamics of the volcanic system (McNutt and Roman, 2015). Seismic activity to the west of Mt. Ruapehu rapidly increased in the year prior to the 1995 major eruption, yet during the eruption we observe no elevated seismic activity, which is consistent with Hurst et al. (2018). There was a return to high levels of seismic activity during a short period in 1997, immediately following the end of the 1995/1996 eruption sequence, which may indicate a response in stress changes following the eruption, due to magma movement, or pressure changes at the summit (Hayes, 2004). After 2001, the region west of Mt. Ruapehu maintained a near constant rate of seismicity, which is consistent with the absence of major volcanic activity during this time. Occasional small spikes in seismic moment have no correlation to the small eruptions which occurred in 2006 and 2007. A larger increase in regional seismic activity occurred during 2015/2016, which correlates with C2-2015 (Figure 5C).

The timing of earthquake sequences detected using DBSCAN are unique to each region, and do not last longer than 100 days. Notably, 75% of the earthquake sequences detected by DBSCAN showed no clear correlation with recorded volcanic events. However, the earthquake sequences within C0-2006 and C5-1995, demonstrated temporal proximity to the 2006 phreatic eruption and the 1995 major eruption (Figure 5F). We deem temporal proximity to be within 6 months, given the history of earthquake swarms relating to eruptions at Mt. Ruapehu (e.g.,

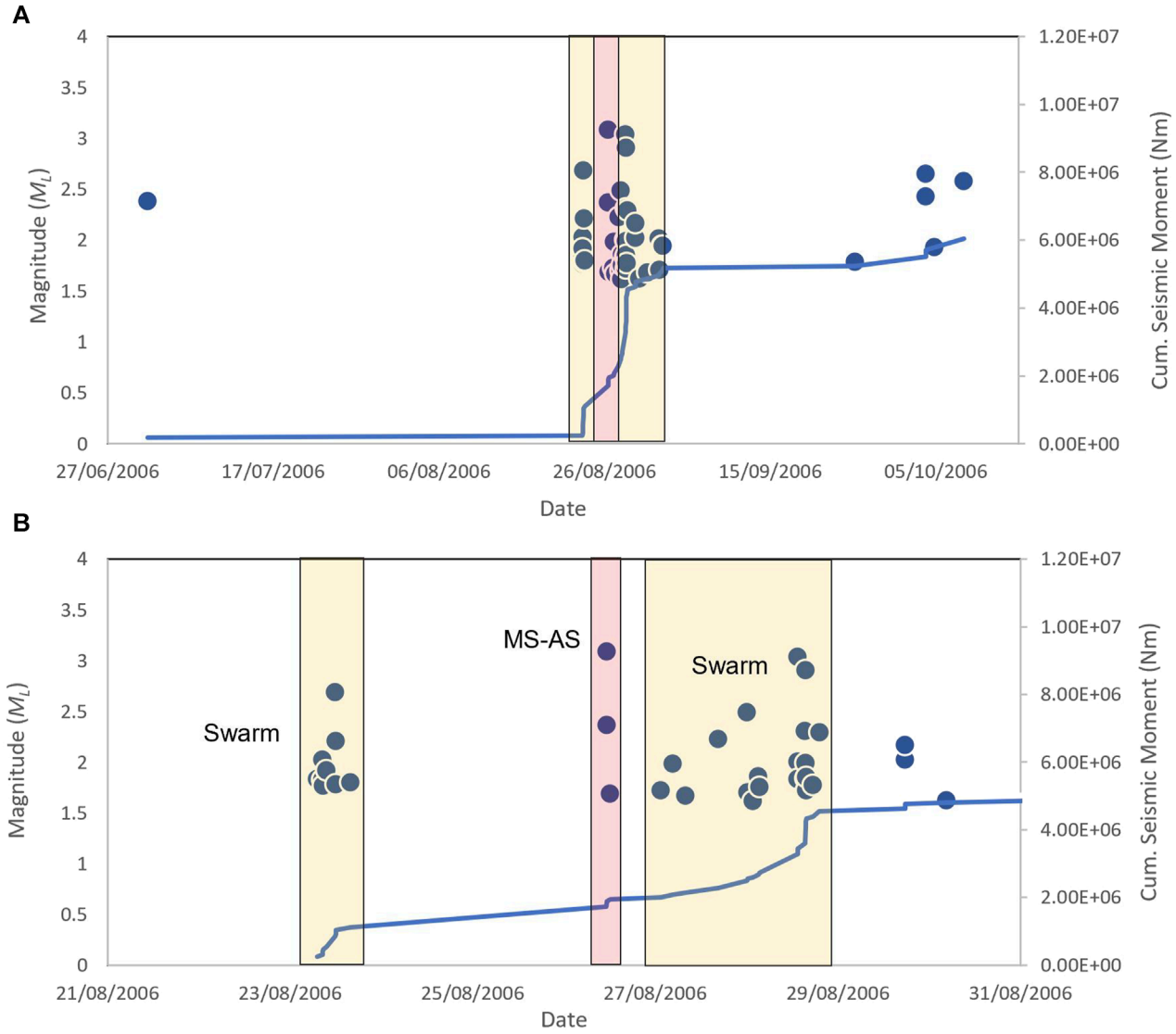


FIGURE 8
Total cumulative M_0 of earthquake events in C0-2006 for the temporal cluster detected by DBSCAN **(A)** and a zoom into the sequence **(B)**, showing the small swarm, highlighted yellow, before the MS-AS event preceding the main swarm, which is highlighted in red. The first event in **(A)** is not part of the DBSCAN sequence and is only included to assist with the visual. The events after the main swarm were detected by DBSCAN.

Hurst et al., 2018). Indeed, C5-1995 (Figure 5F), is the same signal interpreted to be a precursor for the 1995 major eruption by Hurst and McGinty (1999); Hurst et al. (2018).

Cumulative seismic moment with time show similarities between clusters. Broadly, we can group sequence types into the following categorical descriptions, 1) swarm-like sequences yielding a relatively small cumulative seismic moments through time, e.g., C3-2003 and C3-2009, 2) swarms with longer run-up time, maintaining a high seismic rate and a larger cumulative seismic moment, e.g., C2-2015 and C5-1995, 3) Complex sequences with MS-AS events followed immediately by a swarm-like sequence, e.g., C0-2006 and C5-1995, 4) a narrow band of magnitudes typically $\leq \Delta 0.5$ M_L with no significant energy release, e.g., C1-2008 and C4-2005.

B-values were estimated for each temporal cluster, with values ranging from 0.88 to 2.97. An estimation of the b-value in volcanic regions can be a useful indicator for monitoring periods of unrest (e.g., Farías et al., 2023) as they can provide additional information about whether seismicity is of tectonic or volcanic origin (Lanza et al., 2022). In regions with high b-values, there are proportionally more smaller earthquakes than large earthquakes (Gudmundsson, 2020). Naturally, therefore one tends to see higher b-values in volcanic regions than tectonic. However, although it is common to observe high b-values in volcanic areas, b-values have be found to be skewed to larger values particularly for small catalogues (Geffers et al., 2022). This might be the case for the temporal clusters at Mt. Ruapehu, where we can see in general, those sequences with fewer events, tend to have high b-values (1.51–2.97) with the

TABLE 2 High density earthquake sequences identified in the time series using DBSCAN for each HDBSCAN cluster (DBSCAN ε: number of days=30, minimum events=25). The DBSCAN cluster is denoted as a temporal sequence, labelled consecutively in the time series.

Sequence	Events	Max M_L	$M_L > 3$	b-value	Avg. Depth (km)	Start date	No days
C0-2006	45	3.1	2	1.13 ± 0.18	15.2	2006-08-23	42
C1-2008	27	2.5	0	2.97 ± 1.02	16.9	2008-04-06	48
C2-2015	44	3.5	3	0.99 ± 0.16	14.5	2015-07-02	99
C3-2003	34	2.9	0	1.51 ± 0.29	12.8	2003-09-25	86
C3-2009	60	2.7	0	2.05 ± 0.39	14.7	2009-10-30	87
C4-2005	34	2.1	0	2.25 ± 0.27	13.0	2005-01-18	76
C5-1995	39	3.4	5	0.88 ± 0.16	12.2	1995-02-24	97
C5-2001	40	3.4	1	1.67 ± 0.37	11.8	2000-12-30	88

exception of C5-1995, which had proportionally more earthquakes >3 M_L (Table 2). These b-values seem reasonable for Mt. Ruapehu and are similar for b-values estimated in other volcanic regions derived from earthquake catalogues, which can range from 1.4 to 3.5 (Roberts et al., 2015). We acknowledge that there is uncertainty in estimating the b-value given the small sample size and are therefore used to only to aid the interpretation of the temporal clusters.

5.1 The C0 cluster

The earthquake events occurring within C0-2006, situated along the eastern boundary of the Ohakune Fault on the southernmost region of the research domain, coincided closely in time with the phreatic eruption of October 2006. This synchronicity may hold significance due to the surprising nature of the eruption, which seemingly erupted without any warning signs of imminent volcanic activity (Jolly et al., 2010). A cursory examination of the earthquake catalog through visual analysis might not yield any apparent indicators of abnormal seismic activity in this area. Hence, the identification of the earthquake sequence from HDBSCAN and the temporal sequence from DBSCAN, was as an unforeseen revelation. C0-2006 was indicative of a complex sequence, exhibiting both MS-AS and swarm behaviour in a short period of time (Figure 8). The largest event (3.1 M_L) on 26 August 2006 was followed by two successively smaller events following a typical mainshock-aftershock sequence. On 28 August 2006, 16 individual events were recorded with the largest and second largest magnitude of 3.03 M_L and 2.9 M_L. The characteristics of the differences in magnitude and the cumulative seismic moment over a short period of time coupled with a b-value of 1.13, suggests this earthquake sequence is of swarm origin and may be the result of an intrusion of fluids into a dike at shallow depths, which interacted with the hydrothermal system and the crater lake. The sudden intrusion of magma into a crack near the surface may have triggered a rapid expansion and steam driven eruption without necessarily raising the crater temperature.

5.2 The C1 cluster

This cluster of earthquakes is the only set which occurs entirely away from a mapped fault source. C1-2008 is situated >1 km west of the Raurimu fault and is clearly a separate signal to the neighbouring C5-1995 and C5-2001. A near-constant seismic sequence began around the year 2000 at depths >12 km and has continued until 2023, with the most active phase from 2005 to 2012. The earthquake depth is most likely linked to the angle of the subducting plate, which is probably why the earthquakes in C5 occur at a shallower depth on average (e.g., Keats et al., 2011). The b-value for C1-2008 was the highest in the region, but also had the largest uncertainty (Table 2), which is probably driven by the small sample size.

5.3 The C2 cluster

C2-2015, on the northwest extent of the study area, may be an artifact of the timing of the station deployment. The earthquake rate increases to the threshold defined in DBSCAN, to a peak activity from 24 August 2015, where there was an increase in magnitude to the 3.31 M_L event on 26 August. The C2-2015 distribution of events is located away from the fault with a b-value of 0.99 (Figure 7). The temporal evolution of C2-2015 develops from a series of earthquakes with a narrow magnitude band, to a gradual increase in magnitude and rate, with the largest event occurring in the middle of the sequence (Figure 9). Extrapolating from the C2-2015 sequence, there is clear evidence of average earthquake source depths from deep to shallow (Figure 5C), which could be evidence of fluid migration through the mid-crust (Yoshida and Hasegawa, 2018; Yoshida et al., 2023).

5.4 The C3 cluster

The C3 cluster is the most seismically active region in the study area. The near-constant rates with occasional bursts is similar to the seismicity described by Hayes (2004) to the southeast at Waiouru,

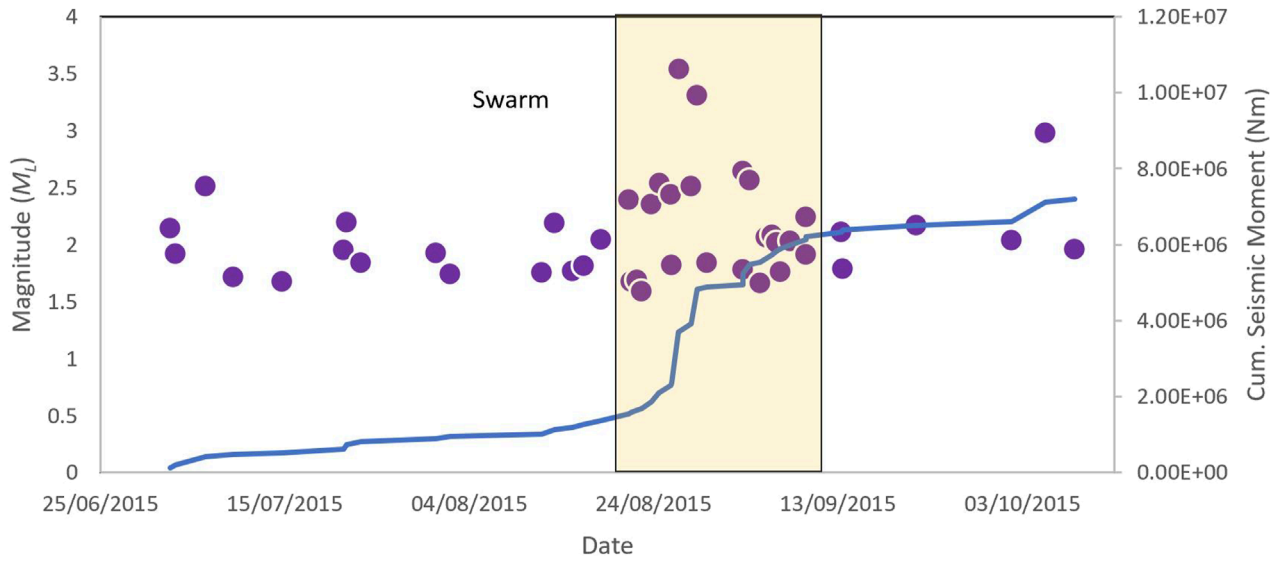


FIGURE 9
Total cumulative M_0 of earthquake events in Cluster 2 for the temporal cluster detected by DBSCAN. The distribution of events is suggestive of a swarm-like sequence.

which is outside of our study area. The first and second temporal bursts in C3-2003 and C3-2009 occur on different trajectories across the National Park fault, yet are similar in geographical extent (Figure 7). C3-2003 and C3-2009 exhibit similar swarm-like behavior (Figure 10), with high b-values of 1.51 and 2.05 respectively. We observe multiple temporal bursts in seismicity, which did not reach the threshold set for DBSCAN, and a lower threshold would certainly yield more swarm events. The temporal evolution of the C3 cluster has similar characteristics to swarm bursts discussed by Petersen and Pankow (2023). Here the variable inter-event bursts are related to variability in fluid migration in a fault-valve model (Sibson, 1992). For C3, changes in pressure from fluids below the surface, coupled with precipitation from the hydrothermal system at Mt. Ruapehu (Christenson and Wood, 1993), can cause the fault to open and close, leading to unpredictable variation in inter-event times (Petersen and Pankow, 2023). The fluid-driven source is supported by relatively high b-values for the region. Further sensitivity analysis into the time series for C3 might offer more insight into the processes we see in this region; however that is beyond the scope of this paper.

5.5 The C4 cluster

C4-2005 was detected in close proximity to the extinct volcano, Hauhungatahi (Cameron et al., 2010), which is made up of >170 ka old eruptive material. The small earthquake sequence detected at the end of 2004 is most dissimilar from the other signals, exhibiting earthquakes with a narrow band of magnitudes typically ≤ Δ0.5 M_L with no significant energy release (Figure 11). C4-2005 has the high estimated b-value of 2.25, which is perhaps driven by a small sample size (e.g., Geffers et al., 2022) coupled with a narrow magnitude band and the majority of earthquakes <3 M_L.

5.6 The C5 cluster

C5, situated along the northern extent of the Raurimu fault, close to the small town of Erua, is a region with earthquake activity which is often referred to in the literature as a precursor to the major 1995 eruption (Hurst and Vandemeulebrouck, 1996; Hurst and McGinty, 1999; Hurst et al., 2018). Although, the cluster is in the orientation, and in close proximity to the Raurimu fault, seismicity is thought to have not originated from the fault (Keats et al., 2011). The first earthquake sequence, C5-1995, began in January 1995 and ended at the end of July 1995. There were two bursts of activity, a small MS-AS sequence occurred mid-March, before the main burst of activity which began early April. The April burst had 16 individual earthquake events with 2 M_L >3 and coincided with a warming of the crater lake along with an increase in Mg2+ ions (Nakagawa et al., 1999), which may be an indication of fresh magma interacting with the hydrothermal system (Hurst and Vandemeulebrouck, 1996). The earthquake rate then waned until the beginning of the major eruption sequence on 17 September 1995 (Figure 12A). Conversely, the b-value of the DBSCAN sequence is 0.88, which is comparatively low for earthquake swarms in the region and is consistent with Hurst and McGinty (1999), and the tectonic mechanisms occurring to the southeast during the same time-frame (Hayes, 2004).

The second burst of activity (C5-2001), which began in January 2001, had a similar orientation, geographical extent and total number of events as the 1995 sequence, however did not precede any kind of volcanic activity. C5-2001 had a different temporal evolution

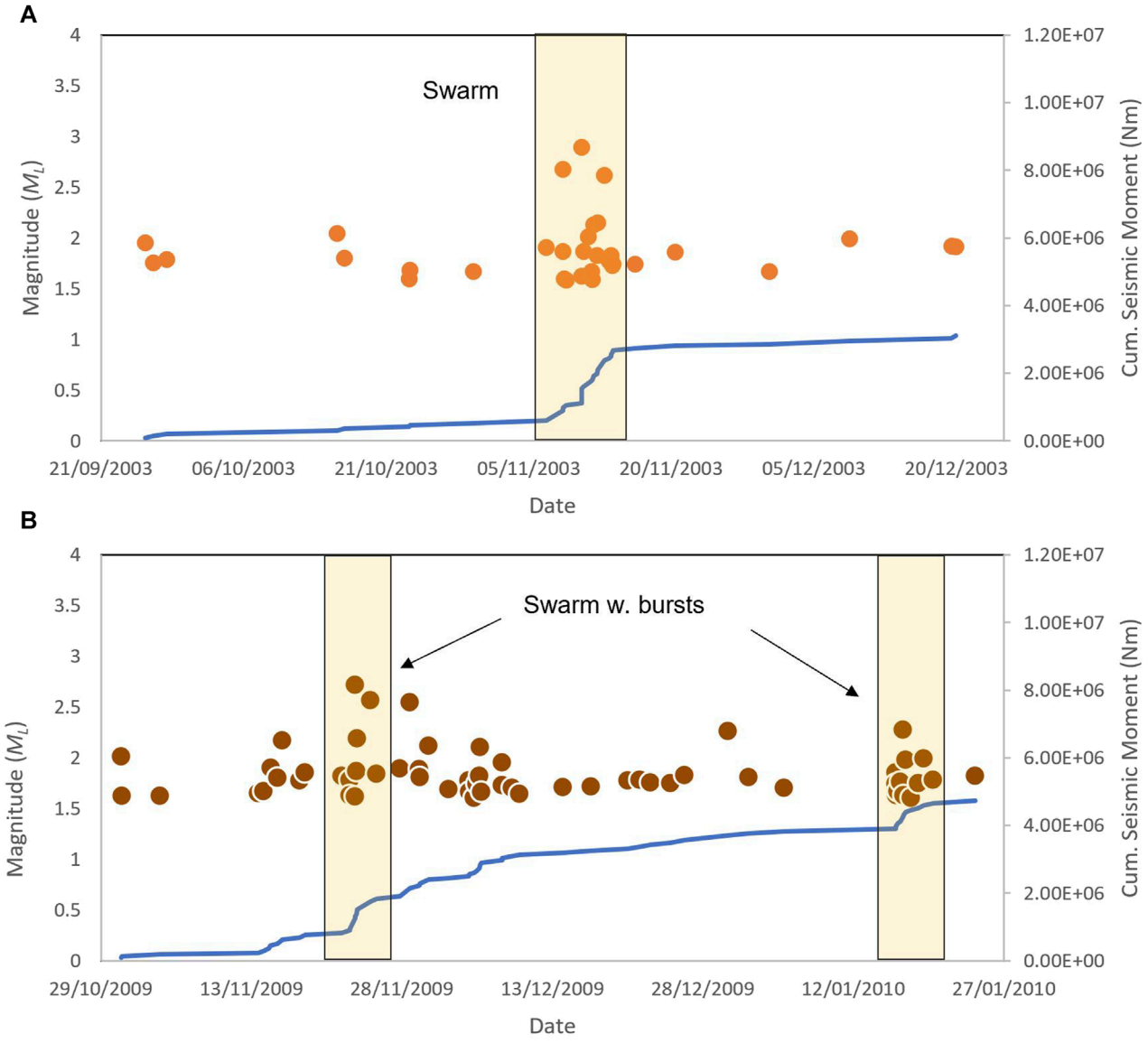


FIGURE 10
Total cumulative M_0 of earthquake events for C3-2003 **(A)** and C3-2009 **(B)** for the temporal cluster detected by DBSCAN. The swarm bursts in C3-2009 last between 3 and 4 days and are separated by 53 days.

to the 1995 sequence, with a buildup of seismic activity preceding the largest event on 3 March 2001, and smaller events following (Figure 12B). The largest event occurring around the middle phase of the temporal evolution is a key characteristic of swarm-like behaviour associated with volcano-tectonic events (Jones, 2005; Pesicek et al., 2018). The difference between the largest two events, 3.45 ± 0.23 and 2.85 ± 0.23, is 0.60 ± 0.46, and perhaps is on the boundary of swarm-like magnitude differences. The b-value of the swarm is calculated as 1.67 which is typical for swarm behaviour driven by volcano-tectonic processes (Gudmundsson, 2020).

5.7 Should C5-2001 have preceded an eruption?

The cumulative seismic moment of an earthquake sequence is a proxy for the total amount of energy released and provides insights into the mechanisms driving the seismicity in the region. If there is a lack of pressure or insufficient volume, magma may not have enough energy to overcome stress barriers and can become arrested (Caricchi et al., 2021), which may provide a reason for why there was not an eruption following the earthquake swarm in 2001.

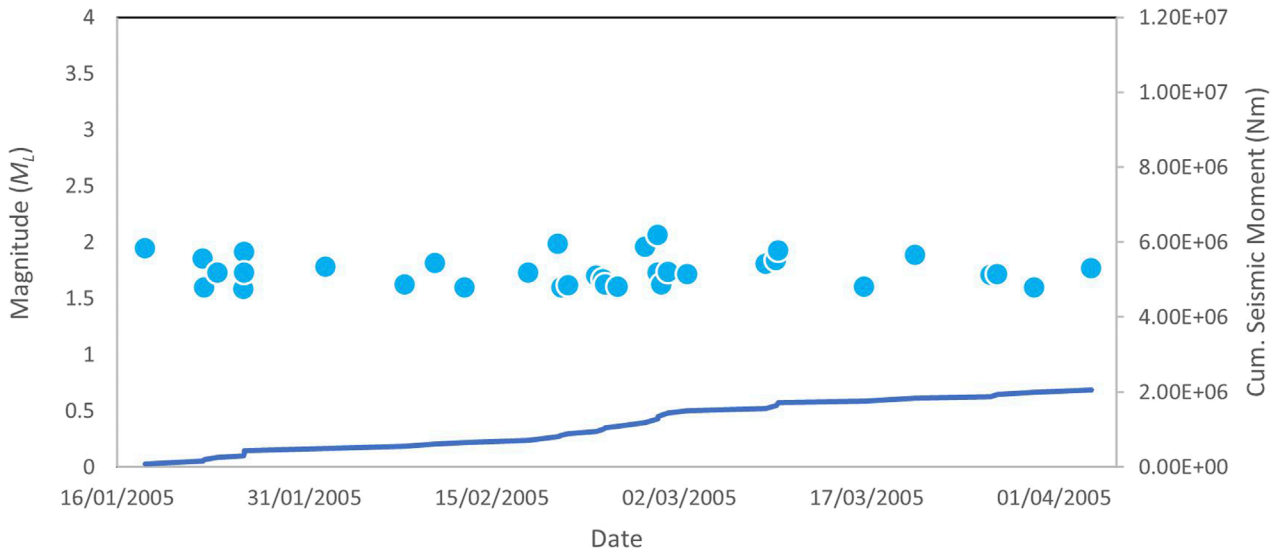


FIGURE 11
Total cumulative M_0 of earthquake events in C4-2005 for the temporal cluster detected by DBSCAN.

5.7.1 Intrusion volume estimates

Earthquake swarms occurring in proximity to a volcano may be a proxy for magma movement or a change in reservoir pressurisation (Ebmeier et al., 2016). The ability to make real-time forecasts based volcano-tectonic earthquake swarms remains a key challenge in volcano monitoring (Roman and Power, 2011; White and McCausland, 2016). Experimental evidence from water injections in deep wells, showed that total seismic moment (ΣM_0) is proportional to a cumulative change in intrusive volume (McGarr, 1976). White and McCausland (2016) translated these properties to infer that the volume *(V m^3)* of a magmatic intrusion should be proportional to the ΣM_0 *(Nm)* equivalent of a volcano-tectonic earthquake swarm:

$$\log_{10} V = 0.77 \times \log_{10} \sum M_0 - 5.32$$

Using this equation, we calculated the intruded volume of magma for the C5-1995 sequence as 0.0012 km^3 for a $\log_{10}$ cumulative moment of 7.02 Nm. C5-2001 had an estimated intruded volume of 0.00085 km^3 for a $\log_{10}$ cumulative moment of 6.81 Nm. A study by Meyer et al. (2021) approximated that seismic moment release should be proportionally larger for swarms that eventually lead to an eruption, which is not the case for C5-1995 and C5-2001. However, the study also stated that uncertainties in the measurements would be too large to make meaningful forecasts. Furthermore, when applied to known magmatic intrusions, it was found that the main controlling factor on the seismic energy release was the background seismicity values for that region, rather than the extent of the dike (Pedersen et al., 2007).

To summarise, a vertically and laterally interconnected system with periodic feeding from a deeper source and magma mixing during migration from intrusion to the chamber is a process suggested in other volcanic arc settings and has been linked to distal earthquake swarms in the months preceding eruptions (e.g., Albino et al., 2019). Studies concerning the plumbing system models at Mt. Ruapehu support the hypothesis described by Hurst et al. (2018), that laterally interconnected intrusions could be feeding a small, shallow reservoir (Kilgour et al., 2013). This may exert pressure on the surrounding rocks causing earthquake swarms at considerable distances from the summit. This implies that an earthquake swarm in the Erua region in April 1995 could indeed be interpreted as a precursory warning sign for the 1995 eruption. In the pursuit of identifying viable volcanic eruption precursors for eruption forecasting using machine learning, Ardid et al. (2022) suggested precursory signals should correlate across multiple eruptions and be absent from non-eruptive repose periods. However, the Erua sequence which preceded the 1995 major eruption is either absent, or not documented, prior to the other magmatic eruptions and the 2006 and 2007 phreatic eruptions, which is consistent with Keats et al. (2011). We can also say with confidence that swarm-like activity has occurred in the same region suggested to be an eruption precursor (Hurst et al., 2018) on more than one occasion, during period of non-eruptive activity, with the later 2001 (C5-2001) swarm not preceding an eruption. As for the uniqueness of the 1995 Erua swarm, C5-1995 and C5-2001 share similarities in number of events, magnitudes and geometry of the sequence with the Raurimu fault. However, it could be argued the temporal evolution of the C5-2001 sequence perhaps exhibits more swarm-like behaviour. We have seen evidence of other seismic sequences that have reoccurred in the seismically active west region of Mt. Ruapehu, which have no correlation to documented volcanic activity. Therefore, it is impractical to rely on such earthquake sequences for meaningful eruption forecasts at Mt. Ruapehu.

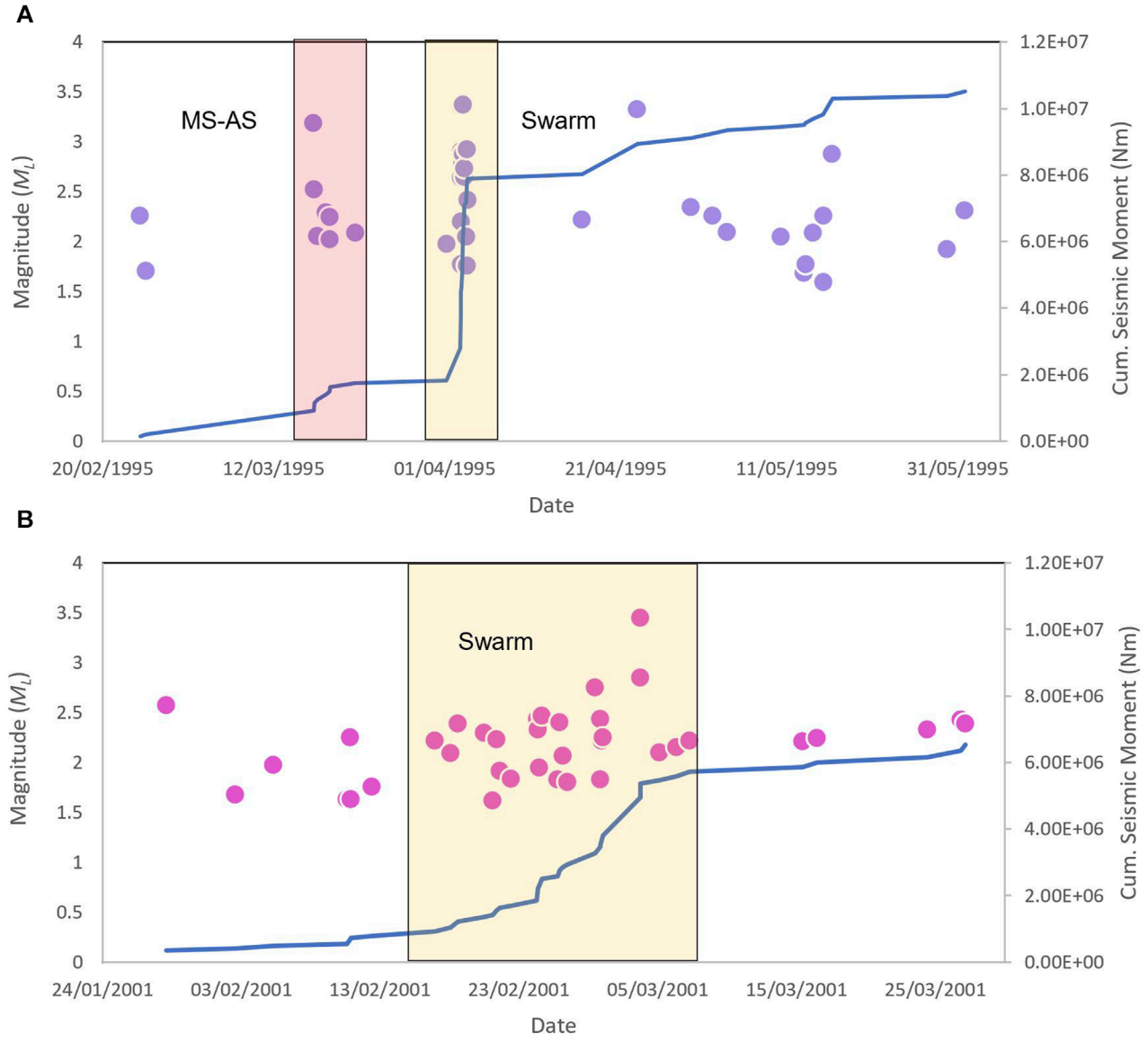


FIGURE 12
Total cumulative M_0 of earthquake events in C5-1995 **(A)** occurred in April 1995 with a maximum magnitude event M_L 3.4 and C5-2001 **(B)** occurred February/March 2001 with a maximum event also M_L 3.4.

6 Summary

This study has successfully employed simple unsupervised machine learning clustering algorithms to detect earthquake sequences at Mt. Ruapehu, without *a priori* knowledge of the system. One of the key motivations for this research was to understand whether the swarm near Erua, which preceded the 1995 major eruption, is a unique signal and one which can be used for mid-term forecasting at Mt. Ruapehu. For this to be true, it should satisfy forecasting constraints, where a signal must be present before an event, but not during periods of no volcanic activity (e.g., Kilburn, 2018; Ardid et al., 2022). Given the evidence in Cluster 5 of a second swarm that did not precede an volcanic activity in 2001, we cannot conclude, based on earthquake swarm characteristics alone, that the Erua swarm is a viable forecasting method at Mt. Ruapehu. We have also discovered evidence for other earthquake sequences, which look similar to the signal detected prior to the 1995 major eruption at Erua during periods of inactivity. We also believe that widening the study area would yield even more earthquake sequences and that these signals are fairly common in the region. Therefore, we have shown that the earthquake swarm near the town of Erua (e.g., Hurst and McGinty, 1999; Hurst et al., 2018) is not a unique signal at Mt. Ruapehu.

6.1 Thoughts going forward

The results of this study have significant implications for understanding the seismic behavior the west of Mt. Ruapehu. By employing density-based clustering algorithms, we have gained

a more comprehensive perspective on the spatial and temporal patterns of earthquake activity. Spatial and temporal clustering has shown promise as a quick and effective way of detecting earthquake swarms, particularly within noisy data sets. Retrospective analysis of earthquake catalogues containing many thousands of events can be filtered into high-density regions using a couple of lines of code to call the HDBSCAN algorithm without the necessity of prior domain knowledge. These regions can then be analysed for temporal bursts in seismic activity by using DBSCAN, using the ε distance function to calculate distance as time. This is useful for detecting swarms; however it requires the statistical calculation of minimum events for a given time window, which can vary for different systems. Furthermore, the parameter selection for the HDBSCAN can be very sensitive, producing very different results for different minimum cluster size parameters. However, the ability to automatically identify spikes in earthquake rate over different periods of time revealed interesting results, which were less sensitive to the initial parameters, and correlated well with the cumulative seismic rate. Further research could explore including magnitude to the temporal analysis to distinguish between swarm-like and MS-AS sequences. This may help focus on further refining our understanding of the relationships between seismic clusters, volcanic processes, and eruption precursors. Additionally, the application of advanced machine learning may provide deeper insights into the complex interplay between seismicity and volcanic activity in the Mt. Ruapehu region.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: https://quakesearch.geonet.org.nz/.

Author contributions

SM: Writing–original draft. JJ: Conceptualization, Supervision, Writing–review and editing. BM: Methodology, Validation, Writing–review and editing, Supervision. JL: Methodology, Validation, Writing–review and editing, Supervision.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. The authors would like to express their gratitude to ARIES and NERC for the financial support provided for this research; grant number NE/S007334/1.

Acknowledgments

The authors would like to thank GeoNet for the open source data which supported the analysis in this research.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/feart.2024.1343874/full#supplementary-material

References

Albino, F., Biggs, J., and Syahbana, D. K. (2019). Dyke intrusion between neighbouring arc volcanoes responsible for 2017 pre-eruptive seismic swarm at Agung. *Nat. Commun.* 10, 748. doi:10.1038/s41467-019-08564-9

Angmo, R., Aggarwal, N., Mangat, V., Lal, A., and Kaur, S. (2021). An improved clustering approach for identifying significant locations from spatio-temporal data. *Wirel. Personal. Commun.* 121, 985–1009. doi:10.1007/s11277-021-08668-w

Ardid, A., Dempsey, D., Caudron, C., and Cronin, S. (2022). Seismic precursors to the Whakaari 2019 phreatic eruption are transferable to other eruptions and volcanoes. *Nat. Commun.* 13, 2002. doi:10.1038/s41467-022-29681-y

Bormann, P., and Di Giacomo, D. (2011). The moment magnitude Mw and the energy magnitude Me: common roots and differences. *J. Seismol.* 15, 411–427. doi:10.1007/s10950-010-9219-2

Bryan, C., and Sherburn, S. (1999). Seismicity associated with the 1995–1996 eruptions of Ruapehu volcano, New Zealand: narrative and insights into physical processes. *J. Volcanol. Geotherm. Res.* 90, 1–18. doi:10.1016/S0377-0273(99)00016-5

Cameron, E., Gamble, J., Price, R., Smith, I., McIntosh, W., and Gardner, M. (2010). The petrology, geochronology and geochemistry of Hauhungatahi volcano, S.W. Taupo Volcanic Zone. *J. Volcanol. Geotherm. Res.* 190, 179–191. doi:10.1016/j.jvolgeores.2009.07.002

Campello, R. J. G. B., Moulavi, D., and Sander, J. (2013). "Density-based clustering based on hierarchical density estimates," in *Advances in knowledge discovery and data mining*. Editors D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, et al. (Berlin, Heidelberg: Springer Berlin Heidelberg), 7819, 160–172. doi:10.1007/978-3-642-37456-2_14

Caricchi, L., Townsend, M., Rivalta, E., and Namiki, A. (2021). The build-up and triggers of volcanic eruptions. *Nat. Rev. Earth Environ.* 2, 458–476. doi:10.1038/s43017-021-00174-8

Carniel, R., Jolly, A. D., and Barbui, L. (2013). Analysis of phreatic events at Ruapehu volcano, New Zealand using a new SOM approach. *J. Volcanol. Geotherm. Res.* 254, 69–79. doi:10.1016/j.jvolgeores.2012.12.026

Carrivick, J. L., Manville, V., and Cronin, S. J. (2008). A fluid dynamics approach to modelling the 18th March 2007 lahar at Mt. Ruapehu, New Zealand. *Bull. Volcanol.* 71, 153–169. doi:10.1007/s00445-008-0213-2

Chouet, B. A., and Matoza, R. S. (2013). A multi-decadal view of seismic methods for detecting precursors of magma movement and eruption. *J. Volcanol. Geotherm. Res.* 252, 108–175. doi:10.1016/j.jvolgeores.2012.11.013

Christenson, B. W., and Wood, C. P. (1993). Evolution of a vent-hosted hydrothermal system beneath Ruapehu crater lake, New Zealand. *Bull. Volcanol.* 55, 547–565. doi:10.1007/BF00301808

Conway, C. E., Leonard, G. S., Townsend, D. B., Calvert, A. T., Wilson, C. J., Gamble, J. A., et al. (2016). A high-resolution 40Ar/39Ar lava chronology and edifice construction history for Ruapehu volcano, New Zealand. *J. Volcanol. Geotherm. Res.* 327, 152–179. doi:10.1016/j.jvolgeores.2016.07.006

De La Cruz-Reyna, S., Tárraga, M., Ortiz, R., and Martínez-Bringas, A. (2010). Tectonic earthquakes triggering volcanic seismicity and eruptions. Case studies at Tungurahua and Popocatépetl volcanoes. *J. Volcanol. Geotherm. Res.* 193, 37–48. doi:10.1016/j.jvolgeores.2010.03.005

Ebmeier, S. K., Elliott, J. R., Nocquet, J.-M., Biggs, J., Mothes, P., Jarrín, P., et al. (2016). Shallow earthquake inhibits unrest near Chiles–Cerro Negro volcanoes, Ecuador–Colombian border. *Earth Planet. Sci. Lett.* 450, 283–291. doi:10.1016/j.epsl.2016.06.046

Ester, M., Kriegel, H.-P., Sander, J., and Xu, X. (1996). "A density-based algorithm for discovering clusters in large spatial databases with noise," in Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, Portland, Oregon, August 2-4, 1996, 226–231.

Farías, C., Lazo, J., Basualto, D., Saavedra, M., Muñoz-Quiroz, F., Zúñiga-Urrea, L., et al. (2023). One decade of b-value variations from volcano-tectonic seismicity as an early indicator of episodes of crisis in a volcano: the case of Copahue, Southern Andes. *Front. Earth Sci.* 11, 1181177. doi:10.3389/feart.2023.1181177

Gamble, J. A., Price, R. C., Smith, I. E., McIntosh, W. C., and Dunbar, N. W. (2003). 40Ar/39Ar geochronology of magmatic activity, magma flux and hazards at Ruapehu volcano, Taupo Volcanic Zone, New Zealand. *J. Volcanol. Geotherm. Res.* 120, 271–287. doi:10.1016/S0377-0273(02)00407-9

Geffers, G.-M., Main, I. G., and Naylor, M. (2022). Biases in estimating *b* -values from small earthquake catalogues: how high are high *b* -values? *Geophys. J. Int.* 229, 1840–1855. doi:10.1093/gji/ggac028

Gudmundsson, A. (2020). *Volcanotectonics: understanding the structure, deformation and dynamics of volcanoes*. 1 edn. Cambridge: Cambridge University Press. doi:10.1017/9781139176217

Hayes, G., Reyners, M., and Stuart, G. (2004). The Waiouru, New Zealand, earthquake swarm: persistent mid crustal activity near an active volcano. *Geophys. Res. Lett.* 31, L19613. doi:10.1029/2004GL020709

Hill, D. P., Pollitz, F., and Newhall, C. (2002). Earthquake–Volcano interactions. *Phys. Today* 55, 41–47. doi:10.1063/1.1535006

Houghton, B. F., Latter, J. H., and Hackett, W. R. (1987). Volcanic hazard assessment for Ruapehu composite volcano, taupo volcanic zone, New Zealand. *Bull. Volcanol.* 49, 737–751. doi:10.1007/BF01079825

Hurst, A., and McGinty, P. (1999). Earthquake swarms to the west of Mt Ruapehu preceding its 1995 eruption. *J. Volcanol. Geotherm. Res.* 90, 19–28. doi:10.1016/S0377-0273(99)00019-0

Hurst, A. W., and Vandemeulebrouck, J. (1996). Acoustic noise and temperature monitoring of the crater lake of mount Ruapehu volcano. *J. Volcanol. Geotherm. Res.* 71, 45–51. doi:10.1016/0377-0273(95)00060-7

Hurst, T., Kilgour, G., and Hamling, I. (2018). Magmatic triggering of earthquakes on distal faults as a potential medium-term warning signal from Ruapehu Volcano. *Geophys. Res. Lett.* 45. doi:10.1029/2018GL080677

Johnston, D. M., Houghton, B., Neall, V., Ronan, K., and Paton, D. (2000). Impacts of the 1945 and 1995-1996 Ruapehu eruptions, New Zealand: an example of increasing societal vulnerability. *Geol. Soc. Am. Bull.* 112, 720–726. doi:10.1130/0016-7606(2000)112<0720:iotare>2.3.co;2

Jolly, A. D., Sherburn, S., Jousset, P., and Kilgour, G. (2010). Eruption source processes derived from seismic and acoustic observations of the 25 September 2007 Ruapehu eruption—north Island, New Zealand. *J. Volcanol. Geotherm. Res.* 191, 33–45. doi:10.1016/j.jvolgeores.2010.01.009

Jones, J. (2005). Mount Hood earthquake activity: volcanic or tectonic origins? *Bull. Seismol. Soc. Am.* 95, 818–832. doi:10.1785/0120040019

Karami, A., and Johansson, R. (2014). Choosing DBSCAN parameters automatically using differential evolution. *Int. J. Comput. Appl.* 91, 1–11. doi:10.5120/15890-5059

Keats, B. S., Johnson, J. H., and Savage, M. K. (2011). The Erua earthquake cluster and seismic anisotropy in the Ruapehu region, New Zealand: the Erua earthquake cluster. *Geophys. Res. Lett.* 38. doi:10.1029/2011GL049014

Kilburn, C. (2012). Precursory deformation and fracture before brittle rock failure and potential application to volcanic unrest: precursors to brittle rock failure. *J. Geophys. Res. Solid Earth* 117. doi:10.1029/2011JB008703

Kilburn, C. R. J. (2018). Forecasting volcanic eruptions: beyond the failure forecast method. *Front. Earth Sci.* 6. doi:10.3389/feart.2018.00133

Kilgour, G., Blundy, J., Cashman, K., and Mader, H. M. (2013). Small volume andesite magmas and melt–mush interactions at Ruapehu, New Zealand: evidence from melt inclusions. *Contributions Mineralogy Petrology* 166, 371–392. doi:10.1007/s00410-013-0880-7

Kilgour, G., Manville, V., Pasqua, F. D., Graettinger, A., Hodgson, K., and Jolly, G. (2010). The 25 September 2007 eruption of Mount Ruapehu, New Zealand: directed ballistics, surtseyan jets, and ice-slurry lahars. *J. Volcanol. Geotherm. Res.* 191, 1–14. doi:10.1016/j.jvolgeores.2009.10.015

Kilgour, G., Saunders, K., Blundy, J., Cashman, K., Scott, B., and Miller, C. (2014). Timescales of magmatic processes at Ruapehu volcano from diffusion chronometry and their comparison to monitoring data. *J. Volcanol. Geotherm. Res.* 288, 62–75. doi:10.1016/j.jvolgeores.2014.09.010

Langridge, R., Ries, W., Litchfield, N., Villamor, P., Van Dissen, R., Barrell, D., et al. (2016). The New Zealand active faults database. *N. Z. J. Geol. Geophys.* 59, 86–96. doi:10.1080/00288306.2015.1112818

Lanza, F., Roman, D. C., Power, J. A., Thurber, C. H., and Hudson, T. (2022). Complex magmatic-tectonic interactions during the 2020 Makushin Volcano, Alaska, earthquake swarm. *Earth Planet. Sci. Lett.* 587, 117538. doi:10.1016/j.epsl.2022.117538

Leonard, G. S., Cole, R. P., Christenson, B. W., Conway, C. E., Cronin, S. J., Gamble, J. A., et al. (2021). Ruapehu and Tongariro stratovolcanoes: a review of current understanding. *N. Z. J. Geol. Geophys.* 64, 389–420. doi:10.1080/00288306.2021.1909080

Manville, V., Hodgson, K. A., and Nairn, I. A. (2007). A review of break-out floods from volcanogenic lakes in New Zealand. *N. Z. J. Geol. Geophys.* 50, 131–150. doi:10.1080/00288300709509826

McGarr, A. (1976). Seismic moments and volume changes. *J. Geophys. Res.* 81, 1487–1494. doi:10.1029/JB081i008p01487

McNutt, S. R. (2002). "25 - volcano seismology and monitoring for eruptions," in *International Geophysics*. Editors W. H. K. Lee, H. Kanamori, P. C. Jennings, and C. Kisslinger (Cambridge, Massachusetts, United States: Academic Press). *Part A*. 383–cp1. doi:10.1016/S0074-6142(02)80228-5

McNutt, S. R. (2005). Volcanic seismology. *Annu. Rev. Earth Planet. Sci.* 33, 461–491. doi:10.1146/annurev.earth.33.092203.122459

McNutt, S. R., and Roman, D. C. (2015). "Volcanic seismicity," in *The encyclopedia of volcanoes* (Amsterdam, Netherlands: Elsevier), 1011–1034. doi:10.1016/B978-0-12-385938-9.00059-6

Meyer, K., Biggs, J., and Aspinall, W. (2021). A Bayesian reassessment of the relationship between seismic moment and magmatic intrusion volume during volcanic unrest. *J. Volcanol. Geotherm. Res.* 419, 107375. doi:10.1016/j.jvolgeores.2021.107375

Nakagawa, M., Wada, K., Thordarson, T., Wood, C. P., and Gamble, J. A. (1999). Petrologic investigations of the 1995 and 1996 eruptions of Ruapehu volcano, New Zealand: formation of discrete and small magma pockets and their intermittent discharge. *Bull. Volcanol.* 61, 15–31. doi:10.1007/s004450050259

Omori, F. (1895). On the aftershocks of earthquakes. *J. Coll. Sci.*

Pavlenko, V. A., and Zavyalov, A. D. (2022). Comparative analysis of the methods for estimating the magnitude of completeness of earthquake catalogs. *Izvestiya, Phys. Solid Earth* 58, 89–105. doi:10.1134/S1069351322010062

Pedersen, R., Sigmundsson, F., and Einarsson, P. (2007). Controlling factors on earthquake swarms associated with magmatic intrusions; Constraints from Iceland. *J. Volcanol. Geotherm. Res.* 162, 73–80. doi:10.1016/j.jvolgeores.2006.12.010

Pesicek, J. D., Wellik, J. J., Prejean, S. G., and Ogburn, S. E. (2018). Prevalence of seismic rate anomalies preceding volcanic eruptions in Alaska. *Front. Earth Sci.* 6, 100. doi:10.3389/feart.2018.00100

Petersen, G. M., and Pankow, K. L. (2023). Small-magnitude seismic swarms in Central Utah (US): interactions of regional tectonics, local structures and hydrothermal systems. *Geochem. Geophys. Geosystems* 24, e2023GC010867. doi:10.1029/2023GC010867

Ramis, R., Garcia, A., Marrero, J., De La Cruz-Reyna, S., Carniel, R., and Vila, J. (2018). Volcanic and volcano-tectonic activity forecasting: a review on seismic approaches. *Ann. Geophys.* 61, 9. doi:10.4401/ag-7655

Rhoades, D. A., Christophersen, A., Bourguignon, S., Ristau, J., and Salichon, J. (2021). A depth-dependent local magnitude scale for New Zealand earthquakes consistent with moment magnitude. *Bull. Seismol. Soc. Am.* 111, 1056–1066. doi:10.1785/0120200252

Ristau, J. (2009). Comparison of magnitude estimates for New Zealand earthquakes: moment magnitude, local Magnitude, and teleseismic body-wave magnitude. *Bull. Seismol. Soc. Am.* 99, 1841–1852. doi:10.1785/0120080237

Roberts, N. S., Bell, A. F., and Main, I. G. (2015). Are volcanic seismic b -values high, and if so when? *J. Volcanol. Geotherm. Res.* 308, 127–141. doi:10.1016/j.jvolgeores.2015.10.021

Roman, D. C., and Power, J. A. (2011). Mechanism of the 1996–97 non-eruptive volcano-tectonic earthquake swarm at Iliamna Volcano, Alaska. *Bull. Volcanol.* 73, 143–153. doi:10.1007/s00445-010-0439-7

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* 20, 53–65. doi:10.1016/0377-0427(87)90125-7

Schaefer, L. N., Kennedy, B. M., Villeneuve, M. C., Cook, S. C., Jolly, A. D., Keys, H. J., et al. (2018). Stability assessment of the Crater Lake/Te Wai-ā-moe overflow channel at Mt. Ruapehu (New Zealand), and implications for volcanic lake break-out triggers. *J. Volcanol. Geotherm. Res.* 358, 31–44. doi:10.1016/j.jvolgeores.2018.06.011

Scitovski, S. (2018). A density-based clustering algorithm for earthquake zoning. *Comput. Geosciences* 110, 90–95. doi:10.1016/j.cageo.2017.08.014

Seropian, G., Kennedy, B. M., Walter, T. R., Ichihara, M., and Jolly, A. D. (2021). A review framework of how earthquakes trigger volcanic eruptions. *Nat. Commun.* 12, 1004–1013. doi:10.1038/s41467-021-21166-8

Sherburn, S., Bryan, C., Hurst, A., Latter, J., and Scott, B. (1999). Seismicity of Ruapehu volcano, New Zealand, 1971–1996: a review. *J. Volcanol. Geotherm. Res.* 88, 255–278. doi:10.1016/S0377-0273(99)00014-1

Shutaywi, M., and Kachouie, N. N. (2021). Silhouette analysis for performance evaluation in machine learning with applications to clustering. *Entropy* 23, 759. doi:10.3390/e23060759

Sibson, R. (1992). Implications of fault-valve behaviour for rupture nucleation and recurrence. *Tectonophysics* 211, 283–293. doi:10.1016/0040-1951(92)90065-E

Starczewski, A., Goetzen, P., and Er, M. J. (2020). A new method for automatic determining of the DBSCAN parameters. *J. Artif. Intell. Soft Comput. Res.* 10, 209–221. doi:10.2478/jaiscr-2020-0014

Thelen, W. A., Malone, S. D., and West, M. E. (2010). Repose time and cumulative moment magnitude: a new tool for forecasting eruptions? REPOSE TIME. *Geophys. Res. Lett.* 37. doi:10.1029/2010GL044194

Vidale, J. E., Boyle, K. L., and Shearer, P. M. (2006). Crustal earthquake bursts in California and Japan: their patterns and relation to volcanoes. *Geophys. Res. Lett.* 33, L20313. doi:10.1029/2006GL027723

Vidale, J. E., and Shearer, P. M. (2006). A survey of 71 earthquake bursts across southern California: exploring the role of pore fluid pressure fluctuations and aseismic slip as drivers. *J. Geophys. Res. Solid Earth* 111, 2005JB004034. doi:10.1029/2005JB004034

Wang, F., Franco-Penya, H.-H., Kelleher, J. D., Pugh, J., and Ross, R. (2017). "An analysis of the application of simplified silhouette to the evaluation of k-means clustering validity," in *Machine learning and data mining in pattern recognition*. Editor P. Perner (Cham: Springer International Publishing), 291–305. Lecture Notes in Computer Science. doi:10.1007/978-3-319-62416-7_21

White, R., and McCausland, W. (2016). Volcano-tectonic earthquakes: a new tool for estimating intrusive volumes and forecasting eruptions. *J. Volcanol. Geotherm. Res.* 309, 139–155. doi:10.1016/j.jvolgeores.2015.10.020

Wiemer, S. (2000). Minimum magnitude of completeness in earthquake catalogs: examples from Alaska, the western United States, and Japan. *Bull. Seismol. Soc. Am.* 90, 859–869. doi:10.1785/0119990114

Yarce, J., Sheehan, A. F., Nakai, J. S., Schwartz, S. Y., Mochizuki, K., Savage, M. K., et al. (2019). Seismicity at the northern hikurangi margin, New Zealand, and investigation of the potential spatial and temporal relationships with a shallow slow slip event. *J. Geophys. Res. Solid Earth* 124, 4751–4766. doi:10.1029/2018JB017211

Yoshida, K., and Hasegawa, A. (2018). Hypocenter migration and seismicity pattern change in the yamagata-fukushima border, NE Japan, caused by fluid movement and pore pressure variation. *J. Geophys. Res. Solid Earth* 123, 5000–5017. doi:10.1029/2018JB015468

Yoshida, K., Uno, M., Matsuzawa, T., Yukutake, Y., Mukuhira, Y., Sato, H., et al. (2023). Upward earthquake swarm migration in the northeastern noto peninsula, Japan, initiated from a deep ring-shaped cluster: possibility of fluid leakage from a hidden magma system. *J. Geophys. Res. Solid Earth* 128, e2022JB026047. doi:10.1029/2022JB026047

frontiers | Frontiers in Earth Science

TYPE Methods
PUBLISHED 26 April 2024
DOI 10.3389/feart.2024.1345104

OPEN ACCESS

EDITED BY
Carmen Benítez,
Universidad de Granada, Spain

REVIEWED BY
Francesco Massimetti,
University of Turin, Italy
Manuel Titos,
Icelandic Meteorological Office, Iceland
Michael Ramsey,
University of Pittsburgh, United States

*CORRESPONDENCE
Pablo Saunders-Shultz,
csaundersshultz@gmail.com

RECEIVED 27 November 2023
ACCEPTED 05 April 2024
PUBLISHED 26 April 2024

CITATION
Saunders-Shultz P, Lopez T, Dietterich H and Girona T (2024), Automatic identification and quantification of volcanic hotspots in Alaska using HotLINK: the hotspot learning and identification network.
Front. Earth Sci. 12:1345104.
doi: 10.3389/feart.2024.1345104

Automatic identification and quantification of volcanic hotspots in Alaska using HotLINK: the hotspot learning and identification network

Pablo Saunders-Shultz[1]*, Taryn Lopez[1], Hannah Dietterich[2] and Társilo Girona[1]

[1]Geophysical Institute, Alaska Volcano Observatory, University of Alaska Fairbanks, Fairbanks, AK, United States, [2]U.S. Geological Survey, Alaska Volcano Observatory, Anchorage, AK, United States

An increase in volcanic thermal emissions can indicate subsurface and surface processes that precede, or coincide with, volcanic eruptions. Space-borne infrared sensors can detect hotspots—defined here as localized volcanic thermal emissions—in near-real-time. However, automatic hotspot detection systems are needed to efficiently analyze the large quantities of data produced. While hotspots have been automatically detected for over 20 years with simple thresholding algorithms, new computer vision technologies, such as convolutional neural networks (CNNs), can enable improved detection capabilities. Here we introduce HotLINK: the Hotspot Learning and Identification Network, a CNN trained to detect hotspots with a dataset of ~3,800 satellite-based, Visible Infrared Imaging Radiometer Suite (VIIRS) images from Mount Veniaminof and Mount Cleveland volcanoes, Alaska. We find that our model achieves an accuracy of 96% (F1-score 0.92) when evaluated on ~1,700 unseen images from the same volcanoes, and 95% (F1-score 0.67) when evaluated on ~3,000 images from six additional Alaska volcanoes (Augustine Volcano, Bogoslof Island, Okmok Caldera, Pavlof Volcano, Redoubt Volcano, Shishaldin Volcano). In comparison with an existing threshold-based hotspot detection algorithm, MIROVA (Coppola et al., Geological Society, London, Special Publications, 2016, 426, 181–205), our model detects 22% more hotspots and produces 12% fewer false positives. Additional testing on ~700 labeled Moderate Resolution Imaging Spectroradiometer (MODIS) images from Mount Veniaminof demonstrates that our model is applicable to this sensor's data as well, achieving an accuracy of 98% (F1-score 0.95). We apply HotLINK to 10 years of VIIRS data and 22 years of MODIS data for the eight aforementioned Alaska volcanoes and calculate the radiative power of detected hotspots. From these time series we find that HotLINK accurately characterizes background and eruptive periods, similar to MIROVA, but also detects more subtle warming signals, potentially related to volcanic unrest. We identify three advantages to our model over its predecessors: 1) the ability to detect more subtle volcanic hotspots and produce fewer false positives, especially in daytime images; 2) probabilistic predictions provide a measure of detection confidence; and 3) its transferability, i.e., the successful application to multiple sensors and multiple volcanoes

without the need for threshold tuning, suggesting the potential for global application.

KEYWORDS

thermal remote sensing, global volcano monitoring, machine learning, neural network, eruption forecasting, VIIRS, MODIS, MIROVA

1 Introduction

Volcanic eruptions pose hazards to human life and society (Loughlin et al., 2015). To mitigate these hazards, volcano monitoring agencies aim to detect signs of unrest and eruption as early as possible. Local monitoring stations and remote satellite observations are commonly used to monitor volcanic unrest (e.g., Dehn et al., 2000; Cameron et al., 2018; Girona et al., 2021). Here we will focus on one satellite-based approach to monitor thermal unrest: detecting localized volcanic heat emissions, also referred to as volcanic hotspots. In a single satellite image, hotspots may be identified as a few pixels of elevated infrared radiance caused by relatively high temperature volcanic features. Hotspots may be produced by various types of volcanic activity, including lava flows (Dehn et al., 2000; Hirn et al., 2009; Blackett, 2013; Harris, 2013; Wright, 2016), explosive and strombolian activity (Harris and Stevenson, 1997; Coppola et al., 2012; Coppola et al., 2014), dome growth (Carter et al., 2007; Ramsey et al., 2012; Coppola et al., 2022), degassing of a hot vent or fumarole field (Oppenhemier et al., 1993; Harris and Stevenson, 1997; Blackett, 2013; Laiolo et al., 2017), or increased surface meltwater in the case of glaciated volcanoes (Pieri and Abrams, 2005; Blackett, 2013; Bleick et al., 2013; Reath et al., 2016). Therefore, monitoring changes in hotspot activity can provide key insights into a volcano's behavior by indicating the presence of thermal volcanic features and characterizing them over time. Due to the utility of these observations, thermal satellite data are used by volcano observatories as part of their daily monitoring operations (Dehn et al., 2000; Dehn et al., 2002; Harris et al., 2016; Harris et al., 2017; Cameron et al., 2018; Coombs et al., 2018; Coppola et al., 2020; Pritchard et al., 2022; Chevrel et al., 2023). Automating the detection and quantification of volcanic hotspots can provide near-real time information to volcano observatory scientists to inform decision-making and provide a mechanism to generate long time series of thermal activity for volcanoes around the world. Time series observations are useful for determining baseline activity, identifying periods of volcanic unrest, characterizing the thermal evolution of ongoing eruptions, and retrospectively studying eruptive histories and processes (Dehn et al., 2002; Wright, 2016; Girona et al., 2021; Chevrel et al., 2023; Coppola et al., 2023).

Surface hotspots will result in increased spectral radiance (Wm^{-2} sr^{-1} μm^{-1}) in both Mid-Infrared (MIR, 3–5 μm) and Thermal-Infrared (TIR, 5–20 μm) wavelengths (Harris, 2013). This behavior is characterized by Planck's Law, which states that as the temperature of a blackbody increases, the spectrum of energy it emits will increase in radiance, and the peak radiance will shift to shorter wavelengths (Planck, 1914). Therefore, a volcanic hotspot can be identified by an elevated TIR radiance above background and an even greater signal above background in MIR radiance (e.g., Blackett, 2013; Blackett, 2017). For especially hot surfaces (>950 K), the peak radiance emission is in the shortwave infrared (SWIR, 1.4–3 μm) part of the spectrum. The distinct features produced by hotspots in MIR and TIR bands have been exploited to automate their detection by different algorithms (Higgins and Harris, 1997; Pergola et al., 2004; Wright et al., 2004; Ganci et al., 2011; Coppola et al., 2016; Gouhier et al., 2016; Lombardo, 2016; Valade et al., 2019; Castaño et al., 2020; Genzano et al., 2020; Layana et al., 2020; Massimetti et al., 2020; Corradino et al., 2023; Ramsey et al., 2023).

One of the first algorithms to automate volcanic hotspot detection, MODVOLC (Wright et al., 2004), applies a threshold to the Normalized Thermal Index (NTI), constructed from radiance values of MIR and TIR bands:

$$NTI = \frac{MIR - TIR}{MIR + TIR} \quad (1)$$

MODVOLC flags nighttime pixels with NTI greater than −0.8, and daytime pixels with NTI greater than −0.55 as hotspots, because of the large impact of solar reflections and heating on daytime images (Wright et al., 2004; Wright, 2016). These thresholds were found by manual analysis of histograms of NTI at 100 locations to minimize false positive detections (Wright et al., 2004). Another popular approach, the MIROVA algorithm, incorporates a new spectral index in addition to NTI, and spatially filters both spectral indices to improve hotspot detections (Coppola et al., 2016, further details on the MIROVA algorithm and its application in this study can be found in Section 2.4). While these and other algorithms define their own band indices, ratios, spatial filters, and corrections in order to accentuate the differences between hotspot and background pixels, each of these approaches use thresholding to automate the flagging of hotspot pixels. The ability of each algorithm to distinguish hotspots from background pixels depends on how successful their index is in separating the two classes and the accuracy and precision of the threshold set for that index. MODVOLC and MIROVA have successfully generated decades long time series of hotspots at volcanoes across the globe, which has allowed for detection and monitoring of eruptions in near-real time and the study of thermal output from different eruptions and volcanic systems (Wright, 2016; Coppola et al., 2023). Still, both datasets contain false detections and missed hotspots, due to the fact that there will inevitably be non-volcanic thermal signals exceeding the set thresholds, and real volcanic signals lower than the detection thresholds.

In this paper, we aim to enhance the automatic detection of volcanic hotspots in infrared satellite data by applying a convolutional neural network (CNN). CNNs are a machine learning technique commonly employed for image analysis (LeCun et al., 2010). They have been applied to numerous problems in the field of computer vision, including to identify cancer cells in MRIs (El Adoui et al., 2019), facial unlock in cellphones (Apple, 2023), and reverse image search algorithms (Wan et al., 2014). In our approach the use of CNNs can be conceptualized as identifying

hotspots based on what they look like, rather than by thresholding a particular thermal index. While previous methods employ human created indices to highlight hotspot pixels, this approach is data-driven—deriving the spectral and spatial characteristics that define hotspots from a large labeled dataset of the hotspots themselves. In this way, the CNN mimics the pattern recognition of a human analyst.

The type of CNN used here is a U-net (Ronneberger et al., 2015). U-nets are a popular architecture for image segmentation, or tasks in which a prediction is made for each pixel in order to both detect and locate features of interest. A U-net was successfully applied to volcanic hotspot detection in data from the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER), achieving a high accuracy (Corradino et al., 2023). In this study, we apply a similar method to data from the Visible Infrared Imaging Radiometer Suite (VIIRS) and Moderate Resolution Imaging Spectroradiometer (MODIS) satellite sensors. Although ASTER has a finer spatial resolution (90 m in TIR bands, used in Corradino et al., 2023) than VIIRS (375 m) and MODIS (1,000 m), we chose to apply this methodology to VIIRS and MODIS data due to their high acquisition rates and MIR and TIR bands. High acquisition rates result in more frequent opportunities to detect and track changes in volcanic unrest. At the time of this writing, VIIRS sensors provide coverage of each Alaska volcano 8–15 times per day, while MODIS sensors provide coverage 1–6 times per day. Volcanoes at higher latitudes are imaged more frequently than those at lower latitudes by the polar-orbiting satellites used here. Detection frequency will increase in the future with the planned launch of additional VIIRS instruments. Although MODIS has a coarser spatial resolution than VIIRS, it has a longer operational history (satellites Terra and Aqua launched in 1999 and 2002, respectively), so it is useful for studying eruptions prior to the launch of VIIRS (Suomi-National Polar-Orbiting Partnership, SNPP, launched in 2011, and National Oceanic and Atmospheric Administration 20, NOAA-20, launched in 2017).

We incorporate data from eight Alaska volcanoes with a wide range of volcanic thermal signals to develop and test our model for broad applicability to many volcanic settings (Table 1). Alaska volcanoes have frequent eruptions, but are very remote, necessitating remote sensing as a primary method for eruption monitoring, forecasting, and response. We use images of Mount Veniaminof (Alaska) acquired between 2018–2019 covering an effusive-explosive eruption, and images of Mount Cleveland (Alaska) between 2017–2018 with coverage of lava dome growth in order to train our model. The Mount Veniaminof eruption captures high temperature basaltic lava flows into a large, ice-filled caldera (Loewen et al., 2021). Mount Cleveland activity consists of explosions, dome growth, and degassing within the summit crater of a stratovolcano (Werner et al., 2017). These volcanoes are quite different in terms of morphology, eruption style, and governing subsurface processes. They also differ in the source of hotspot detections, namely lava surrounded by ice at Mount Veniaminof, *versus* hot rock surrounded by cold rock at Mount Cleveland. These source differences result in hotspots that may differ slightly in intensity and appearance, leading to a more robust model than it would be if trained on just one of these volcanoes alone.

The other six volcanoes in this study are used for model testing, and were chosen to comprise a wide range of edifice morphologies, magma compositions, eruption frequencies, and eruption styles. These include the frequently erupting and typically mafic volcanoes Okmok Caldera, Shishaldin Volcano and Pavlof Volcano, and the less frequently erupting and typically more silicic volcanoes Augustine Volcano, Bogoslof Island, and Redoubt Volcano. Importantly, all have erupted since the launch of the MODIS sensors. Although our development is focused in Alaska, the volcanoes compiled here range widely in terms of the thermal signatures we expect to identify and the meaning of those signatures in terms of eruptive potential. This dataset can help to evaluate the effectiveness of the model across volcanic systems, and inform future application of the model.

We call the final version of our trained U-net model HotLINK: the Hotspot Learning and Identification Network. After testing and training, HotLINK is applied to VIIRS data from 2012–2022 and MODIS data from 2000–2022 for the eight target volcanoes. The result of these analyses are 22 years of hotspot detections for these volcanoes, 10 years of which have both VIIRS and MODIS observations. We also implement an optimized version of the MIROVA algorithm for our target volcanoes to compare the performance of the machine learning and thresholding approaches. We choose to compare our results with MIROVA because it is one of the most widely used algorithms for global volcanic hotspot monitoring, and was already familiar to the authors. Through this work we hope to improve the accuracy of hotspot detections in infrared satellite data and share our methodology so that it can be applied elsewhere. We aim to address the questions: 1) is a CNN approach able to detect volcanic hotspots in infrared data better than a thresholding approach? 2) Can a computer vision model trained on VIIRS data be reasonably applied to MODIS data with a different resolution? 3) What are the limitations of HotLINK in terms of generalizability to other volcanoes, and detection limits for VIIRS and MODIS, night and daytime images? For each detection we calculate radiative power to quantify the heat emissions over the 22-year study period for the target volcanoes. We then discuss the capabilities and limitations of this approach for volcano monitoring.

2 Methodology

Our model takes as input a VIIRS or MODIS image with MIR and TIR bands, and outputs the probability that each pixel in a central region of the scene contains a volcanic hotspot. Once a hotspot is detected we calculate the total volcanic radiative power (RP in Watts) and area (m^2) of the hotspot. The methodology applied here involves the use of four separate VIIRS datasets to: 1) *train* the network, 2) *validate* hyperparameter selection (i.e., tuning parameters that configure the model and training, as opposed to parameters that are used within the model to make predictions), 3) *test* the model's accuracy when applied to new volcanoes, and 4) *analyze* detections and calculate RP for each volcano over an extended time period. Each of these four datasets (with names italicized above) is assembled for the VIIRS sensor, and additional test and analysis datasets are assembled for the MODIS sensor to produce six datasets in total (Table 2).

HotLINK is trained to detect hotspots in VIIRS infrared images on a manually labeled dataset (VIIRS training) of 3,783 images

TABLE 1 Volcanoes used in this study, in order from west to east. Eruption dates and eruption styles are composited from information available on the Alaska Volcano Observatory website (www.avo.alaska.edu/).

Volcano	Eruptive styles	Eruptions within study period (2000–2022)
Mount Cleveland	Explosive, dome-building	2001, 2005, 2006, 2007, 2009, 2010, 2011, 2013, 2014, 2016, 2017, 2019, 2020
Okmok Caldera	Explosive, phreato-magmatic	2008
Bogoslof Island	Phreato-magmatic, explosive, dome-building	2016–2017
Shishaldin Volcano	Effusive, explosive	2004, 2014–2015, 2019–2020
Pavlof Volcano	Explosive, effusive	2007, 2013, 2014, 2016, 2021
Mount Veniaminof	Effusive, explosive	2002, 2004, 2005, 2006, 2008, 2009, 2013, 2018, 2021
Augustine Volcano	Explosive, dome-building	2006
Redoubt Volcano	Explosive, dome-building	2009

TABLE 2 Datasets used in this study.

Dataset	Labeled	Volcanoes (dates)	Number of images
VIIRS Training	By pixel	Mount Veniaminof (2018), Mount Cleveland (2018–2019)	3,783
VIIRS Validation	By pixel	Mount Veniaminof (2018), Mount Cleveland (2018–2019)	1,275
VIIRS Test	By image	Okmok Caldera, Shishaldin Volcano, Augustine Volcano, Redoubt Volcano, Pavlof Volcano, Bogoslof Island (Mar, Jun, Sep, and December 2017)	3,280 (includes 66 ambiguous images moved from the VIIRS validation dataset)
VIIRS Analysis	None	Mount Veniaminof, Mount Cleveland, Okmok Caldera, Shishaldin Volcano, Augustine Volcano, Redoubt Volcano, Pavlof Volcano, Bogoslof Island (2012–2022)	160,497
MODIS Test (Aqua)	By image	Mount Veniaminof (2018)	634
MODIS Analysis (Aqua and Terra)	None	Mount Veniaminof, Mount Cleveland, Okmok Caldera, Shishaldin Volcano, Augustine Volcano, Redoubt Volcano, Pavlof Volcano, Bogoslof Island (2000–2022)	385,426

of Mount Veniaminof and Mount Cleveland volcanoes. We opt for a manual labeling approach because our goal is to create an automated system that simulates the manual hotspot identification which is done on a daily basis by duty satellite scientists at the Alaska Volcano Observatory (AVO). The same training dataset is used to optimize the thresholds of the MIROVA algorithm (Coppola et al., 2016), and results from both the optimized implementation of the MIROVA algorithm and HotLINK are compared using the same validation dataset, which consists of 1,275 images from the same volcanoes. After training and validation, the accuracy of the model is estimated by applying it to the VIIRS test dataset, which is also manually labeled and consists of images from the six other Alaska volcanoes (Figure 1): Okmok Caldera, Shishaldin Volcano, Augustine Volcano, Redoubt Volcano, Pavlof Volcano, and Bogoslof Island.

Although HotLINK is only trained on VIIRS data, we test its applicability to MODIS data simply by inputting the MODIS test dataset into the VIIRS-trained HotLINK model. Data pre-processing for MODIS follows all of the same steps as for VIIRS data (see Section 2.1). Finally, HotLINK is used to detect volcanic hotspots in 10 years of VIIRS data (VIIRS analysis dataset) and 22 years of MODIS data (MODIS analysis dataset) from all eight of the previously mentioned Alaska volcanoes. A subset of the MODIS analysis dataset (MODIS test data, manually labeled for Mount Veniaminof) is reviewed and used to estimate the accuracy of the model when applied to MODIS.

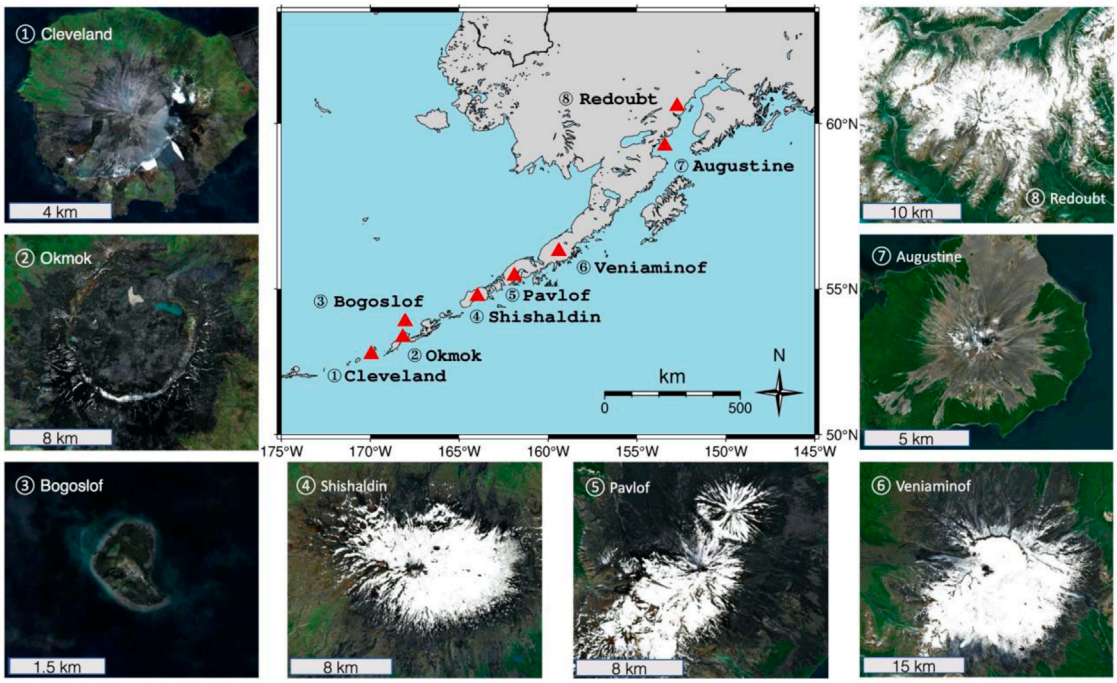


FIGURE 1
Volcanoes used in this study. The map in the center shows all volcano locations in Alaska. Numbered images shows high-resolution satellite data of the volcanoes at various zoom levels, from west to east 1) Mount Cleveland, 2) Okmok Caldera. 3) Bogoslof Island, 4) Shishaldin Volcano, 5) Pavlof Volcano, 6) Mount Veniaminof, 7) Augustine Volcano, and 8) Redoubt Volcano. Satellite data are from Sentinel-2 and composited by CalTopo to provide cloud-free viewing.

2.1 Dataset pre-processing

The pre-processing for all VIIRS and MODIS datasets is the same. First, files containing any of the 8 target volcanoes are downloaded using the Atmosphere Science Investigator-led Processing System API (sips.ssec.wisc.edu) or NASA Earthdata portal (search.earthdata.nasa.gov). Next, terrain and atmospherically corrected radiance data (level 1b) are resampled onto a uniform grid of 64 x 64 pixels centered on the volcano using the nearest neighbor resampling method and the nadir pixel resolution. For VIIRS this corresponds to an area of roughly 24 × 24 km^2 and for MODIS this is an area of 64 × 64 km^2. We use VIIRS image bands I4 (3.55–3.93 μm, MIR) and I5 (10.5–12.4 μm, TIR), and MODIS bands 21 (3.929–3.989 μm, MIR) and 32 (11.77–12.27 μm, TIR). Spectral radiance values have the pixel area (m^2), spectral bandwidth (m), and angular aperture (steradians) factored out of the raw radiative power measurement (W), which allows for direct comparison between data from the two sensors, and normalization using the same factors.

Spectral radiance values (L) are normalized to the minimum (L_{min}) and maximum (L_{max}) possible radiance values for the VIIRS sensor, as determined by scale and offset factors (available in the VIIRS level-1b product user guide; NASA Goddard Space FlightCenter, 2018). Physically, L_{min} and L_{max} represent the limits of the sensor, and possible retrieval values are always within this range. Although the true radiance may be outside this range, the sensor will always return at least L_{min} and will saturate at values greater than L_{max} (NASA Goddard Space Flight Centere, 2018). The equation used to normalize the spectral radiance data is as follows:

$$L_{norm} = \frac{L - L_{min}}{L_{max} - L_{min}} \quad (2)$$

Normalization is important to prevent issues with vanishing or exploding gradients which would make it difficult for the CNN model to converge on a solution (Sola and Sevilla, 1997). We use the same L_{min} and L_{max} for both VIIRS and MODIS data despite the sensors having different minimum and maximum possible spectral radiance values. This is because once the model has been trained on spectral radiance data normalized to a certain range, it must be applied to data normalized in the same way. Lastly, since VIIRS data saturates at a lower spectral radiance than MODIS data, some exceedingly rare MODIS pixels have values higher than one after normalization (<0.002% of pixels in the MODIS test dataset). To remedy this, values are capped at a maximum value of one.

The VIIRS training and validation datasets are assembled by collecting all (day and night) VIIRS data from the SNPP and NOAA-20 satellites with coverage of Mount Veniaminof for the year of 2018 and NOAA-20 VIIRS data (only) with coverage of Mount Cleveland for both 2017 and 2018. These volcanoes and time frames were selected to encompass background non-eruptive behavior, increasing unrest, and eruption. From this dataset, 75% of images are grouped into the VIIRS training dataset, and the remaining 25% are put into the VIIRS validation dataset. The validation dataset is smaller because it is only used to ensure the model is not overfitting, and a representative population is sufficient.

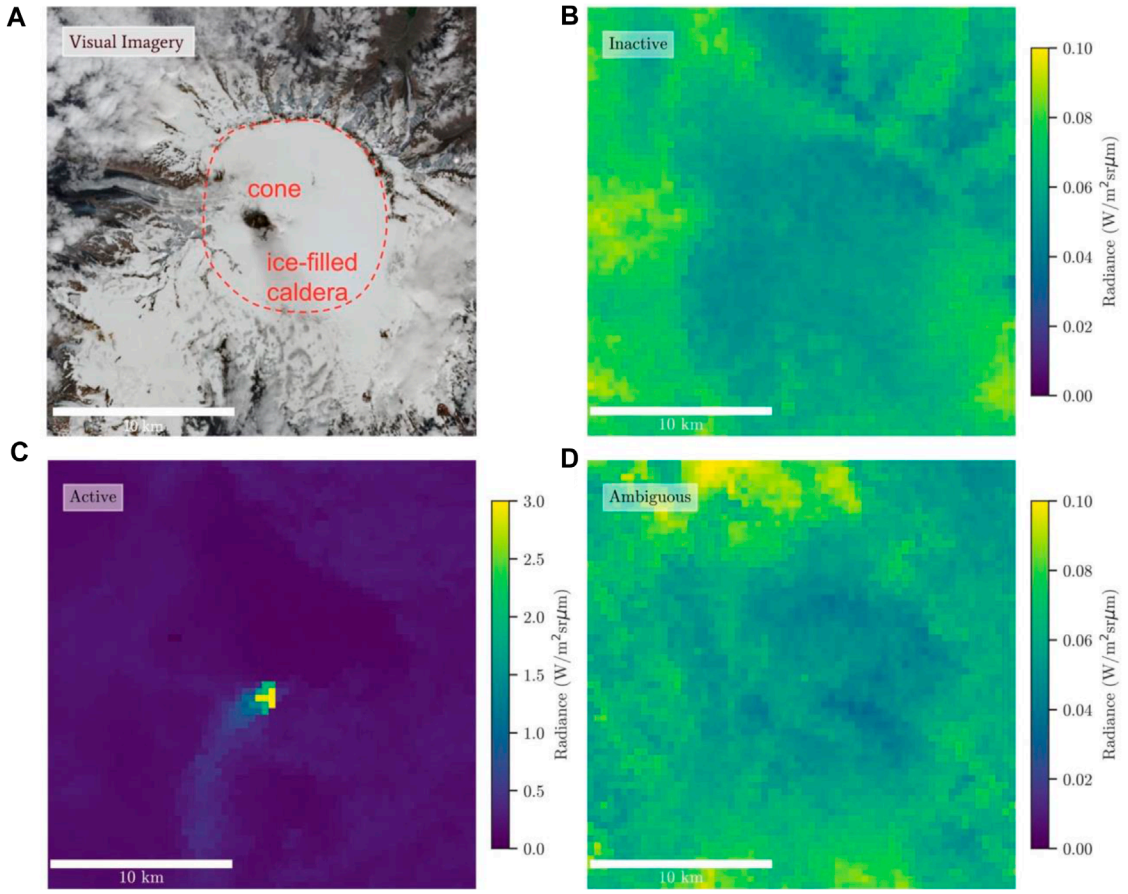


FIGURE 2
Classified example images. **(A)** Sentinel-2 visible RGB image (enhanced natural color visualization) of the ice-filled summit caldera and central cone. The classes of MIR VIIRS images used while training our model shown are **(B)** inactive–not containing a volcanic hotspot as identified by a human analyst, **(C)** active–containing a volcanic hotspot, and **(D)** ambiguous, which could not be confidently categorized into either class. All examples are nighttime images, showing the same cropped region of Mount Veniaminof (24 by 24 km). Note that images C and D have the same color mapping, but image B is scaled differently. Color bars show the range of radiance values in each image.

Whereas the training dataset is larger because data in this group is used to actually train the model, and more data results in better model performance. The grouping between these two datasets is done randomly, with the exception that each image is grouped together with its closest temporal neighbor, since overpasses of SNPP and NOAA-20 satellites can be within −45 min of each other. This prevents having one image in the training dataset and a nearly identical image in the validation dataset.

Images are manually classified into three groups: "active" defined as images containing a volcanic hotspot, "inactive" or images with no volcanic hotspot, and "ambiguous," where we cannot conclusively identify whether or not the image contains a volcanic hotspot (Figure 2). Next, all hotspot pixels within the active-labeled images are identified to construct pixel-wise masks. The ambiguous images are not used for training or validation since we only want images we can characterize with confidence in those datasets. All ambiguous images from the VIIRS validation and training datasets are moved into the VIIRS test dataset, which can have images of any class (66 ambiguous images in total are moved). The final training dataset contains 3,783 images and the final validation dataset contains 1,275 images. In both the VIIRS training and validation datasets, 45% of images are of Mount Veniaminof, 55% are of Mount Cleveland, and 32% of the total are classified as active.

To evaluate how well the model generalizes to other volcanoes not used in training, a test dataset is assembled consisting of 4 months (March, June, September, and December 2017) of VIIRS data for the six additional Alaska volcanoes (Augustine Volcano, Bogoslof Island, Okmok Caldera, Pavlof Volcano, Redoubt Volcano, and Shishaldin Volcano). These months are chosen from throughout the year to capture the full extent of Alaska's seasonal variations. Of our target volcanoes, only Bogoslof Island had an eruption during 2017, so few volcanic hotspots are expected in the VIIRS test dataset. Although choosing data from different volcanoes or times could have yielded more hotspot detections, the volcanoes were chosen with the aim of facilitating future interdisciplinary analysis, and the time period was chosen to ensure standardization across all volcanoes. The resulting dataset is a good indicator of the model's performance when applied to new volcanoes during typical conditions. Images in the test dataset are also manually classified as active, inactive, or ambiguous, but not further classified on a pixel-wise basis. Therefore, the VIIRS test dataset is only used to test the ability of the model to detect images containing hotspots,

not whether it accurately retrieves all of the pixels associated with the hotspot.

The VIIRS analysis dataset consists of the remaining (unlabeled) data, which are analyzed by the trained model and used to generate a hotspot detection time series from 2012–2022 for each of the eight volcanoes in this study. It is the largest VIIRS dataset of our study, consisting of 160,497 individual images of the volcanoes. Note that the VIIRS analysis dataset encompasses data that is already a part of the VIIRS training, validation, and test datasets.

We generate additional MODIS test and analysis datasets in order to test the applicability of our model to MODIS data, compare time series results for VIIRS and MODIS, and extend the time series of detections back to the year 2000. The MODIS test dataset consists of all 2018 MODIS data from the Aqua satellite of Mount Veniaminof classified by image. This volcano and time period were chosen for the MODIS test dataset to encompass a known eruption at Mount Veniaminof that was included in the VIIRS training data. The MODIS analysis dataset consists of all MODIS data from both Aqua and Terra satellites from 2000 to 2022 with coverage of the eight target volcanoes.

2.2 U-net architecture and training

CNNs utilize 3 × 3 (or other sized) matrices, known as convolution kernels, to search for specific patterns within an image (LeCun et al., 2010). The kernel is moved across the image and multiplied with each 3 × 3 subsection to create a new filtered image that shows the degree of correlation between the features of the kernel and the image. This allows the network to identify and locate specific spatial patterns within the image. By stacking multiple layers of convolutions, the network is able to detect increasingly larger and more complex features. At first the network's kernels are populated randomly, but through an iterative training process the kernels are adapted to identify spatial patterns optimized for the task at hand.

Training a CNN involves inputting batches of labeled images into the model. As each image is passed into the model the probabilistic prediction (initially computed by the randomly initialized kernels) is compared to the truth value (the class of each pixel), which is known by prior manual analysis. Then a value, the "loss," is calculated to quantify how well the model prediction compares to the truth value. This is calculated by the "loss function," which, in simple terms, is a quantitative measure of how poorly the model performs—so, a lower loss score indicates better performance. Importantly, the loss function is differentiable with respect to the model—meaning that the gradient of the loss function can be calculated for the entire model. The gradient is very high dimensional, with a value for each trainable parameter of the entire model. By taking a small step in the direction of the gradient, each parameter of the model is adjusted slightly in the optimal direction to decrease the loss, which thereby increases the performance. With each pass over the training dataset, or epoch, each parameter is adjusted slightly, the loss decreases, and the performance of the model improves. This iterative training process is called gradient descent, since the model is descending step-by-step down the gradient of the loss function with the goal of reaching a local minimum. For a more comprehensive explanation of the training, underlying mathematics, and applications of CNNs, see LeCun et al. (2010).

We chose a U-net CNN architecture, because it allows for predictions to be made in the same resolution as the input (Figure 3; Ronneberger et al., 2015). This allows individual pixels to be flagged as hotspots or not. The input for our model is normalized radiance data from the MIR and TIR bands of the VIIRS or MODIS sensor, resampled to uniform resolution and cropped to 64 × 64 pixels centered on the main vent of the volcano of interest (64 × 64 pixels and 2 channels). The output is the probability that each pixel in a central area of the input belongs to one of three classes: background, hotspot, or hotspot-adjacent (24 × 24 pixels and 3 classes). The third class of pixels, hotspot-adjacent, helps the model to train faster; these pixels are considered background pixels during validation and testing. The output region is smaller than the input, due to the fact that convolutions of border pixels are undefined, resulting in a smaller image after each convolution. We consider that a 24 × 24 area of pixels is sufficient for detecting most hotspots (9 × 9 km^2 for VIIRS, and 24 × 24 km^2 for MODIS), but acknowledge that it may miss distal regions of large lava flows, or eruptions which occur far from the main vent.

Many additional parameters can be adjusted in order to alter the architecture, training, or functionality of the model—these are referred to as hyperparameters. We experimented with many of these, selecting the hyperparameters which result in the best performance (as measured by the validation dataset). Parameters that we tested include the random seed and distribution used to initialize the kernels (Glorot and Bengio, 2010), the number of convolutional filters used in each layer (i.e. the width of each rectangle in Figure 3), the gradient descent algorithm (Kingma and Ba, 2014), and the number of training epochs. We also tried many techniques to address the class imbalance in our training dataset. In the VIIRS training dataset approximately 25% of images contain a hotspot, while the remaining 75% do not. We explored several methods to mitigate the effects of the class imbalance, including: oversampling images with hotspots, undersampling the background images, using class weights, and using simple image augmentations to generate more training samples (details in the appendix). Out of all these methods explored, only the image augmentation resulted in an increase in model performance. The rest of this paper only describes the final model, referred to as HotLINK, which uses the best hyperparameters found through dozens of training iterations.

HotLINK is trained on the VIIRS training dataset for 250 epochs, which is the point when the loss ceases to decrease for the validation dataset. During training, input images are augmented using 90° rotations and flips applied randomly after each epoch using the Albumentations library (Buslaev et al., 2020). This produces eight unique orientations for each original input image, which helps the model to learn only the most relevant features for prediction. The model is trained using the Adam optimizer (Kingma and Ba, 2014) with a sparse categorical cross entropy loss function, both of which are a part of the TensorFlow Python library (Abadi et al., 2015). Our U-net took ~2 h to train on a 6-core Intel i7 processor, and after training makes predictions at an average rate of ~5 images per second. Further details on the specific hyperparameters used in the training of the HotLINK model can be found in the code itself, available in the appendix and on GitHub (Saunders-Shultz,

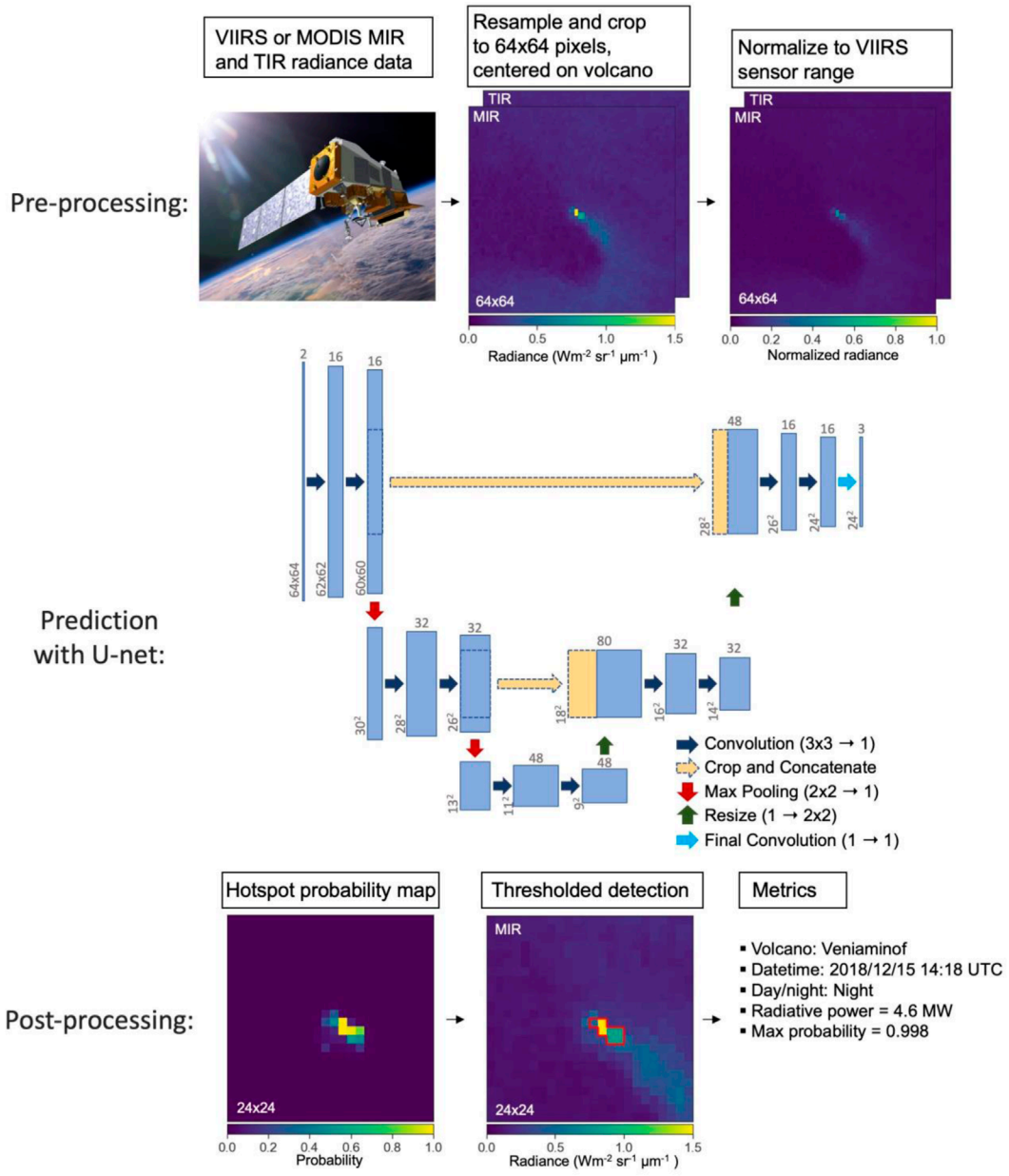


FIGURE 3
Steps of HotLINK processing: pre-processing, prediction with the U-net, and post-processing of a hotspot detection. Blue and tan rectangles of the U-net diagram represent data, the dimensions of which are labeled and denoted by the shape of the rectangles. For example, the input is [64 x 64 x 2] pixels and the output is [24 x 24 x 3] pixels. Note that at each convolution step the height and width of the data are decreased by two, since convolutions on the perimeter pixels are undefined. This progressive loss of perimeter pixels results in a prediction area significantly smaller than the input area. For further description of the motivation and function of the U-net architecture, see Ronneberger et al. (2015).

2023). Although we found these hyperparameters to work best for our problem, they may require modification for other hotspot detection applications.

2.3 Validation and testing

During the training process, we use the validation dataset to try out many different versions of the model in order to test which architectures, model hyperparameters, etc., result in the best hotspot predictions. This process also helps to ensure that the model is learning patterns that are applicable to unseen data and not overfitting. Validation data are also used to tune threshold parameters applied to the output probability maps, and to compare HotLINK and our optimized application of the existing threshold-based algorithm, MIROVA (Coppola et al., 2016). To assess how the trained and validated model performs on new data, we use the test dataset, which is composed

entirely of images from volcanoes the model has not seen during training.

We use two main metrics during validation and testing to evaluate HotLINK and MIROVA's performance: accuracy and F1-score. Accuracy is simply the percentage of images correctly identified by the model. It is defined as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

where TP, TN, FP, and FN refer to the number of true positives (true hotspot detections), true negatives (true background detections), false positives (erroneous hotspot detections), and false negatives (missed volcanic hotspot detections), respectively, generated by the model. However, accuracy may not be the most appropriate metric for imbalanced datasets, which have higher proportions of some classes than others. For example, in this study a high percentage of images do not contain a volcanic hotspot. Therefore, a high accuracy could be achieved simply by predicting no hotspots in any image. A better metric for evaluating model performance in cases with imbalanced datasets is the F1-score (Ferri et al., 2009), defined as:

$$F_1 = \frac{TP}{TP + \frac{1}{2}(FN + FP)} \tag{4}$$

The F1-score rewards true positive results and equally punishes false positives and false negatives, while true negatives have no impact on the score. Although our model predicts whether or not each pixel comprises a hotspot, accuracy and F1-scores are calculated on an image-wise basis. Image-wise metrics are used to evaluate the model's ability to detect a hotspot, because image-wise labelling is faster allowing us to create largervtest datasets (Table 2). The training dataset is labeled for each pixel, since the U-net requires every pixel to be labeled in order to train.

Another way to compare HotLINK and our optimized MIROVA algorithm is by using receiver operating characteristic (ROC) curves, which provide a graphical means to characterize the effectiveness of binary classification models (Figure 4). For a given index or predicted probability, an ROC curve plots the true positive rate against the false positive rate achieved by thresholding at different values. In this way it shows the tradeoff between false positives and true positives. For example, setting a low threshold will achieve a high true positive rate at the expense of more false positives, and setting a high threshold will achieve a low true positive rate while providing fewer false positives. ROC curves plot a model's performance at all possible thresholds, thereby showing a particular model's ability to identify hotspots with low FP and FN rates. The ROC curve comparison of HotLINK and MIROVA is further discussed in Section 3.3.

2.4 MIROVA optimization on the VIIRS training dataset

In order to test the performance of HotLINK, we compare our results to the MIROVA algorithm, which was originally developed for use with MODIS data (Coppola et al., 2016). The MIROVA algorithm has already been applied to VIIRS data (Campus et al., 2022, using moderate resolution bands; Aveni et al., 2023, using the same image bands used here). However, these studies use the

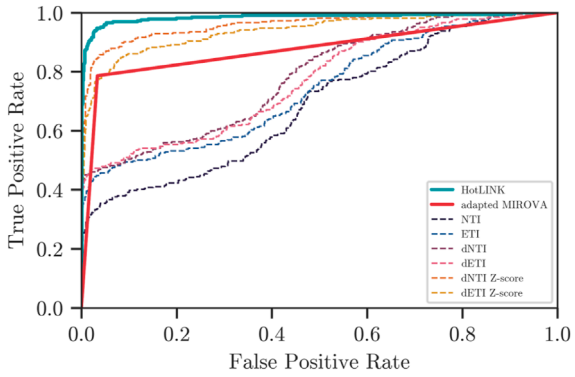


FIGURE 4
Receiver Operating Characteristic (ROC) curve applied to HotLINK and the adapted MIROVA algorithm. HotLINK probabilities are shown in blue, MIROVA prediction is red, and the different indices used in MIROVA are the thinner dashed lines. Preferred classifiers have a high true positive rate (TPR) and low false positive rate (FPR). Note that MIROVA consists of two straight lines because it produces just a binary output.

original thresholds of the MIROVA algorithm that were designed for use with MODIS data. Since VIIRS and MODIS have different spatial resolutions and slightly different spectral bands, it is possible that the original thresholds could be improved for use with VIIRS data. To make a fair comparison between MIROVA's threshold methodology and our model, we optimize the thresholds of the MIROVA algorithm using a grid search over the same VIIRS training dataset that is used to train HotLINK. By using the same training dataset to tune each model and the same validation dataset to evaluate them, we ensure a fair and consistent benchmark between the two approaches. This allows for an unbiased comparison, ensuring that any observed performance differences can be attributed to the inherent capabilities of each model rather than variations in the data they are applied to.

MIROVA employs three thresholds (C1, C2, and K) on multiple indices calculated from the MIR and TIR spectral bands. These indices are the Normalized Thermal Index (NTI), Enhanced Thermal Index (ETI), spatially filtered versions of the first two indices called dNTI and dETI, and the Z-scores of dNTI and dETI. These indices are designed to increase the contrast between hotspot and background pixels, by combining spectral information at each pixel (indices NTI and ETI) with spatial information from surrounding pixels (indices dNTI and dETI) and the scene as a whole (Z_{dNTI} and Z_{dETI}). A full description of the algorithm and definitions of indices are presented in Coppola et al. (2016). In brief, pixels are flagged as active if the index NTI is greater than the threshold K, or if the indices dNTI, dETI, and the Z-scores of both, surpass the C1 and C2 thresholds, respectively:

$$\begin{gathered}(NTI > K)\ or \\ ((\mathrm{dNTI} > C_1)\ \text{or}\ (Z_{dNTI} > C_2))\ \text{and}\ ((\mathrm{dETI} > C_1)\ \text{or}\ (Z_{dETI} > C_2))\end{gathered} \tag{5}$$

In order to optimize MIROVA for use with VIIRS data, we conduct separate grid-searches for nighttime and daytime data to define new threshold values for C1 and C2, which minimize the error

rate on images within the VIIRS training dataset. The daytime grid search is conducted between C1 values of 0.0–0.29 with a stepsize of 0.01, and C2 values of 2.0–11.75 with a stepsize of 0.25. The nighttime samples are more sensitive to the C1 threshold so we use a finer stepsize of 0.005 and smaller range of 0.0–0.095. The C2 range and stepsize remain the same for the nighttime grid search. At each step the accuracy of MIROVA using specific thresholds is calculated. The K threshold was not optimized because it was found to have little effect on the pixel selections made by the algorithm, so it was left as the default value of −0.8 for nighttime images and −0.6 for daytime images. Default MIROVA values for daytime data are C1=0.02 and C2=15, and for nighttime data are C1=0.003 and C2=5. With our grid search we found the highest accuracy using values of C1=0.11 and C2=6.25 for daytime data, and C1=0.075 and C2=5.25 for nighttime data (see Supplementary Figures A.3 and A.4 in the appendix for visualization of both grid searches). The grid searches demonstrate that slight changes to threshold values can result in slight increases in the performance of MIROVA, at least when applied to our particular dataset.

2.5 Hysteresis thresholding and radiative power calculation

Some final considerations for implementing the model are choosing how to threshold pixels in the output probability map (Figure 3), and then calculating useful metrics for each detection to better track changes in volcanic thermal emissions over time. Although each pixel is predicted with an individual probability, we recognize that a pixel is more likely to be a hotspot if it is adjacent to a hotspot pixel. For that reason, we implement hysteresis thresholding, in which a high threshold is used to initialize hotspot detections and a lower threshold is used to continue them. Here, all pixels with a probability greater than 0.5 are classified as hotspots, and pixels with a probability greater than 0.4 are classified as hotspot pixels if they are adjacent to other hotspot pixels. The high threshold is set by optimizing the validation dataset for image F1-score, and then the low threshold is set by optimizing for pixel-wise F1-score. To clarify, these metrics are chosen because only the high threshold determines which images are active, while the low threshold determines which pixels within the image are active.

Once active images are detected and all hotspot pixels within those images are identified, radiative power (RP) is calculated following the method of Wooster, (2003), using the following formula:

$$RP = C \times A_{pix} \times \sum^{n} L_{pix} - L_{BG} \quad (6)$$

where RP is the radiative power measured in Watts, C is a constant of proportionality that is specific to the sensor ($sr^{-1}\mu m^{-1}$, 18.9 for MODIS and 17.34 for VIIRS), A_{pix} is the area of the pixel in kilometers squared (1 km^2 for MODIS, 0.14 km^2 for VIIRS), n is the number of pixels in the hotspot, L_{pix} is the radiance of each hotspot pixel ($Wm^{-2}sr^{-1}\mu m^{-1}$), and L_{BG} is the mean radiance of pixels directly surrounding the hotspot detection ($Wm^{-2}sr^{-1}\mu m^{-1}$, following the established methods of Wooster, 2003). RP is a measure of how much energy is released over the entire hotspot, and includes corrections for pixel size, central wavelength, and background radiance. Since pixel size and central wavelengths are different for VIIRS and MODIS, using RP allows us to make direct comparisons between the two sensors.

3 Results

3.1 Validation and test results

Results on the VIIRS validation dataset (Table 3) show that the final model works well when applied to data that has not been seen during training but comes from the same volcanoes. Specifically, both Mount Veniaminof and Mount Cleveland validation data yield model accuracies >95% and F1-scores >0.9.

On the VIIRS test dataset, which includes data from the six volcanoes that the model has not seen previously, HotLINK achieves a relatively low F1-score of 0.667 (Table 3). This seemingly poor performance is best explained by the lack of true hotspots in the dataset used; out of the six volcanoes, only Bogoslof Island erupted during the sampling period of the test dataset (Table 2). Since F1-score is mainly a function of true positive detections we achieve a poor score on most of the volcanoes since there were not many true hotspots to detect. False negative and false positive rates on all datasets do not exceed 4%, except for the Augustine Volcano false negative rate, which is 7.9%.

3.2 HotLINK results on MODIS test data

The MODIS test dataset consists of all Mount Veniaminof data from the Aqua satellite in 2018, including 634 images in total. HotLINK achieves an accuracy of 98% on the MODIS test dataset, and an F1-score of 0.95 (Table 3). Unexpectedly, this performance is better than the model performs on VIIRS data. In section 4.3 we discuss a possible explanation for this.

3.3 HotLINK and adapted MIROVA algorithm results on the VIIRS validation dataset

The VIIRS validation dataset is used to compare the results of HotLINK and the optimized MIROVA algorithm after both models are trained/optimized with the VIIRS training dataset. On the validation dataset, we find that HotLINK outperforms our implementation of the MIROVA algorithm in all metrics (Table 4). Specifically, HotLINK produces more true positives (fewer missed detections), and more true negatives (fewer false detections) than the MIROVA approach. Both methods score higher on nighttime data than daytime data. The conditions under which each model performs best is further discussed in Section 4.4.

The ROC curve (Figure 4) further demonstrates that HotLINK (blue line) outperforms the MIROVA algorithm implementation (red line) with respect to true and false positives. In this plot, preferred classifiers have a high true positive rate (TPR) and low false positive rate (FPR). So better classifiers are those which plot further into the top left corner. These results show that HotLINK performs better than the overall optimized MIROVA algorithm,

TABLE 3 HotLINK results on training, validation, and test datasets. Metrics shown are: accuracy, F1-scores, ratio of True Negatives, True Positives, False Negatives, and False Positive detections, and the total count of images used to calculate the metrics. Each row shows the average of all volcanoes first, and then specific values for each volcano in the dataset. Note that ambiguous images (195 total) are removed prior to this analysis.

Dataset	Accuracy	F1-score	TN	TP	FN	FP	Count
VIIRS Training	0.952	0.914	0.698	0.254	0.031	0.017	3,781
Mount Cleveland	0.962	0.898	0.795	0.167	0.017	0.021	1,551
Mount Veniaminof	0.945	0.920	0.631	0.314	0.041	0.014	2,230
VIIRS Validation	0.962	0.923	0.731	0.231	0.022	0.016	1,275
Mount Cleveland	0.977	0.933	0.820	0.157	0.011	0.011	527
Mount Veniaminof	0.951	0.919	0.668	0.282	0.029	0.020	748
VIIRS Test	0.947	0.667	0.908	0.049	0.024	0.019	2,956
Augustine Volcano	0.914	0.172	0.901	0.009	0.079	0.011	547
Bogoslof Island	0.955	0.892	0.765	0.189	0.024	0.022	460
Okmok Caldera	0.956	0.512	0.927	0.024	0.024	0.026	468
Pavlof Volcano	0.974	0.723	0.936	0.037	0.008	0.019	483
Redoubt Volcano	0.919	0.608	0.940	0.040	0.002	0.019	530
Shishaldin Volcano	0.979	0.444	0.970	0.009	0.002	0.019	468
MODIS Test (Mount Veniaminof)	0.981	0.954	0.786	0.195	0.019	0.0	646

TABLE 4 Comparison of HotLINK and the adapted MIROVA algorithm on the VIIRS validation dataset. Metrics shown are: accuracy, day/night/combined F1-scores, and ratio of True Negatives, True Positives, False Negatives, and False Positive detections.

Model	Accuracy	F1-score	Night F1-score	Day F1-score	TN	TP	FN	FP
HotLINK	0.962	0.923	0.929	0.916	0.731	0.231	0.022	0.016
Adapted MIROVA algorithm	0.921	0.834	0.894	0.765	0.722	0.198	0.054	0.025

as well as all of the individual indices used by the MIROVA algorithm (thin dashed lines) with respect to TPR and FPR. This indicates that HotLINK is able to better differentiate hotspot and background pixels in comparison with individual indices, regardless of threshold selection. This is due to the CNN's ability to extract additional spatial information compared to manually tuned spatial filters.

3.4 Time series results

After applying HotLINK to the validation and test datasets, we apply HotLINK to the VIIRS and MODIS analysis datasets. This provides 10 years of VIIRS and 22 years of MODIS hotspot detections for the eight target alaska volcanoes. These results can be found in Figure 5. Despite being unlabeled, these results can help provide a qualitative check on the effectiveness of the model when applied to different volcanoes experiencing background, unrest, or eruptive behavior. All detections found in this dataset are plotted as time series in Figure 5, with the Alaska Volcano Observatory (AVO) Aviation Color Code as the background color. In this analysis we use the AVO Aviation Color Code as a proxy for the state of activity of the volcano. A color code of "green" is used to indicate that a volcano is at a background non-eruptive state, "yellow" indicates increasing unrest with the possibility of an eruption in the future, "orange" indicates that effusive or low-level explosive eruptions are occurring or are expected in the immediate future, "red" indicates a significant explosive eruption is occurring or imminent, and

"unassigned" (colored as gray in Figures 5, 7) indicates that there is insufficient ground-based monitoring data to assign a color code (Guffanti and Miller, 2013). While accuracy metrics are useful, the time series plots demonstrate the utility of HotLINK in practical applications. Figure 5 illustrates that HotLINK succeeds at detecting eruptions, which are accompanied by significant increases in the frequency and RP of detected hotspots. This figure also shows patterns of potential false positive detections during non-eruptive periods at all volcanoes, which are discussed in the following paragraphs.

Mount Cleveland erupts frequently, as indicated by many periods of orange color code in the timeline (Figure 5), which represent lava dome eruptions and other elevated activity (e.g., Werner et al., 2017). The Mount Cleveland time series shows numerous hotspot detections, which are much more frequent during periods of orange color code compared to when the color code is unassigned.

Okmok Caldera had only one eruption during our analysis period, in 2008. Only MODIS data is available for this eruption, from which there was one nighttime and three daytime detections during the eruptive period all with RP values >5 MW. Steady detections occur in VIIRS night and daytime data at Okmok Caldera, which we infer may be due to the presence of lakes within the caldera.

At Bogoslof Island we see a strong seasonal trend, in which VIIRS daytime detections and associated RP increase in the summer and decrease during winter. These seasonal trends are observable both before and after the 2017 eruption, but are stronger post-eruption. The 2016–2017 Bogoslof Island eruption is captured well, with VIIRS nighttime detections producing higher RP values than at any other time.

At Shishaldin Volcano, extended eruption periods from 2014–2016 and 2019–2020 are tracked well by HotLINK detections. The onset of these eruptions are accompanied by significant increases in the rate and RP of detections, and the end of eruptions are accompanied by a return to background values.

Pavlof Volcano eruptions are detected well by the HotLINK system, with RP values during eruptive episodes significantly higher than during non-eruptive periods. The 2007 eruption is captured well in MODIS data, and subsequent eruptions are captured well in both VIIRS and MODIS data.

At Mount Veniaminof there have been multiple eruptions that are detected by HotLINK, but there is also a high rate of background detections, which could either be indicative of background heat output or potentially the emissivity and thermal inertia differences between the active cone and surrounding glacier. In Section 4.4 we further discuss the nature of these signals.

Augustine Volcano had one observed eruption in 2006. Augustine Volcano has infrequent VIIRS nighttime detections, but does show a seasonal signal with increased VIIRS daytime detections during winter and increased MODIS daytime detections during summer.

Redoubt Volcano also had only one eruption during our analysis period, in 2009, which was detected well in MODIS data. Since then, no anomalous thermal activity has been detected but there have been frequent hotspot detections in VIIRS nighttime and daytime data, which may be attributed to localized persistent degassing and snow melt on the 2009 lava dome.

4 Discussion

In this section we discuss the time series results at all volcanoes to investigate the strengths and weaknesses of our model. We also discuss the probabilistic output of HotLINK, and our finding that probabilities are well calibrated. Next, we compare VIIRS and MODIS applications of HotLINK, and estimate detection limits for each sensor. Finally, we advance our comparison of HotLINK and the threshold-based MIROVA algorithm by looking at a case study of the Mount Veniaminof time series.

4.1 Analysis of time series results from all volcanoes

Based on the time series of detections at all volcanoes (Figure 5), we find that 1) the HotLINK model, as currently trained, works well for many, but not all volcano morphologies/settings, 2) the VIIRS sensor has a lower detection limit than MODIS due to a finer spatial resolution, which also results in a slightly higher false positive rate for VIIRS, and 3) the RP and relative frequency of daytime and nighttime detections reveals distinct periods in the eruptive chronologies at many volcanoes, which can be used to further discern true and false detections. We discuss how we can discern true and false hotspot detections during non-eruptive periods at volcanoes, why false positive detections appear more often in some volcanoes during certain times of the day and year than others, and how results can be further filtered to remove many of the false detections.

Although HotLINK has a lower false positive rate than MIROVA in the validation dataset (Table 4), in the analysis dataset we still see nearly continuous hotspot detections at all volcanoes even between eruptive periods (Figure 5). Even though HotLINK makes many detections when volcanoes are at "green," or a background state (e.g. Okmok Caldera 2012–2022), that does not mean that all of those detections are false positives as it is common for many volcanoes to be persistently degassing and producing heat at the surface even in absence of an eruption. In this case, increases in the rate and RP of detections, rather than the detection of a single hotspot, may indicate volcanic unrest or eruption. However, as testing shows (Table 3), we expect HotLINK to have a false positive rate ~2%, such that some of the detections during background periods are likely not true volcanic hotspots.

In our analysis of Figure 5, we expect true volcanic hotspot detections to be those which are spaced closely together in time and at higher RP than other detections observed during periods with no eruptive activity. At all volcanoes, likely false positives seem to occur in VIIRS daytime images with RP in the range of ~1–10 MW, and in VIIRS nighttime images with RP ~0–0.5 MW. We determine that most detections with RP above these thresholds are true positives, but that does not preclude the possibility of true (but weak) volcanic hotspot detections within those ranges.

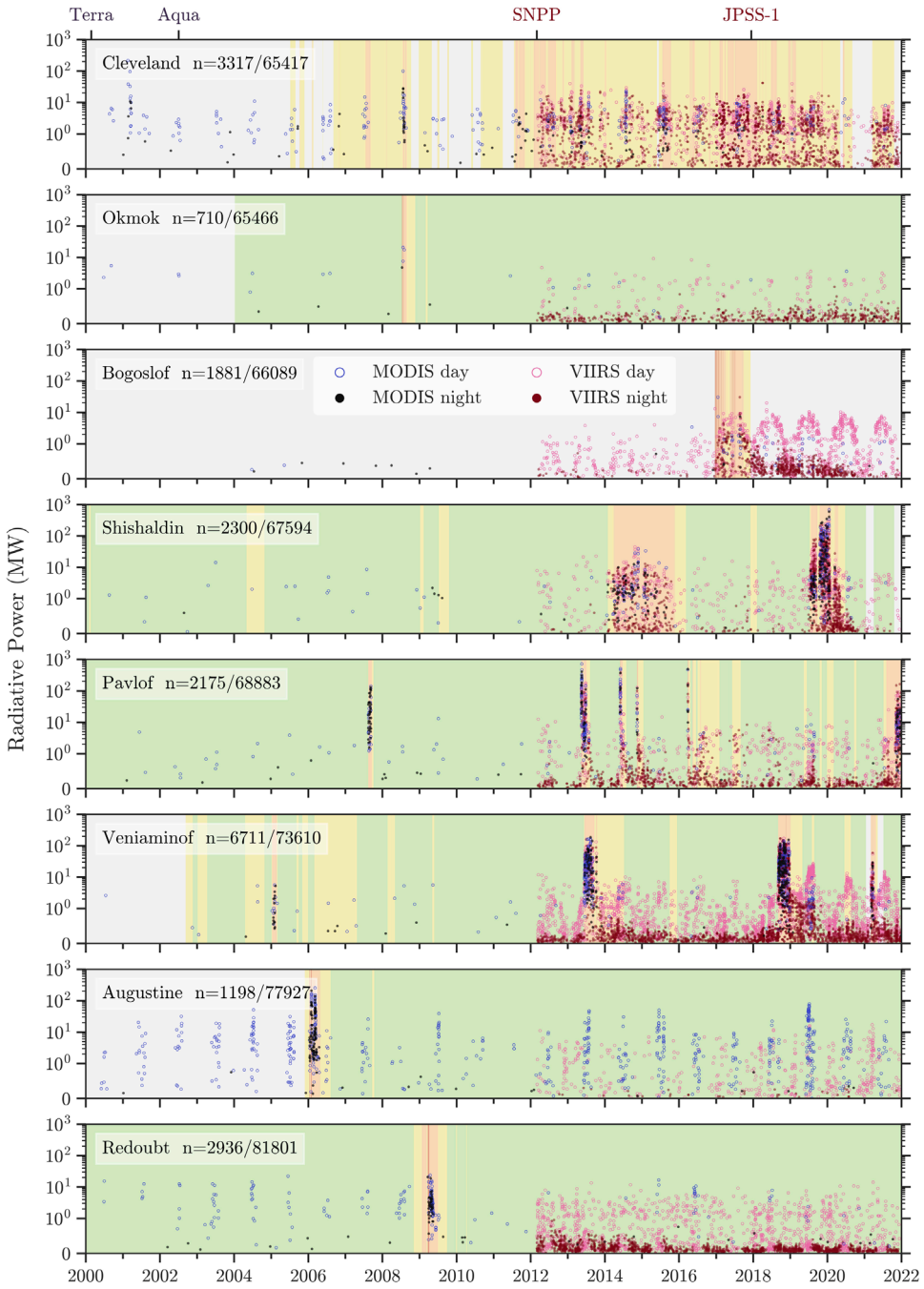


FIGURE 5
Time series results of HotLINK detections and calculated radiative powers for all eight target volcanoes: Mount Cleveland, Okmok Caldera, Bogoslof Island, Shishaldin Volcano, Pavlof Volcano, Mount Veniaminof, Augustine Volcano, Redoubt Volcano. The AVO color code at each volcano is shown as the background color of each figure for general context on the state of activity at the volcano (see Section 3.3 for description of color codes), with gray indicating a period with insufficient monitoring data for AVO to assign a color code ("Unassigned; " Guffanti and Miller, 2013). The RP of individual hotspot detections are shown as points (MODIS = black, VIIRS = red), with solid and open points representing night and daytime acquisitions, respectively. Next to the name of each volcano is the number of total detections at each volcano, and the number of total images for each volcano in both the VIIRS and MODIS analysis datasets. Note that for all plots the *y*-axis scale is linear between 0 and 1 MW, and logarithmic >1 MW. The top axis shows the start of data acquisition from satellites used.

At some volcanoes (Bogoslof Island and Augustine Volcano) there are notable seasonal variations in the number of detections and the RP of those detections. At these volcanoes we believe the source of these detections is primarily from diurnal effects on land/water boundaries. For example, both Bogoslof Island and Augustine Volcano are island volcanoes, which means that during the day the land surface regularly heats up more than the surrounding ocean, creating a temperature difference that is visible in infrared images, centered on the volcano, and thus to our model looks like a volcanic hotspot. Since Bogoslof Island is

−1.5 km in diameter while Augustine Island is −12 km in diameter, Bogoslof Island tends to appear more like a hotspot in daytime VIIRS data while Augustine Volcano Island regularly is identified as a hotspot in daytime, summer, MODIS data (Figure 5). Similarly, clouds frequently develop during the daytime on land, creating localized solar reflections.

A similar effect occurs at volcanoes that have crater lakes/lagoons (e.g., Okmok Caldera and Bogoslof Island). Since water has a higher thermal inertia than land, it preserves solar heat longer into the night than land and is commonly warmer than land at night, particularly when the land is snow-covered. Volcanic lakes are commonly connected to hydrothermal systems and increasing lake temperature can be linked to volcanic activity (Hurst et al., 1991; Rouwet et al., 2014). However, increasing lake temperatures due to volcanic thermal input are difficult to distinguish from increasing temperatures due to diurnal patterns. With that in mind, a hotspot detection of a lake is not necessarily indicative of increased volcanic or hydrothermal activity. By looking at trends in detections and RP over time, however, HotLINK may have the capability to characterize background lake temperatures and thus detect deviations above background. In our data we did find clear examples of diurnal and seasonal cycles in hotspot detections at Okmok Caldera and Bogoslof Island. However, in neither case did we observe clear deviations in the background radiative power that might have been caused by increased volcanic activity. Example images of false detections at Okmok Caldera and Bogoslof Island and comparison with high resolution true color imagery are available in the Supplementary Material. Other common effects producing non-volcanic hotspot detections are snow melting off rocky areas that then become solar-heated (Mount Veniaminof), and clouds or volcanic plumes reflecting solar radiation.

While these non-volcanic sources of apparent hotspots are considered in our study to be false-positives, they highlight the capability of HotLINK to detect subtle warming signals that could be successfully applied to other research problems. Fundamentally there will always be a tradeoff between the sensitivity of the method to detect real volcanic hotspots, and the number of false positives produced. With this in mind, there are simple ways to minimize the occurrence of the false positives in the dataset through filtering. One easy approach is to only use the nighttime data, which is much less susceptible to false positives, especially those occurring on exposed rocks surrounded by snow and ice fields and solar reflection off clouds or plumes. Another way is to set a specific probabilistic threshold. In Figure 5, we calculated radiative power for all images containing any pixels whose probability exceeds 0.5. However, this probability could be adjusted for different contexts. For example, if conducting a long-term historical analysis, it may be better to set a high confidence threshold and remove as many false positives as possible. Conversely, for near-real-time monitoring it may be important to incorporate as many detections as possible, even if a greater percentage of them might be false.

To illustrate the effects of further filtering the data, we look at time series from Bogoslof Island, Okmok Caldera, Redoubt Volcano, and Augustine Volcanoes, each of which only had one eruption during the time period of study. At all four of these volcanoes combined there are 6,725 total detections made out of 291,283 total images analyzed (Figure 5). These statistics yield a combined detection rate of 2.3% (>97% of images are non-detections). However, if we use only night time data and set a probabilistic threshold of 0.75 at the same volcanoes, HotLINK detects 2,661 hotspots out of 168,400 total images, which is a detection rate of 1.6%. So, with a higher threshold and only using nighttime images HotLINK removes >98% of images as non-detections. These statistics also help us estimate an upper bound on the false positive rate of HotLINK at around 2%, which is similar to what we calculated earlier with the VIIRS test dataset. For comparison to detection rates during eruptions see Section 4.3 in which detection rates of VIIRS and MODIS sensors at Mount Veniaminof during eruptive periods are discussed.

4.2 Analysis of HotLINK probability estimates

In order to use probabilistic predictions from HotLINK for filtering hotspot detections, or for future incorporation into forecasting methods, we must verify that the probabilistic predictions of the model are meaningful. This is especially relevant since modern neural networks have shown a tendency to be overconfident (Guo et al., 2017). Although the model outputs a probability prediction for each pixel in the image, we are most interested in whether the image contains a hotspot at all. Therefore, for the purposes of this analysis we refer to "image probability" as the highest probability of all pixels in the image, since it only takes one hotspot pixel for an image to be classified as active. We evaluate our probability outputs using a reliability diagram, adapted from Hamill, 1997; Figure 6A.

For image probabilities to be well calibrated, we want the accuracy of a thresholded prediction to scale with its probability (Hamill, 1997). For example, if a well-calibrated model predicts five images to contain hotspots at a probability of 80%, four of the images would contain hotspots while one would not. While this may seem counterintuitive, we want some images with high probabilities to be wrong in order to confirm that probabilistic predictions are reliable. We find a strong correlation between the probabilities of HotLINK predictions and whether images contain a hotspot, since they align with the ideal distribution (black line) shown in the reliability diagram below (Figure 6A). This demonstrates that the probabilistic output of HotLINK can be considered a well-calibrated estimate.

While the reliability diagram (Figure 6A) demonstrates that probabilities are well calibrated, we can expand our probabilistic analysis by including the ambiguous images identified by human visual inspection. The ambiguous images contained in the VIIRS validation and test datasets present a great opportunity to compare HotLINK's probability predictions to images we could not confidently classify as volcanic or not. Figure 6B shows that ambiguous images are skewed toward low probabilities, with ~50% of ambiguous images predicted at a probability <0.1. However, ambiguous images are proportionally more represented than each other class in all bins from 0.1–0.8. In other words, ambiguous images are much more likely to be predicted at intermediate probabilities than images labeled as inactive or active. This finding supports the idea that CNNs mimic the visual learning of human experts. It also provides more confidence in the quality

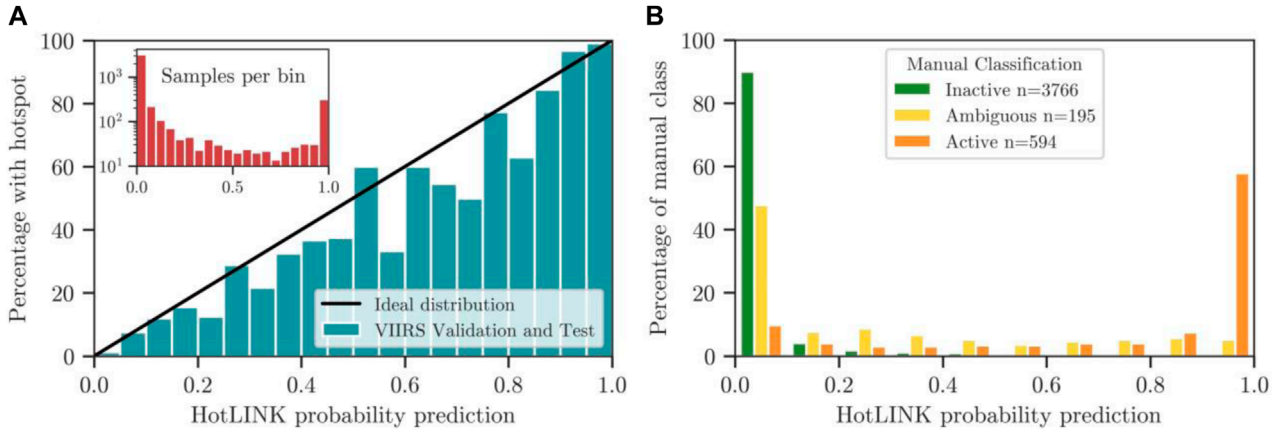


FIGURE 6
Reliability diagram and histogram of VIIRS validation and test datasets. **(A)** Reliability diagram of the HotLINK model applied to the VIIRS training and validation dataset (unambiguous images only). Blue bars represent the proportion of images manually identified as active in 5 percentile bins. The black line represents the ideal probability distribution, indicating that probability predictions are accurate to the true classification. Bars below the black line are overconfident (probability prediction of hotspots is higher than the true probability), and bars above are under-confident (probability prediction of hotspots is lower than true probability). The inset figure shows the number of samples per bin on a logarithmic scale. **(B)** Histogram of the VIIRS validation and VIIRS test datasets, showing the percentage of each class - inactive (green), active (orange), and ambiguous (yellow) - in 10 percentile bins. Ambiguous images are the most represented class at intermediate probabilities (0.1–0.8).

of probabilistic predictions, since images that appear ambiguous to analysts are likely to be predicted at intermediate probabilities by the network.

4.3 Comparison and detection limits of MODIS and VIIRS data

We speculate that the higher accuracy of HotLINK on the MODIS test dataset relative to the VIIRS test and validation datasets is due to the larger pixel size of MODIS preventing small hotspots from being identified by either HotLINK or manual analysis, resulting in an increased number of true negatives for MODIS compared to VIIRS. Similarly, the larger pixel size blurs out smaller scale background variance that is visible in VIIRS data, such that MODIS has a lower false positive rate than VIIRS and a higher F1-score. The larger pixel size of MODIS data results in fewer detections overall than VIIRS.

HotLINK shows a slightly better accuracy on MODIS data than on VIIRS because the MODIS data contains a greater proportion of true negatives and a smaller proportion of false positives. Despite this, the VIIRS data has a higher true positive rate and is able to see smaller and weaker hotspots. To further support this conclusion we compare VIIRS and MODIS detections during three eruptive events at Mount Veniaminof from the analyzed data. From these eruptions we also attempt to quantify a night and daytime detection limit for HotLINK when applied to VIIRS and MODIS data.

Mount Veniaminof had three eruptions between 2012–2022, the time period when both VIIRS and MODIS data are available. These eruptions were effusive-explosive in nature, characterized by lava effusion into and within the intra-caldera glacier, and sporadic ash emissions (Loewen et al., 2021; Waythomas, 2021; Waythomas et al., 2023). Start and end dates for these eruptions are taken from Loewen et al. (2021). During these eruptive periods, both VIIRS and MODIS agree well on RP estimates in our analysis. For the 2013 eruption (June 13—October 17), both MODIS and VIIRS retrieved an average RP of 27.8 MW. During the 2018 eruption (September 4—December 27) MODIS retrieved an average of 27.6 MW and VIIRS 30.2 MW, and for the 2021 eruption (February 28—April 21) MODIS retrieved an average of 6.0 MW and VIIRS 5.0 MW (Figure 7).

Although the average RP retrieved by both sensors is comparable, the VIIRS sensor had a much higher rate of detections during the same eruptive periods. Across all three eruptions, VIIRS had 1,553 detections out of 2,874 total images, for an active percentage of 54%. Meanwhile MODIS had 536 detections out of 1,902 total images, for an active percentage of 28%. We hypothesize VIIRS had a greater active percentage because it was able to capture significantly weaker signals, due to its finer spatial resolution (0.137 km^2 compared to 1 km^2 pixel area at nadir). In future work, this hypothesis could be tested through a more robust analysis of the relative detection rate of VIIRS and MODIS images that are captured at nearly the same time.

To approximate detection limits for both sensors using HotLINK, we use the 5^{th} percentile radiative power of all hotspots detected during the 2013, 2018, and 2021 eruptions at Mount Veniaminof. It is important to acknowledge the possibility of false positives in these data, constituting approximately 2% of samples according to the labeled VIIRS validation and test datasets (Table 3). To mitigate the impact of false positives on the detection limit estimate, we opt for a conservative approach by using the 5th percentile, which is more than twice the estimate for the percentage of false positives in the dataset. This ensures that potential low RP false positives do not artificially lower the detection limit estimate. Still, our estimate for detection limit is not the threshold at which signals are missed, but approximates this by indicating the weakest signals retrieved by HotLINK. This estimate allows us to compare the relative detection limits between sensors. For VIIRS data, we find the 5th percentile of daytime detections to be 0.69 MW, and nighttime detections to be 0.26 MW. For MODIS data, we

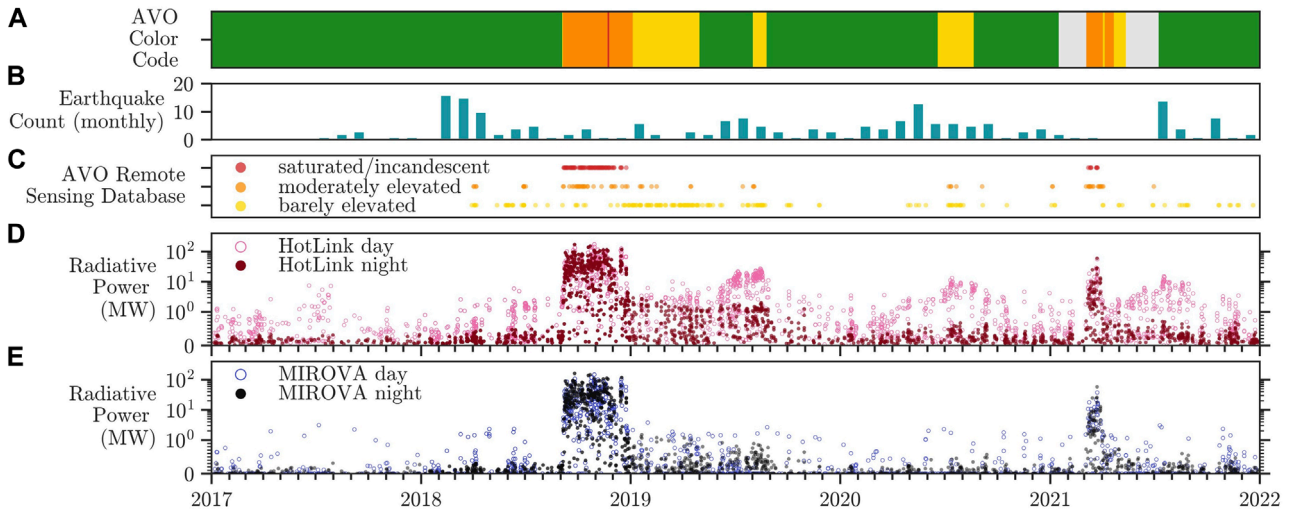


FIGURE 7
Multidisciplinary observations at Mount Veniaminof. Subplots show: **(A)** AVO Color Code timeseries, with color code levels indicated by their respective colors, and gray indicating periods with insufficient monitoring data for AVO to designate a color code (Guffanti and Miller, 2013). **(B)** Monthly earthquake counts within 20 km of Mount Veniaminof, assembled from the USGS ComCat earthquake database (https://earthquake.usgs.gov/earthquakes/search/). **(C)** Analyst flags from the AVO remote sensing database, showing analyst identified thermal signals in VIIRS images, characterized as being "saturated," "moderately elevated," and "barely elevated." **(D)** Mount Veniaminof hotspot detections in VIIRS images from 2017 to 2022 using HotLINK and **(E)** hotspot detections from the adapted MIROVA algorithm.

find the 5th percentile of daytime detections to be 1.4 MW, and nighttime detections to be 0.79 MW. These results demonstrate that HotLINK is 1.8–3 times more sensitive to nighttime observations than daytime observations, and that HotLINK is 2–3x more sensitive when applied to VIIRS data compared to MODIS. To compare with literature values, the MIROVA algorithm applied to MODIS data cites a detection limit of −1 MW irrespective of the time of day (Coppola et al., 2020). This is the first time the authors are aware of a comparison of the detection limits between MODIS and VIIRS I-bands, although the radiative power between MODIS and VIIRS M-bands (750 m at nadir) have been previously compared, finding that the VIIRS M-bands are more sensitive than MODIS bands to thermal signals (Li et al., 2018; Campus et al., 2022). We caution that these detection limits are only approximations, since we are only using one volcano for this analysis and are not looking at the radiative power of missed detections. Detection limits could be more rigorously ascertained by comparing the radiative power of true positive and false negative detections across many volcanoes. Here we only calculated the radiative power for images that were detected as hotspots by HotLINK and statistical analysis of the RP of false negative detections was not done.

4.4 Analysis of HotLINK and adapted MIROVA on the Mount veniaminof time series

Table 4 shows a higher true positive rate of HotLINK relative to our implementation of the MIROVA algorithm, indicating a greater sensitivity to smaller and lower temperature hotspots. Similarly, the high true negative rate of HotLINK relative to this adapted MIROVA indicates that HotLINK is less susceptible to false positive detections. We can expand on this analysis by examining the Mount Veniaminof time series from 2017–2021 to further compare results during eruptive and inter-eruptive periods (Figure 7). During this time period there were two eruptions, one in 2018 and one in 2021. The main difference between HotLINK and the optimized MIROVA detections during this period is that HotLINK detects more hotspots. From an eruption tracking perspective, the MIROVA algorithm does well as it has a similar detection rate as HotLINK during eruptions. In contrast, during non-eruptive periods HotLINK makes a greater number of detections than MIROVA, which may represent volcanic thermal output associated with volcanic unrest, as well as false positives. Therefore, while both models perform well for eruption detection and tracking, HotLINK is able to detect weaker signals that may be relevant for monitoring unrest at Mount Veniaminof.

Figure 7 shows an increase in HotLINK detected RP prior to the 2018 eruption, and more peaks in 2019 and 2020 that are not seen in MIROVA data. These HotLINK detections are consistent with AVO analyst checks of VIIRS MIR images, where analysts observed weakly to moderately elevated surface temperatures qualitatively prior to eruption at Mount Veniaminof, and again during discrete time periods in the summers of 2019 and 2020 (Figures 7C,D; Cameron et al., 2018; Orr et al., 2023). We therefore find that the HotLINK detections are real, capturing weaker, but notable above-background thermal signals as seen in both the rate and radiative power of detections. These HotLINK results also have the advantage of providing quantitative information in comparison to the qualitative AVO remote sensing database classifications of "barely elevated," "moderately elevated," and "saturated/incandescent."

Inspection of these signals through complementary high resolution optical satellite imagery (e.g. Sentinel-2, Maxar) suggests that they comprise a combination of subtle surface heating, potentially due to increased vent degassing behavior at the volcano, as well as a seasonal signature due to the still-warm 2018 lavas

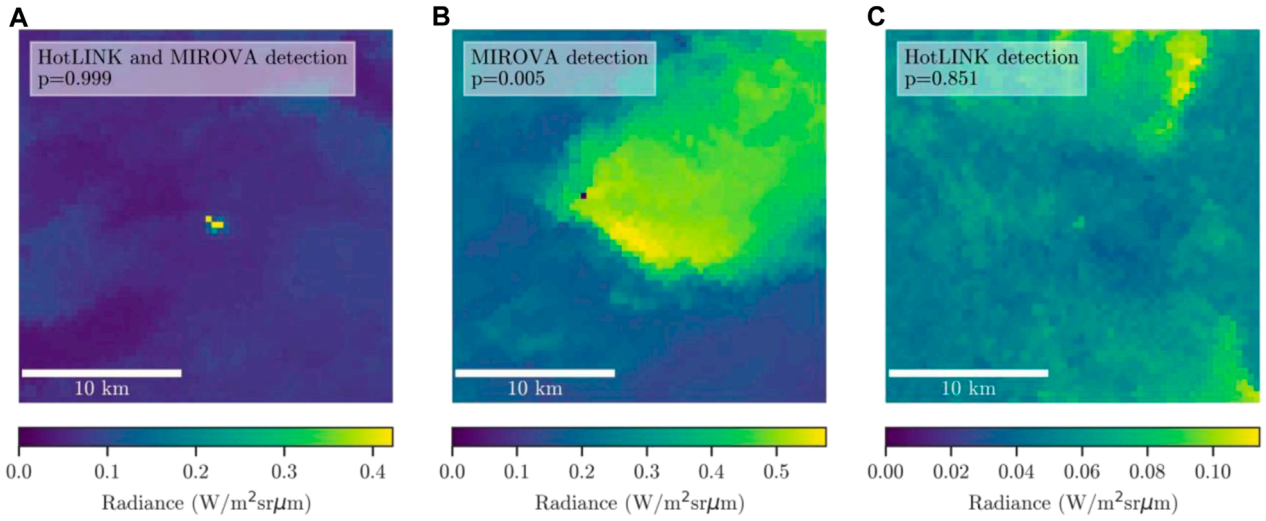


FIGURE 8
Example images from the VIIRS validation dataset. All images show MIR spectral radiance (Wm^{-2} sr^{-1} μm^{-1}) at Mount Veniaminof. **(A)** a true hotspot detection made by both HotLINK and the adapted MIROVA algorithm (nighttime image), **(B)** a false positive detection of a bright cloud made by the adapted MIROVA algorithm (daytime image), and **(C)** a true positive detection of a more subtle hotspot made by HotLINK, which is missed by the adapted MIROVA algorithm (night image). All images are 64 x 64 pixels, or roughly 24 x 24 km. Note that each image has its own colorbar scale in order to show the maximum contrast within each image.

readily melting the overlying snow cover in spring. The 2018 pre-eruptive hotspot signals suggest increased thermal output, perhaps via increased degassing or ground surface temperatures of the active cone (Orr et al., 2023). The 2019 and 2020 peaks in RP coincide with seasonal snow melting that exposed the large and relatively-warm lava flow field, but these signals also coincide with seismic unrest noted by AVO that prompted AVO to raise the color code from green to yellow on 1 August 2019 for 24 days and on 18 June 2020 for 64 days (Cameron et al., 2018; Orr et al., 2023). Further analysis of the detected radiative power in comparison with complementary multiparameter datasets and higher resolution infrared images (e.g., Figure 1) could help tease out the origins and processes associated with these detections.

Our analysis shows that while both HotLINK and MIROVA are able to detect large and high temperature hotspots (e.g. Figure 8A), more subtle hotspots (Figure 8C) are only detected by HotLINK. The MIROVA system struggles to disregard bright and dispersed signals, such as solar reflections off of clouds, which exceed thresholds defined in the algorithm, but are visibly not hotspots in context (Figure 8B). HotLINK is able to detect more subtle hotspots that may be weak but still match the spatial patterns of a discrete thermal signal. The detection capabilities of HotLINK are similar to what an analyst can detect by eye.

5 Conclusion

This study confirms the capability of machine learning, specifically convolutional neural networks, to automate remote sensing tasks usually designated to human experts (Corradino et al., 2023). This technology provides three main improvements relative to threshold-based algorithms: 1) the model is more sensitive to subtle signals and can detect a larger number of hotspots while also detecting fewer false positive hotspots, 2) the probabilistic nature of the detections makes the model useful for different monitoring contexts, and 3) the same model performs well on data from different sensors (MODIS and VIIRS) and different Alaska volcanoes (with some caveats for volcanoes that are islands or have crater lakes).

The ability to detect more and weaker hotspots opens up the possibility of detecting precursory as well as eruptive hotspot signals. Specifically, our network detects subtle increases in volcanic surface temperature from Mount Veniaminof that correspond with both increased number of analyst detections of thermal signals and elevated seismicity. The capability to detect subtle signals associated with volcanic unrest, as well as eruptions, may aid in eruption forecasting efforts. Another advantage of our network is the probabilistic output. This expands the amount of information available to human analysts and will facilitate incorporation into statistical eruption forecasting models.

We found that HotLINK was able to detect hotspots in MODIS data with an even higher accuracy than for VIIRS data. Our model is therefore directly applicable to both VIIRS and MODIS data and is shown to work well on multiple volcanoes, only producing large errors in cases with crater lakes or small island volcanoes, which are especially susceptible to seasonal false detections. These errors could be minimized in the future using a detection threshold that exceeds the seasonal background signals at relevant volcanoes and/or by filtering out daytime images.

In conclusion, with a labeled training dataset of less than 4,000 VIIRS images from two volcanoes we were able to train a model to detect hotspots in both VIIRS and MODIS data that is applicable to many volcanoes. The time series for the eight volcanoes analyzed here captures volcanic unrest and eruption and thus can provide critical input into data-driven volcano monitoring and forecasting

studies, as well as valuable insight into the magmatic and eruptive processes occurring in active volcanic systems across Alaska. The model itself is also readily applicable for near-real-time or historical hotspot detection efforts by volcano observatories.

6 Plain language summary

Volcanoes release heat on their surface, and by monitoring this heat, we can determine if a volcano is erupting or might erupt soon. Heated areas, called hotspots, can be detected by satellite sensors, which generate images from space in infrared wavelengths. Traditionally, volcanologists or simple computer programs would identify the hotspots in infrared images. Now, advanced computer algorithms based on artificial intelligence can accurately identify complex features in images. We used these algorithms to improve the way we detect volcanic hotspots. Our approach detects more subtle heat signals than other algorithms, which is useful for detecting different types of volcanic activity, and may contribute to better forecasting of volcanic eruptions.

Data availability statement

The trained HotLINK model and datasets generated for this study can be found on Github: https://github.com/csaundersshultz/HotLINK.

Author contributions

PS-S: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Software, Validation, Visualization, Writing–original draft, Writing–review and editing. TL: Funding acquisition, Supervision, Writing–review and editing, Conceptualization, Investigation, Project administration, Resources. HD: Conceptualization, Software, Writing–review and editing. TG: Writing–review and editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. Funding for this project came from the National Science Foundation through a Prediction of and Resilience against Extreme Events (PREEVENTS) award (number 1855126), and also through U.S. Geological Survey Cooperative Agreement No. G21AC10384.

Acknowledgments

Thanks to Israel Brewster among others at the University of Alaska Fairbanks who aided in the efforts of the authors.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Author disclaimer

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the United States Government.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/feart.2024.1345104/full#supplementary-material

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., et al. (2015). TensorFlow: large-scale machine learning on heterogeneous systems. Available at: https://www.tensorflow.org/.

Apple (2023). About Face ID advance technology. Available at: https://support.apple.com/en-us/102381 (Accessed November 22, 2023).

Aveni, S., Laiolo, M., Campus, A., Massimetti, F., and Coppola, D. (2023). The capabilities of FY-3D/MERSI-II sensor to detect and quantify thermal volcanic activity: the 2020–2023 Mount Etna case study. *Remote Sens.* 15 (10), 2528. doi:10.3390/rs15102528

Blackett, M. (2013). Review of the utility of infrared remote sensing for detecting and monitoring volcanic activity with the case study of shortwave infrared data for Lascar Volcano from 2001–2005. *Geol. Soc. Lond. Spec. Publ.* 380, 107–135. doi:10.1144/SP380.10

Blackett, M. (2017). An overview of infrared remote sensing of volcanic activity. *J. Imaging* 3 (2), 13. doi:10.3390/jimaging3020013

Bleick, H. A., Coombs, M. L., Cervelli, P. F., Bull, K. F., and Wessels, R. L. (2013). Volcano–ice interactions precursory to the 2009 eruption of Redoubt Volcano, Alaska. *J. Volcanol. Geotherm. Res.* 259, 373–388. doi:10.1016/j.jvolgeores.2012.10.008

Buslaev, A., Iglovikov, V. I., Khvedchenya, E., Parinov, A., Druzhinin, M., and Kalinin, A. A. (2020). Albumentations: fast and flexible image augmentations. *Information* 11 (2), 125. doi:10.3390/info11020125

Cameron, C. E., Prejean, S. G., Coombs, M. L., Wallace, K. L., Power, J. A., and Roman, D. C. (2018). Alaska Volcano Observatory alert and forecasting timeliness: 1989–2017. *Front. Earth Sci.* 6, 1–16. doi:10.3389/feart.2018.00086

Campus, A., Laiolo, M., Massimetti, F., and Coppola, D. (2022). The transition from MODIS to VIIRS for global volcano thermal monitoring. *Sensors* 22, 1713. doi:10.3390/s22051713

Carter, A. J., Ramsey, M. S., and Belousov, A. B. (2007). Detection of a new summit crater on Bezymianny Volcano lava dome: satellite and field-based thermal data. *Bull. Volcanol.* 69 (7), 811–815. doi:10.1007/s00445-007-0113-x

Castaño, L. M., Ospina, C. A., Cadena, O. E., Galvis-Arenas, B., Londono, J. M., Laverde, C. A., et al. (2020). Continuous monitoring of the 2015–2018 Nevado del Ruiz activity, Colombia, using satellite infrared images and local infrasound records. *Earth, Planets Space* 72, 81. doi:10.1186/s40623-020-01197-z

Chevrel, M. O., Villeneuve, N., Grandin, R., Froger, J. L., Coppola, D., Massimetti, F., et al. (2023). Lava flow daily monitoring: the case of the 19 September–5 October 2022 eruption at Piton de la Fournaise. *Volcanica* 6 (2), 391–404. doi:10.30909/vol.06.02.391404

Coombs, M. L., Wech, A. G., Haney, M. M., Lyons, J. J., Schneider, D. J., Schwaiger, H. F., et al. (2018). Short-term forecasting and detection of explosions during the 2016–2017 eruption of Bogoslof volcano, Alaska. *Front. Earth Sci.* 6, 1–17. doi:10.3389/feart.2018.00122

Coppola, D., Cardone, D., Laiolo, M., Aveni, S., Campus, A., and Massimetti, F. (2023). Global radiant flux from active volcanoes: the 2000–2019 MIROVA database. *Front. Earth Sci.* 11. doi:10.3389/feart.2023.1240107

Coppola, D., Laiolo, M., Cigolini, C., Donne, D. D., and Ripepe, M. (2016). Enhanced volcanic hot-spot detection using MODIS IR data: results from the MIROVA system. *Geol. Soc. Lond. Spec. Publ.* 426, 181–205. doi:10.1144/SP426.5

Coppola, D., Laiolo, M., Cigolini, C., Massimetti, F., Delle Donne, D., Ripepe, M., et al. (2020). Thermal remote sensing for global volcano monitoring: experiences from the MIROVA system. *Front. Earth Sci.* 7, 362. doi:10.3389/feart.2019.00362

Coppola, D., Laiolo, M., Delle Donne, D., Ripepe, M., and Cigolini, C. (2014). Hot-spot detection and characterization of strombolian activity from MODIS infrared data. *Int. J. Remote Sens.* 35 (9), 3403–3426. doi:10.1080/01431161.2014.903354

Coppola, D., Piscopo, D., Laiolo, M., Cigolini, C., Delle Donne, D., and Ripepe, M. (2012). Radiative heat power at Stromboli volcano during 2000–2011: twelve years of MODIS observations. *J. Volcanol. Geotherm. Res.* 215, 48–60. doi:10.1016/j.jvolgeores.2011.12.001

Coppola, D., Valade, S., Masias, P., Laiolo, M., Massimetti, F., Campus, A., et al. (2022). Shallow magma convection evidenced by excess degassing and thermal radiation during the dome-forming Sabancaya eruption (2012–2020). *Bull. Volcanol.* 84 (2), 16. doi:10.1007/s00445-022-01523-1

Corradino, C., Ramsey, M. S., Pailot-Bonnetat, S., Harris, A. J. L., and Negro, C. D. (2023). Detection of subtle thermal anomalies: deep learning applied to the ASTER global volcano dataset. *IEEE Trans. Geoscience Remote Sens.* 61, 1–15. doi:10.1109/TGRS.2023.3241085

Dehn, J., Dean, K., and Engle, K. (2000). Thermal monitoring of North Pacific volcanoes from space. *Geology* 28 (8), 755–758. doi:10.1130/0091-7613(2000)028<0755:tmonpv>2.3.co;2

Dehn, J., Dean, K., Engle, K., and Izbekov, P. (2002). Thermal precursors in satellite images of the 1999 eruption of Shishaldin Volcano. *Bull. Volcanol.* 64, 525–534. doi:10.1007/s00445-002-0227-0

El Adoui, M., Mahmoudi, S. A., Larhmam, M. A., and Benjelloun, M. (2019). MRI breast tumor segmentation using different encoder and decoder CNN architectures. *Computers* 8 (3), 52. doi:10.3390/computers8030052

Ferri, C., Hernández-Orallo, J., and Modroiu, R. (2009). An experimental comparison of performance measures for classification. *Pattern Recognit. Lett.* 30 (1), 27–38. doi:10.1016/j.patrec.2008.08.010

Ganci, G., Vicari, A., Fortuna, L., and Del Negro, C. (2011). The HOTSAT volcano monitoring system based on combined use of SEVIRI and MODIS multispectral data. *Ann. Geophys.* 54, 544–550. doi:10.4401/ag-5338

Genzano, N., Pergola, N., and Marchese, F. (2020). A google Earth engine tool to investigate, map and monitor volcanic thermal anomalies at global scale by means of mid-high spatial resolution satellite data. *Remote Sens.* 12, 3232. doi:10.3390/rs12193232

Girona, T., Realmuto, V., and Lundgren, P. (2021). Large-scale thermal unrest of volcanoes for years prior to eruption. *Nat. Geosci.* 14 (4), 238–241. doi:10.1038/s41561-021-00705-4

Glorot, X., and Bengio, Y. (2010). "Understanding the difficulty of training deep feedforward neural networks," in Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, in Proceedings of Machine Learning Research, 249–256. Available at: https://proceedings.mlr.press/v9/glorot10a.html.

Gouhier, M., Guéhenneux, Y., Labazuy, P., Cacault, P., Decriem, J., and Rivet, S. (2016). HOTVOLC: a web-based monitoring system for volcanic hot spots. *Geol. Soc. Lond. Spec. Publ.* 426, 223–241. doi:10.1144/SP426.31

Guffanti, M., and Miller, T. P. (2013). A volcanic activity alert-level system for aviation: review of its development and application in Alaska. *Nat. hazards* 69, 1519–1533. doi:10.1007/s11069-013-0761-4

Guo, C., Pleiss, G., Sun, Y., and Weinberger, K. Q. (2017). "On calibration of modern neural networks," in Proceedings of the 34th International Conference on Machine Learning, USA, August 6 - 11, 2017 (PMLR), 1321–1330.

Hamill, T. M. (1997). Reliability diagrams for multicategory probabilistic forecasts. *Weather Forecast.* 12 (4), 736–741. doi:10.1175/1520-0434(1997)012<0736:rdfmpf>2.0.co;2

Harris, A. (2013). *Thermal remote sensing of active volcanoes: a user's manual*. Cambridge: Cambridge University Press.

Harris, A. J., De Groeve, T., Garel, F., and Carn, S. A. (2016). "Detecting, modelling and responding to effusive eruptions,"*Geol. Soc. Lond*. doi:10.1144/SP426.0

Harris, A. J., and Stevenson, D. S. (1997). Thermal observations of degassing open conduits and fumaroles at Stromboli and Vulcano using remotely sensed data. *J. Volcanol. Geotherm. Res.* 76 (3-4), 175–198. doi:10.1016/S0377-0273(96)00097-2

Harris, A. J., Villeneuve, N., Di Muro, A., Ferrazzini, V., Peltier, A., Coppola, D., et al. (2017). Effusive crises at Piton de la Fournaise 2014–2015: a review of a multi-national response model. *J. Appl. Volcanol.* 6 (1), 11–29. doi:10.1186/s13617-017-0062-9

Higgins, J., and Harris, A. (1997). Vast: a program to locate and analyse volcanic thermal anomalies automatically from remotely sensed data. *Comput. Geosciences* 23, 627–645. doi:10.1016/S0098-3004(97)00039-3

Hirn, B., Di Bartola, C., and Ferrucci, F. (2009). Combined use of SEVIRI and MODIS for detecting, measuring, and monitoring active lava flows at erupting volcanoes. *IEEE Trans. Geoscience Remote Sens.* 47 (8), 2923–2930. doi:10.1109/TGRS.2009.2014224

Hurst, A. W., Bibby, H. M., Scott, B. J., and McGuinness, M. J. (1991). The heat source of Ruapehu crater lake; deductions from the energy and mass balances. *J. Volcanol. Geotherm. Res.* 46, 1–20. doi:10.1016/0377-0273(91)90072-8

Kingma, D. P., and Ba, J. (2014). Adam: a method for stochastic optimization. doi:10.48550/arXiv.1412.6980

Laiolo, M., Coppola, D., Barahona, F., Benítez, J. E., Cigolini, C., Escobar, D., et al. (2017). Evidences of volcanic unrest on high-temperature fumaroles by satellite thermal monitoring: the case of Santa Ana volcano, El Salvador. *J. Volcanol. Geotherm. Res.* 340, 170–179. doi:10.1016/j.jvolgeores.2017.04.013

Layana, S., Aguilera, F., Rojo, G., Vergara, Á., Salazar, P., Quispe, J., et al. (2020). Volcanic Anomalies monitoring System (VOLCANOMS), a low-cost volcanic monitoring system based on Landsat images. *Remote Sens.* 12 (10), 1589. doi:10.3390/rs12101589

LeCun, Y., Kavukcuoglu, K., and Farabet, C. (2010). "Convolutional networks and applications in vision," in Proceedings of 2010 IEEE international symposium on circuits and systems, China, 3-6 May 2010 (IEEE), 253–256. doi:10.1109/ISCAS.2010.5537907

Li, F., Zhang, X., Kondragunta, S., and Csiszar, I. (2018). Comparison of fire radiative power estimates from VIIRS and MODIS observations. *J. Geophys. Res. Atmos.* 123, 4545–4563. doi:10.1029/2017JD027823

Loewen, M. W., Dietterich, H. R., Graham, N., and Izbekov, P. (2021). Evolution in eruptive style of the 2018 eruption of Veniaminof volcano, Alaska, reflected in groundmass textures and remote sensing. *Bull. Volcanol.* 83 (11), 72. doi:10.1007/s00445-021-01489-6

Lombardo, V. (2016). AVHotRR: near-real time routine for volcano monitoring using IR satellite data. *Geol. Soc. Lond. Spec. Publ.* 426, 73–92. doi:10.1144/SP426.18

Loughlin, S. C., Vye-Brown, C., Sparks, R. S. J., Brown, S. K., Barclay, J., Calder, E., et al. (2015). An introduction to global volcanic hazard and risk. *Glob. Volcan. Hazards Risk* 2015, 1–80. doi:10.1017/CBO9781316276273.003

Massimetti, F., Coppola, D., Laiolo, M., Valade, S., Cigolini, C., and Ripepe, M. (2020). Volcanic hot-spot detection using SENTINEL-2: a comparison with MODIS–MIROVA thermal data series. *Remote Sens.* 12 (5), 820. doi:10.3390/rs12050820

NASA Goddard Space Flight Center (2018). NASA visible infrared imaging radiometer suite level-1B product user guide. Available at: https://landweb.modaps.eosdis.nasa.gov/NPP/forPage/NPPguide/NASAVIIRSL1BUGMay2018.pdf.

Oppenheimer, C., Rothery, D. A., and Francis, P. W. (1993). Thermal distributions at fumarole fields: implications for infrared remote sensing of active volcanoes. *J. Volcanol. Geotherm. Res.* 55 (1-2), 97–115. doi:10.1016/0377-0273(93)90092-6

Orr, T. R., Cameron, C. E., Dietterich, H. R., Dixon, J. P., Enders, M. L., Grapenthin, R., et al. (2023). 2019 volcanic activity in Alaska—summary of events and response of the Alaska Volcano Observatory. *U. S. Geol. Surv*. doi:10.3133/sir20235039

Pergola, N., Marchese, F., and Tramutoli, V. (2004). Automated detection of thermal features of active volcanoes by means of infrared AVHRR records. *Remote Sens. Environ.* 93, 311–327. doi:10.1016/j.rse.2004.07.010

Pieri, D., and Abrams, M. (2005). ASTER observations of thermal anomalies preceding the April 2003 eruption of Chikurachki volcano, Kurile Islands, Russia. *Remote Sens. Environ.* 99, 84–94. doi:10.1016/j.rse.2005.06.012

Planck, M. (1914). *The theory of heat radiation*. USA: Blakiston.

Pritchard, M. E., Poland, M., Reath, K., Andrews, B., Bagnardi, M., Biggs, J., et al. (2022). Optimizing satellite resources for the global assessment and mitigation of volcanic hazards—suggestions from the USGS powell center volcano remote sensing working group. *U. S. Geol. Surv*. doi:10.3133/sir20225116

Ramsey, M. S., Corradino, C., Thompson, J. O., and Leggett, T. N. (2023). Statistical retrieval of volcanic activity in long time series orbital data: implications for forecasting future activity. *Remote Sens. Environ.* 295, 113704. doi:10.1016/j.rse.2023.113704

Ramsey, M. S., Wessels, R. L., and Anderson, S. W. (2012). Surface textures and dynamics of the 2005 lava dome at Shiveluch Volcano, Kamchatka. *Bulletin* 124 (5-6), 678–689. doi:10.1130/B30580.1

Reath, K. A., Ramsey, M. S., Dehn, J., and Webley, P. W. (2016). Predicting eruptions from precursory activity using remote sensing data hybridization. *J. Volcanol. Geotherm. Res.* 321, 18–30. doi:10.1016/j.jvolgeores.2016.04.027

Ronneberger, O., Fischer, P., and Brox, T. (2015). INet: convolutional networks for biomedical image segmentation. *IEEE Access* 9, 16591–16603. doi:10.1109/ACCESS.2021.3053408

Rouwet, D., Tassi, F., Mora-Amador, R., Sandri, L., and Chiarini, V. (2014). Past, present and future of volcanic lake monitoring. *J. Volcanol. Geotherm. Res.* 272, 78–97. doi:10.1016/j.jvolgeores.2013.12.009

Saunders-Shultz, P. (2023). Hotspot learning and identification network (HotLINK). Available at: https://github.com/csaundersshultz/HotLINK.

Sola, J., and Sevilla, J. (1997). Importance of input data normalization for the application of neural networks to complex industrial problems. *IEEE Trans. Nucl. Sci.* 44 (3), 1464–1468. doi:10.1109/23.589532

Valade, S., Ley, A., Massimetti, F., D'Hondt, O., Laiolo, M., Coppola, D., et al. (2019). Towards global volcano monitoring using multisensor sentinel missions and artificial intelligence: the MOUNTS monitoring system. *Remote Sens.* 11, 1528–1531. doi:10.3390/rs11131528

Wan, J., Wang, D., Hoi, S. C. H., Wu, P., Zhu, J., Zhang, Y., et al. (2014). "Deep learning for content-based image retrieval: a comprehensive study," in Proceedings of the 22nd ACM international conference on Multimedia, 2014, November (IEEE), 157–166. doi:10.1145/2647868.2654948

Waythomas, C. (2021). Simultaneous effusive and explosive cinder cone eruptions at Veniaminof Volcano, Alaska. *Volcanica* 4 (2), 295–307. doi:10.30909/vol.04.02.295307

Waythomas, C. F., Edwards, B. R., Miller, T. P., and McGimsey, R. G. (2023). Lava-ice interactions during historical eruptions of Veniaminof Volcano, Alaska and the potential for meltwater floods and lahars. *Nat. Hazards* 115 (1), 73–106. doi:10.1007/s11069-022-05523-4

Werner, C., Kern, C., Coppola, D., Lyons, J. J., Kelly, P. J., Wallace, K. L., et al. (2017). Magmatic degassing, lava dome extrusion, and explosions from Mount Cleveland volcano, Alaska, 2011–2015: insight into the continuous nature of volcanic activity over multi-year timescales. *J. Volcanol. Geotherm. Res.* 337, 98–110. doi:10.1016/j.jvolgeores.2017.03.001

Wooster, M. (2003). Fire radiative energy for quantitative study of biomass burning: derivation from the BIRD experimental satellite and comparison to MODIS fire products. *Remote Sens. Environ.* 86 (1), 83–107. doi:10.1016/S0034-4257(03)00070-1

Wright, R. (2016). MODVOLC: 14 years of autonomous observations of effusive volcanism from space. *Geol. Soc. Spec. Publ.* 426, 23–53. doi:10.1144/SP426.12

Wright, R., Flynn, L. P., Garbeil, H., Harris, A. J. L., and Pilger, E. (2004). MODVOLC: near-real-time thermal monitoring of global volcanism. *J. Volcanol. Geotherm. Res.* 135, 29–49. doi:10.1016/j.jvolgeores.2003.12.008

frontiers | Frontiers in Earth Science

TYPE Original Research
PUBLISHED 17 June 2024
DOI 10.3389/feart.2024.1372621

OPEN ACCESS

EDITED BY
Silvia Scarpetta,
University of Salerno, Italy

REVIEWED BY
Yingxiang Liu,
University of Southern California, United States
Füsun Tut Haklıdır,
Istanbul Bilgi University, Türkiye

*CORRESPONDENCE
Michail Giannoulis,
michail.giannoulis@uca.fr
Andrew Harris,
andrew.harris@uca.fr

RECEIVED 18 January 2024
ACCEPTED 28 May 2024
PUBLISHED 17 June 2024

CITATION
Giannoulis M, Pailot-Bonnétat S, Barra V and Harris A (2024), External factors driving surface temperature changes above geothermal systems: answers from deep learning.
Front. Earth Sci. 12:1372621.
doi: 10.3389/feart.2024.1372621

External factors driving surface temperature changes above geothermal systems: answers from deep learning

Michail Giannoulis[1]*, Sophie Pailot-Bonnétat[2], Vincent Barra[1] and Andrew Harris[2]*

[1]Université Clermont-Auvergne, CNRS, Mines de Saint-Étienne, Clermont-Auvergne-INP, LIMOS, Clermont-Ferrand, France, [2]Université Clermont-Auvergne, CNRS, LMV, Clermont-Ferrand, France

Introduction: The surface expression of enhanced geothermal heat fluxes above an active hydrothermal system causes a surface thermal anomaly (ΔT). The thermal anomaly is expressed by the difference between the temperature within the heated zone (T_h) and the temperature of non-heated surfaces (T_0). Given that the resulting thermal anomaly at the surface is of extremely low magnitude (1°C–5°C at Vulcano, Italy), it is extremely sensitive to overprinting by external factors, namely, meteorological influences on surface temperature variation, such as solar heating, wind and rain.

Methods: To test the sensitivity of the surface to external drivers, we installed two surface temperature measurement stations within the Vulcano's Fossa crater, one inside the thermal anomaly and one outside (separation = 50 m), with a weather station co-located with the T_0 station. Time series of T_h and T_0 were collected for 2020, when the Vulcano Fossa hydrothermal system was at a low and stable level of activity so that external drivers would have been the main influences on T_h and T_0, and hence also ΔT. To test for divergence from normality in terms of diurnal and seasonal variations in T_h and T_0, and the role of external factors in causing abnormality, we used the deep learning engine DITAN: a domain-agnostic framework to detect and interpret anomalies in time-series data.

Results: During the year, DITAN found 16 cases of two types of meteorological events: intense low-pressure systems and high-intensity rainstorms (cloudbursts). Passage of 13 abnormal low-pressure systems were detected (10 between February and May, and three in December), with three abnormal rainstorm events (all in December); all three being coincident with the abnormal low pressure events. We find just two abnormalities in the time series for of T_h and T_0, both of which coincide with passage of abnormal low-pressure systems, and neither of which coincide with abnormal rain events. We conclude that diurnal and annual heating and cooling cycles, subject to normal meteorological inputs and at a surface above a geothermal-heated source, are immune to anomalous behaviour to the external (meteorological) variations.

KEYWORDS

deep learning, geothermal system, anomaly detection, external factor, temperature change

1 Introduction

A "geothermal system" is defined by Hochstein and Browne (2000) as a cascading system through which "natural heat transfer within a confined volume of the Earth's crust where heat is transported from a heat source to a heat sink." Geothermal systems are located in areas where heat flow is enhanced and where the structural setting allows for fluid circulation, such as at convergent plate margins, spreading centers, rift systems and mantle hot spots (Stimac et al., 2015). Geothermal systems not associated with a magmatic source have been termed by Nicholson and Nicholson (1993) "non-volcanic geothermal systems" and can be found in active tectonic areas where heat can be produced by deep water circulation in a faulting context. However, if the heat source is provided by magma, we have a "volcanic geothermal system" (Nicholson and Nicholson, 1993). In a volcanic geothermal system, heat and mass are transferred to the surface from a magma reservoir (Hochstein, 2005), where ascending magmatic fluids mix with descending fluids from the near-surface groundwater system to create a "volcanic-hydrothermal system", hereafter termed "hydrothermal system". A hydrothermal system is thus a system with four key components (Figure 1):

- a heat source (the magmatic source),
- an area of recharge (i.e., a "hydrothermal reservoir" of highly permeable rocks),
- an area of discharge (i.e., the "geothermal field" at the surface), and
- circulation of hydrothermal fluids contributing to mass and energy flows.

Diffuse heat fluxes from the hydrothermal system result in widespread and pervasive ground surface heating to cause a low amplitude (typically 1°C–10°C) thermal anomaly at the surface (e.g., Hochstein and Browne, 2000; Chiodini et al., 2005; Lagios et al., 2007; Aubert et al., 2008; Harris et al., 2009; Diliberto, 2011). The difference between the temperature of heated (T_h) and unheated ground (T_0) is here termed the thermal anomaly (Figure 1, $\Delta T = T_h - -T_0$). Geothermally-heated surfaces are relatively cool compared to other active volcanic surfaces such as pyroclastic and lava flows, so that the associated thermal anomalies are just a few degrees centigrade above the ambient background (Figure 1). The low amplitude of thermal anomalies associated with hydrothermal systems make them difficult to detect and handle (Harris and Stevenson, 1997), and are likely to be extremely sensitive to external factors, especially meteorological factors such as rain, wind, humidity and atmospheric pressure. While low amplitude geothermal thermal anomalies are liable to dampening due to the cooling effect of rain and wind, radiative heat fluxes will be controlled by variations in water vapor pressure (e.g., Sekioka and Yuhara, 1974; Bahrami et al., 2019; Ishibashi et al., 2023), which is in turn controlled by humidity and atmospheric pressure. At such a low temperature system, forced or free convection will also be the dominant heat loss (Keszthelyi et al., 2003), where the convective heat transfer coefficient will depend on wind and air temperature (Harris, 2013).

We thus define the magmatic and hydrothermal reservoir components of the volcanic-hydrothermal system as internal sources that drive changes in surface temperature (T_h), and the atmospheric system as an external driver (Figure 3). To date, most studies have focused on the role of the magmatic source and the permeability of the hydrothermal reservoir in driving

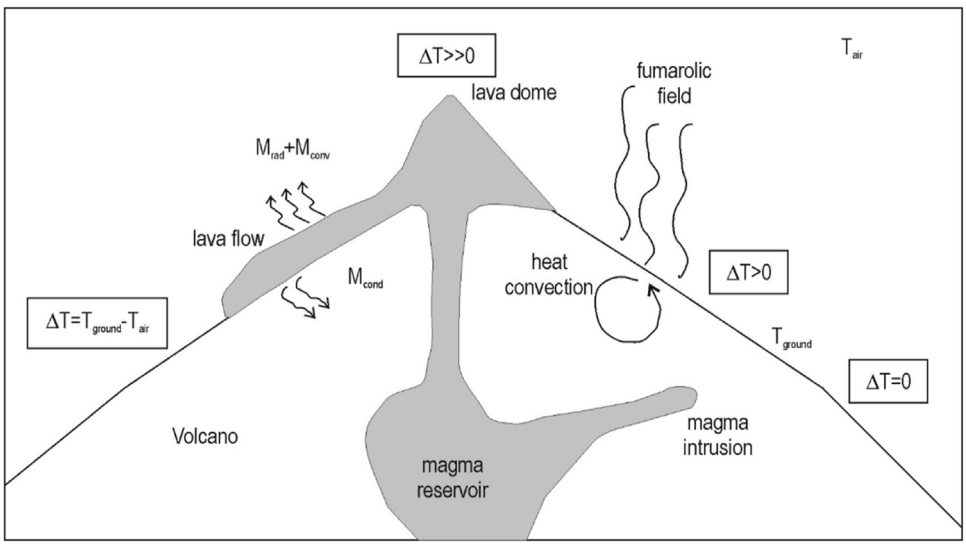


FIGURE 1
Figure 1 from chapter 0 of Harris (2013). Sketch of the main sources of thermal emission at a volcanic hydrothermal system, modified Figure 1 in Bonneville and Gouze (1992) and reproduced by permission of American Geophysical Union. In normal conditions ground (T_{ground}) and air temperature (T_{air}) are approximately equal, so that $\Delta T = T_{ground} - T_{air} \approx 0$. Over a subsurface heat supply, such as a magmatic intrusion above which natural convection in a porous, or fractured, medium carries heat to the surface, ΔT becomes positive. Over a high temperature surface heat source, such as an active lava, ΔT becomes strongly positive. The schematic also shows the main sources of heat loss from an active lava body. These being radiation (M_{rad}), convection (M_{conv}) and conduction (M_{cond}).

changes in temperature recorded at the surface (e.g., Dobson et al., 2003; Chiodini et al., 2005; Aubert et al., 2008; Harris et al., 2012). However, we are aware of no study that has focused purely on the potential role of external drivers on variation in T_h, T_0 and, hence, ΔT (Figure 1). Thus, to provide data capable of defining the role of external (meteorological) drivers on surface temperature variation above a hydrothermal system, we installed two surface-temperature measurement stations (to measure T_h and T_0), plus a weather station co-located with the T_0 station, inside Vulcano's (Aeolian Islands, Italy) Fossa crater (Figure 2). In this regard Vulcano, being the host of a thermal anomaly associated with an active hydrothermal system and a well-monitored site (Chiodini et al., 2005; Harris et al., 2009), is an ideal laboratory to test hypotheses regarding the internal and external effects on the amplidude of DT.

We installed the sensors in 2020, when levels of heat flux were particularly low so that the internal element of the system could be considered minimal and stable, and in a background or baseline state (Mannini et al., 2019; Pailot-Bonnétat et al., 2023). In such a state, temperature variations at the surface should be particularly sensitive to external drivers. However, given the sensitivity of temperature variation to both external and internal drivers (Figure 3), inter-relations between T_h, T_0 and external parameters will be complex, and any correlation will be extremely subtle. Such a multi-variate time series, in which there are unknown interactions between variables, is ideally suited to a machine learning (e.g., Malfante et al., 2018; Carniel and Guzmán, 2020; Watson et al., 2020). In particular, application of a deep learning-based approach to time series data can greatly aid in advancing volcano monitoring, and processing of volcanic geophysical and geochemical signals, especially when searching for patterns and interrelations in multi-sensor time series (e.g., Manley et al., 2022; Corradino et al., 2023; Ferreira et al., 2023).

We thus focus on isolating and defining those external drivers that change the surface thermal state of ground above a hydrothermal system in an abnormal fashion, our research question being:

Can, and if so how, do meteorological factors impinge on surface thermal anomalies above geothermal systems?

To do this we apply the deep learning framework DITAN (Giannoulis et al., 2023) to time series data for surface temperature and meteorological conditions collected *in-situ* at the active hydrothermal system at Vulcano (Pailot-Bonnétat and Harris, 2024), in an experiment designed to answer the stated research question. We begin by describing the characteristics of the study site (Vulcano) and the instrumentation installed there, before reviewing DITAN, its performance and output. In effect, this is the third in a three part series of papers where we have first set-up, tested and cross-validated DITAN in Giannoulis et al. (2023), and then (second) set up the experiment and data collection design in Pailot-Bonnétat and Harris (2024).

2 Background and data

At Vulcano, the magmatic source has been placed at a depth of 2–3 *km* by Ferrucci et al. (1991), and the magmatic contribution to the fumarolic gases has been established from chemical composition (Nuccio et al., 1999). The gas signature at Vulcano thus results from mixing of magmatic and hydrothermal fluids (marine and meteoritic), with the hydrothermal system being described as a biphase (water–vapor) boiling saline solution with a central vapor-monophase zone (e.g., Carapezza et al., 1981; Chiodini et al., 1995; Nuccio et al., 1999). Heat ascends from the mixing zone, which has its depth ≈1 *km* below the Fossa Crater (Nuccio et al., 1999; Alparone et al., 2010) to form a "heat pipe" (White et al., 1971) at the top of which there is a bottom-heated surface zone (Figure 2C). Heat loss from the heated surface is partitioned between the conductive flux through diffuse soil emissions and advection at fumarole vents (Sekioka and Yuhara, 1974; Chiodini et al., 2005; Harris, 2013). Heat flux from the zone of soil emission accounts for 93% ± 2% of the total energy budget, is defined by a thermal anomaly (ΔT) of 1°C–5°C (Harris et al., 2009; Mannini et al., 2019), and is associated with vertical gradients in the temperature profile within 1 *m* of the surface 50–135 K/m (Aubert et al., 2008).

At Vulcano, heat fluxes during the period 2010 to 2020 were particularly low and stable at around 4–12 *MW*, as was ΔT (Mannini et al., 2019). The period 2010–2020 has thus been defined as a baseline or background level for heat flux, against which change in the thermal state of the hydrothermal system can be assessed (Pailot-Bonnétat et al., 2023). We thus chose the year 2020 on which to target our study so as to determine external drivers on ΔT with the system in its background state. At Vulcano, the value of understanding system behaviour during such baseline studies was highlighted by the unrest that followed our period of interest. This new phase of unrest began in September 2021 and continued into 2022, and during which heat fluxes increased to peaks of around 120 MW (Pailot-Bonnétat et al., 2023).

Within the Fossa, in the vertical sense, there are thus two elements to the system that need to be constrained: 1) the immediate subsurface where enhanced geothermal heat flux (internal driver) causes elevated surface temperatures and 2) the atmosphere where meteorological parameters (external drivers) modulate surface temperature (Figure 3). Horizontally, there are also two zones: one of which is heated from below, and one of which is not (Figure 2C). Thus, in our experiment set-up, the surface thermal state of the heated and non-heated zones are tracked by two sensors that monitor surface and air temperature, with external meteorological conditions being measured by a third sensor array (Figure 3).

We installed the sensor network in the Fossa crater in January 2020. The network consists of two temperature stations separated by 50 *m*, with one inside the heated zone and one in the non-heated zone, plus a weather station co-located with the latter station (Figure 2C). The experimental set-up, sensors and data sets used here are fully described in Pailot-Bonnétat and Harris (2024), and are summarized here in Table 1. The two temperature stations consisted of two thermocouples (Onset HOBO TMC1-HD, signal to noise ratio = 0.1°C) measuring surface temperature (T_s) and air temperature (T_a) at a height of 15 *cm* above the surface. The thermocouples were linked to an Onset HOBO U12-008 data logger (storage capacity = 43,000 measurements) with a sampling rate of one record every five-to-ten minutes. The weather station was an Onset HOBO H21-USB which measured atmospheric pressure (S-BPB-CM50), air temperature and relative humidity (S-THC-M002), wind speed (S-WSB-M003), and rainfall using a tipping bucket rain gauge (S-RGF-M002). These meteorological sensors were installed at ground level, and the sampling rate was one record per minute. The chosen period of analysis is 31 January 2020, 12:00:00, to 31 December 2020, 23:00:00, with data gaps existing due to the logger

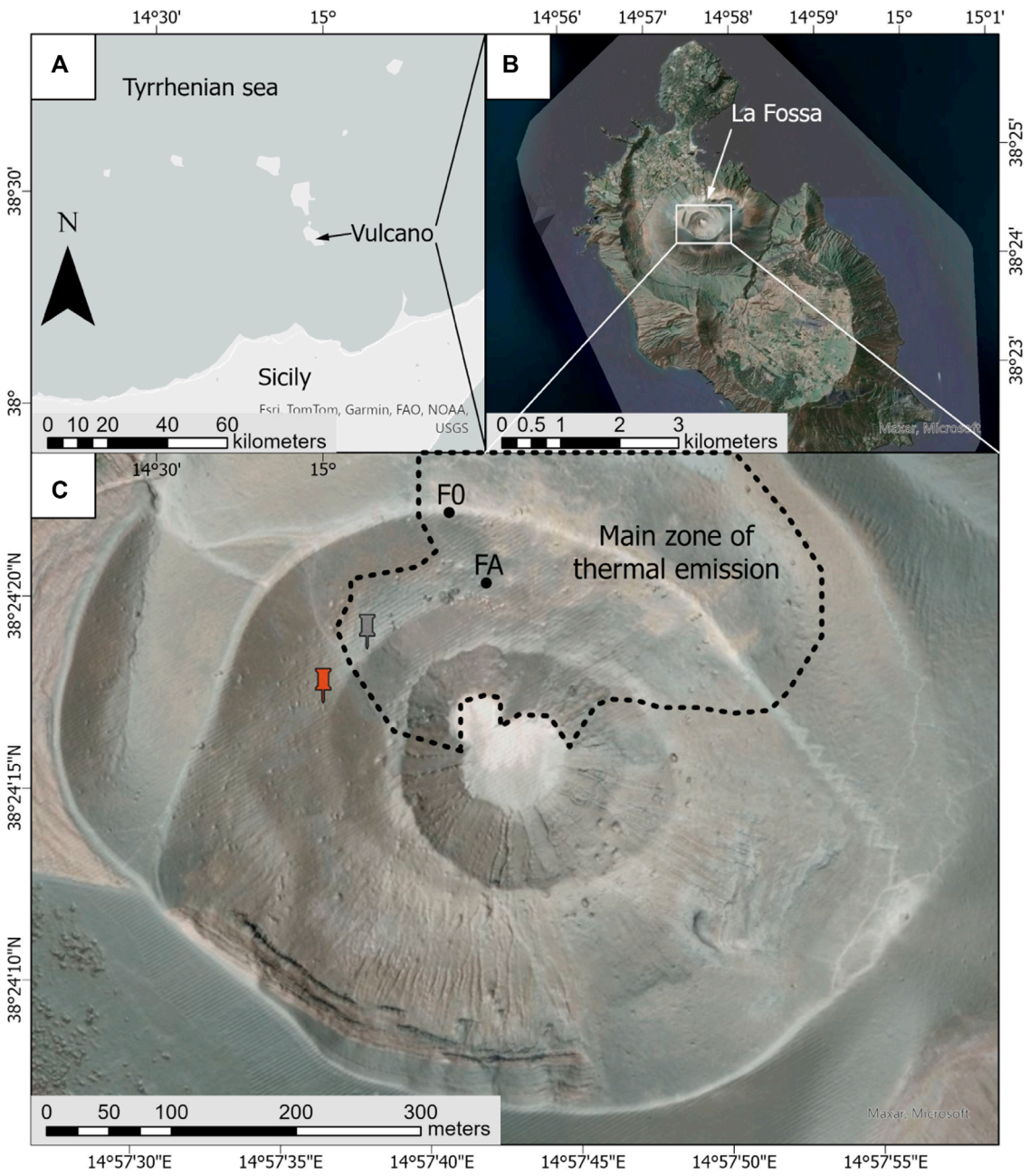


FIGURE 2
(A) Location of Vulcano in the Aeolian Island arc, and **(B)** of the Fossa on Vulcano, with **(C)** location of our measurement stations in respect to the heated zone in the Fossa crater. F0 and FA are the two main fumaroles, gray pin is the temperature station in the heated zone, and red pin is the co-located weather station and temperature station in the non-heated zone.

capacity being reached before download, which was a problem due to mobility restrictions during the COVID-19 pandemic period (Table 2).

3 Methodology

Our sensor network provides a multivariate time series, in which each record is characterized by seven sensors, two sensors from the surface system and five sensors from the atmospheric (external) system (Figure 3). To learn the normal thermal behavior of the surface, and subsequently define externally-driven anomalies, we developed in Giannoulis et al. (2023) DITAN, a domain-agnostic deep learning based framework that is effective in detecting and interpreting temporal-based anomalies. A temporal-based anomaly, or simply anomaly, occurs when one or more sensor values (e.g., pressure and/or wind speed) deviate from their expected "normal" behavior. When this occurs on the full feature set, the anomaly is called full-space anomaly, otherwise sub-space anomaly if it applies to a sub feature set. When an anomaly persists for more

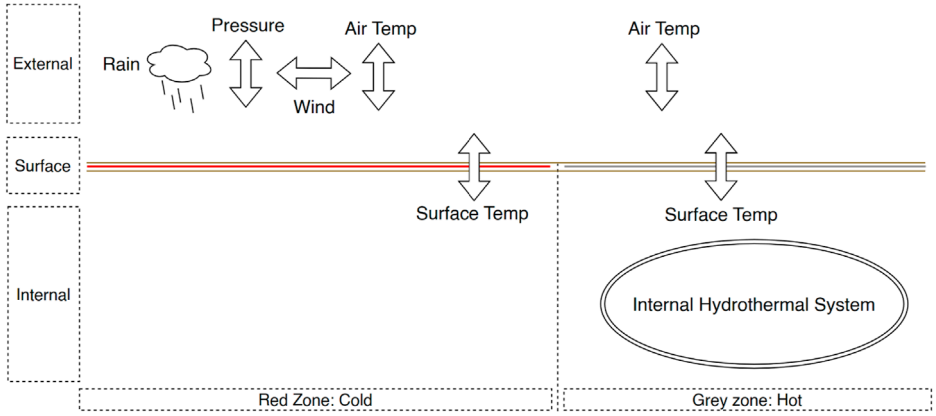


FIGURE 3
A descriptive illustration of the thermal system at Vulcano and its parameters.

TABLE 1 The sensor network of Vulcano in year 2020.

Station	GPS (WGS-84)	Sensors name	Sensors unit
Red (cold)	38.405367, 14.960742	T_s: surface temperature	degrees Celsius (°C)
Grey (hot)	38.405222, 14.960564	T_s: surface temperature	degrees Celsius (°C)
		T_a: air temperature	degrees Celsius (°C)
Weather	38.405367, 14.960742	P: air pressure	millibar (mbar)
		T_a: air temperature	degrees Celsius (°C)
		U: wind speed	meters per second (m/s)
		Rain: rainfall	millimeters (mm)

TABLE 2 Data gaps in the 2020 time series.

Station	Data gaps (mm/dd/yyyy)
Red (cold)	05/31/2020–06/09/2020
	09/04/2020–10/01/2020
	10/20/2020–10/24/2020
Grey (hot)	05/31/2020–06/09/2020
	09/04/2020–10/01/2020
	10/20/2020–10/24/2020
Weather	05/28/2020–06/09/2020
	09/01/2020–10/01/2020

than one record (e.g., two time steps) it is called sub-sequence anomaly. The severity score of an anomaly refers to the intensity of its contamination. Local anomalies typically receive lower scores due to their limited impact on the time series as a whole, being primarily confined to a specific region of the time-series. Global anomalies, conversely, tend to receive higher scores, indicating their widespread impact on the time series and significance across multiple data regions.

DITAN encompasses the integral temporal properties of a time series, built upon three assumptions: 1) the time series is predictable, 2) normality is identical to regularity, and 3) irregular records are temporally less predictable than regular records.

The model has been tuned using Bayesian optimization, to systematically search and exploit the range of values for various hyperparameters, aiming to determine the optimal configuration. Throughout this process, a forward chaining cross-validation was used to ensure that there is no leakage of information between the training and validation sets during the hyperparameter optimization phase.

In Giannoulis et al. (2023), the model has been deeply validated on six multivariate timeseries of varying anomalous types, using several performance metrics including confusion matrices, precision, recall and $F_{0.5}$ score. It has also been favorably compared to state of the art methods on anomaly detection.

DITAN is composed of four modules: 1) pre-processing, 2) domain-agnostic modeling, 3) application of a dynamic threshold with built-in pruning, and 4) numerical interpretation. We briefly describe here these four modules and refer the reader to Giannoulis et al. (2023) for further details.

3.1 Pre-processing

DITAN applies a series of steps to prepare a predictable multivariate time series in a favorable format for deep learning analysis. These steps include an optional decomposition of the sensor values into residuals or trends to focus on short-term or long-term regularities, respectively. Additionally, a mandatory min-max scaling is applied across sensors to maintain a common value range within [−3, 3]. A prediction-based protocol is used to forecast a horizon given a context. In this protocol, a horizon includes the estimation of the records within the context and the prediction of one or more subsequent records. Context-horizon mapping can then be seen as an "IF-THEN" rule, where the "IF" part needs to be estimated along with the prediction of the "THEN" part.

3.2 Domain-agnostic modeling

DITAN learns regularity across the context-horizon mappings using an LSTM encoder-decoder with attention mechanisms (Figure 4). LSTMs (Long Short-Term Memory) (Hochreiter and Schmidhuber, 1997) are a type of recurrent neural network (RNN) specially designed to model long-term dependencies in data sequences. Unlike traditional RNNs, LSTMs incorporate gating mechanisms, including forgetting, input and output gates, which enable them to regulate the flow of information through the network's cells. These mechanisms enable LSTMs to capture and store important information over long periods of time, while mitigating the problems of gradient disappearance or explosion commonly encountered in conventional RNNs. Attention mechanisms (Luong et al., 2015) as for them enable the model to selectively focus on different parts of input data when making predictions. They allow DITAN to weigh the importance of different features or elements within the input, enhancing their ability to capture relevant information for the task at hand. This mechanism typically involves learning attention weights that dynamically allocate resources to different parts of the input.

In DITAN, the context is encoded in such a way that the decoder is responsible for both reconstructing the context and constructing the subsequent records. Each record is represented by an LSTM cell, which helps to memorize short-term relations. Additionally, each LSTM cell in the decoder explicitly attends to the encoder's LSTM cells using a cross-attention layer. Moreover, a masked self-attention layer is employed to assess temporal weights based on the relative positions of the LSTM cells in the decoder. The structure of this architecture is optimized using the Bayes optimizer, which explores and exploits the value range of hyper-parameters with respect to the application features to select an optimal hyper-parameter configuration.

In the training phase, both estimation and prediction errors are utilized. An error e^i is computed as the difference between the forecasted (estimated or predicted) value x^i, and the observed value x^i for each sensor d in D. DITAN employs a configurable error function, which can be defined as either the squared $e^i_d = \|x^i_d - x^{i'}_d\|_2$ or absolute $e^i_d = |x^i_d - x^{i'}_d|$ error for all $d \in D$. The squared, as opposed to absolute, errors are non-linear weighted with respect to the error magnitude, and thus higher errors are magnified. The higher the error on a sensor value, the more abnormal the value is considered. A threshold is required to determine the turning point from normal to abnormal.

Recall that normality is equated to regularity, and irregular records are considered temporally less predictable than regular one. Consequently, during the training phase, the objective function of the model disregards context-horizon patterns that are contaminated by outliers and deemed irregular. These patterns, which constitute a small portion of the total number of patterns, are assigned a low weight and are effectively ignored.

3.3 Dynamic thresholding with built-in pruning

In the testing phase, only prediction errors are used. To determine the level of prediction error that records change from regular (normal) to irregular (abnormal), we use a thresholding methodology (Figure 5). Prediction errors are first smoothed to introduce locality. Next, all critical peaks are computed, which are peak values above the minimum peak height value. To find the minimum peak height value, we downhill the errors in the frequency space until we reach a bin in which variability is introduced. The corresponding error value of that bin is used as the minimum peak height. Then, we expand each critical peak into a critical region, as long as the average error within its expanding region is statistically above the overall average. The boundaries of each critical region are trimmed using a high-pass filter that corresponds to the maximum of error values that are shifting from the clustered errors. Error values within critical regions are considered abnormal (severity scores), and the corresponding records anomalous.

3.4 Numerical interpretation

The magnitude of scores across sensors within anomalous records can be used to provide sufficient information to allow an understanding and troubleshooting of anomalies. The contribution of each sensor d to an anomalous record i, namely, root-cause, is computed using a softmax function across to its corresponding severity scores s:

$$e^{s^i_d} / \sum_j e^{s^i_j}$$

We estimate the similarity between anomalous records in the model-space using the unit dimensionality instead of feature dimensionality. First, we find the internal representations of each record r^i. Then, we apply Gaussian Mixture Modeling (GMM) using an optimal number of components to classify each anomalous record to the component with maximum probability $p(r^i)$. The anomalous records under the same component are considered similar.

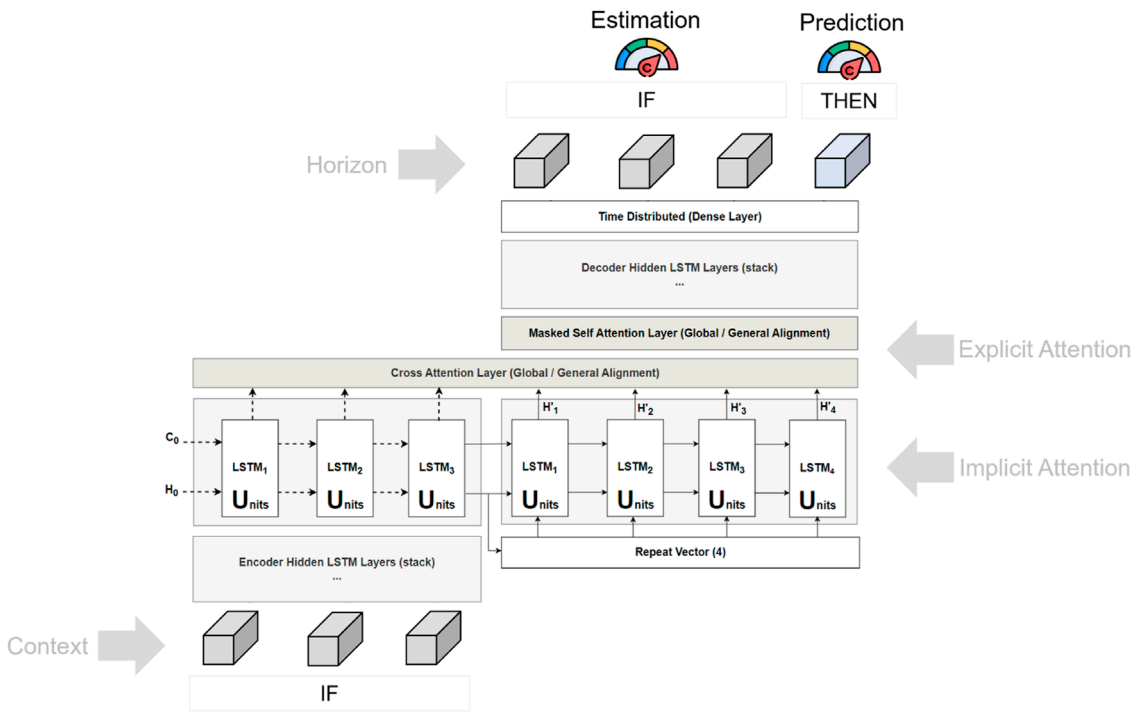


FIGURE 4
A graphical illustration of the DITAN architecture. The model is composed of LSTM layers along with a composite decoder and soft attention mechanisms to capture short-term and long-term normal (regular) patterns. The number of LSTM cells in the encoder (left) is equal to the context size, while the decoder (right) has a context sized number of LSTM cells to reconstruct, as well as a horizon sized number of LSTM cells to construct. Such a composite decoder forces the network to stay attentive to all time steps in the encoder, instead of only the last few steps. A full description of this model is proposed in Giannoulis et al. (2023).

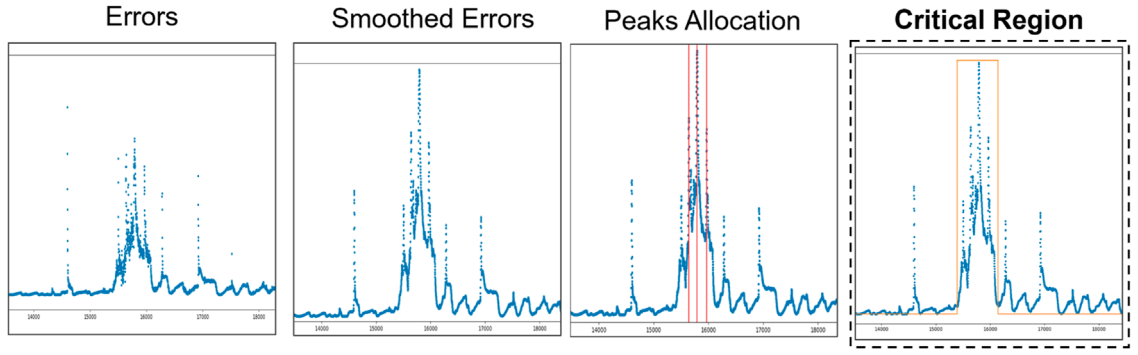


FIGURE 5
DITAN thresholding methodology.

Upon this already validated architecture, we build and introduce here a fifth module (knowledge management), involving physical interpretation of the time series and anomalies. It is a graphical environment which allows to create, read, update and delete (CRUD) knowledge held in the Knowledge Base (KB, see Supplementary Appendix A). The knowledge management module allows complete control over the KB in the form of IF-THEN rules that incorporate temporal constraints or rules. The scope of a rule in our system is to characterize the physical event responsible for the occurrence of a series of conditions.

3.5 DITAN knowledge management module

3.5.1 Conditions

A condition (C) applies to a sensor (S) or physical event (E) that is in an abnormal state. We define five abnormal states: increasing, decreasing, positive, tare and missing values. Although an S-condition can indicate any of these abnormal states, an E-condition can only have a positive abnormal state, meaning that the event has occurred. Therefore, post-condition is always a physical

event, while the preconditions consist of one or more S- and/or E-conditions. A condition occurs when an abnormal state is found to be valid for a certain duration. In a given sequence the validation of abnormal states is thus assessed as it follows:

- Increasing: at least three consecutive values exhibit only an increasing trajectory.
- Decreasing: at least three consecutive values exhibit only a decreasing trajectory.
- Positive: consecutive values exhibit only positive values.
- Tare: the first order difference at each step in the sequence results in the same maximum value.
- Missing value: there is at least one missing time-step in the sequence.

3.5.2 Constraints

Following Giannoulis et al. (2019), the preconditions of a rule are constrained using both logical and temporal operators. The conjunction (AND) logical operator is used to combine conditions, while the negation (NOT) logical operator is used to define the abnormal state of a condition. The conditions are temporally constrained such that the start time of the first condition and the start time of the last condition are at most 60 min, 24 h or 7 days apart, introducing lag resolutions of minutes, hours or days across rules. These constraints ensure that all preconditions occur within a defined time interval. The post-condition of a rule exists from the start time of the first condition to the end time of the last condition.

3.5.3 Inference engine

Following Giannoulis et al. (2019), the rule preconditions guide the inference process by incorporating an event-driven protocol, where DITAN's inference engine is given in Algorithm 1. It is handed the executable rules from the knowledge base (KB) and the anomalous sensor values from critical regions (CR) identified by DITAN in the time series. The objective is to report all post-conditions in the events base (EB), by identifying valid preconditions. This requires support by rule-chains in which the post-condition of a rule lies in the preconditions of another. Thus, in an outer-loop, N executable rules with e events in preconditions are selected from the KB, e being iteratively increased until N becomes zero. At each iteration, the FetchRules function is responsible for selecting and ordering rules. Rules referenced the most frequently are executed first. Given a rule R, REexecute is responsible for validating the occurrences of its conditions across CR (sensor type) or EB (event type). The preconditions of a rule are valid when both temporal and logical constraints are satisfied. The occurrences of a post-condition can also overlap. In such a case, overlapping occurrences are merged. Merging is applied at the end of the algorithm to preserve different event start times which are needed in the chaining process. The final occurrences of R are stored in EB.

```
Require: KB,CR ▷ Knowledge Base (KB), Critical Regions (CR)
Ensure: EB ▷ Events Base (EB)
  e ← 0 ▷ Number of events allowed in rule's precondition
  while True do
     rules ← FetchRules(KB,e) ▷ Fetch rules ordered by the number of events
     if len(rules) == 0 then ▷ Halt if there are no more rules
          break
     end if
     for R ∈ rules do
          occurences ← REexecute(CR,EB,R) ▷ Find occurrences in which R is validated
          if occurences ≠ [] then
               EB.update([R:occurences])▷ Keep only the executed rules
          end if
     end for
     e+ = 1 ▷ Increase number of events for the next iteration
end while
```

Algorithm 1. Inference Engine.

3.5.4 Rule validation

A rule is valid when its logical and temporal constraints are valid. Validation of the logical constraints (negation, conjunction) is straightforward, because each constraint is a well-defined operator. Instead, further analysis is required for handling temporal relations. Once all valid occurrences of the conditions have been examined, the next step is thus to analyze their temporal differences. To accomplish this, DITAN constructs an upper triangular matrix containing all possible pairs of occurrences for all conditions. For instance, if there are four conditions (C1, C2, C3, C4) DITAN will construct six pairs of categories (C1-C2, C1-C3, C1-C4, C2-C3, C2-C4, C3-C4), and the upper triangular matrix allows consideration of each unique pair of conditions without redundancy. Next, DITAN constructs chains between the occurrences of different pairs so as to derive the final temporal relations. A chain is deemed valid if it incorporates all the different pairs of categories. A pair is added to the chain if it does not violate the temporal constraint. Therefore, a rule is linked to the occurrences of its valid chains (if multiple conditions apply) or its valid condition (if there is only a single condition).

3.5.5 Risk factor

The risk of a (valid) rule is related to the intensity of its preconditions. This is usually quantified by experts using fuzzy logic (Leung and Lam, 1988) or certainty factors (Shortliffe, 1976), such as in Giannoulis et al. (2019). However, DITAN quantifies risk by using the severity scores. The severity score of an anomaly refers to the intensity of its contamination. Local anomalies typically receive lower scores due to their limited impact on the time series as a whole, being primarily confined to a specific region in data. Global anomalies, conversely, tend to receive higher scores, indicating their widespread impact on the time series and significance across multiple data regions. The risk associated with an S condition is equal to the average of the severity scores within the partition (duration) j of a critical region: $RF_S = average(CR^j_{scores})$, and the risk of an event E condition is equal to the average risk of its preconditions: $RF_E = average(RF_{S1}, \ldots, RF_{Ek})$. As a result, the risk factor RF provides

TABLE 3 The knowledge defined by Experts in the form of temporal rules.

R1: if decrease on pressure sensor and increase on wind speed sensor started within hours	then low pressure system (meteorological) event
R2: if positive on rain sensor and low pressure system event started within hours	then rainstorm (meteorological) event
R3: if low pressure system event and decrease on red surface temperature and decrease on grey surface temperature started within hours	then decrease on surface temperatures (surface external driver) event
R4: if rainstorm event and decrease on red surface temperature and decrease on grey surface temperature started within hours	then decrease on surface temperatures (surface external driver) event
R5: if low pressure system event and decrease on red surface temperature started within hours	then decrease on red surface temperature (surface external driver) event
R6: if low pressure system event and decrease on grey surface temperature started within hours	then decrease on grey surface temperature (surface external driver) event
R7: if tare on air temperature	then rapid change on air temperature (instrumental) event

a general overview of the predictability offset across conditions within their validated durations. The maximum risk is selected when an E condition occurs multiple times within overlapping durations. Finally, to preserve a relative risk across valid rules, within a probabilistic-like range, max-normalization is applied.

3.6 DITAN: initialization of the knowledge management module for Vulcano

DITAN's "knowledge" is based on rules entered by experts. An example of rule entry via DITAN's graphical interface is given in Supplementary Appendix B. The time scale of "normal" variations in the time series then needs to be defined, before DITAN is trained and anomalies detected.

3.6.1 Rules

For Vulcano, knowledge is expressed in the form of seven temporal rules R as presented in Table 3. The rules were set to describe severe meteorological (external) events, allowing us to assess storm systems as causes of anomalous decreases in surface temperatures, followed by recovery. We define three event types (Table 3):

1. R1 and R2 define a meteorological event, since only external conditions are linked to each other.
2. R3 to R6 characterize anomalous surface temperature events resulting from external drivers (Figure 3).
3. R7 considers sensor failure, thus finding an "instrumental event."

The post-condition of R1 is in the preconditions of R2, R3, R4 and R5, and the post-condition of R2 is in precondition of R4. Instead, R1 and R7 have no events in their preconditions. Therefore, all of the possible inferences are: (a) R1, (b) R1 → R2, (c) R1 → R3, (d) R1 → R2 → R4, (e) R1 → R5, (f) R1 → R6, (g) R7. According to Algorithm 1 executions are divided into three iterations. In the first iteration, (a, g) are executed. In the second iteration, (b, c, e, f) are executed. Finally, in the third iteration, (d) is executed.

3.6.2 Data preparation

There are seven sensors in our Vulcano network (Table 1), for which the frequency of measurements varies between one record every 5 min for the hot and cold stations, and one record per minute for the weather station. Thus, all measurements were sub-sampled to a frequency of 5 min, and then down-sampled (averaged) to a common frequency of one record (time-step) per hour. This results in 6,799 records (for around 283 days), where each record is a vector of seven sensor values. Thus, the temporal resolution of anomalies is expected to be of at least 1 h. An overview of range of values recorded for data set is given in Table 4. There are three main data set characteristics:

1. All temperature sensors exhibit diurnal and annual cycles (normality).
2. Sensors do not have similar value ranges, have scales that differ by six orders of magnitude between rainfall, through temperature, humidity and wind speed, to pressure.
3. There are three significant data gaps in June, September, and October.

3.6.3 Pre-processing

Of the 1,253 missing time-steps (hours), only 73 can be linearly interpolated. The remaining 1,180 missing time-steps were removed, because they formed gaps that were too long to allow interpolation. Since normality is learned as the regularity of the context-horizon mappings, the impact of these data gaps on the frequency of occurrence of regular context-horizon mappings is negligible due to their relatively small size.

Although decomposition is an option, we chose to decompose measurements solely into the residuals component. This decision was based on the understanding that external phenomena manifest themselves as short-term interruptions to normality causing perturbations to the diurnal cycle. In contrast, the internal driver primarily affects the long-term trend. Therefore, the values of each sensor are transformed into residuals by estimating its decomposition type and period. Min-max normalization is then applied across all sensors to introduce a common scale, without biasing any correlations or underlying distributions. The resulting

TABLE 4 Sensor description for Vulcano in the year 2020.

Station	Sensor	Unit	Min	Max	Mean	Standard Deviation
Weather	Pressure	mbar	967	1,002	986	5.8
Weather	Air-temp	°C	4.6	48.3	19.8	8.6
Weather	Wind-speed	m/s	0	4.4	0.3	0.49
Weather	Rain	mm	0	0.5	0.002	0.019
Grey (hot)	tsurf-grey	°C	6.7	45.2	23	7.7
Grey (hot)	tair-grey	°C	6.9	45.2	22.5	7.6
Red (cold)	tsurf-red	°C	5.9	45.3	20.7	8.4

TABLE 5 Pre-processed sensor description for Vulcano in the year 2020.

sensor	Decomposition	Period	min	max	Mean	std
Pressure	Additive	143	−3	3	−0.04	0.6
Air-temp	Additive	24	−3	3	−0.4	1.07
Wind-speed	Additive	22	−3	3	−0.9	0.44
Rain	Additive	78	−3	3	−2.6	0.2
tsurf-grey	Additive	24	−3	3	−0.04	0.83
tair-grey	Additive	24	−3	3	0.004	0.82
tsurf-red	Additive	24	−3	3	−0.3	0.97

value range [−3, 3] allows a good spread in range that is not too broad. This ensures that the presence of outlier (extreme) values are not excessively compressed, and maintains outlier relative positions and magnitudes. The effectiveness of the pre-processing strategy is closely tied to its ability to preserve the actual correlations between sensors. The objective is to convert the data into a format suitable for analysis and input into DITAN, while still preserving the inherent relationships within the data. A summary of pre-processing output is given in Table 5.

3.6.4 Forecasting scenario

The selection of the appropriate size for observation context and forecast horizon needs to be based on the time scale of expected variations. Surface temperatures will exhibit diurnal cycles of 24 h, while also following an annual cycle. In addition, following (Sanders, 1984) major storm systems will develop over hours so that parameters such as wind speed and pressure will evolve over timescales of 6–24 h during high intensity events, such as Medicanes. Thus, following Haque et al. (2021) the context window is set to 24 h to allow a forecast horizon of 6 h.

The training phase of DITAN is thus conducted on data with 24 h context size and 6 h horizon size. The aim of learning is to reduce the differences between forecasted and actual values by identifying suitable model parameters. To ensure equal importance in minimizing all differences, we employ mean absolute error (MAE) as the loss function. By using absolute differences to compute gradients of the loss function, DITAN can mitigate the influence of extreme events such as Mediterranean hurricanes, which would otherwise dominate as the main indicator of normality. Instead, DITAN prioritizes the average understanding of underlying patterns, with patterns appearing more frequently having a greater influence on determining normality.

During the Bayes optimization process, a total of 20 different hyper-parameter configurations were examined. Each configuration was assessed using four expanding time windows applied to the pre-processed data set, resulting into an examination of 80 (=4 × 20) models. The first (initial) configuration is described in Table 6.

We observe that no improvement occurs until the 5^{th} configuration, with a significant improvement being observed during the 6^{th} configuration with relatively small variations occurring up to the 16th configuration. A gradual decrease is then observed between the 16th and 20th configurations, at which point the objective function converges (Figure 6). By changing from the initial to the optimal configuration, the optimization error is decreased from 0.268 to 0.195, resulting in a 27% improvement on the objective function.

TABLE 6 Initial parameter setup.

Parameter	Initial value	Parameter	Initial value
Layers	1 (per network)	Learning rate	0.001
Units	80	Learning scheduler	constant
Unit decay	1.0	Learning patience	10
Dropout	0.3	Batch size	64
reg. strength	0.0		

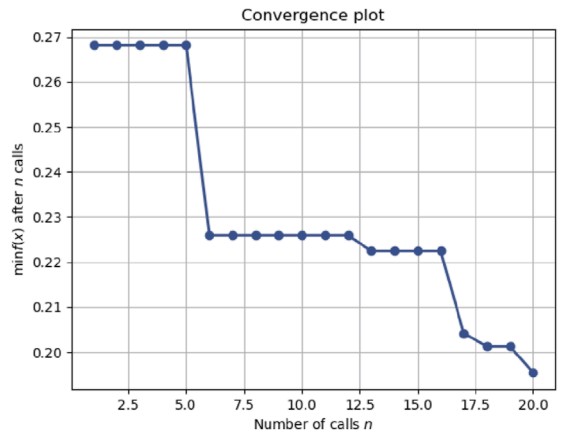


FIGURE 6
Convergence plot of DITAN across configurations using Bayes Optimizer.

The hyper-parameters of the optimal (20th) configuration are reported in Table 7. The resulting model consists of 19.815 parameters. These parameters are updated in batches of 32 consecutive patterns, where each record within these patterns is encoded using 32 units. In addition, use of a larger learning rate of 0.01 means that the convergence process becomes capable at exploring global maxima more effectively throughout all epochs. To maintain stability during training, a step decay factor is used, which gradually reduces the learning rate every 4 epochs. This approach strikes a balance between exploration and stability in the optimization process. The final parameters are selected at epoch 78, at which point the patience of early stopping is exhausted, with the internal validation loss not improving by > 0.0003 for 10 consecutive epochs after epoch 68.

3.6.5 Detecting anomalies

In the detection phase the trained model is used to predict normality across 6,843 (records) hours, resulting in a corresponding number of errors per sensor, as illustrated in Figure 7. Each error is computed as the absolute difference between the predicted and observed value, according to the selected loss function. Each sequence is then smoothed using a simple moving average (SMA) of 6-hour. An important consideration when reducing temporal resolution is to maintain a balanced trade-off between smoothness and introduced lag. The "goodness" of the proposed window size is demonstrated in Figure 8, using a subset of errors from a randomly selected (pressure) sensor. We observe that the smoothed versions of the errors maintain a responsiveness to the raw errors.

4 Results

Our objective is to examine the intensity, duration and type of anomalies caused on surface temperature by meteorological/atmospheric effects at a hydrothermal system. External drivers on the thermal state of the surface are described by air pressure, wind speed, rain, humidity and air temperature. To do this, DITAN detects anomalies across the seven sensors, and then uses the expert rules to interpret them as physical events. Within this framework, two or more anomalies are considered to be linked to each other when their start difference is within a defined time interval.

4.1 Normality

Figure 9 gives key statistics of the preprocessed sensors. We see 3 features of normality: one typical of the meteorological system, and two characteristic of the hydrothermal system:

1. The four sensors of the meteorological station recording the external parameters (pressure, air temperature, wind-speed, and rain), exhibit the expected correlation during the passage of a low-pressure system. That is, as pressure falls so too does air temperature, but wind speed and rainfall increase.
2. The surface temperatures for the hot zone exhibit a higher median than those of the cold zone. This is because the surface temperature of the hot zone is buffered by the enhanced geothermal heat flux associated with the hydrothermal system. The difference between the medians for the two surface temperatures (0.3) thus gives the median normalized magnitude of ΔT for the year 2020.
3. The median air temperature in the hot zone is higher than that at the weather station, again due to buffering by the geothermally heated surface 15 cm below the air temperature sensor, giving a median normalized magnitude for the air temperature anomaly during 2020 of 0.5.

4.2 Anomalies

From the 6,843 records (hours), 1,737 were detected as anomalous. This means around 72 days (or 25% of total records) experience anomalous events. The number of anomalies and critical regions per sensor are given in Table 8. The most critical regions occur on the temperature sensors, but the number of critical regions in the cold zone is slightly higher than in hot zone for air and surface temperatures. This again highlights the buffering effect of the geothermal heat flux in the hot zone. Wind-speed, rain and pressure also exhibit anomalies, indicating that the main source of anomalies on the temperature sensors at Vulcano during 2020, and thus also on ΔT, were due to external drivers.

TABLE 7 The hyper-parameters of the optimal DITAN model.

Hyper-parameters	Values	Hyper-parameters	Values
Layers	1	Learning Rate	0.01
Units	32	Learning Scheduler	step decay
Units Decay	0.910362945353	Learning Patience	10
Dropout	0	Batch Size	32
Regularization Strength	0		

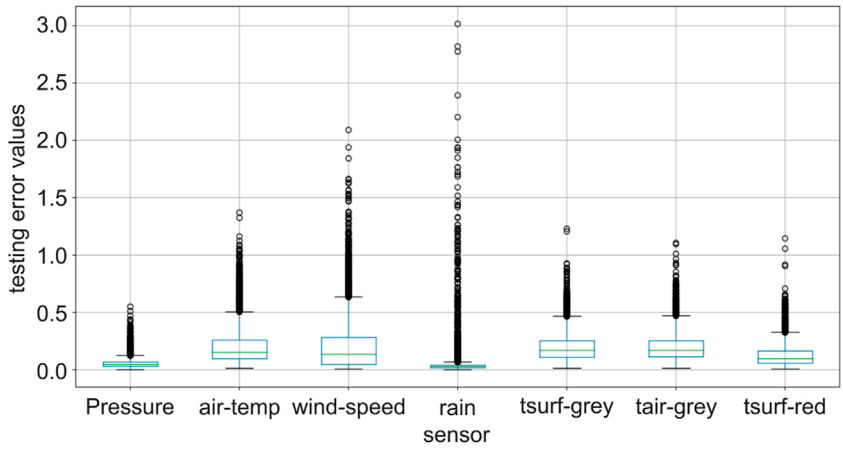


FIGURE 7
Box plots of the prediction errors of DITAN across the 7 sensors. For each sensor, the boxplot consists of a rectangular box which spans the interquartile range (IQR) of the data, the line inside representing the median. "Whiskers" extend from the edges of the box to indicate the range of the data outside the IQR, often with outliers (black points) shown beyond the whiskers. Small median errors were observed for all sensors with small IQRs. Several outliers were observed for the rain sensor, since rain is a rare event in the observed sequence.

4.3 Rule execution

The number of valid executions per rule are given in Table 9. The two meteorological events (rules R1 and R2) are executed seven and three times, respectively. Instead, R4 does not occur, R3 and R6 are executed only once, and R5 twice. This highlights the influence of external drivers on the surface temperature on both the hot and cold zones, but with a higher influence on the cold zone. Rule R7 was executed nine times. This indicates that data were corrupted on nine occasions due to the sensor giving a spurious output, such as during automatic reset, or the presence of recording glitches.

The main anomalous meteorological events occurring at Vulcano in the year 2020 were thus associated with the passage of low pressure systems and rainstorms (cloudbursts), as defined by rules R1 and R2, respectively. To obtain the temporal occurrences of these anomalous meteorological events we need to analyze the critical regions of DITAN within the pressure, rain and wind speed sensors, where the root-causes of the critical regions in these sensors are given in Figure 10. Recall that root causes are probabilistic contributions of each sensor to the anomalous character of a time-step (record), ranging from 0 to 1. During normal time-steps, all root cause values are set to zero. Consequently, the root cause diagrams provides a detailed analysis, at a time-step resolution, of the temporal difference (delay) between anomalous events triggered at various sensors. We observe that most of critical regions for wind speed and pressure are closely correlated in terms of time, contributing to the occurrence of R1. The critical regions for rain are concentrated later in the year, in close proximity to critical regions for wind speed and pressure, resulting in execution of R2 at this time.

4.4 Abnormal meteorological events

Critical regions for the air pressure sensor demonstrate a negative correlation with the wind speed sensor. This implies that when pressure exhibits an abnormal decrease, the anomalous values of the corresponding critical region for wind speed tends to increase. The positive correlation with critical regions of the rain sensor indicate that low pressure systems that were classified as abnormal were also associated with intense rain fall. We here use abnormal in the sense that the pressure, wind speed and rainfall intensity associated with the low pressure system in question was not normal compared with the trends learnt by DITAN for "regular" system behavior.

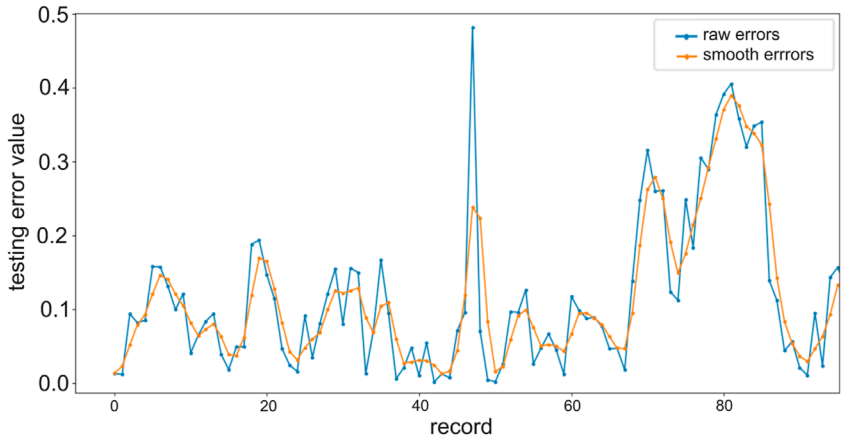


FIGURE 8
A sub-part of the raw and smoothed error sequences on *pressure* sensor.

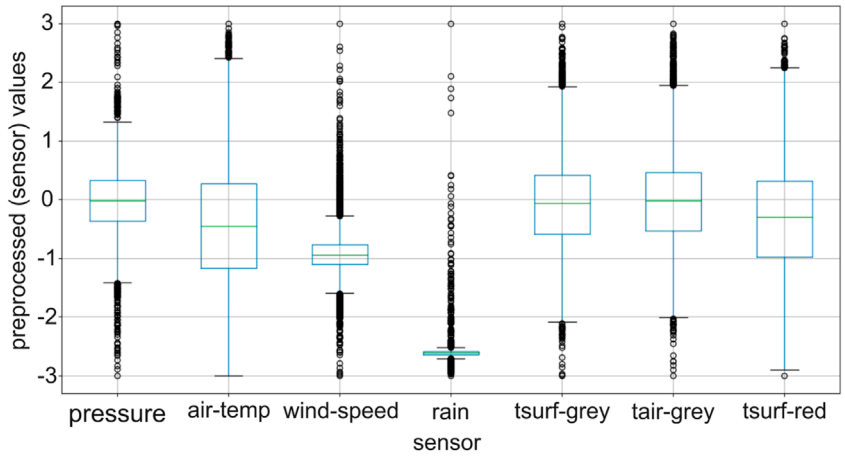


FIGURE 9
Box-plot of the pre-processed data for Vulcano in 2020. Despite the fact that decomposing to only residuals results in more values outside of the interquartile ranges, the sensors are able to preserve their correlations.

Of all the 19 critical regions identified for 2020, nine are for pressure, seven are for wind speed, and two are for rain. Abnormalities in the time series for pressure, wind speed and rainfall, related to passage of a low pressure system (Rule R1), occurred during the first 4 months of 2020, and then again in December. R1 was executed (Figure 11):

- once in early-February (event duration = 2 days);
- once in late-March (event duration = 2 days);
- once in mid-April (event duration = 1 day);
- once in early-May (event duration = 5 days);
- three times during December (combined duration = 4 days).

Summer and autumn lacked abnormal meteorological events, with all sensors maintaining relatively low (normal) levels (Figure 11). Rainfall remained at relatively low (normal) levels during the spring, summer and autumn months of 2020, but in December the rainstorm rule (R2) was executed three times. These abnormal rainstorm events lasted a total of 5 days, and coincided with the three occurrences of abnormalities due to low pressure systems.

4.5 Abnormalities in surface temperatures

The values of the surface temperatures in both the cold and hot zones are positively correlated (Figure 12). However, a partial correlation is observed between their anomalous values within critical regions, since the surface temperature of the hot zone is buffered to the external conditions by the internal driver (i.e., the geothermal heat source). Two critical regions were identified for surface temperature in both the hot zone and cold zone (Figure 13). Anomalies are confined to the winter and spring months, while summer and autumn are free of critical regions. An abnormal decrease in surface temperature in the cold zone (R5) was detected in

TABLE 8 DITAN's detection results per sensor.

Sensor name	Critical regions	Anomalous values
Pressure	11	462
Air-temp	11	99
Wind-speed	11	678
Rain	6	537
tsurf-grey	13	298
tair-grey	10	363
tsurf-red	16	562

TABLE 9 Rules executed using the inference engine.

Rule ID	Event type	Valid executions
R1	Meteorological	7
R2	Meteorological	3
R3	Surface external driver	1
R4	Surface external driver	0
R5	Surface external driver	2
R6	Surface external driver	1
R7	Instrumental	9

early-February and lasted 2 days, although this did not affect the hot zone. An abnormal decrease in surface temperature was recorded for both zones in mid-May, and lasted 6 days. During this period, the surface temperature at the hot zone began to decrease a few hours before the cold zone.

The two-day-long February abnormality in surface temperature began on 3 February, and was characterized by a dampening of the diurnal cycle, but only for the cold zone whose amplitude decreased by 4°C (Supplementary Appendix C). Abnormality began around midnight of 2–3 February, and was coincident with an abnormal decrease in pressure from 996 mbar to 976 mbar until midnight on 4–5 February. The pressure drop at 0.4 mb h^{-1} was matched by an increase in wind speed from 7 km/h to 36 km/h, with no rainfall being recorded (Supplementary Appendix C). DITAN therefore characterized this meteorological event as a low-pressure system, but not a rain storm.

The six-day-long May surface temperature abnormality began on 9 May, continued until 15 May, and affected both the cold and the hot zones. The abnormality was due to a dampening of the diurnal cycles in surface temperature by 2°C–3°C (Supplementary Appendix C). For the 6 days prior to the abnormality diurnal ranges were 10°C (32°C–22°C) in the cold zone and 16°C (36°C–20°C) in the hot zone, whereas for the 6 days of abnormality ranges were 8°C (32°C–26°C) and 13°C (31°C–18°C). The surface temperature abnormality was associated with an abnormality in pressure which fell from 990 mbar to 976 mbar over 48 h beginning at midnight 8–9 May (Supplementary Appendix C). Pressure recovered on 15–16 May, also synchronous with the end of abnormality in surface temperature. The greatest decrease in pressure was at a rate of 1 mbar per hour during 10 May, which classes the event as a "meteorological bomb." Such events are defined as "extra-tropical surface cyclones whose central pressure fall averages at least 1 mbar h^{-1} for 24 h" being a "maritime, cold-season event, with hurricane-like features" (Sanders and Gyakum, 1980). Meteorological bombs are not necessarily associated with rain, but are associated with sustained high winds which increase as the pressure decreases, making them "wind storms" (Sanders and Gyakum, 1980). Indeed, the 6 days of abnormality were characterized by no rain, but winds peaked at 30 km/h which was twice the speed recorded in the 6 days prior to passage of the bomb (Supplementary Appendix C) and, given the sheltered location of the anemometer inside the crater, high. DITAN therefore also characterized this event as an anomalous low-pressure system, but not a rain storm.

Surface temperature abnormalities are thus due to suppressed ranges of the diurnal cycle. This is triggered by decreases in pressure, which would have increased radiative cooling of the surface through its influence on vapor pressure. In addition, forced convection would have been greatly enhanced by the high winds. The combined effect is to dampen the diurnal cycle of surface temperature, making the cycles abnormal.

4.6 Prediction errors

Figure 7 presents the prediction errors across sensors, and based on these errors, the risk associated with the occurrence of abnormal meteorological events is also depicted in Figure 11. The risk of an abnormal low-pressure system varies from 0.33 to 0.43, whereas the risk of an abnormal rainstorm varies from 0.62 to the maximum risk of 1.0. That is because, especially during the early days of December, the prediction offset in rain activity was higher than the prediction offset of wind speed and pressure activities.

The risk of abnormal decrease in cold surface temperature (R5) is 0.31, while the risk of abnormal decrease of hot surface temperature (R6) is higher at 0.43 (Figure 13). This suggests that abnormal surface temperature changes in the hot zone have a higher level of risk of occurrence than in the cold zone. In addition, the risk of abnormal surface temperature decreases at both cold and hot zones (R3) is 0.27, which is lower because it considers all conditions from R5 and R6. This observation suggests that abnormal temperature changes across the entire surface carry slightly less risk of occurrence compared to changes in either the cold zone or hot zone temperatures.

5 Discussion

The critical regions for surface temperatures in the cold and hot zones, as well as wind speed and pressure, are given as a root-causes diagram in Figure 12. The detected physical events associated

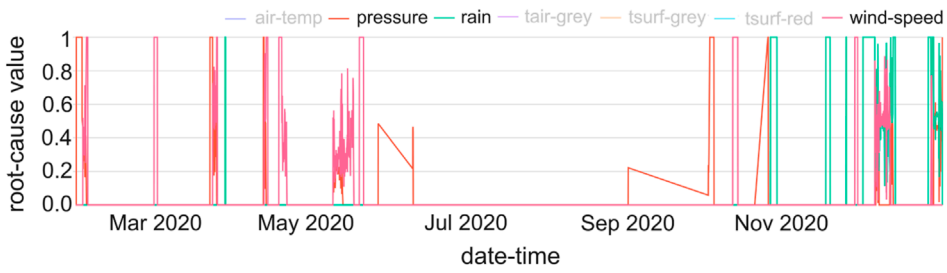


FIGURE 10
The root-causes diagram for pressure, rain and wind speed sensors.

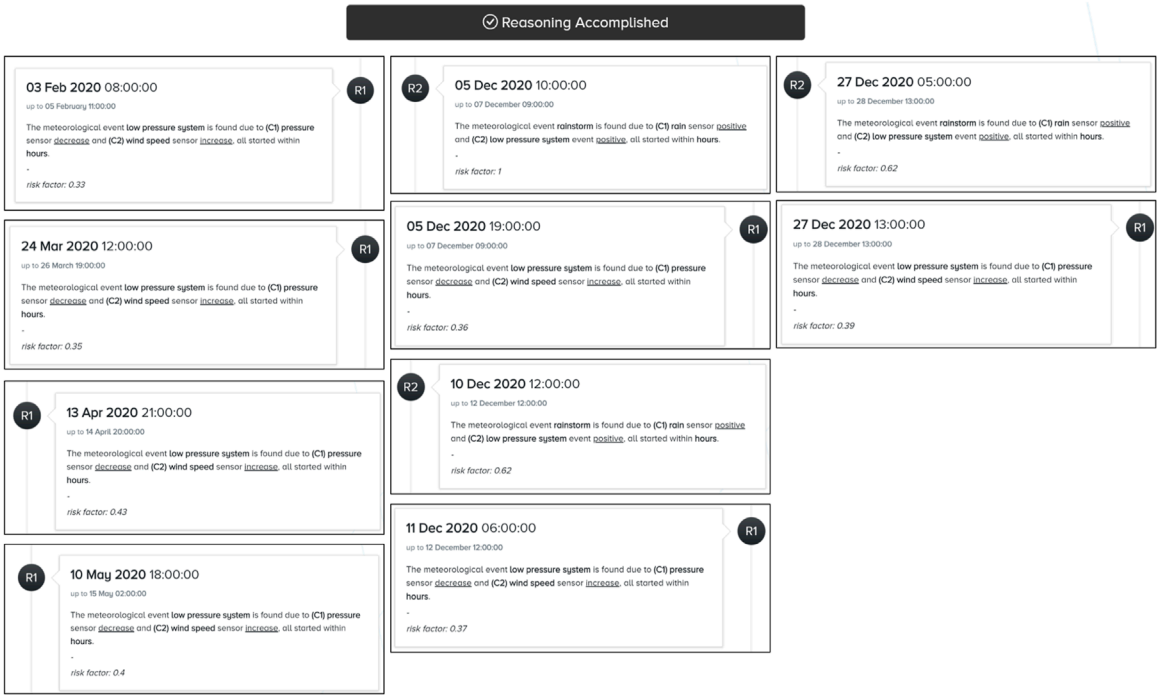


FIGURE 11
Timeline of detected meteorological events (R1, R2).

with anomalies in these critical regions have been checked as true positives (Table 10). This analysis shows that critical regions for surface temperature, wind speed and pressure are closely associated, showing partial or complete overlap. The passage of abnormally intense low pressure systems (meteorological event R1) always caused surface temperature anomalies involving abnormal decreases in temperature. Meteorological event R1 drove decreases in surface temperature for both hot and cold zones. However, the cooling experienced by the hot zone was modulated by the effect of the hydrothermal system heat source. Thus, critical regions were more associated with passage of a low pressure system for the cold zone than the hot zone.

This has implications for the effect of external drivers on the surface thermal anomaly (ΔT). In early-February 2020, a low-pressure system passed over Vulcano, persisting for 2 days. During this period, the cold zone experienced an abnormal decrease in surface temperature. However, no abnormality was recorded in the hot zone. This, thus, drove the thermal anomaly upwards, but increasing ΔT was the result of an external rather than an internal driver. However, the effect was limited to just 2 days. A second low pressure system passed over Vulcano in late-March and lasted 2 days. This was followed a one day-long period of low pressure system conditions in mid-April. In both of these cases, there were no abnormalities in the surface temperatures at either the cold or hot zones, meaning that the thermal anomaly was unaffected. However, in early-May, a low-pressure system persisted for approximately 5 days. Its influence decreased the surface temperature in the hot zone and the cold zone, disrupting normality across the entire surface. Abnormal rainstorm events were detected throughout December, and were associated with abnormalities in wind speed and pressure. However, surface temperatures retained their normality in both the cold and hot zones.

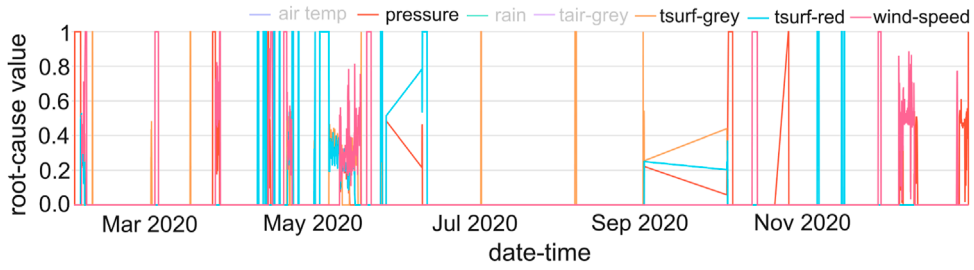


FIGURE 12
The root-causes diagram for pressure, wind speed and (red, grey) surface temperature sensors.

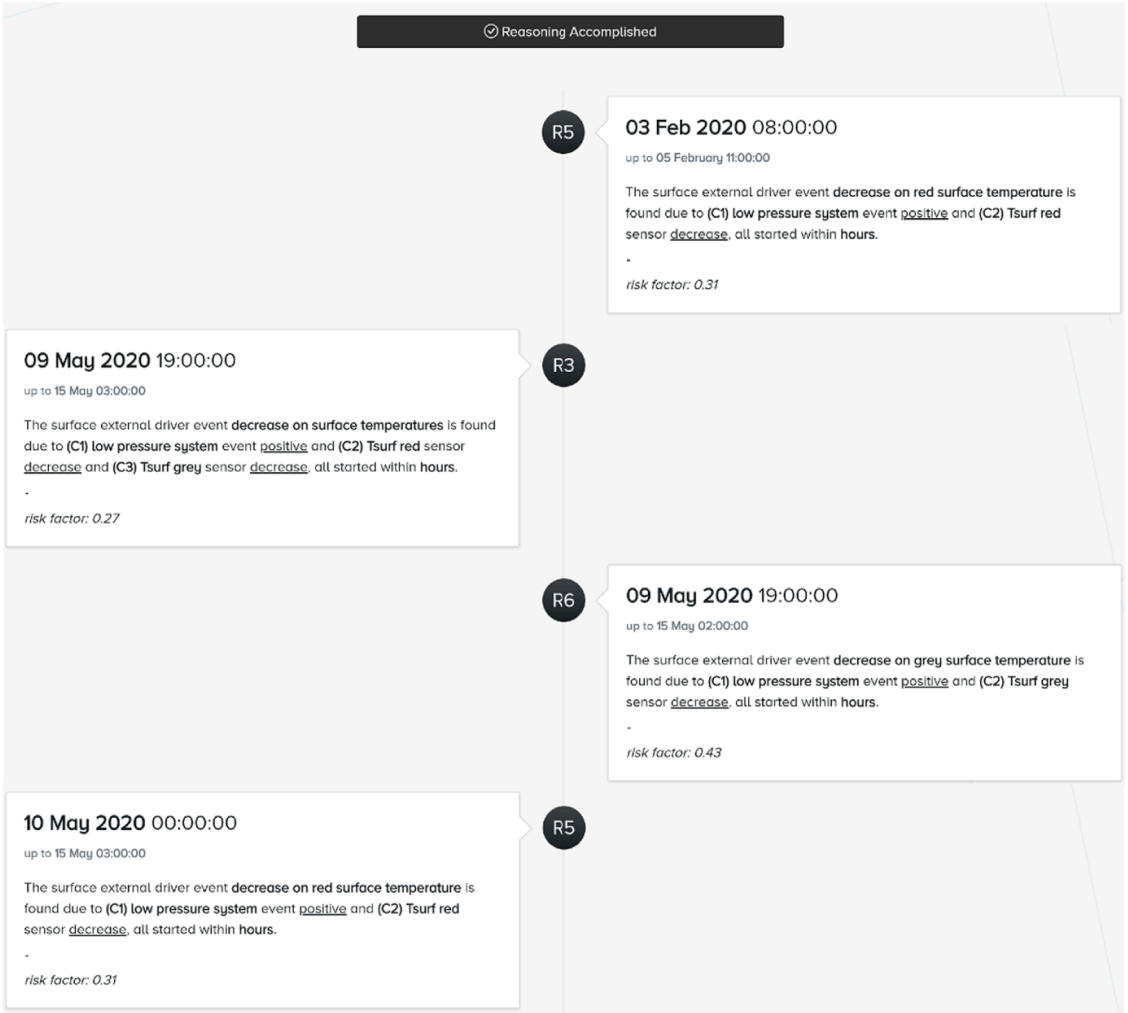


FIGURE 13
The timeline of the detected surface external driver events (R3, R5, R6).

This means that, even when at low, baseline levels, external drivers have a minimal role in causing abnormalities in surface temperature in the hot and cold zones, and hence also the surface thermal anomaly. Passage of intense low pressure systems with high winds was the main external driver for abnormal variations in surface temperature. However, even the most intense, cloudburst events did not force abnormality on the surface temperature time series, meaning that when the system is at baseline levels, surface temperature normality is immune to rain. However, this conclusion applies only to surface temperatures (Pailot-Bonnétat et al., 2023; Pailot-Bonnétat and Harris, 2024).

TABLE 10 The detected meteorological and surface external driver events.

Rule ID	Physical event	Average risk	Verified
R1	Low pressure system	0.38	✓
R2	Rainstorm	0.75	✓
R3	Decrease on surface temperatures	0.27	✓
R5	Decrease on cold zone surface temperature	0.31	✓
R6	Decrease on hot zone surface temperature	0.43	✓

5.1 Physical link between intensity of meteorological abnormality and surface cooling

At Vulcano, heat produced by condensation of ascending hydrothermal fluids has been shown to heat the surface (Chiodini et al., 2005) buffering the surface temperature so that surfaces are warmer than in an ambient scenario. Both heated and non-heated surfaces undergo diurnal and annual variation. Thus, because the heated surface is just a few degrees greater than ambient it is subject to the same external factors, such as solar heating, wind and rain, that drive surface heating and cooling. However, temperature cycles for the heated zone are at a higher level, thus creating the thermal anomaly (ΔT). For 2020 at Vulcano we found that the only external factor that could drive an abnormal behavior in the diurnal cycles was particularly low atmospheric pressures and wind speeds, rain not playing a role.

5.2 Abnormality in the thermal anomaly due to external drivers

Only particularly intense and long lasting low-pressure systems associated with meteorological bombs can dampen the diurnal cycles at both heated and non-heated surfaces. In such a case, because both T_h and T_0 behave in the same way, there is no abnormality in ΔT. However, passage of less intense systems of shorter duration only dampen the cycles in T_0, thereby causing abnormality in ΔT. The buffering provided by the internal heat source is sufficient to protect the hot surface from abnormal cooling. Thus, we consider passage of moderate intensity systems the only external driver on ΔT, where the effect of abnormal decrease in T_0, but not T_h, will increase ΔT for the period of the event.

6 Conclusion

Deep learning is beginning to be applied to understand abnormalities in diurnal cycles of surface temperature in non-volcanic environments, such as at urban heat islands (Qi et al., 2023), but not yet at low temperature thermal anomalies associated with hydrothermal systems. At bottom-heated surfaces above hydrothermal systems, due to the low amplitude of the anomaly we expect the form and amplitude of diurnal cycles in surface temperature to have both external drivers (the meteorological system) and internal drivers (the magmatic and hydrothermal systems). However, the role of the meteorological system in driving abnormality in surface temperatures at both geothermally heated and non-heated ground, in a vegetation-free crater, is poorly constrained. Most studies focus instead on the role of rainfall in influencing meteoric inflow into the hydrothermal mixing zone (Carapezza et al., 1981; Chiodini et al., 1995; Nuccio et al., 1999). We have thus applied DITAN to a year of meteorological data for the geothermally-heated zone in the Vulcano Fossa crater to understand the role of external (meteorological) factors in heating and cooling the surface.

We intentionally chose the year 2020, as heat fluxes were particularly low and stable allowing us to assess the background state of the system. Defining baselines is fundamental in volcano surveillance (McGuire et al., 1995). In the baseline state, the system appears to be remarkably robust and relatively immune to external drivers. Divergence from normality in surface temperature only occurred twice, and was associated with the passage of particularly intense low-pressure systems with durations of 2–5 days. Abnormal rainfall events did not cause divergence from normality in surface temperature cycles. Thus, during low levels of heat flux, external factors play a minimal role in driving the system away from normality in terms of diurnal and seasonal cycles of surface heating and cooling.

When applied to meteorological and surface temperature time series for surfaces above an active hydrothermal system, and configured with rules to define physical events, DITAN allows anomalies driven by external factors to be detected and classified. It could thus be applied to any geothermal system where appropriate data are available. If, as here, applied during a period when external drivers are variable, but internal drivers are stable and at background levels, all anomalies will be characteristic of a "stable" system whose surface temperature abnormalities are only driven by meteorological events. Once this is defined, such externally-driven abnormalities can be identified, cleaned and removed from periods when the internal drivers become variable and, hence, the hydrothermal system becomes unstable.

We consider this work as test and starting point for a model that cleans data set of externally driven thermal anomalies, isolating abnormalities in thermal anomalies due to internal

drivers, such as recharge of the magmatic system or changes in system permeability. During our period of study, no such events occurred. However, recharge of the magmatic system or changes in system permeability can be the prelude to eruption, or enhanced CO_2 soil degassing, so that isolating such internal changes is fundamental. That the thermal anomaly is so subtle, and that influences on it are multivariate, makes this task challenging, and hence requiring of a machine-learning based approach. Thus, defining the key internal drivers for, and their association and temporal interplay with, surface temperature abnormalities as the system moves into unrest will be our next step.

Data availability statement

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

Author contributions

MG: Conceptualization, Methodology, Software, Writing–original draft, Writing–review and editing, Investigation. SP-B: Data curation, Resources, Validation, Writing–original draft, Writing–review and editing, Investigation. VB: Conceptualization, Formal Analysis, Supervision, Writing–original draft, Writing–review and editing. AH: Data curation, Funding acquisition, Supervision, Validation, Writing–original draft, Writing–review and editing, Conceptualization.

Funding

The authors declare that financial support was received for the research, authorship, and/or publication of this article. ANR-19-CE04-0014-01 DIRE.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/feart.2024.1372621/full#supplementary-material

References

Alparone, S., Cannata, A., Gambino, S., Gresta, S., Milluzzo, V., and Montalto, P. (2010). Time-space variation of volcano-seismic events at La Fossa (Vulcano, Aeolian Islands, Italy): new insights into seismic sources in a hydrothermal system. *Bull. Of Volcanol.* 72, 803–816. doi:10.1007/s00445-010-0367-6

Aubert, M., Diliberto, S., Finizola, A., and Chébli, Y. (2008). Double origin of hydrothermal convective flux variations in the Fossa of Vulcano (Italy). *Bull. Of Volcanol.* 70, 743–751. doi:10.1007/s00445-007-0165-y

Bahrami, A., Safavinejad, A., and Amiri, H. (2019). Spectral radiative entropy generation in a non-gray planar participating medium including H2O and CO2. *J. Of Quantitative Spectrosc. And Radiat. Transf.* 227, 32–46. doi:10.1016/j.jqsrt.2019.01.024

Bonneville, A., and Gouze, P. (1992). Thermal survey of Mount Etna volcano from space. *Geophys. Res. Lett.* 19, 725–728. doi:10.1029/92gl00580

Carapezza, M., Nuccio, P., and Valenza, M. (1981). Genesis and evolution of the fumaroles of Vulcano (Aeolian Islands, Italy): a geochemical model. *Bull. Volcanol.* 44, 547–563. doi:10.1007/bf02600585

Carniel, R., and Guzmán, S. (2020). *Chapter machine learning in volcanology: a review*. London, UK: InTechOpen.

Chiodini, G., Cioni, R., Marini, L., and Panichi, C. (1995). Origin of the fumarolic fluids of Vulcano Island, Italy and implications for volcanic surveillance. *Bull. Of Volcanol.* 57, 99–110. doi:10.1007/bf00301400

Chiodini, G., Granieri, D., Avino, R., Caliro, S., Costa, A., and Werner, C. (2005). Carbon dioxide diffuse degassing and estimation of heat release from volcanic and hydrothermal systems. *J. Of Geophys. Res. Solid Earth* 110. doi:10.1029/2004jb003542

Corradino, C., Ramsey, M., Pailot-Bonnétat, S., Harris, A., and Del Negro, C. (2023). Detection of subtle thermal anomalies: deep learning applied to the ASTER global volcano dataset. *IEEE Trans. Geoscience And Remote Sens.* 61, 1–15. doi:10.1109/tgrs.2023.3241085

Diliberto, I. (2011). Long-term variations of fumarole temperatures on Vulcano Island (Italy). *Ann. Of Geophys.*

Dobson, P., Kneafsey, T., Hulen, J., and Simmons, A. (2003). Porosity, permeability, and fluid flow in the Yellowstone geothermal system, Wyoming. *J. Of Volcanol. And Geotherm. Res.* 123, 313–324. doi:10.1016/s0377-0273(03)00039-8

Ferreira, A., Curilem, M., Gomez, W., and Rios, R. (2023). Deep learning and multi-station classification of volcano-seismic events of the Nevados del Chillán volcanic complex (Chile). *Neural Comput. And Appl.* 35, 24859–24876. doi:10.1007/s00521-023-08994-z

Ferrucci, F., Gaudiosi, G., Milano, G., and Nercessian, A. (1991). Seismological exploration of Vulcano (aeolian islands): case history. *Acta Vulcanol.*

Giannoulis, M., Harris, A., and Barra, V. (2023). DITAN: a deep-learning domain agnostic framework for detection and interpretation of temporally-based multivariate ANomalies. *Pattern Recognit.* 143, 109814. doi:10.1016/j.patcog.2023.109814

Giannoulis, M., Kondylakis, H., and Marakakis, E. (2019). Designing and implementing a collaborative health knowledge system. *Expert Syst. Appl.* 126, 277–294. doi:10.1016/j.eswa.2019.02.010

Haque, E., Tabassum, S., and Hossain, E. (2021). A comparative analysis of deep neural networks for hourly temperature forecasting. *IEEE Access* 9, 160646–160660. doi:10.1109/access.2021.3131533

Harris, A. (2013). *Thermal remote sensing of active volcanoes: a user's manual*. Cambridge, England: Cambridge University Press.

Harris, A., Alparone, S., Bonforte, A., Dehn, J., Gambino, S., Lodato, L., et al. (2012). Vent temperature trends at the Vulcano Fossa fumarole field: the role of permeability. *Bull. Of Volcanol.* 74, 1293–1311. doi:10.1007/s00445-012-0593-1

Harris, A., Lodato, L., Dehn, J., and Spampinato, L. (2009). Thermal characterization of the Vulcano fumarole field. *Bull. Of Volcanol.* 71, 441–458. doi:10.1007/s00445-008-0236-8

Harris, A., and Stevenson, D. (1997). Thermal observations of degassing open conduits and fumaroles at Stromboli and Vulcano using remotely sensed data. *J. Of Volcanol. And Geotherm. Res.* 76, 175–198. doi:10.1016/s0377-0273(96)00097-2

Hochreiter, S., and Schmidhuber, J. (1997). Long short-term memory. *Neural Comput.* 8, 1735–1780. doi:10.1162/neco.1997.9.8.1735

Hochstein, M. (2005). "Heat transfer by hydrothermal systems in the east african rifts," in Proceedings Of The World Geothermal Congress 2005, Antalya, Turkey.

Hochstein, M., and Browne, P. (2000). Surface manifestations of geothermal systems with volcanic heat sources. *Encycl. Of Volcanoes*, 835–855.

Ishibashi, H., Tomabechi, R., Nishidate, K., Osaka, N., Shimomura, T., Yamada, S., et al. (2023). Evaluation of radiative absorption effect to estimate mean radiant temperature in environments with high water vapor concentration such as in a sauna. *Build. And Environ.* 243, 110684. doi:10.1016/j.buildenv.2023.110684

Keszthelyi, L., Harris, A., and Dehn, J. (2003). Observations of the effect of wind on the cooling of active lava flows. *Geophys. Res. Lett.* 30. doi:10.1029/2003gl017994

Lagios, E., Vassilopoulou, S., Sakkas, V., Dietrich, V., Damiata, B., and Ganas, A. (2007). Testing satellite and ground thermal imaging of low-temperature fumarolic fields: the dormant Nisyros Volcano (Greece). *ISPRS J. Of Photogrammetry And Remote Sens.* 62, 447–460. doi:10.1016/j.isprsjprs.2007.07.003

Leung, K., and Lam, W. (1988). Fuzzy concepts in expert systems. *Computer.* 21, 43–56. doi:10.1109/2.14346

Luong, T., Pham, H., and Manning, C. D. (2015). "Effective approaches to attention-based neural machine translation," in Proc. of the 2015 Conf. on Empirical Methods in Natural Language Processing, Lisbon, Portugal, September 17-21, 2015, 1412–2142. doi:10.18653/v1/d15-1166

Malfante, M., Dalla Mura, M., Mars, J., Métaxian, J., Macedo, O., and Inza, A. (2018). Automatic classification of volcano seismic signatures. *J. Of Geophys. Res. Solid Earth* 123, 10–645. doi:10.1029/2018jb015470

Manley, G., Mather, T., Pyle, D., Clifton, D., Rodgers, M., Thompson, G., et al. (2022). A deep active learning approach to the automatic classification of volcano-seismic events. *Front. Earth Sci.* 10, 807926. doi:10.3389/feart.2022.807926

Mannini, S., Harris, A., Jessop, D., Chevrel, M., and Ramsey, M. (2019). Combining ground-and ASTER-based thermal Measurements to Constrain fumarole field heat budgets: the case of Vulcano Fossa 2000–2019. *Geophys. Res. Lett.* 46, 11868–11877. doi:10.1029/2019gl084013

McGuire, B., Kilburn, C., and Murray, J. (1995). *Monitoring active volcanoes: strategies, procedures and techniques*. 1st ed. (England, UK: Routledge).

Nicholson, K., and Nicholson, K. (1993). Geothermal systems. *Geotherm. Fluids Chem. And Explor. Tech.*, 1–18. doi:10.1007/978-3-642-77844-5_1

Nuccio, P., Paonita, A., and Sortino, F. (1999). Geochemical modeling of mixing between magmatic and hydrothermal gases: the case of Vulcano Island, Italy. *Earth And Planet. Sci. Lett.* 167, 321–333. doi:10.1016/s0012-821x(99)00037-0

Pailot-Bonnétat, S., and Harris, A. (2024). A thermal record for unrest at Vulcano 2020–2022: *in situ* meteorological data and soil temperature recorded at high temporal resolution. *Bull. Of Volcanol.* 86, 13. doi:10.1007/s00445-023-01696-3

Pailot-Bonnétat, S., Rafflin, V., Harris, A., Diliberto, I., Ganci, G., Cappello, A., et al. (2023). *Others Anatomy of thermal unrest at a hydrothermal system: case study of the 2021-2022 crisis at Vulcano*.

Qi, W. Q., Wang, X., Meng, Y., Zhou, Y., and Wang, H. (2023). Exploring the impact of urban features on the spatial variation of land surface temperature within the diurnal cycle. *Sustain. Cities Soc.* 91, 104432. doi:10.1016/j.scs.2023.104432

Sanders, F. (1984). *The meteorological "bomb"; an explosive maritime cyclone. OCEANS 1984 318-323*.

Sanders, F., and Gyakum, J. (1980). Synoptic-dynamic climatology of the bomb. *Mon. Weather Rev.* 108, 1589–1606. doi:10.1175/1520-0493(1980)108<1589:sdcot>2.0.co;2

Sekioka, M., and Yuhara, K. (1974). Heat flux estimation in geothermal areas based on the heat balance of the ground surface. *J. Of Geophys. Res.* 79, 2053–2058. doi:10.1029/jb079i014p02053

Shortliffe, E. (1976). Books: computer-based medical consultations: MYCIN. *J. Of Clin. Eng.* 1, 69. doi:10.1097/00004669-197610000-00011

Stimac, J., Goff, F., and Goff, C. (2015). Intrusion-related geothermal systems. *Encycl. Of Volcanoes*, 799–822. doi:10.1016/b978-0-12-385938-9.00046-8

Watson, A., Schuster, U., Shutler, J., Holding, T., Ashton, I., Landschützer, P., et al. (2020). Revised estimates of ocean-atmosphere CO2 flux are consistent with ocean carbon inventory. *Nat. Commun.* 11, 4422. doi:10.1038/s41467-020-18203-3

White, D., Muffler, L., and Truesdell, A. (1971). Vapor-dominated hydrothermal systems compared with hot-water systems. *Econ. Geol.* 66, 75–97. doi:10.2113/gsecongeo.66.1.75

frontiers | Frontiers in Earth Science

TYPE Original Research
PUBLISHED 26 June 2024
DOI 10.3389/feart.2024.1342468

OPEN ACCESS

EDITED BY
Nick Varley,
University of Colima, Mexico

REVIEWED BY
Mario La Rocca,
University of Calabria, Italy
Gabriela Alejandra Badi,
National University of La Plata, Argentina
Francesca Bianco,
National Institute of Geophysics and Volcanology (INGV), Italy

*CORRESPONDENCE
Pablo Rey-Devesa,
pablord@ugr.es

RECEIVED 21 November 2023
ACCEPTED 30 May 2024
PUBLISHED 26 June 2024

CITATION
Rey-Devesa P, Carthy J, Titos M, Prudencio J, Ibáñez JM and Benítez C (2024), Universal machine learning approach to volcanic eruption forecasting using seismic features. *Front. Earth Sci.* 12:1342468. doi: 10.3389/feart.2024.1342468

Universal machine learning approach to volcanic eruption forecasting using seismic features

Pablo Rey-Devesa[1,2]*, Joe Carthy[3,4], Manuel Titos[3,4], Janire Prudencio[1,2], Jesús M. Ibáñez[1,2] and Carmen Benítez[3,4]

[1]Department of Theoretical Physics and Cosmos, Science Faculty, Avd. Fuentenueva s/n, University of Granada, Granada, Spain, [2]Andalusian Institute of Geophysics, University of Granada, Granada, Spain, [3]Department of Signal Theory, Telematics and Communication, University of Granada, Informatics and Telecommunication School, Granada, Spain, [4]Center for Information and Communication Technologies Research (CITIC), University of Granada, Granada, Spain

Introduction: Volcano seismology has successfully predicted several eruptions and includes many reliable methods that have been adopted extensively by volcanic observatories; however, there are several problems that still lack solutions. Meanwhile, the overwhelming success of data-driven models to solve predictive complex real-world problems positions them as an effective addition to the monitoring systems deployed in volcanological observatories.

Methods: By applying signal processing techniques on seismic records, we extracted four different seismic features, which usually change their trend when the system is approaching an eruptive episode. We built a temporal matrix with these parameters then defined a label for each temporal moment according to the real state of the volcanic activity (Unrest, Pre-Eruptive, Eruptive). To solve the remaining problem developing early warning systems that are transferable between volcanoes, we applied our methodology to databases associated with different volcanic systems, including data from both explosive and effusive episodes, recorded at several volcanic scenarios with open and closed conduits: Mt. Etna, Bezymianny, Volcán de Colima, Mount St. Helens and Augustine.

Results and Discussion: This work proposes the use of Neural Networks to classify the volcanic state of alert based on the behaviour of these features, providing a probability of having an eruption. This approach offers a Machine Learning tool for probabilistic short-term volcanic eruption forecasting, transferable to different volcanic systems. This innovative method classifies the state of volcanic hazard in near real-time and estimates a probability of the occurrence of an eruption, resulting in a period from at least hours to several days to forecast an eruption.

KEYWORDS

volcano seismology, forecasting, deep learning, neural networks, geophysics, signal processing

Introduction

Volcanic eruptions are powerful and complex natural phenomena that can cause significant impact to the environment, human lives and infrastructure. Being able to forecast and understand these events is crucial to enable the instigation of preventive measures and to mitigate the associated risks. For decades, the scientific community has been developing reliable volcanic eruption monitoring methods and early warning tools. Volcanoes are dynamic systems with multiple correlated geophysical variables, like seismicity, temperature anomalies, ground deformation or gas emission (Fee et al., 2020; Girona et al., 2021; Angarita et al., 2022; Grapenthin et al., 2022; Torrisi et al., 2022; Shreve et al., 2023). The analysis and interpretation of these time-series remains a difficult task, requiring an interdisciplinary approach.

Between these disciplines, some of the most used tools in warning systems come from volcano seismology. Volcanic processes generate seismic activity that can be recorded as time series and studied to understand how the volcanic system is evolving. Through the extraction of different parameters of this signal, like the energy and the frequency content, several tools have been developed capable of finding possible precursory indicators whenever an eruption is approaching. Some of these methods are based on the study of the temporal evolution of these seismic parameters (Boué et al., 2015; Boué et al., 2016; Caudron et al., 2021; Ardid et al., 2022), while others are based on the classification of different types of seismic signals, that seismologists have been associating to different eruptive processes occurring in the volcano (Chouet and Matoza, 2013; McNutt and Roman, 2015; Girona et al., 2019).

Recently, machine learning (ML) emerged as an interesting and promising tool for pattern recognition and volcanic eruption forecasting (Curilem et al., 2009; Malfante et al., 2018a; Malfante et al., 2018b; Manley et al., 2020). Due to its capability of fast processing, the problem of large databases is being solved, and ML is providing itself to be a robust real-time monitoring tool, allowing continuous surveillance of volcanic processes and a rapid identification of significant changes in volcano-seismic patterns, not evident with traditional methods (Titos et al., 2018a; Manley et al., 2021), as well as improved classification of volcano seismic events associated with different volcanic processes (Titos et al., 2018b; Titos et al., 2019; Tan et al., 2023).

However, ML solutions still have unsolved challenges and limitations (Carniel and Guzmán, 2020; Whitehead and Bebbington, 2021). Performance of ML models depends on both the quality and the size of the available dataset. If the conditions of the volcanic system change abruptly within the training data, or if new volcanic phenomena appear, the classification capability of the model might not be very precise. Another limitation of ML models is that they fail to generate evidence to enable the formulation of an underlying physical model supported by geophysical and geochemical data to explain the results they are obtaining. Thus, human supervision and experts are still needed for interpreting the results obtained at observatories.

Besides these limitations, signal processing techniques have enabled the development of volcanic early warning tools, demonstrating its capability to detect significant changes in volcanic and seismic activity (Rey-Devesa et al., 2023b; Ardid et al., 2022; Caudron et al., 2020; Dempsey et al., 2020). This allows better hazard evaluation policy and protection for the population living in hazardous regions. These techniques take data from multiparametric forecasting systems and can provide a better understanding of the underlying volcanic processes. They have great potential to detect different type of changes in the system, their performance being independent of the type of seismicity recorded (Rey-Devesa et al., 2023a; Steinke et al., 2023).

TABLE 1 Seismic Features and formula used to calculate them.

Seismic feature	Formula
Energy	$\sum_{(i=1)}^{n} S[i]^2$
Frequency Index	$\log_{10}\left(\frac{E_{highfreq}}{E_{lowfreq}}\right)$
Kurtosis	$\frac{1}{n}\sum_{(i=1)}^{n}\left(\frac{S[i]-\bar{S}}{\sigma_S}\right)^4$
Shannon Entropy	$-\sum_{(i=1)}^{n} P(S[i])\log_2(P(S[i]))$

In this work we show the forecasting potential of a machine learning volcano-independent classification tool based on the temporal evolution of 12 different seismic features, before and after volcanic eruptions. We defined a label associated to the eruptive stage as a function of the volcanic activity recorded on a set of different volcanoes (unrest period, pre-eruptive activity and eruptive volcano). Then we applied signal analysis techniques to the seismic records of these volcanoes in order to extract underlying seismic features. A moving window of 10 minutes was used with an overlap of 50%, generating a vector of features for each window, considering these as temporal units. After this, we associated the volcanic stage label with each temporal unit of the volcanoes studied, each of them with their corresponding values of the seismic features calculated. We used different volcanic systems to test the transferability and universality of the method. The studied volcanoes are associated with different explosive mechanisms, pre-eruptive activities and source processes; we worked with data from Augustine (United States), Mount St. Helens, (United States), Volcán de Colima (Mexico), Mt. Etna (Italy), and two different eruptive periods of Bezymianny (Russia), recorded in 2007 and 2017 respectively. This selection was dependent upon data availability. To make the values obtained for the different volcanoes comparable we normalized them using a z-score normalization. Then we implemented a Neural Network classification model and tested its performance for determining the volcanic stage of the system, based on the values of the seismic features. This model classifies the volcanic state of activity, generating a risk label in near real-time. In addition, it estimates the probability of being in each of the possible states. In this work we have seen that this probabilistic numeric value of having an imminent eruption increases whenever the volcanic eruption is approaching, independently of the volcanic system studied. This allows the surveillance team to consider the temporal evolution of a probabilistic numeric value as a reliable forecasting tool. This machine learning model is straightforward and allows its implementation in small and less powerful devices; the algorithm focusses on Tiny Machine Learning applications, to achieve the minimum of computation power and time required for analysing the data (Immonen and Hämäläinen, 2022).

The results obtained in this work evidence the reliability of this method as an automatic tool capable of forecasting volcanic eruptions with great potential, and it is shown to be transferable to

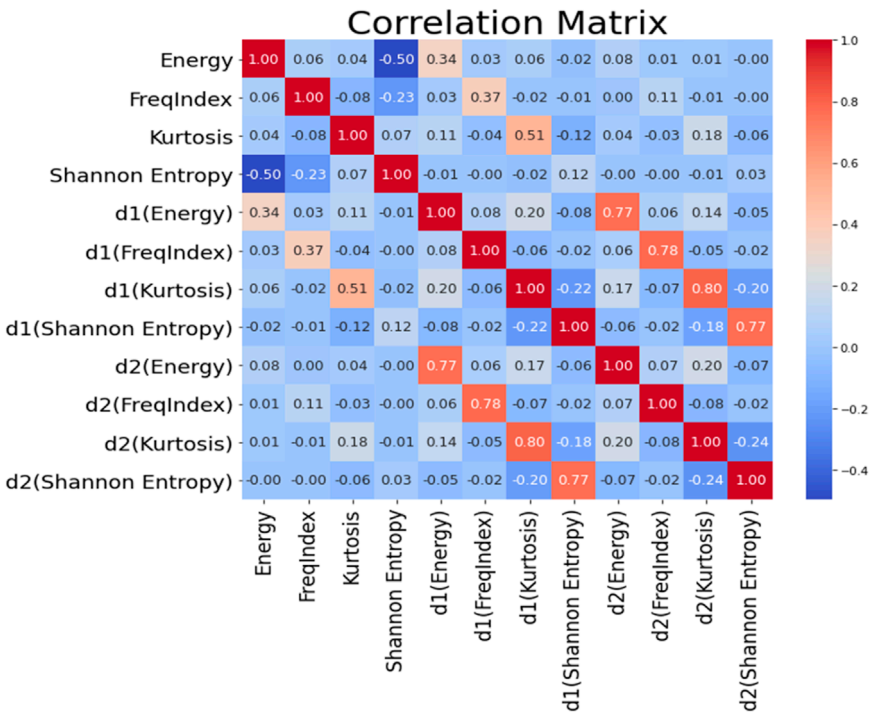


FIGURE 1
Correlation Matrix showing the relation between the temporal evolution of the 12 different features extracted from the seismic signal and used as the input for the neural network.

different volcanic systems around the world. The method could be implemented in volcanic observatories with a consideration of not only seismic datasets. We are able to define a volcanic risk alert state in near real-time based on the experience of previous case studies.

Method and materials

Characterizing seismic signals is possible by using a small set of underlying features. Based on the work of Rey-Devesa et al. (2023a), and by using four seismic features (Energy, softened Shannon Entropy, Kurtosis and Frequency Index) most properties of an evolving volcanic system can be understood (Table 1). Traditionally, the energy of the seismicity and the source process of the seismic activity, defined through the waveform and spectrum of the seismic records, were used to understand the behaviour of the volcanic system and characterize its activity (McNutt and Roman, 2015). The study of parameters like Kurtosis and Frequency Index reflects the type of activity present in the volcano; kurtosis characterizes the presence of spectral peaks in temporal series, so impulsive high frequency signals like volcanic tectonic events will induce changes in the Kurtosis values (Cortés et al., 2015); on the other hand, the frequency index indicates changes in the spectral tendency of the signal, so energetic tremor or a swarm of long period events, which are both low frequency signals, will produce a displacement of the Energy content to the lower frequency bands, reflected in the trend of the Frequency Index (Bueno et al., 2019; Rey-Devesa et al., 2023a). In addition, Shannon Entropy is a measure of the uncertainty, or the amount of information, of a dataset, which provides a quantitative value of the predictability of the system; Shannon Entropy decreases whenever the volcanic seismic signals are homogeneous; thus, the changes of a volcano self-organizing prior to an eruption are reflected in a decreasing trend of the temporal evolution of the Shannon Entropy to minimum values (Shannon, 1948; Delgado-Bonal and Marshak, 2019; Rey-Devesa et al., 2023b). With these seismic features, we built a temporal matrix that involves data from several volcanoes; the data correspond to both eruptive and non-eruptive phases, involving seismicity associated to the noise prior of the eruption, the pre-eruptive activity, and the explosions. We want to highlight that we are working with databases from different volcanoes, each one of them with its own type of explosive activity, magma characteristics, source processes, pre-eruptive seismicity and conduit opening (Acocella et al., 2024).

In Table 1 we can see the formulas used to estimate the values of our features in each window of analysis. Notice that $S[i]$ is the value of the filtered seismic signal at i. In the estimation of the Frequency Index, we have used the energy measured in the frequency band between 1 and 5.5 Hz for the low frequencies, and for the high frequency band the energy between 6 and 16 Hz, following the results of previous works (Rey-Devesa et al., 2023a). In the formula of Shannon Entropy, the P stands for probability density function.

Several authors have used these set of features among others with the idea of classifying different volcanic-seismic signals (Malfante et al., 2018a; Malfante et al., 2018b; Titos et al., 2018a; Titos et al., 2018b). In this work we use this experience to take a step forward and automatically classify the volcanic alert level, defining three labels of volcanic activity (unrest, pre-eruptive and eruptive)

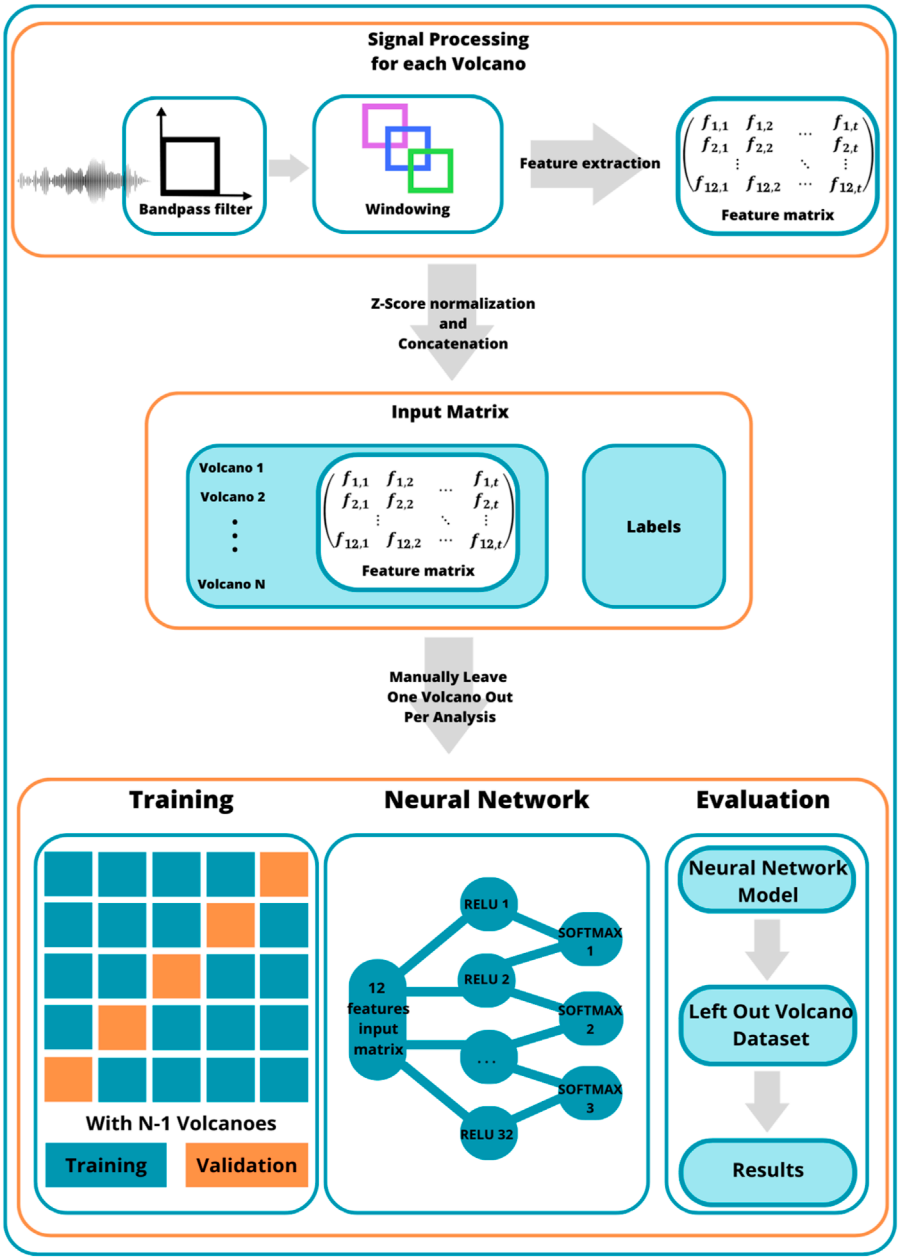


FIGURE 2
Flowchart of the methodology designed for this analysis. In first place we apply signal processing techniques over the seismic records of each volcanic dataset, creating a matrix with the temporal evolution of the seismic features. We use a z-score normalization for each matrix and create a unique matrix involving all the datasets, which now are comparable. We label this matrix in function of the volcanic state of the system (unrest, pre-eruptive and eruptive). We do the analysis independently for each dataset, so we leave one dataset out every time we run the model. Finally, we create a neural network model with a hidden dense layer of 32 units and a ReLU activation function and a second layer with a softmax function for classification. We use a k-fold cross validation model for the training; each of the 5 iterations use 80% of the data for training and 20% for validation. The model obtained is run over the excluded volcanic dataset.

and associating them to the values of these features. We consider as unrest a deviation of the typical background noise, that might or might not prelude an eruption (Phillipson et al., 2013). When the recorded seismic activity increases significatively, reflecting an imminent eruption, we define the pre-eruptive state (or impending eruption). This defines a database useful to compare with new datasets and, based upon that comparison, establish a real-time definition of the volcanic state.

Building the matrix

First, we analysed the vertical component of the seismic records associated with each volcanic system separately. We started by preprocessing the signal with a Butterworth bandpass filter to eliminate low frequencies below 1 Hz and high frequencies over 16 Hz. This is done to remove anthropogenic noise, weather condition noise and tidal noise (Almendros et al., 2000). Then we

TABLE 2 Number of data associated with the volcanoes and the eruptive stages.

	Unrest	Pre-eruptive	Eruptive	Total
Augustine	212	188	102	502
Bezymianny 07	2000	2,756	245	5,001
Bezymianny 17	260	182	19	461
Colima	2,640	1,259	882	4,781
Mt. Etna	2,222	913	705	3,840
Saint Helens	282	231	292	805
Total	**7,616 (49.5%)**	**5,529 (35.9%)**	**2,245 (14.6%)**	**15390**

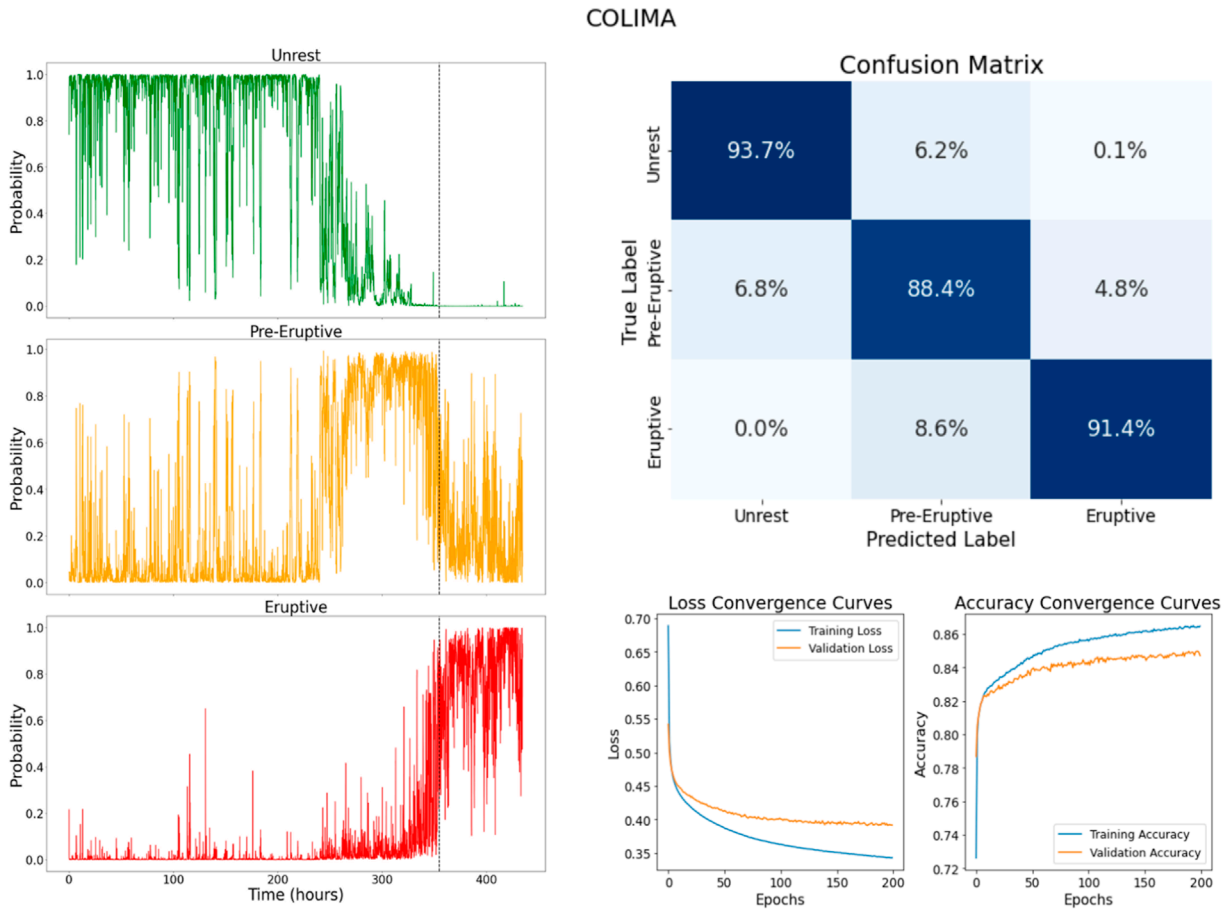


FIGURE 3
Results of the model for Volcán de Colima dataset. Left: we can see the temporal evolution of the probability of being on each state (green for unrest, yellow for pre-eruptive and red for eruptive); the vertical dashed line represents the onset of the eruption of 11 July 2015. Top right: the confusion matrix of the classification system. Bottom right: the training and validation convergence curves for loss and accuracy, showing that the model is not overfitting.

defined a moving window for the analysis, which is displaced along the filtered seismic record. The length of this window was 10 min, since we were working with very large databases, and has an overlap of 50%. We analyse the record of that 10 minutes window and estimate the corresponding value of the four features of interest, then move to the next window and repeat the process. This way, we build a temporal vector for every feature, creating a matrix associated with the temporal evolution of each volcanic system, in

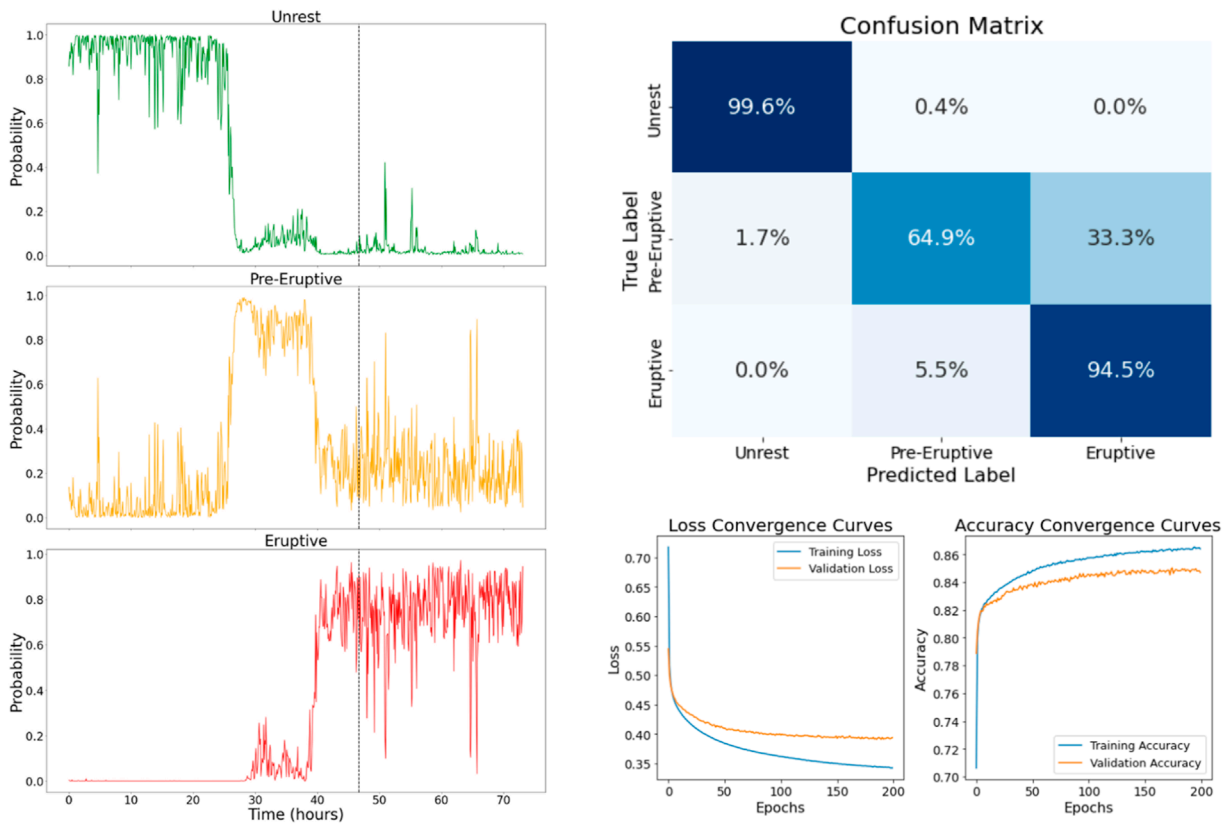


FIGURE 4
Results of the model for Mount St. Helens volcano dataset. Left: we can see the temporal evolution of the probability of being on each state (green for unrest, yellow for pre-eruptive and red for eruptive); the vertical dashed line represents the onset of the eruption of 01 October 2004. Top right: the confusion matrix of the classification system. Bottom right: the training and validation convergence curves for loss and accuracy, showing that the model is not overfitting.

which the rows represent the temporal unit and the columns the feature. We want to capture temporal fluctuations and trends in the behaviour of our seismic features; this way, the neural network can learn from the present information, and also from how the signal is changing according to the previous window. This would make our matrix more sensitive to changes that indicate an evolution of the volcanic activity. To make our matrix consistent over time, we also added the first and the second derivates of every feature, so we have 12 parameters in our matrix. The correlation matrix of the used features (Figure 1) shows that there is no redundant information from our four original input variables, which could negatively affect the performance of our model.

Based on the catalogues and the activity during the studied periods, we associated a volcanic stage label to each temporal unit (references in the Databases section). We defined three different labels: unrest, pre-eruptive and eruptive periods.

In order to make this system universal and transferable to every volcano, we built a unique large database involving the temporal matrix of each studied volcano. To ensure comparability among the matrices of different volcanoes, we normalized the temporal evolution values of each feature within its relative volcanic matrix using z-score normalization. This standardization offers several advantages in the context of machine learning analysis. It eliminates scale differences between features, prevents certain features from dominating the analysis, facilitates comparison, and reduces the risk of numerical problems, like convergence difficulties. To test the comparability of the different databases we generated the correlation matrix of the features for each studied volcano and found that all of them exhibit high similarity, both among themselves and to the reference Figure 1 (Supplementary Figure S1). After normalizing the temporal evolution of the features, we assembled the six matrices into a single matrix. We labelled this matrix with the corresponding label for each temporal unit. The full process is described in Figure 2.

Neural network

Volcano seismology advances during last decades have been driven by the improvement of deep learning and machine learning techniques (Benítez et al., 2006; Cortés et al., 2021; Titos et al., 2022), particularly neural networks (Falsaperla et al., 1996; Scarpetta et al., 2005; Titos et al., 2018a; Titos et al., 2018b). Neural Networks are a computational model inspired by the structure of the human brain. These powerful tools have demonstrated

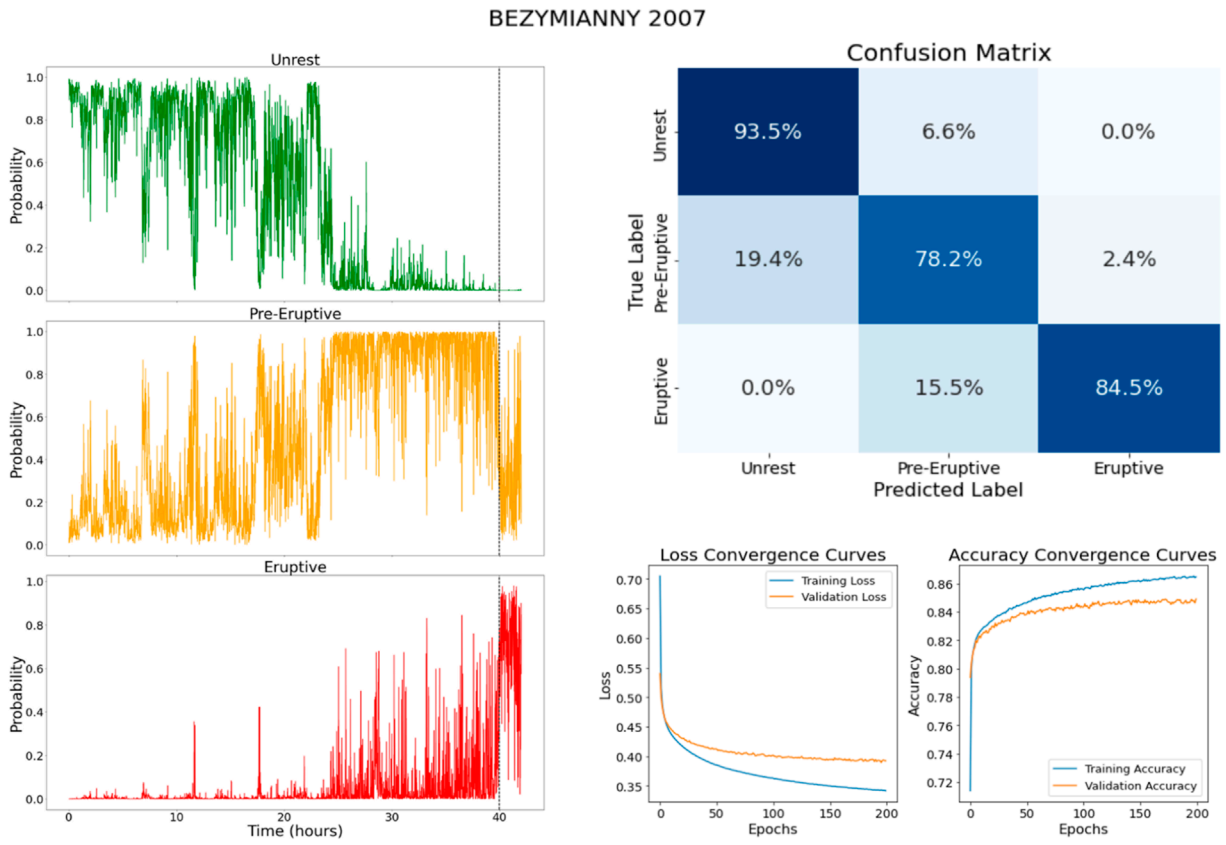


FIGURE 5
Results of the model for the Bezymianny 2007 dataset. Left: we can see the temporal evolution of the probability of being on each state (green for unrest, yellow for pre-eruptive and red for eruptive); the vertical dashed line represents the onset of the eruption of 14 October 2007. Top right: the confusion matrix of the classification system. Bottom right: the training and validation convergence curves for loss and accuracy, showing that the model is not overfitting.

exceptional efficacy in identifying underlying patterns and non-linear relationships in complex seismic datasets (Simpson, 1992; Bengio et al., 2007; Deng and Dong, 2014; LeCun et al., 2015), transforming monitoring and forecasting approaches of volcanic eruptions. The integration of higher sampling frequency data and extensive time series has led to more robust predictive models, capable of anticipating changes in volcanic activity with an unprecedented level of detail. These models can analyse raw seismic data in near real-time, identifying anomalous patterns that may indicate imminent eruptive activity. In addition, deep learning techniques have enabled the integration of data from multiple sources, including satellite imagery and volcanic gas measurements, for a holistic assessment of volcanic threat (Shoji et al., 2018; Martínez et al., 2021; Amato et al., 2023).

Our neural network approach is designed with a dense layer of 32 units, followed by a Rectified Linear Unit (ReLU) activation function, used to pass positive values unchanged and convert negative values to zero; this makes the extraction of crucial features easier (Boob et al., 2022). Then, a second dense layer with a softmax function; this function transforms a vector of real values into a vector of probabilities, enabling the classification of the three different volcanic states (Wang et al., 2018). The model compilation uses the categorical cross-entropy loss function to measure the difference between the predicted probability distribution and the actual distributions of the labels, making the training of the model more accurate; we used the Adam optimizer for updating the network weights during training, fine-tuning the model for multi-class classification challenge (Mehta et al., 2019; Mao et al., 2023). We implemented a K-Fold cross-validation model to improve the training and validation method of our analysis (Bengio and Grandvalet, 2003). We selected a specific volcano dataset of our matrix, leaving it out of the training set, and used it after the model was built for evaluating its performance. This process is repeated for each volcano. We used the remaining volcanoes for creating the models. 80% of this remaining dataset was used for training and the 20% for validation, permuting in each one of our 5 K-Fold iterations, ensuring that each part is used for both training and validation.

Databases

We have selected data from six eruptive periods associated to five different volcanoes. For each eruptive period we used one seismic station, chosen due to the completeness of the dataset and the distance to the volcanic crater; it has been observed that depending on the magnitude of the eruption, the location of the seismic

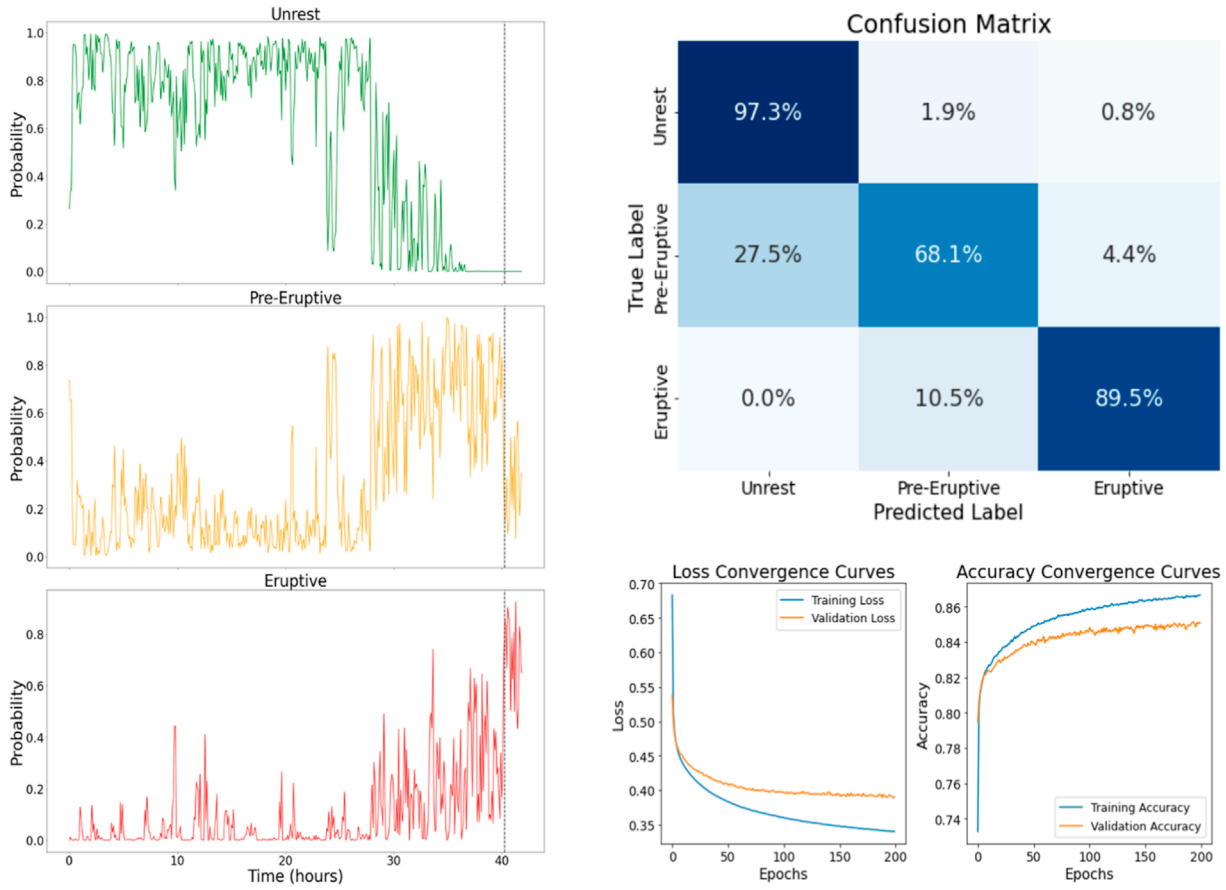


FIGURE 6
Results of the model for the Bezymianny 2017 dataset. Left: we can see the temporal evolution of the probability of being on each state (green for unrest, yellow for pre-eruptive and red for eruptive); the vertical dashed line represents the onset of the eruption of 20 December 2017. Top right: the confusion matrix of the classification system. Bottom right: the training and validation convergence curves for loss and accuracy, showing that the model is not overfitting.

station is relevant (Rey-Devesa et al., 2023a); we suggest distances below 5 km. Each dataset corresponds to the seismic signal recorded during several days or weeks prior to the eruptive episodes studied. These eruptive periods are: Augustine 2006 (AUH station, managed by Alaska Volcano Observatory); Bezymianny 2007 and 2017 (BELO and BZ01 stations respectively, by Kamchatkan Branch of Geophysical Survey); Volcán de Colima, involving a pyroclastic flow in July 2015 (SOMA station, by Centro Universitario de Estudios Vulcanológicos, University of Colima); Mt. Etna eruptive period in November 2013 (EBEM, by Instituto Nazionale di Geofisica e Vulcanologia); Mount St. Helens 2004 (SHW, by Pacific Northwest Seismic Network, University of Washington). The exact location of the seismic stations can be seen in Rey-Devesa et al., 2023a; Rey-Devesa et al., 2023b. Table 2 shows the classification of the total dataset according to the volcanic stage label and the eruptive period.

These volcanoes are representative of different types of volcanic region and eruptions, with and without relevant hydrothermal systems. The case studies are representative of open conduit, like Mt. Etna and Volcán de Colima (Chaussard et al., 2013; Zuccarello et al., 2022), semi-open conduit, like Bezymianny 2007 (West, 2013) and closed conduit systems like Augustine, Bezymianny 2017 and Mount St. Helens (De Angelis et al., 2013; Mania et al., 2019; Schlieder et al., 2022) respectively. Here we describe some differences between the type of eruptive activity, the seismicity generated by the eruptions and the composition of the lavas emitted.

The cases of study involve: a phreatic eruption at Mount St. Helens (De Siena et al., 2014; Gabrielli et al., 2020), pyroclastic flows and VT swarms at Volcán de Colima (Arámbula-Mendoza et al., 2019; Dávila et al., 2019), four episodes of lava fountaining and Strombolian activity at Mt. Etna (Bonaccorso et al., 2014), dome growth activity at Bezimianny (Thelen et al., 2010; Girina et al., 2013; Koulakov et al., 2021) and a Vulcanian eruption at Augustine (Manley et al., 2021; Zhan et al., 2022). In addition, there are variations on the characteristics of the seismicity representing pre-eruptive activity: tremor at Mt. Etna, VT swarms at Bezimianny, and mixed activity at Mount St. Helens (Spampinato et al., 2019; Rey-Devesa et al.,

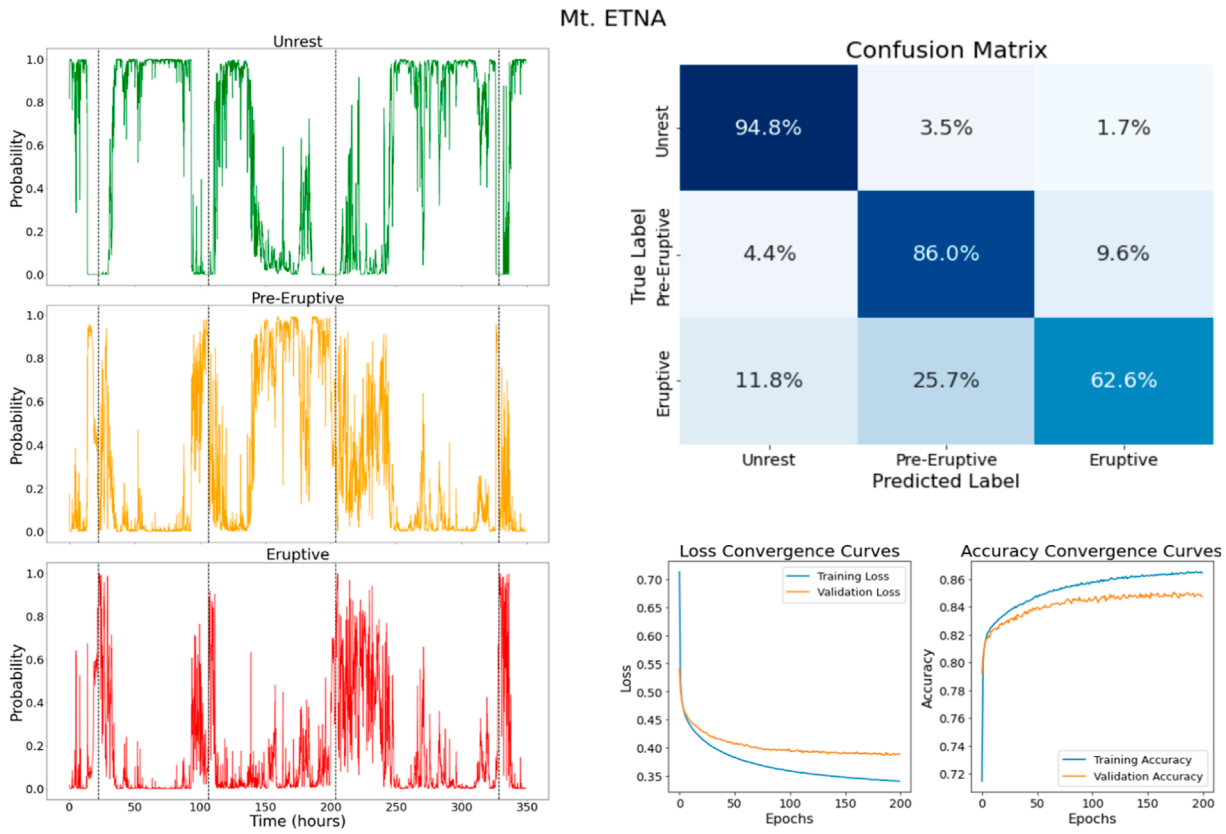


FIGURE 7
Results of the model for Mt. Etna. Left: we can see the temporal evolution of the probability of being on each state (green for unrest, yellow for pre-eruptive and red for eruptive); every vertical dashed line represents the onset of the eruptive paroxysms of November 2013. Top right: the confusion matrix of the classification system. Bottom right: the training and validation convergence curves for loss and accuracy, showing that the model is not overfitting.

2023a). Moreover, the volcanoes studied are representative of different lava types: basaltic lavas at Mt. Etna (Polacci et al., 2019; Zuccarello et al., 2022), andesitic at Bezymianny (Davydova et al., 2022) and dacitic-rhyolitic at Augustine (Wasser et al., 2021).

Results

In Figures 3–8 we show the temporal evolution of the unrest, pre-eruptive and eruptive state probabilities over time for the six different volcanic databases. Confusion matrices provide a detailed view of the model performance for each volcanic database. In this case, an excellent correspondence is observed between the model predictions and the actual states. The loss and accuracy curves offer crucial insights into the model performance during training. The curves indicate that there is no evidence of overfitting. The loss consistently decreases during training, while accuracy increases, suggesting that the model generalizes well to unseen data. These results are very promising and suggest that the model has effectively captured the characteristic patterns associated with the different eruptive phases of these case studies.

Discussion

These results show the temporal evolution of the probability of being in each volcanic state (unrest, pre-eruptive and eruptive) for each of the six databases studied. One key observation is that the unrest state is the easiest to determine for the neural network, being almost never confused with other eruptive states. The probability of being in the unrest state remains almost 100% for every case of study when the volcanic systems are in an unrest period. To complement this observation, we applied the model trained with all databases to 1 month of seismic data recorded at Bezymianny during an unrest period in September 2017. These data were not included in our training dataset. The performance was >95% of data interpreted as unrest and the remaining <5% as a pre-eruptive state, which might be due to VT swarms recorded during that period (Titos et al., 2023). We can affirm that our model hardly ever displays false alarms when it comes to determine that the system is in unrest, displaying great robustness.

The results for determining the probability of being in pre-eruptive and eruptive states are also outstanding. However, confusion between states increases slightly when it comes to determine the difference between pre-eruptive or eruptive state scenarios. One case study that stands out is Augustine volcano. This

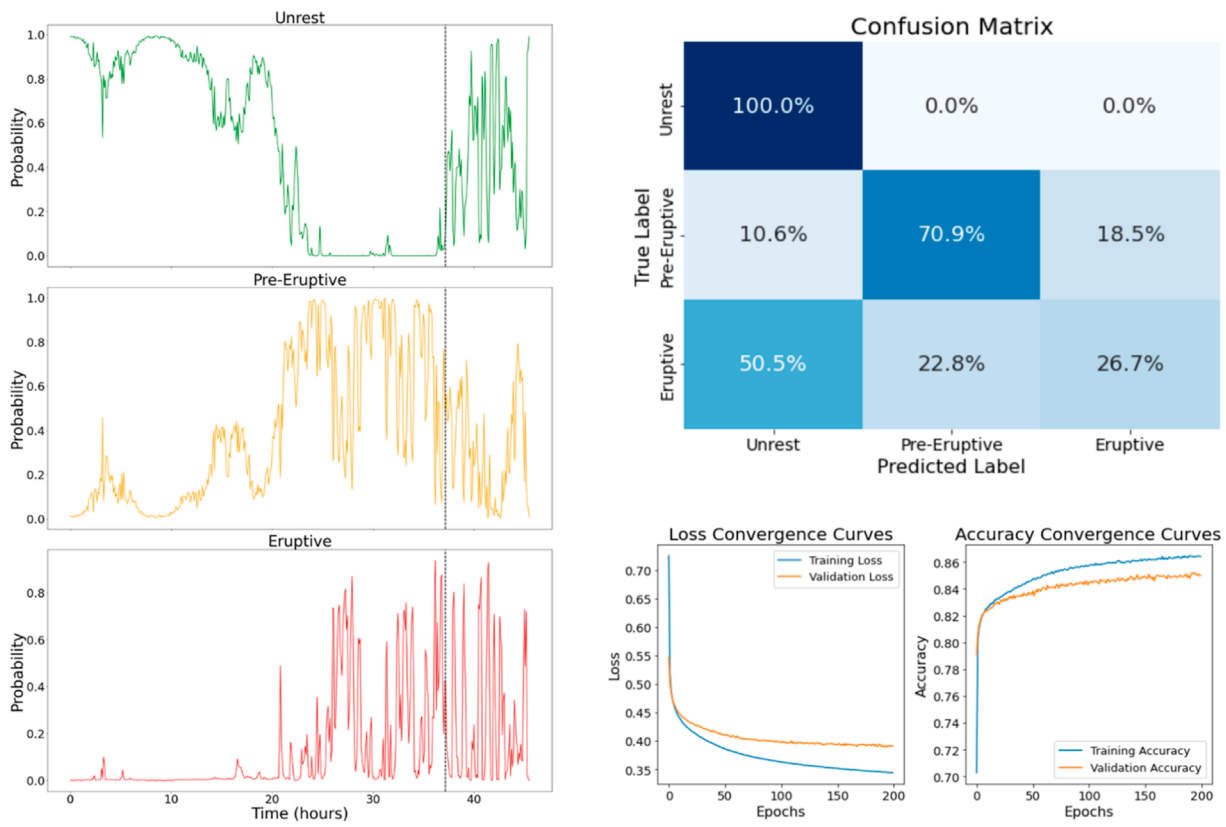


FIGURE 8
Results of the model for Augustine. Left: we can see the temporal evolution of the probability of being on each state (green for unrest, yellow for pre-eruptive and red for eruptive); the vertical dashed line represents the onset of the eruption of 11 January 2006. Top right: the confusion matrix of the classification system. Bottom right: the training and validation convergence curves for loss and accuracy, showing that the model is not overfitting.

can be explained by considering that a volcanic eruption can evolve, or even be temporarily interrupted, implying that the eruptive state is an unstable scenario. The Augustine eruption combined eruptive paroxysms with intervals of low seismicity and quiescence (Global Volcanism Program, 2006).

We want to highlight an observed characteristic pattern, which is relevant to the development of early warning systems for volcanic eruption forecasting. We observe that the pre-eruptive probability starts to increase days before the eruption, dominating the other categories. This is sufficient to establish in real-time that a certain process is ongoing in the volcanic system. Furthermore, it is particularly interesting to note that before entering the period in which the eruptive state dominates, the probability of that state shows a gradual increase. Notice that our model might confuse the beginning of the pre-eruptive state with the unrest, but not with the eruptive, which hast a probability around 0%; however, when the pre-eruptive state is about to finish, the model might confuse it with the eruptive state, and the unrest state probability decays until 0%.

Upon observing our case studies, we notice that the model determines the beginning of the pre-eruptive state between at least 10–15 h before the beginning of the eruption in the worst case, which is the lava fountains at Mt. Etna, and several days in the best case, like Volcán de Colima or Mount St. Helens; these time intervals calculated for the lava fountains at Mt. Etna complement the results of previous seismicity and infrasound forecasting studies (Ripepe et al., 2018; Rey-Devesa et al., 2023a). Moreover, the methodology determines the eruptive state of the volcanic system in near real-time. Identifying the increasing probability of being in an eruptive state is highlighting the possibility of an imminent eruption; this is a reliable numeric value for affirming the probability of a volcanic eruption to happen in the short-term and would improve the surveillance of a volcanic system and the capability of forecasting an eruption. The probability of being in the eruptive state usually goes up to 80% whenever a volcanic eruption is about to start, demonstrating that the methodology has potential as a universal surveillance tool. This method could allow the team responsible of making choices to take preventive measures, thus reducing the volcanic risk and the impact to the surrounding communities.

The relationship between the duration of the pre-eruptive stage and features like the degree of conduit opening (open, closed, semi-open), or the explosivity of the eruption, is crucial for understanding volcanic eruption dynamics and its forecasting. In closed conduit and semi-open conduit systems, a longer pre-eruptive stage results in significant magma accumulation and pressure buildup, leading to explosive eruptions, as seen with Mount St. Helens in 2004, Augustine in 2006 and Bezymianny in 2007 and 2017. Open

conduits typically have shorter pre-eruptive stages, allowing magma to ascend more freely and resulting in effusive eruptions, as observed at Etna in 2013. However, in some cases like Colima in 2014, a prolonged pre-eruptive stage can still be associated with an open conduit situation; this might be induced by the presence of a dome covering the crater of the volcano (Reyes-Dávila et al., 2016). This correlation highlights the results of Acocella et al. (2024), noticing that longer pre-eruptive durations are generally associated with closed conduits and more explosive activity, while shorter durations usually correspond to open conduits and predominantly effusive eruptions.

Conclusion

Efforts to enhance the predictive capabilities of volcanic eruptions have led to the development of sophisticated machine learning applications in volcano seismology. Despite its successes, there remain challenges that require innovative solutions. In this context, data-driven models have emerged as an effective alternative to complement monitoring systems in volcanic observatories. This study introduces a tool designed to establish eruption risk alerts and estimate associated probabilities using a neural network. Signal analysis techniques were applied to seismic records, enabling the extraction of 12 distinct seismic features. These features, typically stable during periods of inactivity, revealed significant trends in the lead-up to eruptive episodes (Rey-Devesa et al., 2023a; b). Through a temporal matrix of parameters linked to the volcanic system and labels corresponding to real states (Unrest, Pre-Eruptive, Eruptive), a robust analytical foundation was established. However, for testing the exportability of the model we needed several databases corresponding to various volcanic systems, including both explosive and effusive eruptive episodes. In this regard, a meticulous analysis of seismic data was conducted in diverse volcanic scenarios: Mount Etna, Bezymianny (2007 and 2017 paroxysms), Volcán de Colima, Mount St. Helens, and Augustine.

We showed the results of the applications of Machine Learning to classify the volcanic alert state based on the behaviour of these features and provide an estimation of the probability of future eruptions. Evaluation was carried out through a neural network approach on the selected volcanoes, using the previously constructed feature databases, demonstrating its transferability by building a model that incorporates data from various case studies. These scenarios represent a diverse set of eruptive activities (explosive and effusive) as well as different seismic precursory patterns (volcano-tectonic events, tremor, mixed activity, etc.) and different conduit systems.

We propose a preliminary tool that can be easily implemented in near real-time at volcanic observatories for complementing the decision-making systems used nowadays for establishing volcanic alerts. The methodology is simple since the features used for training the network have been chosen in a precise way due to the relevant information that they offer about the temporal evolution of the volcanic activity. The extraction of features like Kurtosis and Frequency Index from the continuous seismic signal evaluates the variation of the number of highly impulsive events, like earthquakes, or changes in the frequency band in which the more energetic seismic activity is being recorded, respectively; on the other hand, previous studies on Differential Shannon Entropy give an insight on the role of this statistical parameter as an indicator of the uncertainty of the volcanic state (Rey-Devesa et al., 2023a; b). Thus, this system searches non-linear combinations of these features that allow the classification of three possible states of the volcanic system: Unrest, indicating a deviation from the background noise; Pre-Eruptive, indicating significant changes of the trend of the seismic features calculated that preludes eruptive activity; and the Eruptive state, in which different types of eruptive activity might be ongoing. This approach could be improved including more seismic features in the input matrix.

Our study has been developed in different volcanic scenarios with widely studied eruptive processes. Our neural network approach is not complex; the architecture is provided with 2 dense layers. As the convergence curves show, the model is not overtrained. In the Databases section and Table 2, the number of events and volcanic processes required for the training and for used these experiments are described. During the experiments we have used the 80% of the data for training and the 20% for validation, repeating the procedure 5 times, permuting the data used for each process during the iterations; in every experiment, the database used for evaluation was left out before the training.

The aim of our work is determining the probability of a transition from an apparently quiescence or unrest state of the volcanic system to another state, rather than determining the internal processes of the volcano. Our approximation using statistical features might be not enough for the development of the physical model explaining the underlying processes of the volcanic system. In order to improve the method reliability and performance, the incorporation of real-time datasets of other monitored parameters, like geodesy, gas emissions or temperature measurements would have enormous potential; due to the feasibility of working with the time interval of the windows of analysis, their variety in the sampling frequency is not a problem. Moreover, we have been working with datasets from only one seismic station for every case of study, but to make the methodology more applicable for volcanic observatories it would be interesting to consider training the neural network with data recorded in the largest possible number of available seismic stations at each volcano.

These results represent a step further in the development of effective non-biased volcanic forecasting tools. In addition to this we can offer a quantification of the increasing probability of an eruption as the activity is approaching. The early identification of the pre-eruptive activity offers sufficient time for the decision making before the situation escalates, allowing the application of the appropriate measures in case of a volcanic crisis.

Despite the promising results achieved, we want to acknowledge that the model reliability is associated to the quality and availability of seismic datasets. Thus, human supervision and interpretation of the evolution of the probability is still a crucial step.

In summary, the results of this machine learning approach suggest that the methodology shows great promise as an effective and quantitative method for volcanic eruption forecasting across a variety of volcanic systems.

Data availability statement

The datasets analysed for this study can be found in the ZENODO and IRIS network: (https://doi.org/10.5281/zenodo.7732898; https://doi.org/10.5281/zenodo.6821530; https://doi.org/10.5281/zenodo.6849621; https://ds.iris.edu/mda).

Author contributions

PR-D: Conceptualization, Data curation, Formal Analysis, Methodology, Software, Visualization, Writing–original draft, Writing–review and editing. JC: Methodology, Software, Writing–review and editing. MT: Methodology, Software, Writing–review and editing. JP: Writing–review and editing. JI: Funding acquisition, Project administration, Writing–review and editing. CB: Conceptualization, Funding acquisition, Methodology, Project administration, Software, Visualization, Writing–review and editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article.

This work has been funded by the Spanish Project PID2022-143083NB-I00, "LEARNING", funded by MICIU/AEI /10.13039/501100011033 and by FEDER, UE. The authors have been partially funded by the projects PROOF-FOREVER (EUR2022.134044), ALARM (C.EXP. 022.UGR23) and IMPROVE-- Innovative Multi-disciPlinary European Research training netwOrk on VolcanoEs. PR-D was funded by the Ministerio de Ciencia e Innovación del Gobierno de España (MCIN), Agencia Estatal de Investigación (AEI), Fondo Social Europeo (FSE), and Programa Estatal de Promoción del Talento y su Empleabilidad en I+D+I Ayudas para contratos predoctorales para la formación de doctores 2020 (PRE2020-092719). This work has been partially funded by the Spanish Project: PLEC2022-009271 "DigiVolCa", funded by MCIN/AEI, funded by MCIN/AEI/10.13039/501100011033 and by EU «NextGenerationEU/PRTR», 10.13039/501100011033.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/feart.2024.1342468/full#supplementary-material

References

Acocella, V., Ripepe, M., Rivalta, E., Peltier, A., Galetto, F., and Joseph, E. (2024). Towards scientific forecasting of magmatic eruptions. *Nat. Rev. Earth Environ.* 5 (1), 5–22. doi:10.1038/s43017-023-00492-z

Almendros, J., Ibáñez, J. M., Alguacil, G., Morales, J., Del Pezzo, E., La Rocca, M., et al. (2000). A double seismic antenna experiment at Teide volcano: existence of local seismicity and lack of evidences of volcanic tremor. *J. Volcan. Geotherm. Res.* 103 (1–4), 439–462. doi:10.1016/s0377-0273(00)00236-5

Amato, E., Corradino, C., Torrisi, F., and Del Negro, C. (2023). A deep convolutional neural network for detecting volcanic thermal anomalies from satellite images. *Remote Sens.* 15 (15), 3718. doi:10.3390/rs15153718

Angarita, M., Grapenthin, R., Plank, S., Meyer, F. J., and Dietterich, H. (2022). Quantifying large-scale surface change using SAR amplitude images: crater morphology changes during the 2019–2020 shishaldin volcano eruption. *J. Geophys. Res. Solid Earth* 127 (8), e2022JB024344. doi:10.1029/2022JB024344

Arámbula-Mendoza, R., Reyes-Dávila, G., Domínguez-Reyes, T., Vargas-Bracamontes, D., González-Amezcua, M., Martínez Fierros, A., et al. (2019). "Seismic activity associated with Volcán de Colima," in *Volcán de Colima* (Berlin, Heidelberg: Springer), 195–218. doi:10.1007/978-3-642-25911-1_1

Ardid, A., Dempsey, D., Caudron, C., and Cronin, S. (2022). Seismic precursors to the Whakaari 2019 phreatic eruption are transferable to other eruptions and volcanoes. *Nat. Commun.* 13 (1), 2002–2009. doi:10.1038/s41467-022-29681-y

Bengio, Y., and Grandvalet, Y. (2003). No unbiased estimator of the variance of k-fold cross-validation. *Adv. Neural Inf. Process. Syst.* 16.

Bengio, Y., Lamblin, P., Popovici, D., and Larochelle, H. (2007). "Greedy layer-wise training of deep networks," in *Advances in neural information processing systems* (Cambridge, MA: The MIT Press), 153–160.

Benítez, M. C., Ramírez, J., Segura, J. C., Ibanez, J. M., Almendros, J., García-Yeguas, A., et al. (2006). Continuous HMM-based seismic-event classification at deception island, Antarctica. *IEEE Trans. Geoscience Remote Sens.* 45 (1), 138–146. doi:10.1109/TGRS.2006.882264

Bonaccorso, A., Calvari, S., Linde, A., and Sacks, S. (2014). Eruptive processes leading to the most explosive lava fountain at Etna volcano: the 23 November 2013 episode. *Geophys. Res. Lett.* 41 (14), 4912–4919. doi:10.1002/2014GL060623

Boob, D., Dey, S. S., and Lan, G. (2022). Complexity of training ReLU neural network. *Discrete Optim.* 44, 100620. doi:10.1016/j.disopt.2020.100620

Boué, A., Lesage, P., Cortés, G., Valette, B., and Reyes-Dávila, G. (2015). Real-time eruption forecasting using the material Failure Forecast Method with a Bayesian approach. *J. Geophys. Res. Solid Earth* 120 (4), 2143–2161. doi:10.1002/2014JB011637

Boué, A., Lesage, P., Cortés, G., Valette, B., Reyes-Dávila, G., Arámbula-Mendoza, R., et al. (2016). Performance of the 'material Failure Forecast Method' in real-time situations: a Bayesian approach applied on effusive and explosive eruptions. *J. Volcanol. Geotherm. Res.* 327, 622–633. doi:10.1016/j.jvolgeores.2016.10.002

Bueno, A., Benitez, C., De Angelis, S., Moreno, A. D., and Ibáñez, J. M. (2019). Volcano-seismic transfer learning and uncertainty quanti fication with Bayesian neural networks. *IEEE Trans. Geoscience Remote Sens.* 58 (2), 892–902. doi:10.1109/tgrs.2019.2941494TGRS.2019.2941494

Carniel, R., and Guzmán, S. (2020). Chapter machine learning in Volcanology: a review. doi:10.5772/intechopen.94217

Caudron, C., Chardot, L., Girona, T., Aoki, Y., and Fournier, N. (2020). Editorial: towards improved forecasting of volcanic eruptions. *Front. Earth Sci.* 8, 45. doi:10.3389/feart.2020.00045

Caudron, C., Girona, T., Jolly, A., Christenson, B., Savage, M. K., Carniel, R., et al. (2021). A quest for unrest in multiparameter observations at Whakaari/White Island volcano, New Zealand 2007–2018. *Earth, Planets Space* 73 (1), 195–221. doi:10.1186/s40623-021-01506-0

Chaussard, E., Amelung, F., and Aoki, Y. (2013). Characterization of open and closed volcanic systems in Indonesia and Mexico using InSAR time series. *J. Geophys. Res. Solid Earth* 118 (8), 3957–3969. doi:10.1002/jgrb.50288

Chouet, B. A., and Matoza, R. S. (2013). A multi-decadal view of seismic methods for detecting precursors of magma movement and eruption. *J. Volcanol. Geotherm. Res.* 252, 108–175. doi:10.1016/j.jvolgeores.2012.11.013

Cortés, G., Benitez, M. C., García, L., Álvarez, I., and Ibanez, J. M. (2015). A comparative study of dimensionality reduction algorithms applied to volcano-seismic signals. *IEEE J. Sel. Top. Appl. Earth Observations Remote Sens.* 9 (1), 253–263. doi:10.1109/jstars.2015.2479300

Cortés, G., Carniel, R., Lesage, P., Mendoza, M. Á., and Della Lucia, I. (2021). Practical volcano-independent recognition of seismic events: VULCAN. Ears project. *Front. Earth Sci.* 8, 616676. doi:10.3389/feart.2020.616676

Curilem, G., Vergara, J., Fuentealba, G., Acuña, G., and Chacón, M. (2009). Classification of seismic signals at Villarrica volcano (Chile) using neural networks and genetic algorithms. *J. Volcanol. Geotherm. Res.* 180 (1), 1–8. doi:10.1016/j.jvolgeores.2008.12.002

Dávila, N., Capra, L., Ferrés, D., Gavilanes-Ruiz, J. C., and Flores, P. (2019). Chronology of the 2014–2016 Eruptive Phase of Volcán de Colima and Volume Estimation of Associated Lava Flows and Pyroclastic Flows Based on Optical Multi-Sensors. *Remote Sens.* 11 (10), 1167. doi:10.3390/rs11101167

Davydova, V. O., Shcherbakov, V. D., Plechov, P. Y., and Koulakov, I. Y. (2022). Petrological evidence of rapid evolution of the magma plumbing system of Bezymianny volcano in Kamchatka before the December 20th, 2017 eruption. *J. Volcanol. Geotherm. Res.* 421, 107422. doi:10.1016/j.jvolgeores.2021.107422

De Angelis, S. H., Larsen, J., and Coombs, M. (2013). Pre-eruptive magmatic conditions at Augustine Volcano, Alaska, 2006: evidence from amphibole geochemistry and textures. *J. Petrology* 54 (9), 1939–1961. doi:10.1093/petrology/egt037

Delgado-Bonal, A., and Marshak, A. (2019). Approximate entropy and sample entropy: a comprehensive tutorial. *Entropy* 21 (6), 541. doi:10.3390/e21060541

Dempsey, D. E., Cronin, S. J., Mei, S., and Kempa-Liehr, A. W. (2020). Automatic precursor recognition and real-time forecasting of sudden explosive volcanic eruptions at Whakaari, New Zealand. *Nat. Commun.* 11 (1), 3562–3568. doi:10.1038/s41467-020-17375-2

Deng, L., and Dong, Y. (2014). Deep learning: methods and applications. *Found. Trends Signal Process* 7 (3–4), 197–387. doi:10.1561/2000000039

De Siena, L., Thomas, C., Waite, G. P., Moran, S. C., and Klemme, S. (2014). Attenuation and scattering tomography of the deep plumbing system of Mount St. Helens. *J. Geophys. Res. Solid Earth* 119 (11), 8223–8238. doi:10.1002/2014JB011372

Falsaperla, S., Graziani, S., Nunnari, G., and Spampinato, S. (1996). Automatic classification of volcanic earthquakes by using multi-layered neural networks. *Nat. Hazards* 13, 205–228. doi:10.1007/bf00215816

Fee, D., Lyons, J., Haney, M., Wech, A., Waythomas, C., Diefenbach, A. K., et al. (2020). Seismo-acoustic evidence for vent drying during shallow submarine eruptions at Bogoslof volcano, Alaska. *Bull. Volcanol.* 82, 2–14. doi:10.1007/s00445-019-1326-5

Gabrielli, S., Spagnolo, M., and De Siena, L. (2020). Geomorphology and surface geology of Mount St. Helens volcano. *J. Maps* 16 (2), 585–594. doi:10.1080/17445647.2020.1790048

Girina, O. A. (2013). Chronology of Bezymianny volcano activity, 1956–2010. *J. Volcanol. Geotherm. Res.* 263, 22–41. doi:10.1016/j.jvolgeores.2013.05.002

Girona, T., Caudron, C., and Huber, C. (2019). Origin of shallow volcanic tremor: the dynamics of gas pockets trapped beneath thin permeable media. *J. Geophys. Res. Solid Earth* 124 (5), 4831–4861. doi:10.1029/2019JB017482

Girona, T., Realmuto, V., and Lundgren, P. (2021). Large-scale thermal unrest of volcanoes for years prior to eruption. *Nat. Geosci.* 14 (4), 238–241. doi:10.1038/s41561-021-00705-4

Global Volcanism Program (2006). "Report on augustine (United States)," in *Bulletin of the global volcanism network*. Editor Wunderman, R. (Washington, DC: Smithsonian Institution), 31, 1. doi:10.5479/si.GVP.BGVN200601-313010

Grapenthin, R., Cheng, Y., Angarita, M., Tan, D., Meyer, F. J., Fee, D., et al. (2022). Return from Dormancy: rapid inflation and seismic unrest driven by transcrustal magma transfer at Mt. Edgecumbe (L'ús Shaa) Volcano, Alaska. *Geophys. Res. Lett.* 49 (20), e2022GL099464. doi:10.1029/2022GL099464

Immonen, R., and Hämäläinen, T. (2022). Tiny machine learning for resource-constrained microcontrollers. *J. Sensors* 2022, 1–11. doi:10.1155/2022/7437023

Koulakov, I., Plechov, P., Mania, R., Walter, T. R., Smirnov, S. Z., Abkadyrov, I., et al. (2021). Anatomy of the Bezymianny volcano merely before an explosive eruption on 20.12. 2017. *Sci. Rep.* 11 (1), 1758–1812. doi:10.1038/s41598-021-81498-9

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521 (7553), 436–444. doi:10.1038/nature14539

Malfante, M., Dalla Mura, M., Mars, J. I., Métaxian, J. P., Macedo, O., and Inza, A. (2018a). Automatic classification of volcano seismic signatures. *J. Geophys. Res. Solid Earth* 123 (12), 10645–10658. doi:10.1029/2018JB015470

Malfante, M., Dalla Mura, M., Métaxian, J. P., Mars, J. I., Macedo, O., and Inza, A. (2018b). Machine learning for volcano-seismic signals: chal lenges and perspectives. *IEEE Signal Process. Mag.* 35 (2), 20–30. doi:10.1109/MSP.2017.2779166

Mania, R., Walter, T. R., Belousova, M., Belousov, A., and Senyukov, S. L. (2019). Deformations and morphology changes associated with the 2016–2017 eruption sequence at Bezymianny Volcano, Kamchatka. *Remote Sens.* 11 (11), 1278. doi:10.3390/rs11111278

Manley, G. F., Mather, T. A., Pyle, D. M., Clifton, D. A., Rodgers, M., Thompson, G., et al. (2021). Machine learning approaches to identifying changes in eruptive state using multi-parameter datasets from the 2006 eruption of Augustine Volcano, Alaska. *J. Geophys. Res. Solid Earth* 126 (12), e2021JB022323. doi:10.1029/2021JB022323

Manley, G. F., Pyle, D. M., Mather, T. A., Rodgers, M., Clifton, D. A., Stokell, B. G., et al. (2020). Understanding the timing of eruption end using a machine learning approach to classification of seismic time series. *J. Volcanol. Geotherm. Res.* 401, 106917. doi:10.1016/j.jvolgeores.2020.106917

Mao, A., Mohri, M., and Zhong, Y. (2023). "Cross-entropy loss functions: theoretical analysis and applications," in Proceedings of the 40 th International Conference on Machine Learning, Honolulu, Hawaii (PMLR), 23803–23828.

Martínez, V. L., Titos, M., Benítez, C., Badi, G., Casas, J. A., Craig, V. H. O., et al. (2021). Advanced signal recognition methods applied to seismo-volcanic events from planchon peteroa volcanic complex: deep neural network classifier. *J. S. Am. Earth Sci.* 107 (1–12), 103115. doi:10.1016/j.jsames.2020.103115

McNutt, S. R., and Roman, D. C. (2015). "Volcanic seismicity," in *The encyclopedia of volcanoes*. 2nd edn, Editor H. Sigurdsson (Amsterdam, Netherlands: Elsevier Inc.), 1011–1034. doi:10.1016/b978-0-12-385938-9.00059-6

Mehta, S., Paunwala, C., and Vaidya, B. (2019). "CNN based traffic sign classification using Adam optimizer," in *2019 international conference on intelligent computing and control systems (ICCS)* (Piscataway, NJ: IEEE), 1293–1298.

Phillipson, G., Sobradelo, R., and Gottsmann, J. (2013). Global volcanic unrest in the 21st century: an analysis of the first decade. *J. Volcanol. Geotherm. Res.* 264, 183–196. doi:10.1016/j.jvolgeores.2013.08.004

Polacci, M., Andronico, D., de'Michieli Vitturi, M., Taddeucci, J., and Cristaldi, A. (2019). Mechanisms of ash generation at basaltic volcanoes: the case of Mount Etna, Italy. *Front. Earth Sci.* 7, 193. doi:10.3389/feart.2019.00193

Rey-Devesa, P., Benítez, C., Prudencio, J., Gutiérrez, L., Cortés-Moreno, G., Titos, M., et al. (2023a). Volcanic early warning using Shannon entropy: multiple cases of study. *J. Geophys. Res. Solid Earth* 128, e2023JB026684. doi:10.1029/2023JB026684

Rey-Devesa, P., Prudencio, J., Benítez, C., Bretón, M., Plasencia, I., León, Z., et al. (2023b). Tracking volcanic explosions using Shannon entropy at Volcán de Colima. *Sci. Rep.* 13 (1), 9807. doi:10.1038/s41598-023-36964-x

Reyes-Dávila, G. A., Arámbula-Mendoza, R., Espinasa-Pereña, R., Pankhurst, M. J., Navarro-Ochoa, C., Savov, I., et al. (2016). Volcán de Colima dome collapse of July, 2015 and associated pyroclastic density currents. *J. Volcanol. Geotherm. Res.* 320, 100–106. doi:10.1016/j.jvolgeores.2016.04.015

Ripepe, M., Marchetti, E., Delle Donne, D., Genco, R., Innocenti, L., Lacanna, G., et al. (2018). Infrasonic early warning system for explosive eruptions. *J. Geophys. Res. Solid Earth* 123 (11), 9570–9585. doi:10.1029/2018jb015561

Scarpetta, S., Giudicepietro, F., Ezin, E. C., Petrosino, S., Del Pezzo, E., Martini, M., et al. (2005). Automatic classification of seismic signals at Mt. Vesuvius volcano, Italy, using neural networks. *Bull. Seismol. Soc. Am.* 95 (1), 185–196. doi:10.1785/0120030075

Schlieder, T. D., Cooper, K. M., Kent, A. J., Bradshaw, R., and Huber, C. (2022). Thermal storage conditions and origin of compositional diversity of plagioclase crystals in magmas from the 1980 and 2004–2005 eruptions of Mount Saint Helens. *J. Petrology* 63 (8), egac064. doi:10.1093/petrology/egac064

Shannon, C. E. (1948). A mathematical theory of communication. *Bell Syst. Tech. J.* 27 (3), 623–656. doi:10.1002/j.1538-7305.1948.tb00917.x

Shoji, D., Noguchi, R., Otsuki, S., and Hino, H. (2018). Classification of volcanic ash particles using a convolutional neural network and probability. *Sci. Rep.* 8 (1), 8111. doi:10.1038/s41598-018-26200-2

Shreve, T., Zhan, Y., Le Mével, H., Roman, D., and Moussallam, Y. (2023). Two distinct magma storage regions at Ambrym volcano detected by satellite geodesy. *Geophys. Res. Lett.* 50 (15), e2023GL102925. doi:10.1029/2023GL102925

Simpson, P. K. (1992). "Fundations of neural Network," in *Artificial neural networks, paradigms, applications and hardware implementations*. Editor E. Sanchez-Sinencio, and C. Lau (Piscataway, NJ: IEEE Press). 3–24. ISBN 0-87942-289-0.

Spampinato, S., Langer, H., Messina, A., and Falsaperla, S. (2019). Short-term detection of volcanic unrest at Mt. Etna by means of a multi-station warning system. *Sci. Rep.* 9 (1), 6506–6510. doi:10.1038/s41598-019-42930-3

Steinke, B., Jolly, A. D., Carniel, R., Dempsey, D. E., and Cronin, S. J. (2023). Identification of seismo-volcanic regimes at whakaari/white island (New Zealand) via

systematic tuning of an unsupervised classifier. *J. Geophys. Res. Solid Earth* 128 (3), e2022JB026221. doi:10.1029/2022JB026221

Tan, D., Fee, D., Hotovec-Ellis, A. J., Pesicek, J. D., Haney, M. M., Power, J. A., et al. (2023). Volcanic earthquake catalog enhancement using integrated detection, matched-filtering, and relocation tools. *Front. Earth Sci.* 11, 1158442. doi:10.3389/feart.2023.1158442

Thelen, W., West, M., and Senyukov, S. (2010). Seismic characterization of the fall 2007 eruptive sequence at Bezymianny Volcano, Russia. *J. Volcanol. Geotherm. Res.* 194 (4), 201–213. doi:10.1016/j.jvolgeores.2010.05.010

Titos, M., Bueno, A., Garcia, L., and Benitez, C. (2018a). A deep neural networks approach to automatic recognition systems for volcano-seismic events. *IEEE J. Sel. Top. Appl. Earth Observations Remote Sens.* 11 (5), 1533–1544. doi:10.1109/jstars.2018.2803198JSTARS.2018.2803198

Titos, M., Bueno, A., García, L., Benítez, C., and Segura, J. C. (2019). Classification of isolated volcano-seismic events based on inductive transfer learning. *IEEE Geoscience Remote Sens. Lett.* 17 (5), 869–873. doi:10.1109/LGRS.2019.2931063

Titos, M., Bueno, A., García, L., Benítez, M. C., and Ibañez, J. (2018b). Detection and classification of continuous volcano-seismic signals with recur rent neural networks. *IEEE Trans. Geoscience Remote Sens.* 57 (4), 1936–1948. doi:10.1109/TGRS.2018.2870202

Titos, M., García, L., Kowsari, M., and Benítez, C. (2022). Toward knowledge extraction in classification of volcano-seismic events: visualizing hidden states in recurrent neural networks. *IEEE J. Sel. Top. Appl. Earth Observations Remote Sens.* 15, 2311–2325. doi:10.1109/JSTARS.2022.3155967

Titos, M., Gutiérrez, L., Benítez, C., Rey Devesa, P., Koulakov, I., and Ibáñez, J. M. (2023). Multi-station volcano tectonic earthquake monitoring based on transfer learning. *Front. Earth Sci.* 11, 1204832. doi:10.3389/feart.2023.1204832

Torrisi, F., Amato, E., Corradino, C., Mangiagli, S., and Del Negro, C. (2022). Characterization of volcanic cloud components using machine learning techniques and SEVIRI infrared images. *Sensors* 22 (20), 7712. doi:10.3390/s22207712

Wang, M., Lu, S., Zhu, D., Lin, J., and Wang, Z. (2018). "A high-speed and low-complexity architecture for softmax function in deep learning," in 2018 IEEE asia pacific conference on circuits and systems (APCCAS) (Piscataway, NJ: IEEE), 223–226.

Wasser, V. K., Lopez, T. M., Anderson, K. R., Izbekov, P. E., and Freymueller, J. T. (2021). Multidisciplinary constraints on magma compressibility, the pre-eruptive exsolved volatile fraction, and the H2O/CO2 Molar ratio for the 2006 Augustine eruption, Alaska. *Geochem. Geophys. Geosystems* 22 (9), e2021GC009911. doi:10.1029/2021GC009911

West, M. E. (2013). Recent eruptions at Bezymianny volcano—a seismological comparison. *J. Volcanol. Geotherm. Res.* 263, 42–57. doi:10.1016/j.jvolgeores.2012.12.015

Whitehead, M. G., and Bebbington, M. S. (2021). Method selection in short-term eruption forecasting. *J. Volcanol. Geotherm. Res.* 419, 107386. doi:10.1016/j.jvolgeores.2021.107386

Zhan, Y., Le Mével, H., Roman, D. C., Girona, T., and Gregg, P. M. (2022). Modeling deformation, seismicity, and thermal anomalies driven by degassing during the 2005-2006 pre-eruptive unrest of Augustine Volcano, Alaska. *Earth Planet. Sci. Lett.* 585, 117524. doi:10.1016/j.epsl.2022.117524

Zuccarello, L., Burton, M. R., Saccorotti, G., Bean, C. J., and Patanè, D. (2013). The coupling between very long period seismic events, volcanic tremor, and degassing rates at Mount Etna volcano. *J. Geophys. Res. Solid Earth* 118 (9), 4910–4921. doi:10.1002/jgrb.50363

Zuccarello, L., De Angelis, S., Minio, V., Saccorotti, G., Bean, C. J., Paratore, M., et al. (2022). Volcanic tremor tracks changes in multi-vent activity at Mt. Etna, Italy: evidence from analyses of seismic array data. *Geophys. Res. Lett.* 49 (22), e2022GL100056. doi:10.1029/2022gl100056

frontiers | Frontiers in Earth Science

TYPE Original Research
PUBLISHED 16 October 2024
DOI 10.3389/feart.2024.1440967

OPEN ACCESS

EDITED BY
Pierdomenico Romano,
National Institute of Geophysics and Volcanology (INGV), Italy

REVIEWED BY
Manuel Titos,
Icelandic Meteorological Office, Iceland
Alexander Yates,
Université Libre de Bruxelles, Belgium

*CORRESPONDENCE
Andrea Di Benedetto,
andrea.dibenedetto@ingv.it

RECEIVED 30 May 2024
ACCEPTED 23 September 2024
PUBLISHED 16 October 2024

CITATION
Di Benedetto A, Figlioli A, D'Alessandro A and Lo Bosco G (2024) Grid-search method for short-term over long-term average parameter tuning: an application to Stromboli explosion quakes.
Front. Earth Sci. 12:1440967.
doi: 10.3389/feart.2024.1440967

Grid-search method for short-term over long-term average parameter tuning: an application to Stromboli explosion quakes

Andrea Di Benedetto[1,2]*, Anna Figlioli[3,4], Antonino D'Alessandro[1] and Giosue' Lo Bosco[2,5]

[1]Osservatorio Nazionale Terremoti, Istituto Nazionale di Geofisica e Vulcanologia, Roma, Italy, [2]Dipartimento di Matematica e Informatica, Università degli Studi di Palermo, Palermo, Italy, [3]Sezione di Milano, Istituto Nazionale di Geofisica e Vulcanologia, Milano, Italy, [4]Dipartimento di Scienze della Terra e del Mare, Università degli Studi di Palermo, Palermo, Italy, [5]Dipartimento di Scienze per l'Innovazione Tecnologica, Istituto Euro-Mediterraneo di Scienza e Tecnologia, Palermo, Italy

The collection of a significant catalogue of seismo-volcanic data involves the selection of relevant parts of raw signals, which can be automatised by using the short-term over long-term average (STA/LTA) method. The STA/LTA method employs the "Characteristic Function" to describe a section of a seismic record in terms of trace amplitude and first-time difference. This function is calculated in a short-term and long-term window; the ratio between the two windows defines a quantity that is controlled through threshold values, i.e., trigger on and trigger off. These threshold values indicate whether there is an increase in the energy in the seismic signal compared to the background noise. The common approach to the selection of the STA/LTA values is the adoption of literature-suggested ones. This could be a limitation as there may be cases in which a choice adapted to a specific raw signal may significantly help in the extraction of the relevant parts. To overcome the possible drawbacks of a non-adaptive choice imposed by such standard literature values, in this study, we propose a methodology for the automatic selection of STA/LTA values that can optimise the extraction of explosion quakes (EQs) from a seismo-volcanic raw signal. The values are obtained through a grid search over an index named quality–numerosity index (QNI) that measures the accordance in the automatic cuts and the consequent number of triggered seismo-volcanic events with the ones suggested by a human expert. The method was applied in the volcano domain for the specific application of the explosion quake signal extraction at Stromboli volcano. The experiments were conducted by selecting a subset of the dataset as training where to search for the best values, which were subsequently adopted in a test set. The results prove that the values suggested by our approach significantly improve the quality of the relevant part compared to the one extracted by adopting the values indicated in the literature. The methodology presented in this study can be applied to a wider typology of signals of volcanic, seismic, and other

10.3389/feart.2024.1440967

origin, potentially becoming a widely used approach in parameter optimisation processes.

KEYWORDS

short-term over long-term average method, machine learning, parametrization tuning, grid search, seismo-volcanic signals, explosion quakes, Stromboli volcano

1 Introduction

The issue of detecting seismo-volcanic events and their waveform extraction from raw seismic signals is a key problem of volcanic seismology (Sosa et al., 2024; Journeau et al., 2020; Soubestre et al., 2018). With the continuous growth of available data over time, due to the expansion of seismic networks, meeting this request with the help of a human operator can be laborious and time-consuming. Thus, computational methods need to be adopted, such as machine learning methods, for near real-time event detection and waveform extraction, especially for rapid risk assessment related to potential destructive events (Makus et al., 2024; Konstantinou, 2023; Lara et al., 2020). Machine learning methodologies also emerge in managing post-event intervention (Cannioto et al., 2017). The scientific community is focused on identifying and analysing the seismic signals generated by volcanic activity, for characterising potential precursors that may serve as early indicators of eruptions. This is especially crucial for the Stromboli volcano, where paroxysms—sudden and intense eruptions—pose the greatest danger to populations living in the surrounding areas (Andronico et al., 2021; Metrich et al., 2021; Giudicepietro et al., 2020). The Stromboli volcano (926 m) is part of the Aeolian archipelago in the Tyrrhenian Sea (Italy) and is renowned for its persistent explosive activity, often referred to as "Strombolian eruptions" (Giudicepietro et al., 2020). This volcanic behaviour is characterised by rhythmic bursts of gas and pyroclasts, driven by the degassing of magma (Chouet, 1996). The study of Stromboli's volcanic signals, including seismic activity, ground deformation, and gas emissions, provides crucial insights into the underlying magmatic processes and potential eruption forecasts. Seismic activity signals, such as volcanic tremors and explosions, are particularly significant as they reflect the movement of magma and gas within the volcanic conduit. Ground deformation, monitored through techniques such as GPS and InSAR, offers valuable data on the magma's movement beneath the surface (Schaefer et al., 2019). Gas emissions, especially the flux of sulphur dioxide (SO_2), serve as key indicators of volcanic activity and magma ascent (Aiuppa et al., 2010). Together, these indicators constitute a comprehensive framework for understanding the Stromboli volcano's dynamic nature and assessing the associated risks. Stromboli is an open-conduit volcano, with three summit craters (Figure 1), with persistent Strombolian activity. The explosions occur every 15–20 min. Generally, its volcanic activity is classified as follows: normal activity (specifically "explosion"), major explosion, and paroxysm (Chouet, 1996; Wassermann, 2012; Ripepe et al., 2021b). To distinguish them, the variation of frequency and energy of activity must be calculated (Calvari et al., 2021). The permanent seismic network of Istituto Nazionale di Geofisica e Vulcanologia (INGV) records seismic signals of volcanic nature, which are as follows: very long period (VLP), landslides, tornillos,

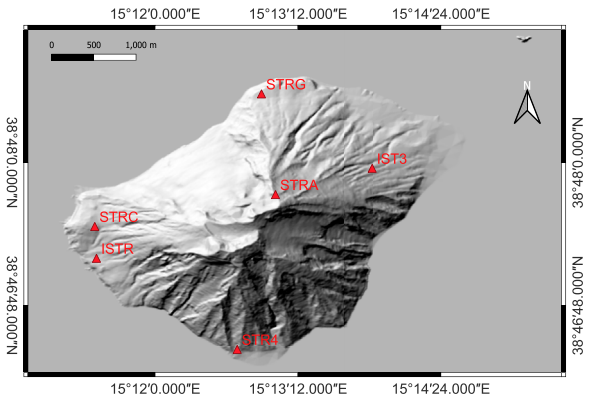


FIGURE 1
Permanent seismic network of Istituto Nazionale di Geofisica e Vulcanologia (INGV) on Stromboli. Map created using QGIS Development Team (2024).

explosion quakes (EQs), and many others (Wassermann, 2012). EQs along a seismic signal are generally clearly visible to the human eye. Their features are based on the variation in amplitude and frequency content, whose range is approximately 10–25 Hz. Another characteristic is that the EQs are preceded by VLPs, which are important for identifying the earthquakes themselves (Legrand and Perton, 2021; Giudicepietro et al., 2019). This feature allows for the description of the phenomena that occur in the plumbing system. Specifically, the mechanism that activates the VLP and subsequently EQs is a progressive degassing magma on the conduit. When the intensity increases, the conduit goes into resonance with the wall and produces seismic waves. After the magma starts migrating from the vent to the crater, it begins to produce resonant events such as VLPs (Konstantinou, 2023; Ripepe et al., 2021a; Liang et al., 2020; Ripepe et al., 2017; Ripepe and Harris, 2008; Chouet et al., 2003). As soon as it reaches the crater, Strombolian activity begins. Seismic stations record the EQs and produce raw signals that can be analysed. Stromboli's volcanic system is characterised by two magma reservoirs: a shallow reservoir located 3–4 km below the surface and a deeper reservoir located approximately 11 km below the surface (Petrone et al., 2022; Mattia et al., 2008; Harris and Ripepe, 2007).

The analysis of seismological data is crucial for understanding and monitoring the volcanic activity, and the short-term average/long-term average (STA/LTA) method is one of the most widely used techniques for detecting seismic events within continuous waveform data. STA/LTA is a ratio-based approach that compares the average signal amplitude over a short time window with the one over a longer time window. When a seismic event

occurs, the short-term window will capture the strong amplitudes associated with the event, causing a significant increase in the STA/LTA ratio. This increase serves as a detection threshold, and when the ratio exceeds this threshold, it triggers an event detection. This method is particularly effective in identifying the onset of seismic events, such as volcanic tremors, explosions, and microseismicity, by highlighting abrupt changes in the amplitude, which indicate the start of an event (Allen, 1978). In volcanic environments, where seismic signals are often complex and embedded within noisy data, STA/LTA provides a robust mechanism for real-time event detection. The sensitivity of the STA/LTA algorithm can be adjusted by tuning the window lengths, making it adaptable to different types of seismic signals and noise levels (Withers et al., 1999). This adaptability is crucial for monitoring diverse volcanic phenomena, where the nature of seismic signals can vary significantly depending on the type of volcanic activity. Selecting STA/LTA window lengths can be an iterative process and ought to be based on real data analysis. This optimisation process may be time-consuming for human operators. The application of STA/LTA in the context of volcanic seismology has proven invaluable for the early detection of eruptive activity, allowing for timely alerts and the implementation of mitigation strategies. Furthermore, the integration of STA/LTA with other signal processing techniques enhances the overall reliability of volcanic monitoring systems (Hagerty et al., 2000). Automatic picking and cutting techniques for seismograms are essential tools in seismic data analysis, particularly in monitoring volcanic activity. These methods involve the automatic detection of seismic phases, such as P-waves and S-waves, and the precise segmentation of relevant seismic events from continuous waveform data. The automation of these processes is crucial in volcanology, where the rapid analysis of large datasets is necessary for timely eruption forecasts and hazard assessments (Beyreuther et al., 2010). Advanced algorithms, such as those based on machine learning and neural networks, have significantly improved the accuracy and efficiency of phase picking, even in the presence of noise, which is common in volcanic environments (Ross et al., 2018). Additionally, the development of techniques for automatic cutting or windowing of seismograms enables researchers to isolate specific seismic events, such as volcanic tremors or explosions, facilitating a detailed analysis of their characteristics (Hammer et al., 2012). These automated processes not only enhance the speed and reliability of seismic monitoring but also reduce the potential for human error in interpreting complex seismic signals, thereby improving the overall understanding of volcanic processes and aiding in the mitigation of volcanic risks.

In this work, we have implemented a system to perform the automatic detection and waveform extraction of a seismo-volcanic event from raw seismic signals such as EQs, using the STA/LTA method. The data were provided by the Osservatorio Vesuviano (OV)-INGV. The time range selected for the analysis is from 01 June 2019 to 14 June 2019, before the occurrence of the double paroxysm of the Stromboli volcano (Andronico et al., 2021). We have chosen the first days of the dataset as the learning set on which to perform the training of our method, owing to the large number of EQs detected.

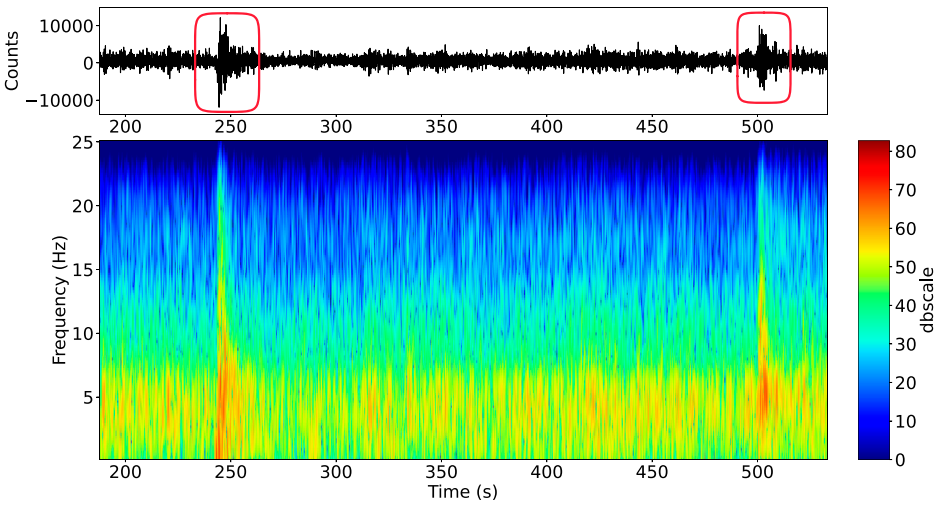
This study is divided into four main sections:

- Methodology for parameter selection: it provides a description of the method and the measures designed to evaluate the extraction of the EQs.
- Experiments and results: it shows the training phase to search for the optimum parameter's combination of STA/LTA and the testing phase where the results of the training were applied on a test set.
- Discussion: it shows the interpretation of results.
- Conclusions and future improvements.

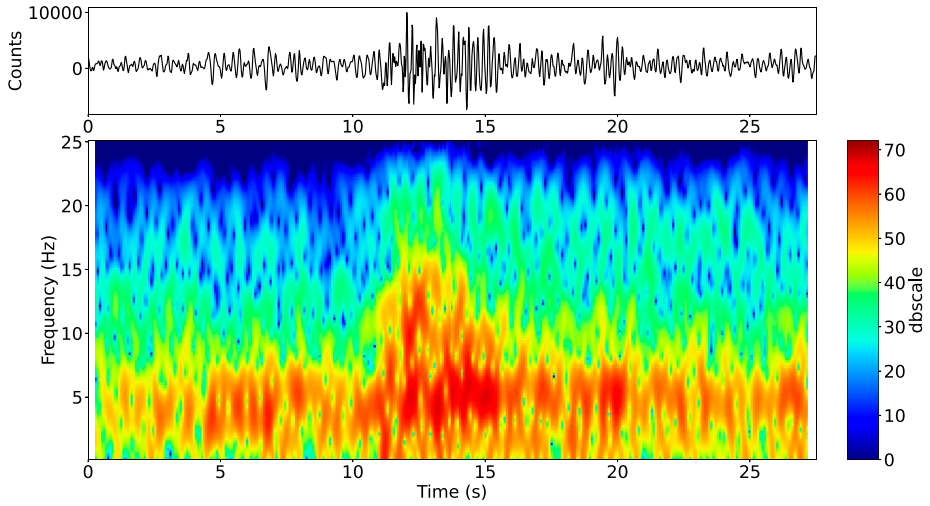
2 Methodology for parameter selection

STA/LTA is based on the analysis of the ratio between short-term and long-term averages of the seismic signal amplitude. This method provides an efficient way to discriminate between seismic events and background noise in seismogram data. With the STA/LTA method, short-term window and long-term window lengths are defined to compute the average amplitudes of seismic signals. The short-term window typically spans a few seconds, capturing the immediate variations in the signal caused by seismic waves. The long-term window, on the other hand, is usually several times longer, capturing the overall background noise level. The STA/LTA ratio is calculated by dividing the average amplitude of the short-term window by the average amplitude of the long-term window. Thus, the STA/LTA is a parametric approach where two basic parameters are STA and LTA window lengths. The ratio between the two windows defines a quantity that is controlled through two other parameters, i.e., trigger on and trigger off. Selection of these parameters can be an iterative process based on real data analysis, where one must remember to continuously monitor the performance of our detection system and make the necessary updates to adapt it to the evolution of the observed phenomenon. We conducted an exploratory research to find the combination of parameters of the STA/LTA method that automatically cut at best EQs, compared to the cuts made by the expert operator. We used our tool developed for our active learning approach (D'Alessandro et al., 2022, see Section 2) to manually cut the EQs. As an example of events present in the dataset, Figure 2A shows the case of two EQs in raw signals, with a zoom on the last one (Figure 2B). To extract the spectrogram, a short-time Fourier transform was calculated using 0.5-s sliding time windows with 90% time overlapping. Figure 3 shows the STA/LTA ratio (bottom) calculated on the zoomed EQ of Figure 2 and the triggers on the signal (top): the red bar consists of trigger on threshold and the blue bar consists of trigger off threshold. When the slope of the curve exceeds the value of the trigger threshold, both for trigger on and trigger off, the red and blue bars are applied on the plot, respectively.

The concept underlying our approach is based on the start and end of the time interval when the event occurs. The start and end times suggested by our approach can be compared with the selections performed by an expert operator by defining a specific measure. In particular, we have proposed two measures: the quality index and the numerosity index. The product of these two is used to define an overall measure called quality–numerosity index (QNI). As a first approach, the characteristic function (CF) E_k [classic one



(a)



(b)

FIGURE 2
This figure presents a representative case of explosion quakes observed in raw seismo-volcanic signals. Each subfigure consists of two components: the raw signal displayed at the top and the corresponding spectrogram at the bottom. The raw signal is plotted with time on the x-axis, while the spectrogram illustrates the representation of this raw signal in the frequency–time domain. The side colour bar indicates frequency values in decibels. In subfigure **(A)**, multiple explosion quakes from the Stromboli volcano's signal are marked with red circles. In subfigure **(B)**, a detailed view of the second explosion quake is provided.

from Allen (1978)] is used for STA/LTA and is defined as follows:

$$E_k = x_k^2 + \left(x_k'\right)^2 + C_k, \tag{1}$$

where x_k is the seismic trace, x_k' its derivative, and C_k (Equation 2) is an empirical weighting constant described as follows:

$$C_k = \frac{\sum_{j=1}^{k} |x_j|}{\sum_{j=1}^{k} |x_j - x_{j-1}|}, \tag{2}$$

to underline the importance of the amplitude and derivative.

For each raw signal considered, the STA/LTA method outputs a list of triggered events, characterised by start time and end time. Those values are compared individually with the ones chosen by the operator. For this comparison, the absolute deviation in terms of the temporal distance is computed. If this value does not exceed a certain residual value k, both for the start time and the end time of the event, then the cut is deemed correct. This comparison enables the assessment of the quality of the cut and the numerosity of the triggered events (see Section 2.1 for details on the quality and numerosity indices). The combination of the two indices is used to obtain an overall index of the triggered events. A list of triggered events can be generated by varying the combination of the following

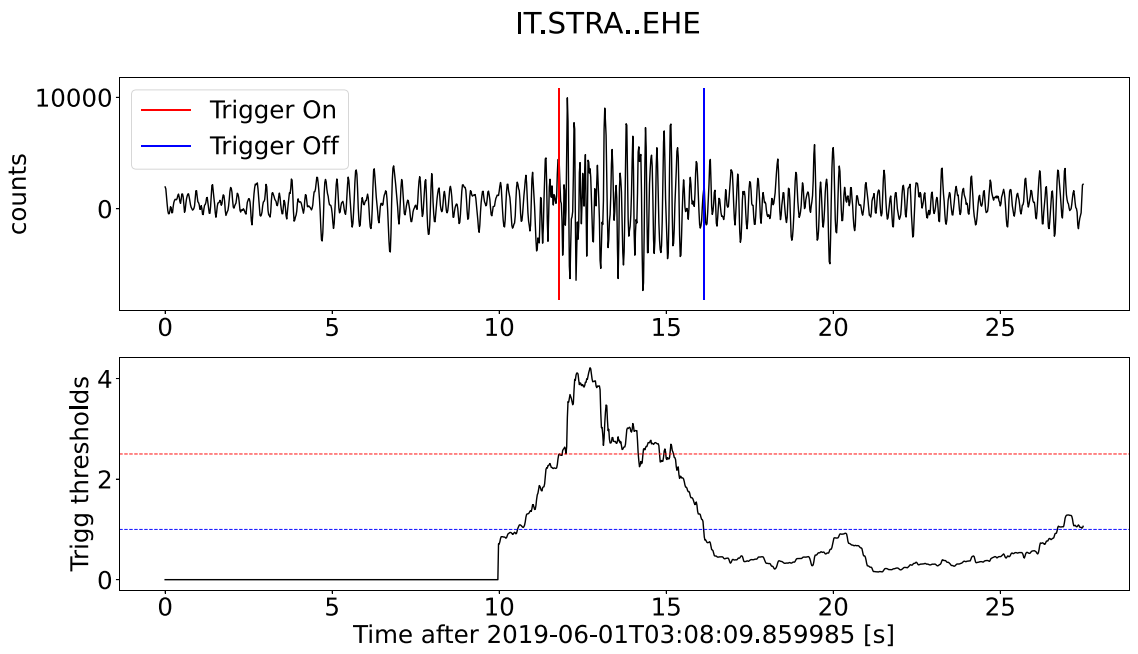


FIGURE 3
STA/LTA performed on the zoomed EQ of Figure 2. Red bar is trigger on and blue bar is trigger off. The figure above illustrates the raw seismo-volcanic signal, while the figure below represents time along the x-axis and trigger threshold values along the y-axis. In this example case, the chosen values of the parameters for this example were STA 1 s, LTA 10 s, trigger on 2.5, and trigger off 1.

four parameters: STA window size (in seconds), LTA window size (in seconds), trigger on threshold, and trigger off threshold. From now on, the four parameters will be indicated with the term "quadruple." A grid search is performed on a quadruple set, by considering the overall index computed on the related triggered events.

2.1 Evaluation measures

The *quality index* of a cut is designated as *qi* and *numerosity index* of a cut as *ni*. These measures were designed to analyse different phenomena represented by time series; in this case, they are used for EQs. The qi measures the degree of greater temporally precise cut performed by the STA/LTA compared to one performed by the human operator. Let *m* be the *mean* of all the temporal deviations computed between the STA/LTA cuts and the operator cuts and *k* be the residual value as *threshold* in seconds. The qi is defined as follows:

$$qi = 1 - (m/k). \tag{3}$$

k is an arbitrarily defined constant, dependent on the reference dataset. In this case, a constant has been empirically set with a value of 10 (preset value. Reported on the repository published on GitHub, see Section 5), based on the average duration of the events (in this case, EQs). This constant falls within the definition of a finite space of values, which in this case are the local events. Therefore, it can also be included if the reference dataset is composed of local seismic events. If, however, regional, teleseismic, anthropic, and landslide (or other types of) events are also included in the dataset, the parameter changes as the finite space in which these events fall varies. Every time STA/LTA outputs a list of triggers, a check is performed to see whether the start time (also indicated as t_{on}, trigger on) and end time (also indicated as t_{off}, trigger off) of every trigger are temporally close to the start and end times of the EQs extracted by the expert. When a match is found (correct cut), the absolute value of the temporal distance is calculated either with the two start times and end times being within *k* or that trigger is not considered, but it is simply taken into account when counting the triggers for the *ni*. If no match is found, the trigger is not considered but is still accounted for when counting triggers for *ni*. Figure 4 shows a representation of this process.

All the computed deviations are stored in a list, and the mean deviation *m* is computed and then normalised by *k*. The *qi* is finally computed as the complement of the ratio m/k so that it is defined in a range between 0 and 1, where 1 means perfect agreement among the automatic cuts and the expert cuts.

The ni measures the agreement between the number of events triggered by the automatic approach and the number of cuts selected by the human operator.

Let *enq* be the *Experimental EQs*, namely, the cardinality of the set of event trigger list produced by the STA/LTA method, and *tnq* be the *Theoretical EQs*, i.e., the number of the cuts performed by the human operator. The ni is thus defined:

$$ni = \begin{cases} enq/tnq, & \text{if } enq < tnq \\ \dfrac{tnq - mod(enq, tnq)}{tnq}, & \text{if } tnq \le enq < 2^{*}tnq \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

The *ni* takes into account the discrepancy between *enq* and *tnq*. If *enq* is lower than *tnq*, or *enq* is between *tnq* and twice *tnq*, then the *ni* will result in a number in the range [0, 1]. Otherwise,

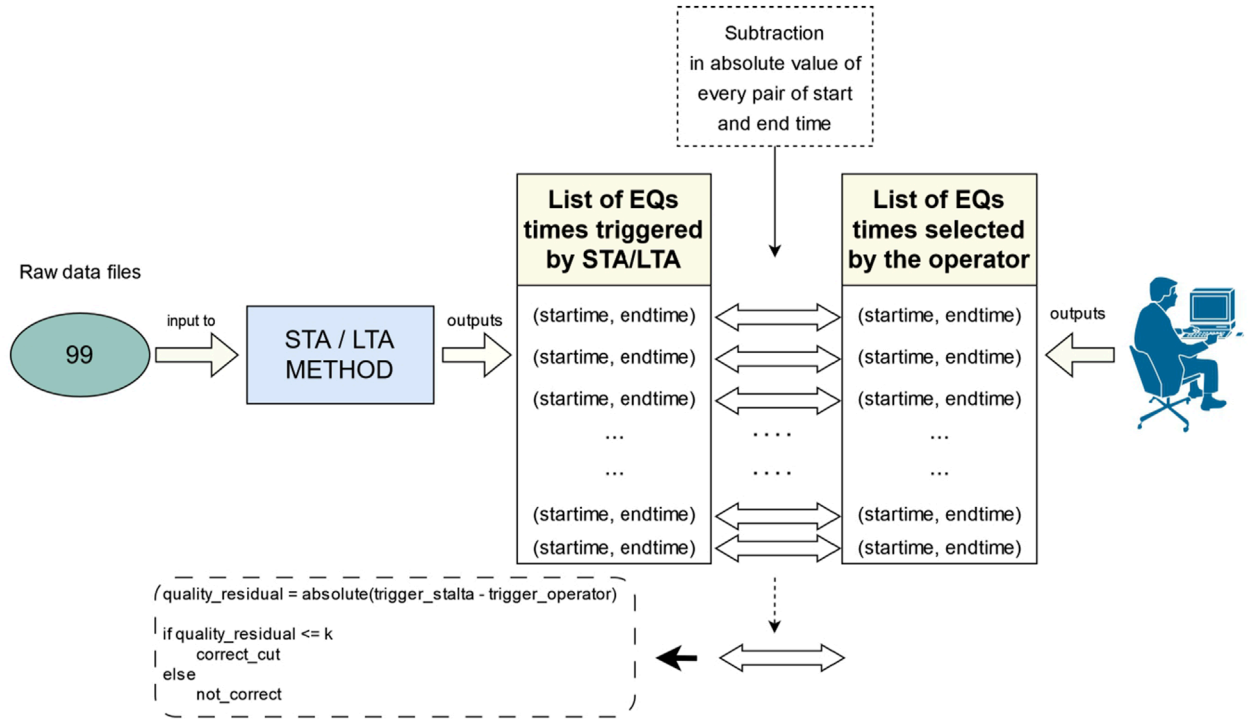


FIGURE 4
Selection of the trigger's start time and end time temporally close to the times of the EQs.

it is set to 0. This latter case occurs when *enq* exceeds *tnq* by at least twice its value, resulting in an *enq*'s value being out of range. The choice of twice the value is justified by limiting the number of false events triggered by STA/LTA with the quadruple considered.

Finally, the combination of the number of events and their temporal precision selected by the automatic process, compared to human experts, makes the QNI the overall measure. This overall measure determines the effectiveness of the cut made by STA/LTA and is defined as follows (Equation 5):

$$qni = qi * ni. \tag{5}$$

The *qni* ranges between 0 and 1 and can be converted into percentages. These measures are mainly dependent on the results of the STA/LTA method (based on its few parameters) and the expertise of the operator because of the manual cut. The range of window values from STA and LTA and the threshold values for trigger on and trigger off can be determined based on the type of event one wants to detect and, therefore, cut. For instance, if one wants to detect teleseismic events compared to local ones, a wide range of STA and LTA window values must be set to ensure the expansion of the grid and improve the search for the optimum. In the beginning of Section 3.2, an in-depth analysis was conducted in this regard. On the other hand, Jones and Baan (2015) used an STA/LTA adaptive method based on the hidden Markov model. This method is independent from data, meaning that it requires only minimal configuration by the user. The goal of this methodology is to determine the probability that a term y(t) is an outlier compared to the noise population. The term y(t) corresponds to the CF of a data point from the seismogram x. STA and LTA windows are composed with these probability levels. Thus, by using this model, the objective of this work is to detect and select a seismic event. Even though this method is adaptive, it is necessary to determine values for STA and LTA window length and threshold adjustment in the initial state.

2.2 Grid-search technique

Grid search is a widely used technique in machine learning and algorithm parameter optimisation. It is used to search for the optimal combination of parameters for a model or algorithm while varying multiple parameters simultaneously. In the present case, a grid search enables an exhaustive search of the quadruples that correspond to the optimum QNI values. Figure 5 shows an example of a representative scheme of the search.

The grid is composed of STA window sizes represented in rows and LTA window sizes represented in columns. Every cell is also a grid, where in abscissa the t_{on} and in ordinate the t_{off} are shown. QNI values are expressed as a percentage and represented as coloured circles: the darker the colour, the higher the value. We started from a basic grid (grid on the left) with a few quadruples (STA window size (in seconds), LTA window size (in seconds), t_{on}, and t_{off}) and gradually expanded to a grid (grid on the right) still containing the previous grid. The cardinality of the grid is determined by the variation of the quadruples by one unit based on their order of magnitude. As an increase in the value of QNI is measured, the grid is expanded until the optimum

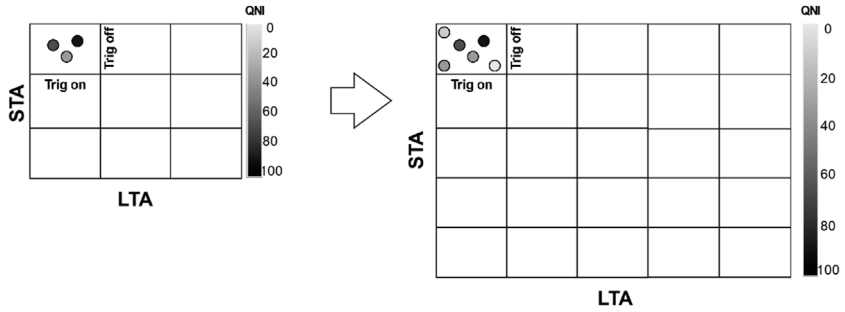


FIGURE 5
Grid-search technique for QNI values.

is found. The process is repeated until the QNI values do not improve further and reach a maximum value. To formalise this concept, a recursive exploratory grid-search algorithm is proposed (Algorithm 1). Choosing to use a recursive algorithm is based on experimenting with all the possible combinations of the quadruples (STA, LTA, t_{on}, and t_{off}) until the best result of QNI is found. As input, the algorithm requires the range of values to construct all possible quadruples (*get_combinations* function), including a *step* value for each parameter (only one is needed for both trigger thresholds). The *step* determines the numerical distance between one value and its next value. After all possible combinations of quadruples have been calculated, the value of the QNI is determined for each of them through the *compute* function; if this value is greater than the previously calculated QNI, then it is classified as the *best_qni* value (the quadruple associated with *best_qni* is also stored). The algorithm ends when the *best_qni* found in lines 5–16 exceeds the threshold value; in this case, the *best_quintuple* list is returned. Otherwise, the algorithm is called recursively by subtracting and adding the *step_parameter* (*step_sta*, for instance) associated with the individual parameter quantities. A series of checks are performed for each parameter to verify that the lower bound of each one is respected. The lower bound is determined by the minimum size of the windows used for the detection, which are usually approximately 1 s for STA and 5 s for LTA. The upper bound for STA and LTA is determined for the search for local and regional events. This limit can, however, be varied based on one's needs (types of events sought) (Küperkoch et al., 2010; Gentili and Michelini, 2006; Earle and Shearer, 1994). If no QNI value exceeds the threshold value, the programme returns the best value of the QNI. In our experiments, 20 iterations were set as thresholds by trial and error.

3 Experiments and results

Starting from 14 days of raw data, our recently developed tool was used (D'Alessandro et al., 2022) to extract the EQ dataset thanks to the expertise of our operator. The entire dataset extracted contains 1,506 EQs. The most significant number of EQs are in the first 4 days of July, i.e., 743 EQs. This set was used as a training set. The subdivision of the training and test datasets is described in Table 1.

3.1 Training phase

A large grid was explored to better view the distribution of the QNI values. We found the densest grid in the training phase with the highest QNI values by combining a range of window values of STA and LTA, respectively, from 2 s to 16 s and 20 s to 220 s, in steps of 2 s for STA and 20 s for LTA. The same representation of the parameter values as in Figure 5 was used. For each combination of windows, t_{on} and t_{off} threshold values were combined, respectively, from 1 to 7, in steps of 0.5 for both. Every cell is a 12*x*12 matrix. Figure 6 shows a screenshot of the terminal grid (8 × 11) or the result of this experiment. At the beginning, the grid was 5 × 5 and step values for STA and LTA were 2 s and 20 s, respectively. After six iterations, an 8 × 11 grid was obtained, according to the STA, LTA, and lower and upper bounds defined in Algorithm 1.

In the lower-left region (highlighted by a red circle), where STA is between 12 s and 16 s and LTA is between 20 s and 60 s, most of the QNI are 0 (no visible circles), and only a few QNI values are approximately 20, which means that this parameter's combination is not suitable for detecting the EQs efficiently. A dashed red line was outlined to show a direction where the QNI values are increasing. The highest values of QNI are found in the central region from left to right, where STA is between 6 s and 10 s, LTA between 60 s and 220 s, t_{on} between 5 and 7, and t_{off} between 2 and 5. Neither QNI, in the first and last rows of the grid, shows an improvement, while the continuous shift toward the right region shows a saturation of the values. This means that increasing the LTA values does not improve the search for local events.

From the training phase, we extracted the list of the QNI values in descending order and generated a plot, shown in Figure 7. An index identifying the value of the QNI is indicated on the abscissa, while the value of the QNI is indicated on the ordinate. The QNI decreases with a moderate slope up to the value 30 and then rapidly decreases to 0. As a consequence, we decided to extract the quadruple associated with the highest QNI value to carry out the test phase, i.e., the one with value 0.78 resulting from the quadruple: 6, 80, 7, and 2. As can be seen, the quadruple is positioned exactly in the distribution indicated by the dashed red line in Figure 6. The selected t_{on} and t_{off} define a good balance between the quality index and numerosity index (see Equations 3, 4) of the detected EQs. This is because a lower t_{on} than 7 can increase the number of false positives, and higher t_{off} can worsen the time's precision of the extraction.

```
procedure GRID_SEARCH(min_sta, max_sta, step_sta,
min_lta, max_lta, step_lta,min_trig_on,
max_trig_on, step_trig_on, min_trig_off,
max_trig_off, step_trig_off, num_iterations)
 quadruples ← get_combinations,
 (min_sta,max_sta,step_sta,
 min_lta,max_lta,step_lta,
 min_trig_on,max_trig_on,
 min_trig_off,max_trig_off, step_trig)
 max ← 0
 best_quintuple ← []
 for all quadruple ∈ quadruples do
  qni ← compute(quadruple)        ▷ QNI
  calculation value
  if qni > max then
     max ← qni
     best_sta ← sta
     best_lta ← lta
     best_trig_on ← trig_on
     best_trig_off ← trig_off
     best_qni ← qni
     best_quintuple ←, best_sta,best_lta,
     best_trig_on,best_trig_off,best_qni
  end if
 end for
 if best_qni > threshold or num_iterations > 20
  then       ▷ For instance: threshold=80
  return best_quintuple
 else
  min_sta ← min_sta − step_sta
  if min_sta < 1 then
   min_sta ← 1
  end if
  max_sta ← max_sta + step_sta
  if max_sta > 16 then
   max_sta ← 16
  end if
  min_lta ← min_lta − step_lta
  if min_lta < 5 then
   min_lta ← 5
  end if
  max_lta ← max_lta + step_lta
  if max_lta > 220 then
   max_lta ← 220
  end if
  min_trig_on ← min_trig_on − step_trig
  if min_trig_on < 0.5 then
   min_trig_on ← 0.5
  end if
  min_trig_off ← min_trig_off − step_trig
  if min_trig_off < 1 then
   min_trig_off ← 1
  end if
  return GRID_SEARCH(min_sta, max_sta + step_sta,
  step_sta, min_lta, max_lta + step_lta,
  step_lta,
  min_trig_on, max_trig_on + step_trig,
  min_trig_off, max_trig_off + step_trig,
  step_trig,
  num_iterations + 1)
 end if
end procedure
```

Algorithm 1 Pseudocode for recursive exploratory grid-search algorithm.

TABLE 1 Dataset subdivision for training and test.

	Num. of EQs extracted	Phase
First 4 days	743	Training
Days 5–14	763	Test
Total	1,506	

3.2 Testing phase

The common approach to the selection of the STA/LTA values is the adoption of literature-suggested ones. Specifically, these values are STA 1 s and LTA 10 s. Regarding the t_{on} and t_{off} values, they were set at 7 and 2, respectively, such as the ones found through the training phase. The QNI computed is indicated using these values as the literature quadruple. The test results are shown in Table 2. In particular, every row corresponds to a range of days in which a certain number of EQs have occurred. This number is indicated in the "Num. EQs extracted" column; in the other columns, the QNI values are reported after an experiment with the associated days was performed with the corresponding quadruples: training quadruple (third column, our quadruple) and literature quadruple (fourth column, quadruple extracted from the literature). An overlap of the days for testing purposes was carried out. In the first row, the testing result is shown using the same dataset used in the training phase. This first comparison was made to show our result compared to the use of the literature quadruple.

4 Discussion

The literature has shown that there is no one single strategy to search the STA and LTA moving windows to select events based on triggers (Earle and Shearer, 1994). The lengths for the STA and LTA windows depend on the frequency content of the seismogram. Long-period records require larger averaging windows than short-period records, which require shorter averaging windows. Among different approaches, we have chosen the work by Küperkoch et al. (2010) as the base case comparison owing to the completeness of this work, focussing on P-phase arrival time, where several CFs for STA/LTA

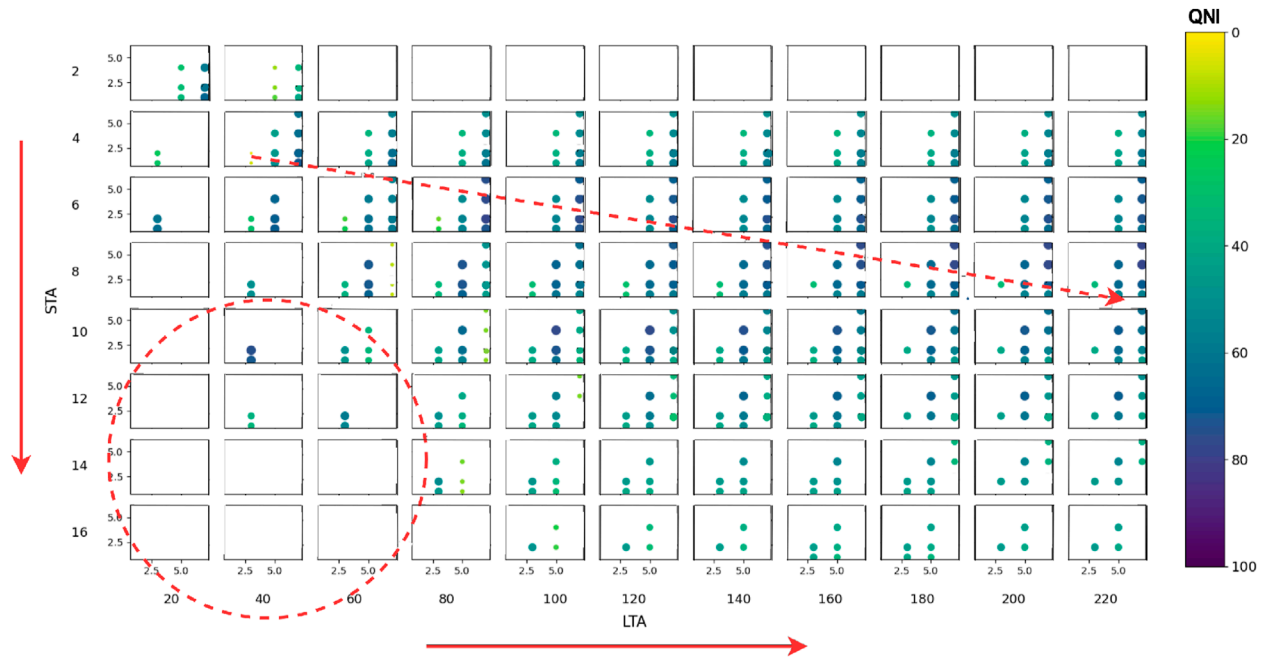


FIGURE 6
Grid-search exploration for STA windows (rows) and LTA windows (columns), represented by the red rows. Both STA and LTA windows are expressed in seconds. Trigger on/off thresholds, respectively, in abscissa and ordinate, are shown in every cell of the matrix. The red circle shows a region where most of the QNI values are 0. The red dashed line indicates a direction where the QNI values are increasing.

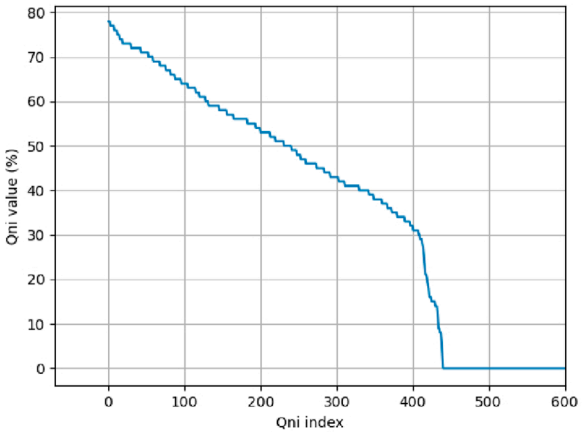


FIGURE 7
QNI value distribution.

were used to compare their method. They have implemented an algorithm based on higher-order statistics (HOS) for automatic P-phase arrival time determination for local and regional seismic events. The algorithm was applied to a large dataset with very heterogeneous qualities of P-onsets. They calculated several CFs by evaluating higher-order statistical moments, like skewness, kurtosis, mean, and variance. In our case, we decided to use the CF reported in Equation 1, which is the base case determined by Allen (1978). When choosing a "literature quadruple," there is no clear standard choice for the types of events analysed, yet Küperkoch et al. (2010) values can provide a useful comparison. The results presented in the previous paragraph indicate that, as we move further away from the training set, the QNI appears to vary over time. This is due to the rapid evolution of the volcanological phenomenon, and therefore the EQs generated by it. As is well known (Andronico et al., 2021), the duration, amplitude, and frequency content of the EQs can vary rapidly as the volcanic process evolves. The results show that on average, our training quadruple produced a QNI 0.24 higher than the literature quadruple. Another strength of our approach is the tuning of the evaluation measures: for instance, *qi* is one of the measures that can be adapted based on the data and objectives one wants to obtain; in this case (where we look for EQs in the raw data), we use the mean of the deviations to evaluate the quality of the cut, but one can consider using different evaluations such as median, mode, or kurtosis. This possibility can lead to extending the research to the general volcano domain and also in the seismic domain in tectonic areas.

Certainly, the main limitation of our study is the low cardinality of the dataset. Based on the time-consuming process of extracting the seismo-volcanic events from real raw data, future improvements will mainly concern the variability of the dataset; extending not only to other types of seismo-volcanic events, such as the following: VLP, landslides, volcanic tremors, and others (Wassermann, 2012), but also to local, regional, and teleseismic events. To achieve this, we aim to replace the human operator by automating the validation process through the use of multiple seismic stations recording the same signal within the same area. This approach replicates the current method used by operators but will be enhanced by deploying seismic stations positioned at approximately equal distances around the crater (Fenner et al., 2022), ensuring more uniform coverage and

TABLE 2 Testing results.

	Num. of EQs extracted	(QNI) training quadruple 6 s 80 s 7 2	(QNI) literature quadruple 1 s 10 s 7 2
First 4 days (train data)	743	0.78	0.59
Days 5–8	395	0.65	0.36
Days 6–9	425	0.64	0.38
Days 7–10	373	0.56	0.33
Days 8–11	397	0.54	0.26
Days 9–12	291	0.48	0.21
Days 10–13	271	0.51	0.3
Last 4 days (11–14)	170	0.5	0.3

reducing potential sources of error. Another limitation is specific to the grid-search technique. In general, grid search is a powerful technique for optimising algorithm parameters, but it ought to be used judiciously as it may be time-consuming when there are many parameter combinations to evaluate. Other techniques such as random search or Bayesian optimisation may be more efficient alternatives in some cases. The choice depends on the specific problem domain and available resources. Stromboli remains a case study, but the method is applicable to any type of seismo-volcanic signal and can therefore be used on other volcanoes as well. For instance, it is also suitable for earthquakes in non-volcanic areas.

5 Conclusion and future improvements

In this scientific study, we have explored the potential of using a grid-search method to study the STA/LTA parameters to select seismo-volcanic events, starting from raw signals, with a particular focus on the volcanic activity of Stromboli. Through this application, we have proved the ability to efficiently detect local events, such as explosion quakes. The results showed a more accurate choice of parameters, compared to what was proposed in the literature (Küperkoch et al., 2010), for searching local events, such as EQs. As a first example for the approach, we exploited the constant presence of EQs before the occurrence of the double paroxysms of Stromboli volcano (Andronico et al., 2021).

With this method, one can collect seismo-volcanic events that can be used from the machine learning perspective, such as classification or regression problems, where a certain amount of data is needed for the dataset (Zhu and Beroza, 2018; Mousavi et al., 2020). We decided to use, among possible characteristic functions for STA/LTA, the classic one by Allen (1978) as the first approach. It is also possible to test other CFs as Küperkoch et al. (2010) did for their method. To compare our quadruple, we referred to the work by Küperkoch et al. (2010) to find the baseline quadruple in the literature for the detection of local events (classifying the explosion quakes as local events).

In summary, the integration of our approach is a compelling way to simplify the acquisition of labelled data in seismic–volcanic and more generally of seismic research. By synergising active learning (D'Alessandro et al., 2022) with robust deep learning algorithms and a large dataset, a path towards greater accuracy, effectiveness, and a comprehensive analysis of seismic–volcanic phenomena can be achieved.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository and accession number(s) can be found at: https://github.com/Andry92/grid-search-sta-lta.

Author contributions

ADB: conceptualization, data curation, formal analysis, investigation, methodology, project administration, software, supervision, validation, visualization, writing–original draft, and writing–review and editing. AF: data curation, formal analysis, investigation, resources, software, writing–original draft, and writing–review and editing. AD'A: conceptualization, formal analysis, investigation, methodology, supervision, validation, visualization, writing–original draft, and writing–review and editing. GLB: conceptualization, formal analysis, investigation, methodology, software, supervision, validation, visualization, writing–original draft, and writing–review and editing.

Funding

The authors declare that no financial support was received for the research, authorship, and/or publication of this article.

Acknowledgments

The authors would like to acknowledge Dario Delle Donne from Istituto Nazionale di Geofisica e Vulcanologica—Osservatorio

Vesuviano (INGV-OV), for helping us to extract the data from STRA seismic station.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Aiuppa, A., Bertagnini, A., Métrich, N., Moretti, R., Di Muro, A., Liuzzo, M., et al. (2010). A model of degassing for stromboli volcano. *Earth Planet. Sci. Lett.* 295, 195–204. doi:10.1016/j.epsl.2010.03.040

Allen, R. V. (1978). Automatic earthquake recognition and timing from single traces. *Bull. Seismol. Soc. Am.* 68, 1521–1532. doi:10.1785/BSSA0680051521

Andronico, D., Del Bello, E., D'Oriano, C., Landi, P., Pardini, F., Scarlato, P., et al. (2021). Uncovering the eruptive patterns of the 2019 double paroxysm eruption crisis of stromboli volcano. *Nat. Commun.* 12, 4213. doi:10.1038/s41467-021-24420-1

Beyreuther, M., Barsch, R., Krischer, L., Megies, T., Behr, Y., and Wassermann, J. (2010). ObsPy: a Python toolbox for seismology. *Seismol. Res. Lett.* 81, 530–533. doi:10.1785/gssrl.81.3.530

Calvari, S., Giudicepietro, F., Di Traglia, F., Bonaccorso, A., Macedonio, G., and Casagli, N. (2021). Variable magnitude and intensity of strombolian explosions: focus on the eruptive processes for a first classification scheme for stromboli volcano (Italy). *Remote Sens.* 13, 944. doi:10.3390/rs13050944

Cannioto, M., D'Alessandro, A., Lo Bosco, G., Scudero, S., and Vitale, G. (2017). Brief communication: vehicle routing problem and uav application in the post-earthquake scenario. *Nat. Hazards Earth Syst. Sci.* 17, 1939–1946. doi:10.5194/nhess-17-1939-2017

Chouet, B. (1996). Long-period volcano seismicity: its source and use in eruption forecasting. *Nature* 380, 309–316. doi:10.1038/380309a0

Chouet, B., Dawson, P., Ohminato, T., Martini, M., Saccorotti, G., Giudicepietro, F., et al. (2003). Source mechanisms of explosions at stromboli volcano, Italy, determined from moment-tensor inversions of very-long-period data. *J. Geophys. Res. Solid Earth* 108. doi:10.1029/2002jb001919

D'Alessandro, A., Di Benedetto, A., Lo Bosco, G., and Figlioli, A. (2022). "An active learning approach for classifying explosion quakes," in 2022 IEEE International Conference on Evolving and Adaptive Intelligent Systems (EAIS), Larnaca, Cyprus, May 25–26, 2022 (IEEE), 1–6. doi:10.1109/EAIS51927.2022.978751910

Earle, P. S., and Shearer, P. M. (1994). Characterization of global seismograms using an automatic-picking algorithm. *Bull. Seismol. Soc. Am.* 84, 366–376. doi:10.1785/BSSA0840020366

Fenner, D., Rümpker, G., Li, W., Chakraborty, M., Faber, J., Köhler, J., et al. (2022). Automated seismo-volcanic event detection applied to stromboli (Italy). *Front. Earth Sci.* 10. doi:10.3389/feart.2022.809037

Gentili, S., and Michelini, A. (2006). Automatic picking of P and S phases using a neural tree. *J. Seismol.* 10, 39–63. doi:10.1007/s10950-006-2296-6

Giudicepietro, F., Calvari, S., Alparone, S., Bianco, F., Bonaccorso, A., Bruno, V., et al. (2019). Integration of ground-based remote-sensing and *in situ* multidisciplinary monitoring data to analyze the eruptive activity of stromboli volcano in 2017–2018. *Remote Sens.* 11, 1813. doi:10.3390/rs11151813rs11151813

Giudicepietro, F., López, C., Macedonio, G., Alparone, S., Bianco, F., Calvari, S., et al. (2020). Geophysical precursors of the July-August 2019 paroxysmal eruptive phase and their implications for Stromboli volcano (Italy) monitoring. *Sci. Rep.* 10, 10296. doi:10.1038/s41598-020-67220-1

Hagerty, M., Schwartz, S., Garcés, M., and Protti, M. (2000). Analysis of seismic and acoustic observations at arenal volcano, Costa Rica, 1995–1997. *J. Volcanol. Geotherm. Res.* 101, 27–65. doi:10.1016/s0377-0273(00)00162-11016/S0377-0273(00)00162-1

Hammer, C., Beyreuther, M., and Ohrnberger, M. (2012). A seismic-event spotting system for volcano fast-response systems. *Bull. Seismol. Soc. Am.* 102, 948–960. doi:10.1785/0120110167

Harris, A., and Ripepe, M. (2007). Synergy of multiple geophysical approaches to unravel explosive eruption conduit and source dynamics – a case study from Stromboli. *Chem. Erode - Geochem.* 67, 1–35. doi:10.1016/j.chemer.2007.01.0032007.01.003

Jones, J., and Baan, M. (2015). Adaptive sta–lta with outlier statistics. *Bull. Seismol. Soc. Am.* 105, 1606–1618. doi:10.1785/0120140203

Journeau, C., Shapiro, N., Seydoux, L., Soubestre, J., Ferrazzini, v., and Peltier, A. (2020). Detection, classification, and location of seismovolcanic signals with multi-component seismic data, example from the piton de la fournaise volcano (la reunion, france). doi:10.1002/essoar.10501605.2

Konstantinou, K. (2023). A review of the source characteristics and physical mechanisms of very long period (vlp) seismic signals at active volcanoes. *Surv. Geophys.* 45, 117–149. doi:10.1007/s10712-023-09800-0

Küperkoch, L., Meier, T., Lee, J., Friederich, W., and Group, E. (2010). Automated determination of p-phase arrival times at regional and local distances using higher order statistics. *Geophys. J. Int.* 181, 1159–1170. doi:10.1111/j.1365-246X.2010.04570.x

Lara, F., Lara-Cueva, R., Larco, J., Carrera, E., and Leon, R. (2020). A deep learning approach for automatic recognition of seismo-volcanic events at the Cotopaxi volcano. *J. Volcanol. Geotherm. Res.* 409, 107142. doi:10.1016/j.jvolgeores.2020.1071421016/j.jvolgeores.2020.107142

Legrand, D., and Perton, M. (2021). What are vlp signals at stromboli volcano? *J. Volcanol. Geotherm. Res.* 421, 107438. doi:10.1016/j.jvolgeores.2021.107438

Liang, C., Karlstrom, L., and Dunham, E. (2020). Magma oscillations in a conduit-reservoir system, application to very long period (vlp) seismicity at basaltic volcanoes: 1. theory. *J. Geophys. Res. Solid Earth* 125. doi:10.1029/2019JB017437

Makus, P., Denolle, M., Sens-Schönfelder, C., Kopfli, M., and Tilmann, F. (2024). Analyzing volcanic, tectonic, and environmental influences on the seismic velocity from 25 years of data at Mount St. Helens. *Seismol. Res. Lett.* 95, 2674–2688. doi:10.1785/0220240088

Mattia, M., Aloisi, M., Di Grazia, G., Gambino, S., Palano, M., and Bruno, V. (2008). Geophysical investigations of the plumbing system of Stromboli volcano (Aeolian islands, Italy). *J. Volcanol. Geotherm. Res.* 176, 529–540. doi:10.1016/j.jvolgeores.2008.04.022

Metrich, N., Bertagnini, A., and Pistolesi, M. (2021). Paroxysms at Stromboli volcano (Italy): source, genesis and dynamics. *Front. Earth Sci.* 9, 593339. doi:10.3389/feart.2021.593339feart.2021.593339

Mousavi, S., Ellsworth, W., Weiqiang, Z., Chuang, L., and Beroza, G. (2020). Earthquake transformer—an attentive deep-learning model for simultaneous earthquake detection and phase picking. *Nat. Commun.* 11, 3952. doi:10.1038/s41467-020-17591-w1038/s41467-020-17591-w

Petrone, C., Mollo, S., Gertisser, R., Buret, Y., Scarlato, P., Del Bello, E., et al. (2022). Magma recharge and mush rejuvenation drive paroxysmal activity at Stromboli volcano. *Nat. Commun.* 13, 7717. doi:10.1038/s41467-022-35405-z

QGIS Development Team (2024). Qgis geographic information system

Ripepe, M., Delle Donne, D., Legrand, D., Valade, S., and Lacanna, G. (2021a). Magma pressure discharge induces very long period seismicity. *Sci. Rep.* 11, 20065. doi:10.1038/s41598-021-99513-4

Ripepe, M., and Harris, A. (2008). Dynamics of the 5 April 2003 explosive paroxysm observed at stromboli by a near-vent thermal, seismic and infrasonic array. *Geophys. Res. Lett.* 35. doi:10.1029/2007gl0325332007GL032533

Ripepe, M., Lacanna, G., Pistolesi, M., Silengo, M., Aiuppa, A., Laiolo, M., et al. (2021b). Ground deformation reveals the scale-invariant conduit dynamics driving explosive basaltic eruptions. *Nat. Commun.* 12, 1683. doi:10.1038/s41467-021-21722-2s41467-021-21722-2

Ripepe, M., Pistolesi, M., Coppola, D., Delle Donne, D., Genco, R., Lacanna, G., et al. (2017). Forecasting effusive dynamics and decompression rates by magmastatic model at open-vent volcanoes. *Sci. Rep.* 7, 3885. doi:10.1038/s41598-017-03833-3s41598-017-03833-3

Ross, Z. E., Meier, M., Hauksson, E., and Heaton, T. H. (2018). Generalized seismic phase detection with deep learning. *Bull. Seismol. Soc. Am.* 108, 2894–2901. doi:10.1785/0120180080

Schaefer, L. N., Di Traglia, F., Chaussard, E., Lu, Z., Nolesini, T., and Casagli, N. (2019). Monitoring volcano slope instability with synthetic aperture radar: a review and

new data from Pacaya (Guatemala) and Stromboli (Italy) volcanoes. *Earth-Science Rev.* 192, 236–257. doi:10.1016/j.earscirev.2019.03.009

Sosa, Y. M., Molina, R. S., Spagnotto, S., Melchor, I., Nuñez Manquez, A., Crespo, M. L., et al. (2024). Seismic event detection in the copahue volcano based on machine learning: towards an on-the-edge implementation. *Electronics* 13, 622. doi:10.3390/electronics13030622

Soubestre, J., Shapiro, N., Seydoux, L., de Rosny, J., Droznin, D., Droznina, S., et al. (2018). Network-based detection and classification of seismovolcanic tremors: example from the klyuchevskoy volcanic group in kamchatka. *J. Geophys. Res. Solid Earth* 123, 564–582. doi:10.1002/2017JB014726

Wassermann, J. (2012). *Volcano seismology, IASPEI new manual of seismological observatory practice 2 (NMSOP-2)*. Potsdam: Deutsches GeoForschungsZentrum GFZ, 1–77.

Withers, M., Aster, R., and Young, C. (1999). An automated local and regional seismic event detection and location system using waveform correlation. *Bull. Seismol. Soc. Am.* 89, 657–669. doi:10.1785/bssa0890030657BSSA08900 30657

Zhu, W., and Beroza, G. C. (2018). PhaseNet: a deep-neural-network-based seismic arrival-time picking method. *Geophys. J. Int.* 216, 261–273. doi:10.1093/gji/ggy423

Frontiers in Earth Science

Investigates the processes operating within the major spheres of our planet

Advances our understanding across the earth sciences, providing a theoretical background for better use of our planet's resources and equipping us to face major environmental challenges.

Discover the latest Research Topics

See more →

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact