RESEARCH ARTICLE

VOLCANICA

A generalized deep learning model to detect and classify volcano seismicity

David Fee*[α], Darren Tan[α], John Lyons[β], Mariangela Sciotto[γ], Andrea Cannata[γ,δ], Alicia Hotovec-Ellis[ε], Társilo Girona[α], Aaron Wech[β], Diana Roman[ζ], Matthew M. Haney[β], and Silvio De Angelis[η]

[α] Alaska Volcano Observatory, Geophysical Institute, University of Alaska Fairbanks, Fairbanks, AK, USA.
[β] U.S. Geological Survey, Volcano Science Center, Alaska Volcano Observatory, Anchorage, AK, USA.
[γ] Istituto Nazionale di Geofisica e Vulcanologia, Osservatorio Etneo – Sezione di Catania, Catania, Italy.
[δ] Dipartimento di Scienze Biologiche, Geologiche e Ambientali - Sezione di Scienze della Terra, Università degli Studi di Catania, Catania, Italy.
[ε] U.S. Geological Survey, California Volcano Observatory, Moffett Field, CA, USA.
[ζ] Earth and Planets Laboratory, Carnegie Institution for Science, Washington DC 20015, USA.
[η] School of Environmental Sciences, University of Liverpool, Liverpool, England, UK.

Abstract

Volcano seismicity is often detected and classified based on its spectral properties. However, the wide variety of volcano seismic signals and increasing amounts of data make accurate, consistent, and efficient detection and classification challenging. Machine learning (ML) has proven very effective at detecting and classifying tectonic seismicity, particularly using Convolutional Neural Networks (CNNs) and leveraging labeled datasets from regional seismic networks. Progress has been made applying ML to volcano seismicity, but efforts have typically been focused on a single volcano and are often hampered by the limited availability of training data. We build on the method of Tan et al. [2024] (10.1029/2024JB029194) to generalize a spectrogram-based CNN termed the VOlcano Infrasound and Seismic Spectrogram Neural Network (VOISS-Net) to detect and classify volcano seismicity at any volcano. We use a diverse training dataset of over 270,000 spectrograms from multiple volcanoes: Pavlof, Semisopochnoi, Tanaga, Takawangha, and Redoubt volcanoes (Alaska, USA); Mt. Etna (Italy); and Kīlauea, Hawai'i (USA). These volcanoes present a wide range of volcano seismic signals, source-receiver distances, and eruption styles. Our generalized VOISS-Net model achieves an accuracy of 87 % on the test set. We apply this model to continuous data from several volcanoes and eruptions included within and outside our training set, and find that multiple types of tremor, explosions, earthquakes, long-period events, and noise are successfully detected and classified. The model occasionally confuses transient signals such as earthquakes and explosions and misclassifies seismicity not included in the training dataset (e.g. teleseismic earthquakes). We envision the generalized VOISS-Net model to be applicable in both research and operational volcano monitoring settings.

Keywords: Volcano Seismicity; Machine Learning; Tremor; Explosion.

1 Introduction

Volcanoes produce a wide range of seismic signals [e.g. McNutt and Roman 2015]. These seismic signals are used to monitor volcanic unrest in real-time and to study subsurface and subaerial processes [e.g. Matoza and Roman 2022]. At many volcanoes, seismic data are the backbone of the monitoring network and the various types of volcano seismicity are often used to infer volcanic processes. For example, volcano-tectonic (VT) events have been used to infer both the location of dikes and ascending magma and fluids [e.g. Sigmundsson et al. 2015; Power and Roman 2024]. Volcanic tremor is often linked to subsurface magmatic and hydrothermal fluids, and can be a key indicator of a pending volcanic eruption [McNutt 1996] or ongoing eruption in the case of "eruption tremor" [e.g. McNutt and Nishimura 2008; Fee et al. 2017]. Similarly, discrete long-period (LP) seismic events provide key information on fluid movement in various parts of the crust [e.g. Hotovec-Ellis et al. 2018]. Volcanic explosions can also be detected in seismic data and provide valuable information on eruption processes [e.g. Nishimura 1998].

Automated detection and classification of volcano seismicity has long been a focus of geophysical research. In addition to traditional methods of earthquake detection and location [e.g. Lahr 1989; Lomax et al. 2000; Klein 2002], seismologists often classify the various types of volcano seismicity by both their spectral characteristics and signal duration. This can be challenging due to the diverse seismicity present at volcanoes, variable spectral characteristics as a function of distance and direction from the volcano, and the subjective interpretation of the seismic analyst. Additionally, increasing amounts of data add workload challenges to the analyst.

Machine learning (ML), in particular deep learning, has recently facilitated substantial progress in image classification, as well as becoming nearly ubiquitous in technology. Marked advances have been made in seismology from ML [Mousavi and Beroza 2022], including those that use spectrograms as images in classifying the type of seismicity [e.g. Linville et al. 2019; Kong et al. 2022; Maguire et al. 2024]. Although unsupervised learning has been successful in some applications of volcano seismology [e.g. Steinke et al. 2023; Zali et al. 2024], supervised learning allows trained experts to define the type of signal to be detected and classified, facilitating interpreta-

* dfee1@alaska.edu

Submitted: 2024-12-19 | Accepted: 2025-05-05 | Published: 2025-06-12

Editor: C. Montagna | Typesetter: J. Farquharson

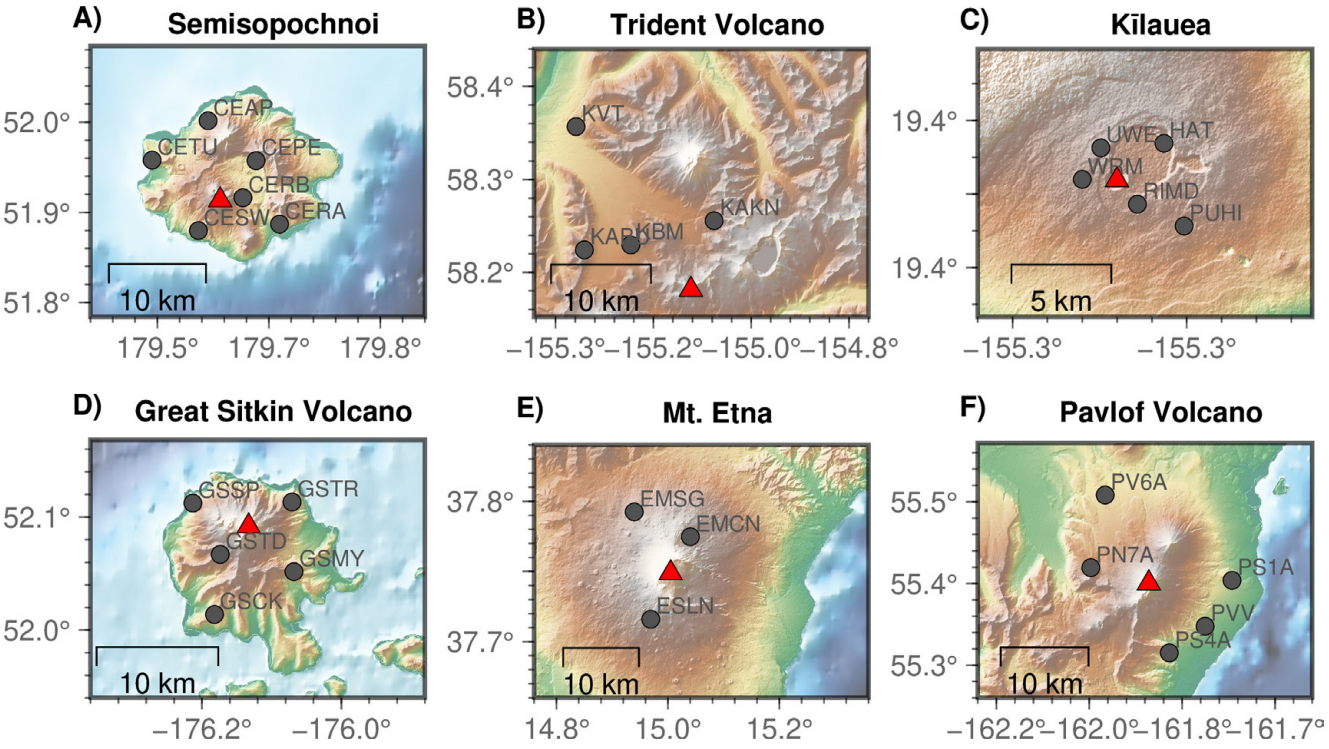

Figure 1: Map of the six volcanoes and seismic stations used as case studies. [A] Semisopochnoi, Alaska; [B] Trident Volcano, Alaska; [C] Kīlauea, Hawai'i; [D] Great Sitkin Volcano, Alaska; [E] Mt. Etna, Italy; [F] Pavlof Volcano, Alaska. Gray circles indicate the locations of the stations we use, and the red triangles are the recent eruptive vents. The figure was created using PyGMT [Uieda et al. 2021].

tion and impact. One of the challenges of supervised ML is the need for large, representative datasets to train the model. Until recently these have been lacking for volcano seismicity, with only a few examples [e.g. Malfante et al. 2018; Titos et al. 2019; Zhong and Tan 2024]. ML-based determination of quiescence versus unrest using extensive feature sets also shows promise for identifying elusive eruptions precursors [Ardid et al. 2022]. Previous work primarily applied neural networks to cataloged volcano seismic events rather than continuous data. A recent study by Tan et al. [2024] utilized deep learning and tens of thousands of analyst-labeled spectrograms to construct a model to detect and classify volcano seismicity and infrasound at Pavlof Volcano, Alaska. Their model, termed the VOlcano Infrasound and Seismic Spectrogram Network (`VOISS-Net`), defined six classes of seismic and infrasound signals which we build upon and define in Section 2. A human analyst labeled thousands of spectrograms for each of these classes, which were then used to train a Convolutional Neural Network (CNN) that was focused primarily on detecting volcanic tremor. `VOISS-Net` can then be applied to continuous seismic or infrasound data and produces both individual station and network-wide classifications.

In this study we build on the work of Tan et al. [2024], hereafter referred to as T24, and develop an ML model that aims to automatically detect and classify the main types of volcano seismicity at any volcano. Our "generalized" `VOISS-Net` model uses a similar architecture and labeled dataset to the original `VOISS-Net` model, but similar to Lyons et al. [2025] adds in an LP class, tens of thousands more spectrogram labels for training from multiple volcanoes, and swaps out the T24 earthquake labels for local earthquakes. Our generalized model also contains some key differences to the spectrograms and image normalization strategies. After introducing the methods and data selection, we present six case studies applying `VOISS-Net` to a diverse suite of seismic signals recorded at volcanoes (Figure 1), and discuss some of the strengths and limitations of applying the generalized model. Our model is designed to ease seismic analyst burden and provide repeatable, fast methods to automatically detect and classify volcano seismicity.

2 Data and Methods

2.1 Data

2.1.1 *Data processing and spectrogram construction*

We construct the spectrograms using vertical component seismic data. We do not use all three components of the seismic data as some stations we have trained and applied the model to are only vertical component. All seismic data were resampled to a 50 Hz sampling rate using `ObsPy` [Beyreuther et al. 2010]. The spectrograms are constructed by removing the instrument response and then using a 10 s Hann window with 90 % overlap. We convert the spectral amplitudes to dB relative to 1 $\mathrm{m\,s^{-2}\,Hz^{-1}}$. The spectrograms are then cropped between 0.5–10 Hz. This frequency band was selected as it encompasses the majority of volcano seismic energy and filters

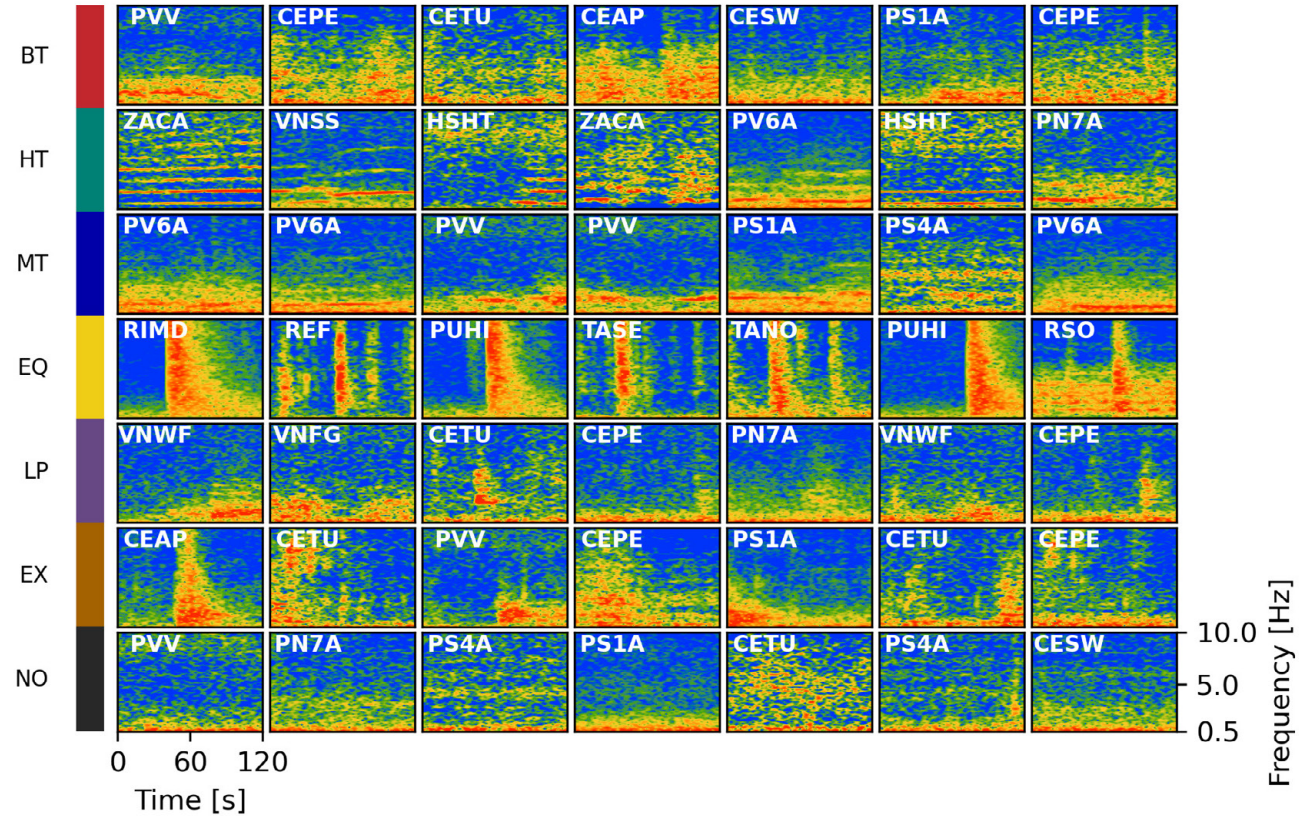

Figure 2: Example test set spectrogram labels for each of the seven seismicity classes. Each row corresponds to a seismicity class, indicated by its abbreviated name. Seven random spectrogram examples are plotted for each class, with the station name indicated in the upper left. Class labels on the left are abbreviated as: Broadband Tremor (BT), Harmonic Tremor (HT), Monochromatic Tremor (MT), Earthquake (EQ), Long Period (LP), Explosion (EX), and Noise (NO).

out lower frequency microseisms and some higher frequency noise sources such as wind. The spectrograms are then split into 2-minute windows and normalized between 0 and 1.

2.1.2 *Seismic spectrogram classes*

We build off T24 and define seven volcano seismicity classes that are broadly defined by their spectral and temporal characteristics:

1. Broadband Tremor (BT): long-duration (>2 min) signals with a broad (>1 Hz) frequency peak below 5 Hz.

2. Harmonic Tremor (HT): long-duration signals (>2 min) with multiple narrow-band (<1 Hz) overtones with energies comparable to that of the fundamental tone.

3. Monochromatic Tremor (MT): long-duration (>2 min) signals with a narrow-band fundamental tone and no energetic overtones.

4. Earthquake (EQ): Local earthquakes with an impulsive broadband onset and a lower frequency coda. Unlike T24, we only use catalog-derived VT earthquakes.

5. Long Period event (LP): Transient (<2 min) signals with frequencies confined between ~0.5–5 Hz.

6. Explosion (EX): transient signals (seconds) which are reliably determined to be volcanic explosions either via (a) air-ground coupled airwaves evident in seismic data, or (b) retrospective infrasound source location analysis [Fee et al. 2021]. Explosions often have a low frequency (<5 Hz) seismic onset followed by a higher frequency air-ground coupled acoustic wave.

7. Noise: catch-all class for any non-systematic variations in seismic spectral signature and instances of quiescence where no notable signal is observed.

These class definitions were used to guide analyst labeling and spectrogram generation. In this work we added the LP class, as LP events are a common and often valuable volcano seismic signal [McNutt 1996; Matoza and Roman 2022]. T24 noted how multiple LP swarms at Pavlof were misclassified, and they recommended an LP class be considered in future work. Lyons et al. [2025] subsequently added an LP class to a `VOISS-Net` model that successfully classified LP seismicity at Semisopochnoi volcano. We acknowledge the aforementioned definitions do not encompass all volcano seismicity characteristics and temporal and spectral variability exists within and between classes. Figure 2 shows seven randomly

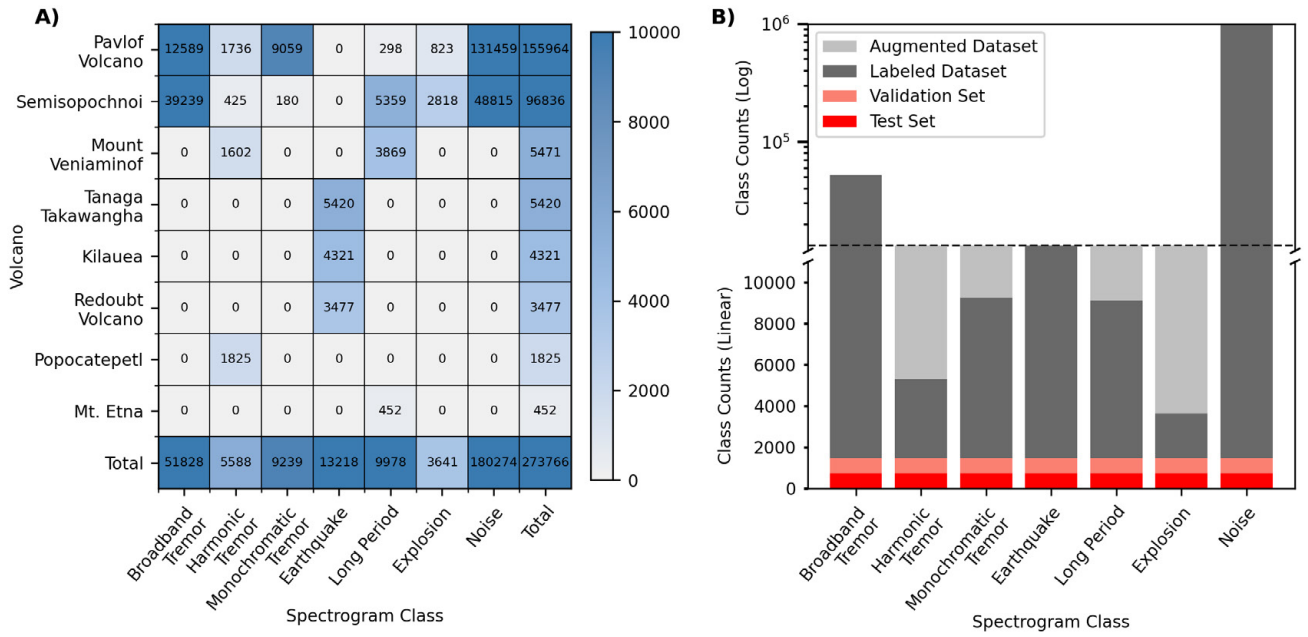

Figure 3: Spectrogram class counts. [A] Matrix of class labels listed by volcano and our defined spectrogram classes. The rightmost column is the total number of spectrogram for each volcano and bottom row the total number of spectrograms per class. [B] Bar chart of class counts and balance between the training, test, and validation sets. Light gray portions indicate augmented data. See text for details.

selected spectrograms from our test set for each seismicity class.

2.1.3 *Spectrogram data*

Our labeled dataset consists of 273,766 2-minute spectrograms from a variety of volcanoes and stations. Figure 3A shows a heatmap of the number of spectrograms for each seismicity class and volcano. The rightmost column represents the total amount of labels for each volcano and bottom row the total class counts for each seismicity class. Supplementary Material 1 Figure S1 displays a map of the volcanoes and stations used to construct the labeled dataset. As a baseline, we use the labeled dataset of T24 from Pavlof Volcano's 2021–22 eruption, which had 35,382 4-minute spectrogram labels. We take their labeled spectrogram pixel bounds and set the individual spectrogram duration to 2 minutes (rather than 4). We choose 2 minute spectrograms to better capture transient signals, as T24 was focused on longer duration tremor. See T24 for a description of the manual spectrogram labeling procedure. We supplement these labels with a combination of analyst-derived and automated spectrogram generation detailed below. Our preference would be to have equal amounts of robust, labeled signals for each class from each volcano. However, this is not feasible due to the uniqueness of each volcano around the world and the type of signals commonly observed. For example, some volcanoes like Great Sitkin produce extensive VTs but relatively few instances of tremor or explosions.

We integrate the substantial number of spectrogram labels from Semisopochnoi volcano, Alaska, from Lyons et al. [2025]. Semisopochnoi experienced unrest and eruption intermittently between 2018–2022, and this long-lived phreatomagmatic eruption produced a wide variety of seismoacoustic signals recorded on a local monitoring network (Figure 1A). Lyons et al. [2025] labeled seismic data into the same seven volcano seismicity classes between 0.5–20 Hz on the east channel to create 97,456 4-minute spectrogram labels. They used a novel implementation of the `REDPy` repeating event detector [Hotovec-Ellis and Jeffries 2016] to detect and catalog explosions, and, similar to T24, they used infrasound location algorithms [Fee et al. 2021] to confirm `REDPy` infrasound explosion detections and guide labeling of explosions in seismic data. Here we incorporate the Lyons et al. [2025] labels but trim the spectrograms to 2 minutes, 0.5–10 Hz, and use the vertical channel to ensure consistency with our other labels.

The labeled LP class consists of 5359 spectrograms from Semisopochnoi, 3869 from Mount Veniaminof, 452 from Mt. Etna, and 298 more from Pavlof Volcano (Figure 3A). All but the Mt. Etna LPs were hand-labeled by an analyst on spectrograms. For Mt. Etna, we use LP events detected using the methodology of Cannata et al. [2013] that had high signal-noise ratio (SNR) at stations EMFO and ECPN, and select a random two minute window surrounding the LP onset. We enforce the start time to be at least 20 s from the end of the spectrogram to fully capture the waveform coda. The Mount Veniaminof LPs were labeled in data from September 2018.

We construct the VT earthquake spectrogram labels from a diverse set of volcanoes and stations. Since we have robust, validated seismic catalogs [e.g. Power et al. 2019], we use these to help build our spectrogram labels rather than hand-labeling spectrograms. The earthquake class in T24 primarily consisted of regional seismicity from the nearby M8.2 Chignik earthquake and aftershocks. For this study we are

interested in training the model to detect spectral characteristics of VT earthquakes occurring near the volcano, so we disregard the T24 earthquake labels. Additionally, we discard the earthquake labels from Lyons et al. [2025], as we found many of them to contain somewhat atypical spectra, perhaps representing hybrid seismic events [Lahr et al. 1994]. Instead we use data from three volcano seismic networks and time periods (Supplementary Material 1 Figure S1): three short-period stations (REF, RDN, RSO) near Redoubt Volcano, Alaska, between January 1–March 22, 2009; five broadband stations (UWE, HAT, WRM, RIMD, PUHI) near Kīlauea Volcano, Hawai'i, between January 1–December 31, 2020; and two broadband stations (TANO, TASE) near Tanaga and Takawangha (TT) Volcanoes, Alaska, between November 18, 2022–July 18, 2023. We use earthquakes with depths between 0–30 km, within a 15 km radius from the volcano (20 km from the mean location of Tanaga and Takawangha Volcanoes), and magnitudes above M_l 0.0 for Redoubt Volcano and Kīlauea and M_l 1.0 for Tanaga and Takawangha Volcanoes. Similar to LPs for Mt. Etna, we randomize the start time of the 2-minute spectrogram window and leave at least 20 s after the origin time to encompass the earthquake. To ensure sufficient signal-noise, we compute the Short-Term Average/Long-Term Average (STA/LTA) on filtered data between 0.8–5 Hz using values of 1 and 10 s, respectively. We only keep earthquakes with STA/LTA above 4, which produce relatively high signal-noise ratio (SNR) earthquake spectra. Many of our selected earthquakes occur within two minutes of each other. In this situation we keep the entire overlapping window but do not create unique spectrograms for each event. Multiple earthquakes in a 2-minute window is a relatively common occurrence at volcanoes [McNutt and Roman 2015] so we want the ML model to be able to detect these in other data. This process results in 13,218 total Earthquake spectrogram labels, including 3477 for Redoubt Volcano, 4321 for Kīlauea, and 5420 for Tanaga and Takawangha Volcanoes (Figure 3A). These Earthquake labels are by no means comprehensive, but the variable seismicity rates, magnitudes, depths, source-receiver distances, instrumentation, etc. provide a reasonable sample of VT seismicity recorded at volcanoes worldwide (Figure 2). The earthquake catalogs were produced by the Alaska Volcano Observatory (AVO) [Power et al. 2019] and Hawaiian Volcano Observatory (HVO) as part of the Advanced National Seismic System (ANSS) Comprehensive Earthquake Catalog (ComCat) [U.S. Geological Survey 2017]. The earthquakes were all hand-picked and located by experienced analysts on 3 component seismometers. We exclude traditional LP events identified by the analyst. Notably, the three Redoubt Volcano stations are all short-period stations while the others are all broadband.

T24 had relatively sparse numbers of labeled Explosions (846) and Harmonic Tremor (886) spectrograms. This resulted in relatively small test and validation sets (169 per class) (see T24 Supplementary Material 1 Figure S3) which can bias the ML model to a small number of examples and result in overfitting. We therefore supplement both of these classes. The addition of spectrogram labels from Semisopochnoi from Lyons et al. [2025] added 2818 explosions. In total we have 3641 explosion spectrograms from Pavlof Volcano and Semisopochnoi (Figure 3A). We sought out additional examples of harmonic tremor from previous studies. We selected Mount Veniaminof, Alaska [De Angelis and McNutt 2007; Cameron et al. 2023] and Popocatepetl, Mexico [Roman 2017], which both contained clear, high SNR harmonic tremor in data that are publicly available. We hand-labeled 6 days of harmonic tremor and LPs from 7 stations (VNSS, VNWF, VNFG, VNHG, VNNF, VNKR, VNSW) at Mount Veniaminof from September 2018, resulting in 1602 Harmonic Tremor labels. We also labeled 3 months of harmonic tremor from 4 stations (ZACA, NEXP, HSHT, BIBI) near Popocatepetl, guided by the harmonic tremor catalog of Roman [2017], to produce an additional 1825 Harmonic Tremor labels (Figure 2, 3). This labeling, along with 425 labels from Semisopochnoi and 1736 from Pavlof Volcano after splitting the spectrogram duration to 2 minutes, resulted in 5280 Harmonic Tremor labels.

Broadband Tremor, Monochromatic Tremor, and the Noise class labels all come from Pavlof Volcano and Semisopochnoi. The respective class counts are 51,828, 9239, and 180,274 (Figure 3).

2.2 Methods

2.2.1 *Model architecture and training*

Our ML model architecture and training closely follows that of T24. The architecture is fairly standard for a CNN and consists of three convolution and max-pooling layer pairs, a flattening operation, and three dense layers, the last of which is a softmax classification layer (Figure 4). Since we choose to use 2-minute-long spectrograms, each spectrogram has a shape of 94×120 pixels. The convolutional layers have 32, 64, and 128 filters and a 3×3 kernel size and 1×1 kernel stride. We implement dropout layers during training [Srivastava et al. 2014] at a rate of 0.2 for the flattened layer and 0.5 after each dense layer to reduce overfitting. The model has 298,119 total and trainable parameters, which is within the same order of magnitude as our number of labeled spectrogram samples.

Prior to model training, we balance our training and validation classes as there is moderate imbalance within our training dataset (Figure 3B), with many more Broadband Tremor and Noise samples than the other classes. We train our model using a 60/20/20 training, test, and validation split. The Explosion class has the lowest number of samples (3641), therefore we randomly select 728 (20 % of 3641) from each seismicity class to obtain balanced test and validations sets. This procedure results in 5096 total samples for the balanced test and validation sets respectively, in contrast to 1014 total samples in T24 (169 per class). Next, similar to T24, we augment spectrograms in the Harmonic Tremor, Monochromatic Tremor, Long Period, and Explosion classes to correct for imbalance. We do so by summing 80 % and 20 % of the spectrogram amplitudes from random training and noise spectrograms, respectively, in an element-wise manner. This is presumably adding in random but real noise. T24 augmented with 35 % noise. The augmentation is performed until each class size meets the Earthquake class size of 13,218 (Figure 3B, dashed horizontal line). We initiate training using the Categorical Cross Entropy loss function [Goodfellow et al. 2016] in Tensor-

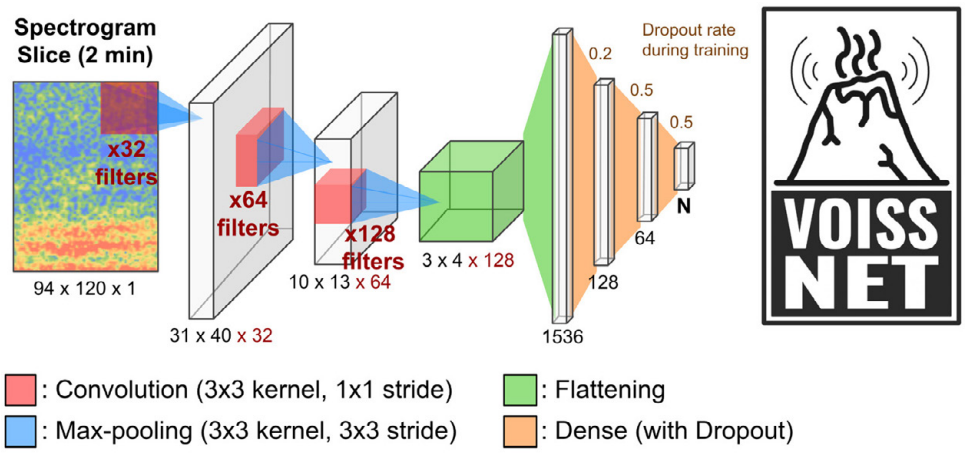

Figure 4: Generalized `VOISS-Net` architecture. Two minute spectrograms are input into a 3 layer Convolutional Neural Network with max-pooling and flattening layers. Dropout is only implemented during training.

flow [Abadi et al. 2016]. We use the Adam optimizer [Kingma and Ba 2017] with a learning rate of 0.0002. We explored different learning rates but did not find noticeable improvements in model performance. During training, spectrograms are input into the model in batches of 100 and we implement early stopping after 20 epochs without a decrease in validation loss (Figure 5A). A notable difference in our training compared to T24 is that we do not standardize each spectrogram with a running mean and variance [Ioffe and Szegedy 2015] prior to min-max scaling. We view spectrogram standardization as a strategy best reserved for models dedicated to a single volcanic network, as it removes site-specific spectral features prior to classification. Our generalized model is transferred between numerous volcanic networks, and we find that standardization impedes model performance instead. We therefore opt to only min-max scale each spectrogram's amplitude values to 0 to 1 prior to forward feeding, similar to Maguire et al. [2024]. Table 1 lists our model hyperparameters in comparison to T24. We trained our model using an NVIDIA L40S GPU card, and it took approximately 33 minutes to converge.

Our model converged with a test set accuracy of 87 % after 39 epochs. Precision, recall, and F1 scores are all similar. Figure 5 shows the validation accuracy and loss curves, as well as the confusion matrix on the test set. The validation accuracy and loss curves show a gradual, smooth increase and decrease, respectively, with each additional epoch, suggesting that we are not over- or under-fitting the data significantly. The confusion matrix displays the true label (y-axis) versus the predicted label (x-axis) ratio for each class within the test set. For a perfect classification, the diagonal values would all be at 1.0 (100 %). Our confusion matrix shows relatively high classification accuracies for all classes but with a somewhat bimodal distribution. Examining the diagonal of Figure 5C we see the Harmonic Tremor (91 %), Monochromatic Tremor (92 %), Earthquake (97 %), and Noise (92 %) classes have high prediction accuracies, while the Broadband Tremor (78 %), LP (83 %), and Explosion (75 %) classes have lower prediction accuracies. Supplementary Material 1 Table S1 presents the precision, recall, and accuracy for each class. Notably, Broadband Tremor has a low precision (0.76). This suggests it has a higher false positive rate, perhaps related to other classes (e.g. Noise) with similar characteristics. The Explosion class also has a low recall of 0.75 we discuss in Section 4.1.

We evaluated model convergence and performance by adding in 1 and 2 additional convolutional layers, changing the kernel size to 2×2 and 2×3, implementing global average pooling rather than flattening, and exploring different dropout rates. None of these changes produced notable improvements in convergence, accuracy, or qualitative classification improvements when the subsequent model was applied to real data. As expected, we found that the most notable changes in model performance are related to the spectrogram labels, and hence that is where we focus much of our efforts (Section 2.1).

2.2.2 *Model implementation*

Once trained, we can apply the `VOISS-Net` model to seismic data from any number of stations at a volcano. We nominally read in data from an International Federation of Digital Seismograph Networks (FDSN) web service client. The data for each station is classified independently by first being split into 2-minute windows with a specified overlap (50 % in this study). We compute the spectrogram in the same manner as for the labels in Section 2.1. `VOISS-Net` then predicts one of the seven classes for each station's spectrogram for each time step. Each prediction has an associated softmax probability between 0–1, which provides a measure of confidence in the classification. Similar to T24, we then use a network "voting" scheme that 1) computes the average probabilities for each class for the "network" of stations and then 2) chooses the class with the highest average probability as the "network vote". We have found the network vote process often removes station outliers and poor classifications. The network-averaged probability, or P_{norm}, is used as a confidence and threshold metric. In this study, we discard the network classification if P_{norm} is below 0.4 for that time segment, which often corresponds to disagreement across network stations or model unfamiliarity with the observed signal. We choose a threshold of 0.4 based on manual testing and evaluation of the results,

Table 1: Comparison of hyperparameters between our generalized model and T24. Differences between the hyperparameters are highlighted in bold.

Hyperparameter	Generalized Model	T24
Spectrogram duration	**120 s**	240 s
Image size	**94 × 120**	94 × 240
Batch size	100	100
Patience	20	20
Number of convolutional layers	3	3
Kernel Size	3 × 3	3 × 3
Stride	1 × 1	1 × 1
Activation function	ReLU	ReLU
Learning Rate	**0.0002**	0.0005
Dropout	0.2 (flattened), 0.5 (dense)	0.2 (flattened), 0.5 (dense)
Optimizer	Adam	Adam
Image standardization	**Normalized**	Normalized and removed running mean and variance

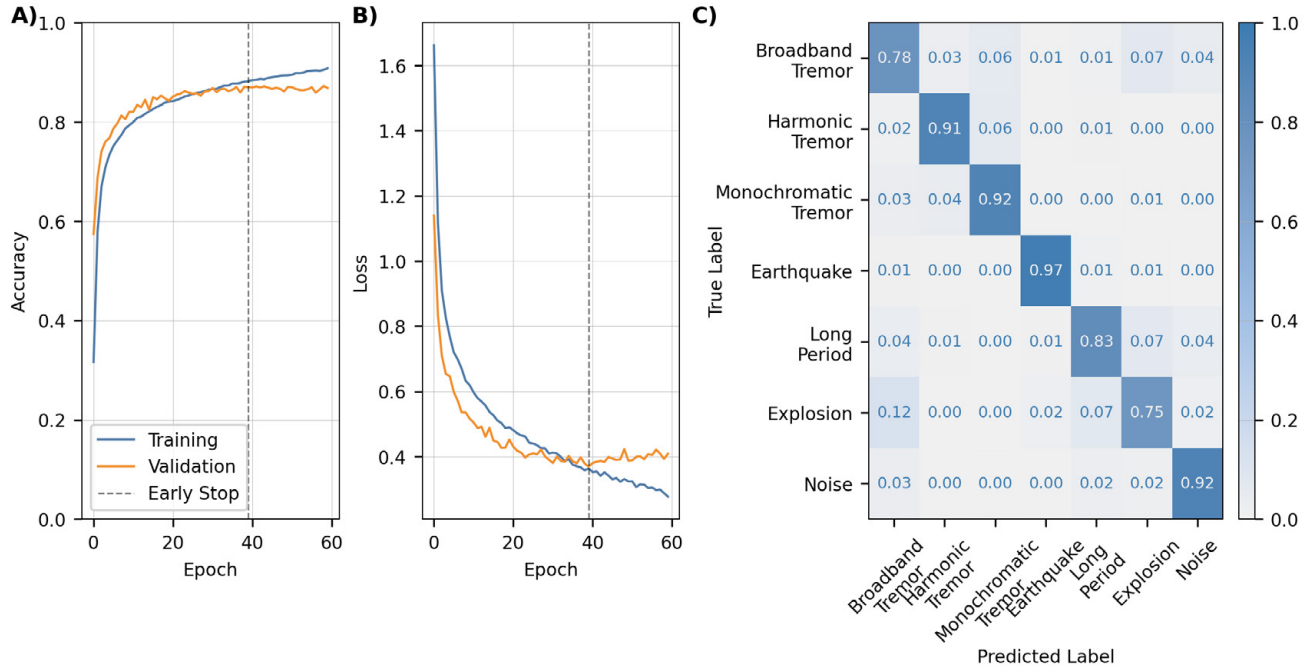

Figure 5: Model training results. [A] Validation accuracy and loss curves and [B] test set confusion matrix. The accuracy and loss curves display gradual, smooth increases and decreases, respectively, with each additional epoch. Early stopping is indicated by the dashed vertical line. [C] The confusion matrix shows the true versus predicted label for the test set. Diagonals indicate the true positive ratio. Overall the model does well at classifying all the classes, with lower predication accuracy for the Broadband Tremor, Long Period, and Explosion classes.

but acknowledge that this threshold will depend on the application and user preference. This process is then repeated for each of the subsequent overlapping time windows. The model can be applied to real data relatively quickly. For example, 4 hours of data from 5 stations split into 2-minute windows with 50 % overlap can be classified in approximately 15 s using a standard laptop computer. Reading in the seismic data from an FDSN server is included in that time estimate, and often takes longer than ML processing.

Similar to T24, we compute reduced displacement, or D_R [Aki and Koyanagi 1981], for each time period analyzed by `VOISS-Net`. Like RSAM [Endo and Murray 1991], D_R provides a sense of seismic amplitude over time, but is by definition comparable from one volcano to another regardless of source-station distance. Reduced displacement is computed using a single reference station at each volcano and in 2-minute windows with 50 % overlap to match the `VOISS-Net` classifications. We assume that the signal is primarily composed of

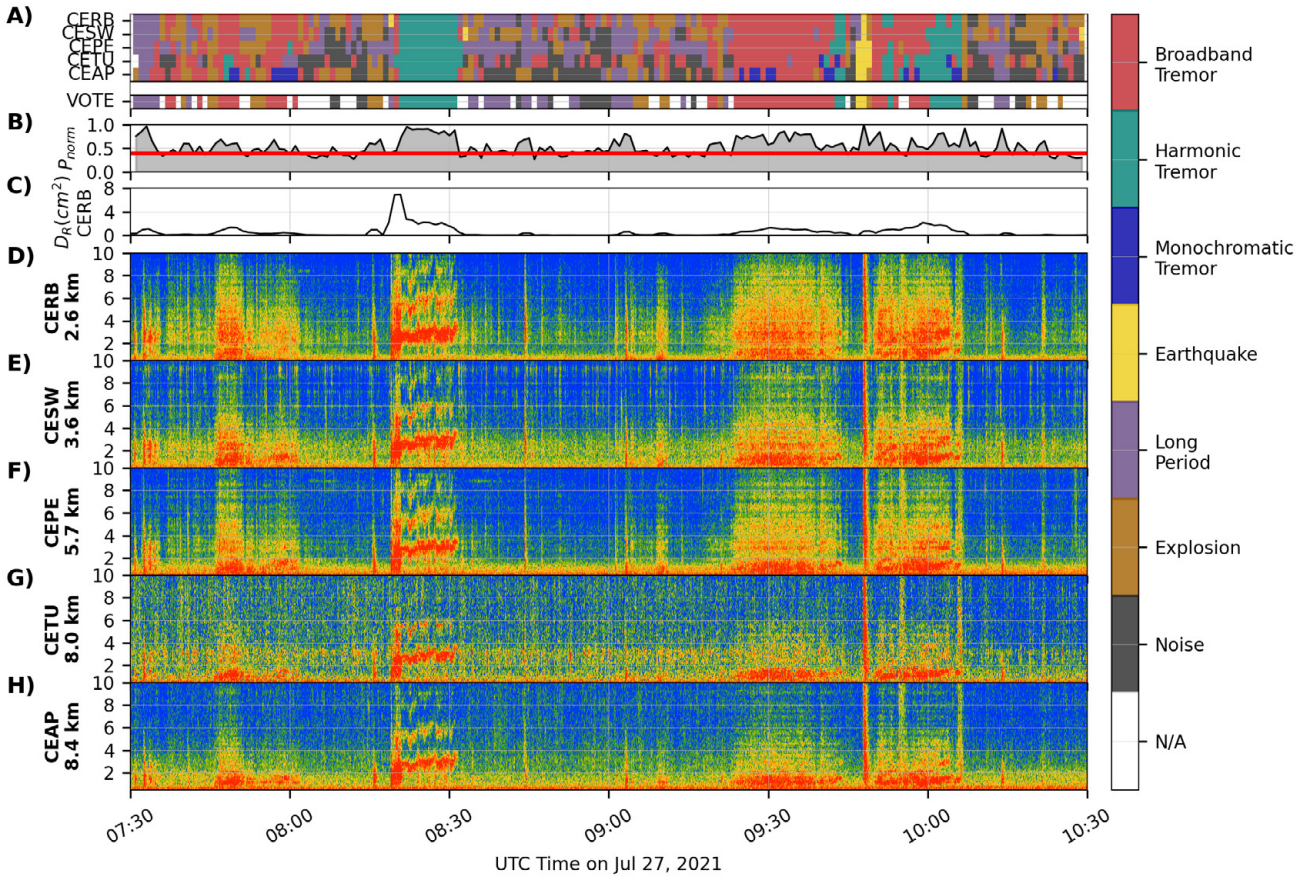

Figure 6: Generalized `VOISS-Net` model applied to Semisopochnoi volcano seismic data between July 27, 2021 07:30–10:30. [A] Single-station and network classifications colored by seismicity class. [B] Network vote probability (P_{norm}), with our threshold of 0.4 indicated by the red line. [C] Reduced displacement for the reference station CERB. D-H) Spectrograms for the five selected stations, with their distance to the vent listed under the station name on the right. This diverse sequence features multiple types of tremor and transient events. `VOISS-Net` classifications generally agree with our interpretations.

surface waves at a dominant frequency of 2 Hz and that the source is at the eruptive vent indicated by the red triangle in Figure 1.

We typically show hours to days-long `VOISS-Net` classifications in a "timeline" fashion, such as the 3-hour-long example from Semisopochnoi in Figure 6. The top panel shows the `VOISS-Net` prediction as a function of time for each station, with the seismicity class indicated by a specific color. Below this is the network vote panel, followed by P_{norm} and D_R. Network votes with a $P_{norm} < 0.4$ are discarded. The spectrograms for the selected stations are shown in the lower panels.

3 Results

We apply the generalized `VOISS-Net` model to six different volcanoes to analyze its effectiveness and limitations on a variety of different types of seismicity. The six volcanoes and associated time periods we choose are: Semisopochnoi volcano, Alaska, in July 2021; Trident Volcano, Alaska, in December 2023; Kīlauea Volcano, Hawai'i, in June 2023; Great Sitkin Volcano, Alaska, in May 2021; Mt. Etna Volcano, Italy, in December 2018; and Pavlof Volcano, Alaska, in 2021–22 (Figure 1). These volcanoes and time periods provide a mix of eruptive and non-eruptive seismicity with different styles and durations of unrest and seismicity. We acknowledge most of these volcanoes are in Alaska, but choose them since this is where we have the most experience applying `VOISS-Net` and also because of the varied, extensive volcano seismicity in the region. Additionally, the data we choose are open source and accessible. Further details on the volcanoes and associated seismicity are given below. Note we only provide brief details of the associated volcanic activity for each of the periods as our focus here is on the `VOISS-Net` classifications and not on the volcano itself. All times listed are in UTC.

3.1 Semisopochnoi, Alaska, July 2021

First, we apply `VOISS-Net` to seismic data from Semisopochnoi volcano, Alaska. Semisopochnoi is a remote volcano in the western Aleutian Islands (Figure 1A). It most recently erupted between 2018–2023 and the activity consisted of intermittent phreatomagmatic eruptions from the north cone of Mt. Young (red triangle in Figure 1A) that produced mostly low-level ash and gas plumes. The seismicity from this eruption typically consisted of various

types of tremor, LP events, and explosions. Semisopochnoi had few VT earthquakes during this eruption period [Lyons et al. 2025].

Figure 6 shows the VOISS-Net classifications and spectrograms between July 27, 2021 07:30–10:30 for five Semisopochnoi stations (CERB, CESW, CEPE, CETU, CEAP) (Figure 1A). The spectrograms during this period show a wide range of spectral signatures that are generally consistent with the VOISS-Net classifications. VOISS-Net detects LP signals from 7:30–7:35, which are apparent as multiple transient low-frequency pulses in the spectrograms. The next ~45 minutes have a period of low P_{norm} classifications of Broadband Tremor, Explosions, and an LP, consistent with the spectrograms that have energy concentrated below 5 Hz and occasional weak broadband transients. An energetic transient around 08:16 is classified as an Explosion. This low-amplitude event was only weakly recorded on the local infrasound network and not in the explosion catalog of Lyons et al. [2025]. A few minutes later an approximately 12 minutes long, energetic sequence of Harmonic Tremor is detected well across all stations and shows up clearly as multiple spectral lines in the spectrograms. This period also has the highest amplitudes with D_R up to 7 cm^2. Intermittent Noise, LPs, and weak explosive events are then detected with low P_{norm} until ~09:20 when a more energetic Broadband Tremor pulse is detected until 09:43. An impulsive, broadband signal at 09:48 is classified as an earthquake. Another tremor pulse starts at 09:50 and is classified as a mix of Broadband and Harmonic tremor. Weak LPs, Explosions, and a few periods of Noise are then detected.

The five stations selected for this example all show generally similar spectral characteristics, which is consistent with their individual VOISS-Net station classifications generally being in agreement (Figure 6A). We suggest this is likely related to the overall high SNR. The different station distances to the active vent (2.6–8.4 km) and associated path and site effects likely account for spectral differences.

3.2 Trident Volcano, Alaska, December 2023

We apply VOISS-Net to a sequence of non-eruptive, but somewhat difficult-to-classify seismicity near Trident Volcano, Alaska. Trident Volcano is a stratovolcano in the Katmai volcanic cluster, (Figure 1B). Its most recent eruption was a long-lasting, intermittent eruption between 1953–1974 that consisted of lava effusion and sporadic explosive activity [Hildreth et al. 2001]. Elevated seismic unrest began at Trident Volcano in August 2022 and lasted for ~1.5 years. The seismicity during this period consisted of VT earthquakes, mostly at shallow depths, along with periods of deep event sequences with mixed frequency content and durations of tens of seconds to minutes [Orr et al. 2025]. Occasional "tornillos"", or discrete events with monochromatic frequency content and a decaying coda, are known to occur near Trident Volcano as well [De Angelis 2006].

Here we apply VOISS-Net to four stations near Trident Volcano (KAKN, KBM, KABU, KVT) (Figure 1B) from December 4, 2023 04:45–07:15 (Figure 7). The stations are relatively far from Trident Volcano at 7.3–19.9 km distance. During this period there are multiple types of local and teleseismic seismicity, some of which are difficult to quantify. The majority of the time period is classified as Noise across all stations, which is consistent with the general level of seismicity (D_R < 0.2 cm^2) and the fact that this volcano is not erupting or in a state of heightened unrest. Two sets of broadband/mixed frequency events are apparent from ~05:05–05:12 and 05:48–05:55. VOISS-Net classifies the first sequence as a mix of an LPs, an Explosion, and Broadband Tremor, and the second as a series of LPs and an Earthquake. No Explosions or subaerial activity occurred at this time, but otherwise these classifications are consistent with our interpretations of the spectrograms. AVO seismic analysts were able to locate a couple events within these sequences and they had deep hypocenters between 30–35 km depth and weak S waves. A low-amplitude tornillo occurs at 06:26 and is classified as Monochromatic Tremor. These events are typically difficult to automatically detect. VOISS-Net also detects multiple other Earthquakes and LP signals between 06:55–07:15. These also generally agree with our interpretation and are consistent with AVO seismic analyst picks. Note multiple low frequency teleseismic events are incorrectly classified as LPs and Explosions, such as the ones around 05:20 and 06:22. Teleseismic events are not included in our training dataset, which is focused on volcanic seismicity.

3.3 Kīlauea Volcano, Hawai'i, June 2023

We apply VOISS-Net to the build-up and subsequent eruption of Kīlauea, Hawai'i, on June 7, 2023. Kīlauea is a well-monitored, frequently active shield volcano (Figure 1C). In recent years it has had eruptions from both the summit region and East Rift Zone. The June 2023 sequence consisted of elevated VT seismicity and ground deformation near the summit, and eventual eruption of lava from multiple fountains within the Halema'uma'u crater in Kīlauea's summit caldera that commenced at 14:44. The seismicity from recent Kīlauea summit eruptions typically consists of pre-eruptive VT swarms followed by sustained tremor.

Figure 8 shows the VOISS-Net classifications and spectrograms between June 7, 2023 14:00–17:00 for five broadband seismic stations around Kīlauea's summit caldera (RIMD, WRM, UWE, HAT, PUHI) (Figure 1C). The stations are between 1.6–4.1 km from Halema'uma'u crater. The seismicity between 14:00 to around 14:52 is classified as earthquakes with high P_{norm} and D_R, which is generally consistent with the spectrograms and HVO earthquake catalog. Clear broadband transient signals are apparent in the spectrograms during this period, with some of the events having longer duration codas. VOISS-Net classifications change to a mix of Broadband Tremor and Earthquakes from 14:52–15:15. Notably, the eruption onset at 14:44 (red arrow in Figure 8A) is not clearly detected by a change in VOISS-Net classifications despite broadband, sustained energy being visible in the spectrograms below ~4 Hz. This is likely related to earthquakes and tremor occurring concurrently during this period and overlapping in the spectrograms, with VOISS-Net selecting the Earthquake class or returning a low P_{norm} due to mixed classifications. The first network-wide Broadband Tremor classifica-

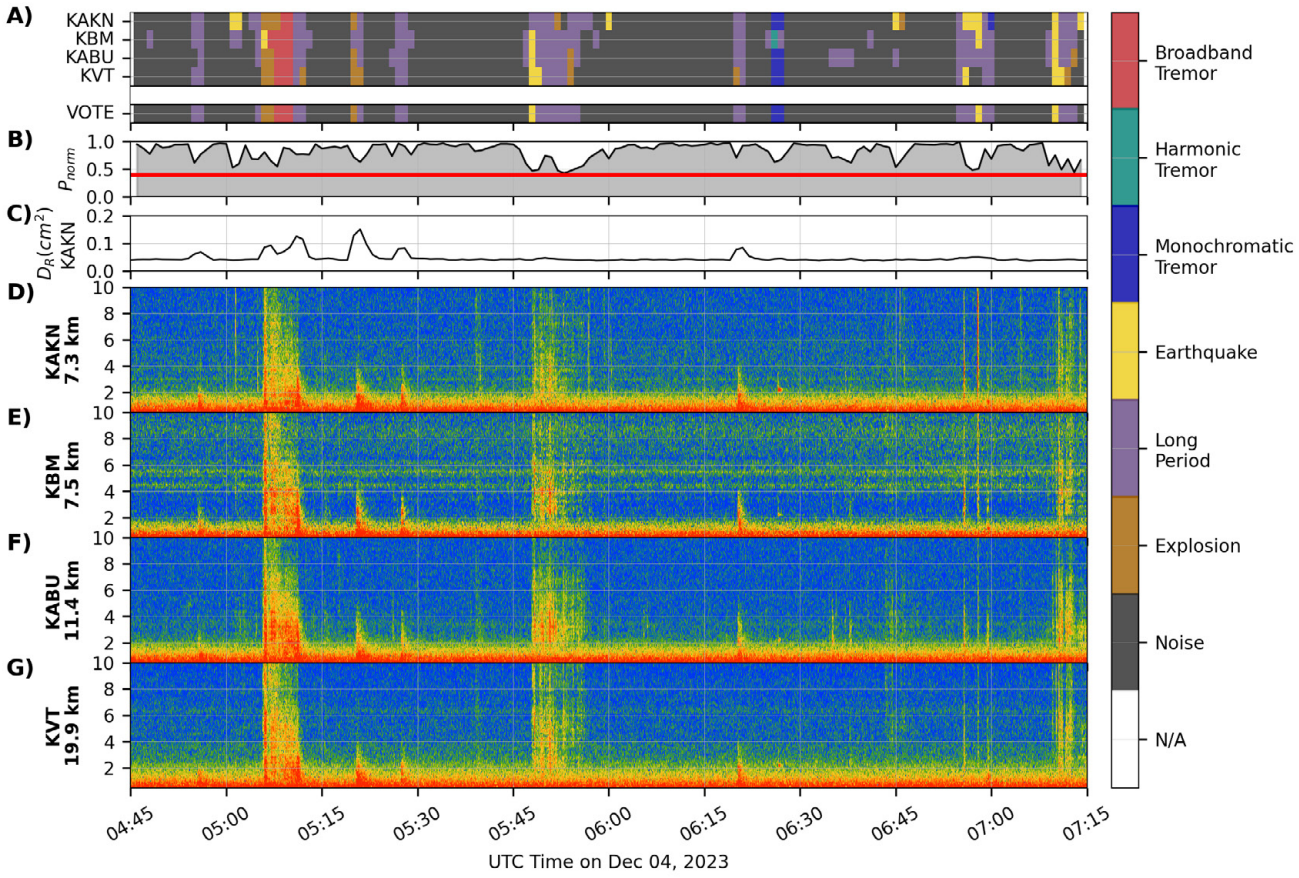

Figure 7: Generalized `VOISS-Net` model applied to Trident Volcano seismic data between December 4, 2023 04:45–07:15. Figure layout is the same as Figure 6. Spectrograms for four stations are shown for this period with a mix of VTs, regional earthquakes, a tornillo, and two episodes of mixed frequency events.

tion is at 14:54, and this is where tremor begins to dominate in the spectrograms. From 15:15–17:00 the seismicity is primarily classified as Broadband Tremor with a couple Earthquakes and LPs. P_{norm} is relatively low during this period due to some mixed classifications between stations. There is some weak harmonic structure in some of the stations, particularly WRM and UWE from ~15:45–16:30. Overall the transition from VT earthquakes to broadband tremor is captured fairly well by `VOISS-Net`, although the tremor onset is masked by some coincident earthquakes. We performed a similar analysis with some other Kīlauea stations and found no major changes in results. `VOISS-Net` also preformed well for other recent Kīlauea summit eruptions with similar characteristics.

3.4 Great Sitkin Volcano, Alaska, May 2021

Here we apply `VOISS-Net` to the build-up and subsequent eruption of Great Sitkin Volcano, Alaska, on May 26, 2021. Great Sitkin is an andesitic composite stratovolcano in the Aleutian Islands that has erupted eight times since 1792 (Figure 1D). It is monitored by AVO with a five station local seismic network and one infrasound station. Earthquakes rates began to increase above background at Great Sitkin in early May 2021. Elevated SO_2 and summit temperatures were also observed during this time [Orr et al. 2024]. In the 24 hours leading up to the eruption, Power and Roman [2024] note that the earthquake frequency content is relatively low and the events have poorly developed S-waves, and they classify the events as LPs. We note that the spectral content of most of these events more closely resembles our earthquake than LP class, as they have relatively impulsive, broadband onsets (Figure 9). The earthquake rate increased until 05:03, when a single vulcanian explosion occurred at the summit that produced notable infrasound, ballistics, and an ash plume to 4.5 km above sea level [Orr et al. 2024].

`VOISS-Net` generally detects and classifies the precursory seismicity, explosion, and return to background of the May 26, 2021 Great Sitkin eruption. Figure 9 shows the `VOISS-Net` classifications and spectrograms between May 26, 2021 04:00–07:00 for five stations around Great Sitkin (GSTD, GSTR, GSSP, GSMY, GSCK) (Figure 1D). The stations range from 3.3–7.8 km from the eruptive vent. The time period from 04:00-05:00 is classified primarily as Earthquakes across all stations, with D_R being relatively low. We note that these Earthquake classifications from `VOISS-Net` are in contrast to the LP classifications of Power and Roman [2024], and we discuss this disagreement in Section 4.1. A few LP and 4 Explosion classifications are also returned. These Explosions are erroneous classifications and all have low P_{norm} of <0.5, and

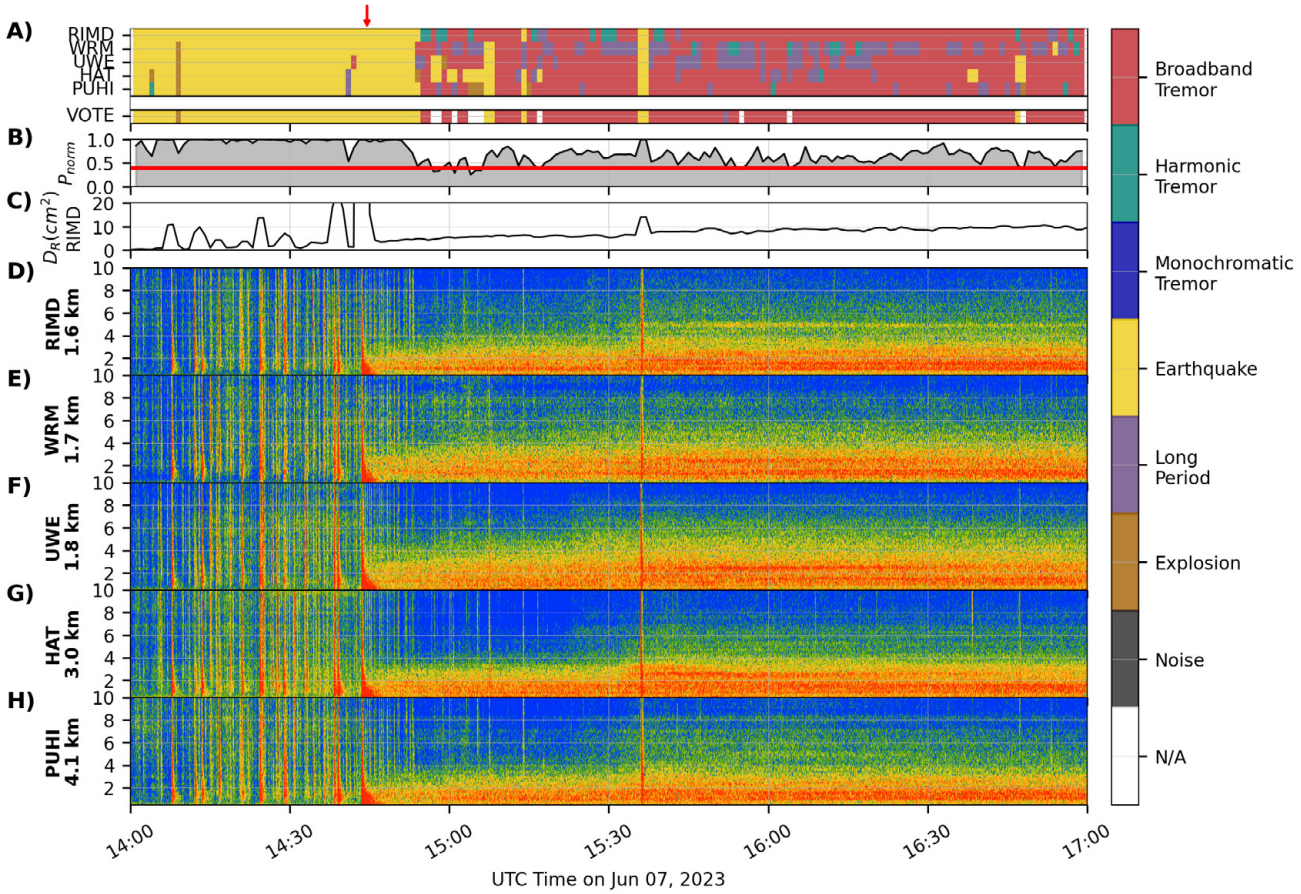

Figure 8: Generalized `VOISS-Net` model applied to Kīlauea seismic data between June 7, 2023 14:00–17:00. Figure layout is the same as Figure 6. The red arrow at the top indicates the eruption onset at 14:44. Many of the earthquakes have very high amplitudes, so the D_R panel in [C] is clipped at 20 cm^2 to permit viewing of lower level tremor and impulsive earthquakes.

may be related to earthquakes with particularly sharp onsets. A high P_{norm} Explosion classification is returned from 05:03-05:05, consistent with visual and infrasound observations of an explosive event [Orr et al. 2024] (red arrow in Figure 9A). This event has the highest D_R as well at over 30 cm^2. After the Explosion, a couple LP classifications are returned, but the majority of the seismic data are classified as Noise, which is also consistent with reported activity.

3.5 Mt. Etna Volcano, Italy, December 2018

We next examine how `VOISS-Net` characterizes an energetic fissure eruption at Mt. Etna Volcano, Italy, that began on December 24, 2018. Mt. Etna is a well-studied, large, frequently active basaltic volcano that produced a range of eruption styles and seismicity [Patanè et al. 2013; Andronico et al. 2021] (Figure 1E). Eruptions occur within both a central crater region and also frequently as flank eruptions. The volcano is well-monitored by the National Institute of Geophysics and Volcanology (INGV) with 16 seismometers, in addition to a number of other multi-parameter sensors. On December 24, 2018, VT earthquake activity at Mt. Etna increased around 08:30, the result of dike intrusion [Cannavo' et al. 2019]. A fissure eruption began at 11:11 and eruptive vents propagated SSE over the following ~25 minutes. The eruption produced intense tremor and multiple large (M_l 4+) earthquakes. Eruptive activity consisted of lava fountaining, an ash-rich plume, and lava flows [Cannavo' et al. 2019].

Applying `VOISS-Net` to 36 hours of data around the Mt. Etna fissure eruption shows the model performs well at classifying pre- and syn-eruptive seismicity. Figure 10 displays `VOISS-Net` classifications and spectrograms between December 24, 2018 00:00–December 25, 12:00 for three stations on Mt. Etna (EMCN, ESLN, EMSG) (Figure 1E). These stations are at distances of 5.3–8.5 km from the summit. Note many more seismometers are deployed on Mt. Etna but these three stations are broadband and publicly available. `VOISS-Net` predicts a mix of LPs and Broadband Tremor between 00:00–08:30 on December 24. D_R values are low during this period. At ~08:30 the Earthquake class is classified near-continuously until ~11:11 when a mix of classes are predicted between the stations, resulting in generally low P_{norm} values until ~15:00. D_R increases to high levels during this period with a mix of transient and sustained increases. From December 24 ~15:00–December 25 ~03:00 `VOISS-Net` predominantly predicts Earthquakes, as these are dominant on stations ESLN and EMCN. Occasional Broadband Tremor classifications are made on EMSG where it appears strongest, likely related to its greater distance from the earthquake hypocenters. Network-

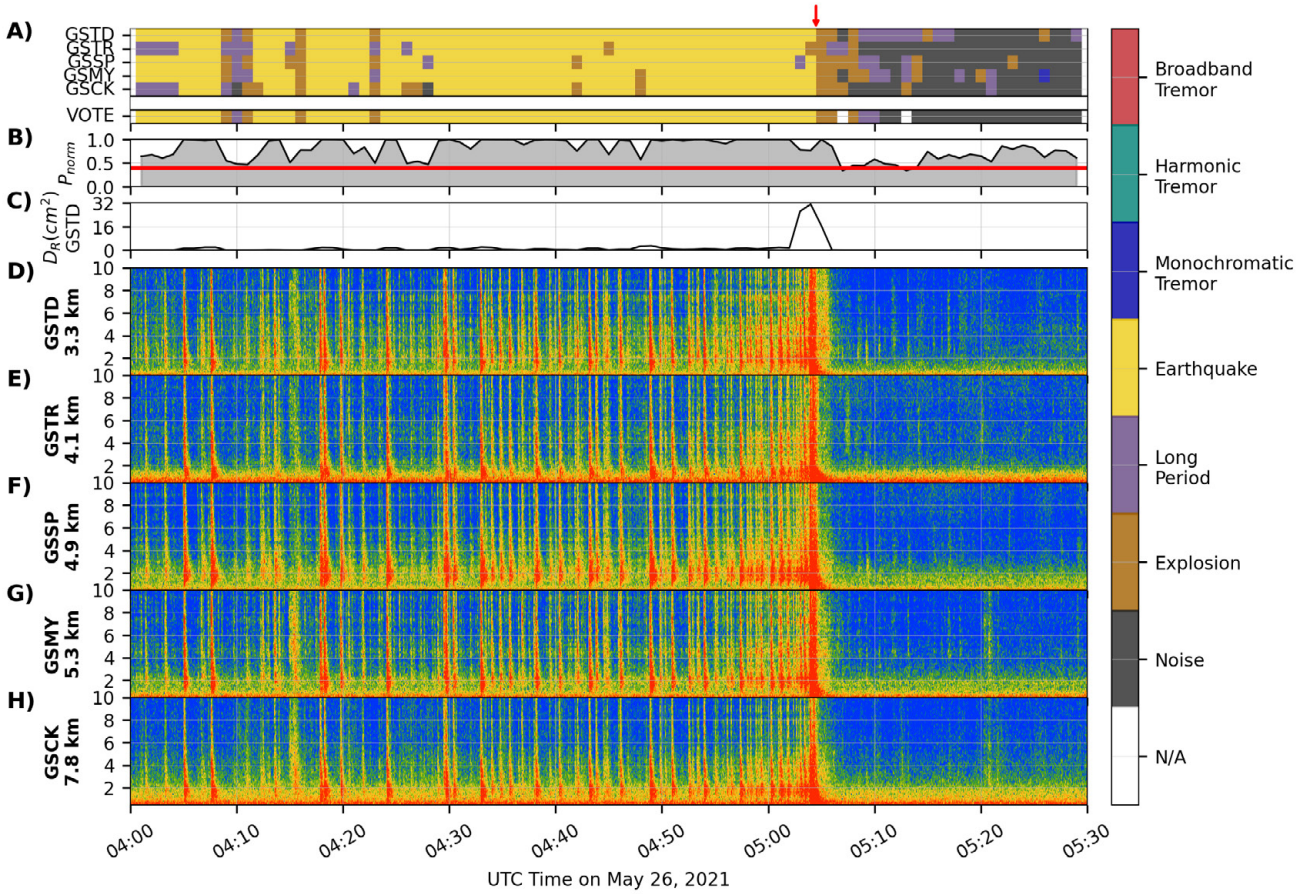

Figure 9: Generalized `VOISS-Net` model applied to Great Sitkin volcano seismic data between May 26, 2021 04:00-05:30. Figure layout is the same as Figure 6. The red arrow at the top indicates the eruption at 05:03.

wide Broadband Tremor classifications become more common from December 25 03:00–12:00, with tremor strongest again on EMSG.

The `VOISS-Net` classifications up until ~11:11 agree well with our visual spectrogram interpretation and that reported in Cannavo' et al. [2019]. However, the onset of energetic tremor around 11:11 does not coincide with a clear tremor prediction by `VOISS-Net`. Similar to the Kīlauea case study, this is due to the multiple types of seismic signals (e.g. tremor, VTs, etc.) occurring near-simultaneously. The complex mix of tremor and VT earthquakes over the next ~24 hours results in disagreement in individual station classification, but is overall consistent with the spectrogram characteristics.

3.6 Pavlof Volcano, Alaska, 2021–2022

Here we analyze how the generalized `VOISS-Net` model performs over an ~2 year period at Pavlof Volcano, and compare our results with that from T24. Pavlof Volcano is a frequently active volcano that has a range of eruption styles, including low-level explosions, lava fountains and occasional lava flows, as well as energetic subplinian eruptions [Waythomas et al. 2014; Fee et al. 2017; Waythomas et al. 2017]. Our analysis period encompasses an ~1.5 year low-level eruption at Pavlof Volcano. The eruptive activity during this period consisted primarily of weak explosions producing small ash plumes and fountain-fed lava flows from a vent on the upper southeast flank of the volcano [Tan et al. 2024]. Extensive, variable seismic signals were recorded and detailed in T24.

The generalized `VOISS-Net` model provides an improved seismic catalog for the 2021–22 Pavlof Volcano eruption. Figure 11 shows the AVO Aviation Color Code, spectrogram for station PS1A, and binned `VOISS-Net` classifications for the generalized and T24 models between January 1, 2021–March 1, 2023. Five stations (PVV, PN7A, PS4A, PV6A, PS1A) at distances of 6.8–11.1 km are used (Figure 1E). The GREEN AVO color code corresponds to background activity, YELLOW to elevated unrest, and ORANGE corresponds to eruption. The spectrogram is computed with a 500 s window length to capture broad spectral features over this long time period. The `VOISS-Net` 2 minute classifications are binned into 1 hour intervals and separated into class-specific rows in Figure 11C and 11D. The plotted bin opacity is defined by the ratio of the class occurrence relative to the total number of time steps per hour. The most striking difference between the performance of the two models is the relative lack of earthquakes detected in the generalized model. This is consistent with the lack of VT seismicity at Pavlof Volcano for this eruption, and is likely due to the different Earthquake class labels between the generalized and T24 models. Many more events are classified as LPs in the generalized model, which is broadly consistent with

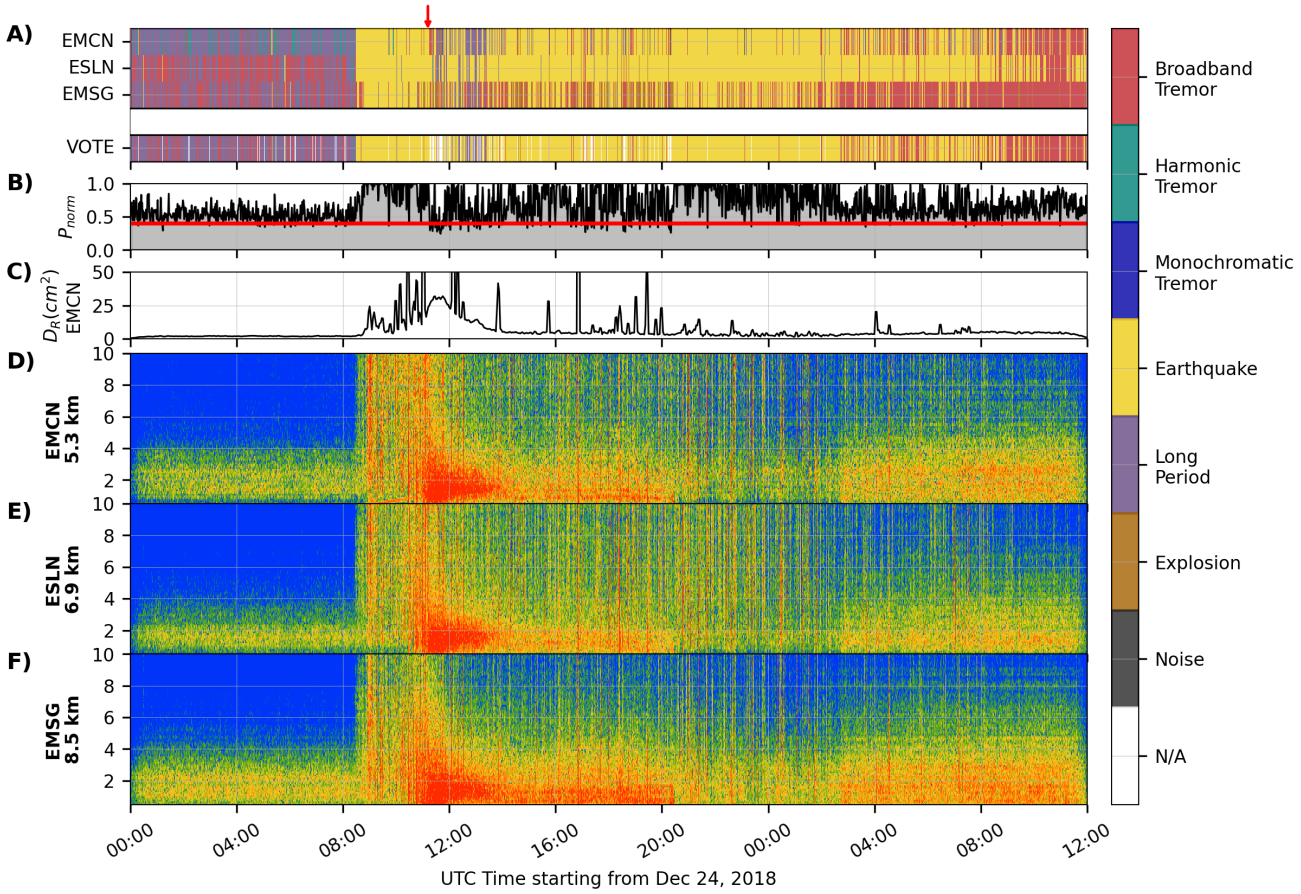

Figure 10: Generalized VOISS–Net model applied to Mt. Etna Volcano seismic data between December 24, 2018 00:00–December 25, 2018 12:00. Figure layout is the same as Figure 6. The D_R panel in [C] is clipped at 50 cm^2. The red arrow at the top indicates the fissure eruption onset at 11:11.

Pavlof Volcano seismicity. LP swarms, such as in March 2021 and January 2023 are mostly classified as LPs by the generalized model, while the T24 model typically classified these as Explosions. Note that some of the elevated regional seismicity, particularly related to the M8.2 Chignik earthquake in late July 2021, is likely being misclassified as LPs and Explosions in the generalized model. The generalized model also seems to detect more episodes of Monochromatic Tremor in late 2021, whereas some of those periods are classified as Broadband Tremor by T24. Classifications during the main portion of the eruption generally agree between the two models. Noise also dominates in both models before and after the eruption, which is also consistent with our interpretation, and the generalized model appears to have fewer misclassifications outside the eruption period. Lastly, our P_{norm} thresholding appears to reduce misclassifications particularly before and after the eruption begins (i.e. less unexpected Broadband Tremor detections during background activity).

4 Discussion

4.1 VOISS–Net classifications

Overall, our generalized VOISS–Net model does well at classifying seismic spectrograms both in our test set (87 % accuracy) and for the six volcano case studies we present here. It also performs well in the individual volcano studies of T24 and Lyons et al. [2025]. In addition to the high classification accuracy for most classes in the test set (Figure 5C), we find the model classifies the main types of volcano seismicity in a spectrogram similar to how we interpret them, with the added benefits that it is automated and repeatable. We find the main limitations of the model to be 1) classifying signals that overlap within a 2-minute spectrogram, 2) differentiating between transients that contain similar spectral characteristics, and 3) classifying signals that lie outside the training dataset or the seven predefined classes. These challenges also exist for human analysts classifying seismicity in spectrograms.

Overlapping seismic signals often occur during periods of heightened volcanic activity (e.g. Figure 8, 10). VOISS–Net is constructed to return one classification per time step, and in cases where different classes occur in that time step, we prioritize the transient class (Earthquake, LP, Explosion). This choice was made as prioritizing long-duration sources such as tremor and noise would often not permit coincident transients to be detected. Adjusting the overlap between consecutive windows or reducing the duration of the spectrogram may alleviate some issues with concurrent signals, but could also lead to additional misclassifications. We have found that a

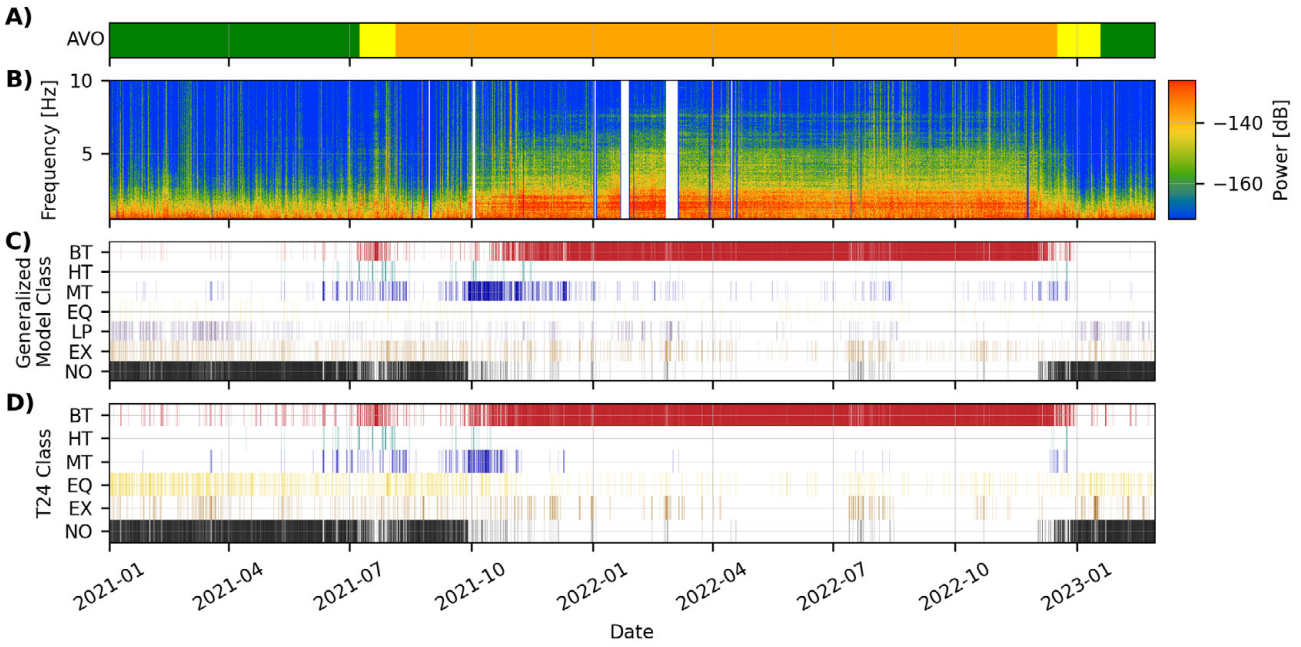

Figure 11: Comparison of the generalized `VOISS-Net` model with that of T24 for the 2021–2022 eruption of Pavlof Volcano. [A] AVO color code roughly indicating the level of unrest. [B] Spectrogram for station PS1A. The associated `VOISS-Net` classifications using the [C] Generalized and [D] T24 models binned into 1 hr intervals for comparison over this long time period. Overall the generalized model seems to characterize the volcanic seismicity more accurately, including the relative lack of earthquakes detected in the generalized model, which is consistent with Pavlof Volcano's seismicity. The generalized model also detect more LPs prior to the eruption and appears to have less false positive classification prior to the eruption.

2-minute spectrogram is a good trade-off to capture transient events of a few to tens of seconds and longer duration signals. We envision that "image segmentation" [Minaee et al. 2021], a popular method in ML image classification for the detection of different classes within a single image, could be utilized to alleviate this issue. However, this would require very careful manual labeling of thousands of spectrograms, a very time-consuming process and potentially prone to mislabeling.

Transient signals, such as LPs and explosions, may also have similar spectral content to each other. Unlike the other classes, our Explosion class and labels are not defined by their time-frequency spectral characteristics but actually by its source process and often by its acoustic wave. We find that there is quite a bit of variability in explosion signal spectrograms (e.g. Figure 2) and lower classification accuracies than the other classes (Figure 5C). The Explosion class also has a low recall of 0.75 (Supplementary Material 1 Table S1), suggesting it has a relatively high false negative rate. Occasionally, a high-frequency air-ground coupled acoustic wave will be present after the explosion seismic signal [e.g. Fee et al. 2016], but this is not always apparent. Acoustic-seismic coupling is complex and can depend on many factors [e.g. Madshus et al. 2005; Wills et al. 2022], thus the associated air-ground coupled wave should not be the characteristic feature of explosion seismic signals. We have found that colocated infrasound sensors or a nearby sensor network are often critical for determining explosions. Future work could explore further the spectral properties of explosions in seismic data and improve this class, and leverage colocated infrasound sensors when available.

As expected for supervised learning, our model also struggles to classify signals that are not present in the training data or that do not fall into one of the predefined classes. The Trident Volcano case study illustrates this well (Figure 7). The mixed-frequency events are classified into multiple classes as a function of time rather than a single, consistent class. Although this may seem incorrect, it is consistent with the evolving/mixed spectral content of the spectrograms. The regional and teleseismic earthquakes in this example are also misclassified as LP events. Regional and telesesimic earthquakes are relatively easily detected with other methods, and it would be straightforward to integrate existing earthquake catalogs alongside `VOISS-Net` to remove these misclassifications. Importantly, we also note that our training dataset consists of spectrograms from stations deployed approximately 2–15 km from the volcanic vent. Stations closer and further than this distance will likely show different spectra, as path effects are known to be significant for LP and other volcano seismic events [e.g. Bean et al. 2008; Titos et al. 2018]. We in fact observe variable spectra as a function of distance in multiple examples (e.g. Figure 6, 8, 10). Future work could take path effects into account for classification, perhaps giving stations more weight near the source than those further away [Titos et al. 2023]. Our current implementation was trained on stations at different distances and volcanoes to decrease the reliance on path effects and focus just on the spectrogram itself. Integrating all three components of the seismometer could also

help with event type characterization, although this assumes three components will always be present in the data.

The Great Sitkin earthquake classifications in Figure 9 are in disagreement with the LP classifications of Power and Roman [2024]. They used Frequency Index (FI) [Buurman and West 2006] and Average Peak Frequency (APF) [Ketner and Power 2013] to differentiate between earthquakes and LPs in the AVO catalog. The events leading up to the May 26 explosion had relatively low FI and APF, along with poorly developed S-waves. In Supplementary Material 1 Figure S2 we compute and plot FI for the same data and time period as Figure 9. Similar to Power and Roman [2024] we find these events to have FI between approximately −0.5 and −2 using lower and upper frequency bands of 1–5 and 5–10 Hz, respectively, albeit with a much longer evaluation window of 120 s. However, the overall event spectral content (onset and coda) is more similar to our VT labels than the LP labels (Figure 2), which is why our model classifies them as VTs. Our LP labels are generally devoid of a clear transient onset like those in Figure 9. The Great Sitkin events may fit the "hybrid" class introduced by Lahr et al. [1994]. FI, APF, and other metrics could be a useful tool to aid `VOISS-Net` in transient event classification, and we integrate FI computations in our code for comparison. However, here we choose not to integrate it directly into the classification scheme as they are useful but imperfect event classification metrics. First, FI has a magnitude bias. Large events will naturally have a lower frequency content. Second, FI is generally computed on 5–7 s of data around the P-wave, while `VOISS-Net` computes a classification over a 2-minute-long window. The long duration, low frequency content of LP events is missed in traditional FI computation but key for `VOISS-Net` classification. Lastly, FI is only applicable for VTs and LPs, and its utility for tremor and the Explosion classes is unclear. Users are encouraged to implement more sophisticated transient classification measures as they deem appropriate.

We also note that `VOISS-Net` does not provide an exact event onset, but rather just the time of the detection window. In our examples with 50 % window overlap, the time window resolution is 1 minute. Other techniques would likely need to supplement the `VOISS-Net` classifications if precise event onsets are desired.

Lastly, we point out that our model does not classify the volcano seismic source process, but rather the properties of the spectrogram. `VOISS-Net` first classifies the station spectrograms and then determines a most likely network-wide class, along with a normalized probability. Source features can then be determined via modeling and other signal analysis after the classification. Future work could also integrate other spectrogram-like quantities, such as time-frequency polarization analysis of three-component seismic data [Haney et al. 2020] or network covariance [Soubestre et al. 2018] to help elucidate source processes.

4.2 Class and label selection

Not all volcano seismicity will fit into our seven predefined classes. Of note is the fact that we only consider signals between 0.5–10 Hz, thus Very Long Period (VLP) [Matoza and Roman 2022] events are out of band and will not be classified. Some events, such as VTs and explosions, often have considerable energy above 10 Hz. We considered incorporating other seismic classes (e.g. hybrid events, tornillos, regional earthquakes), but that would require additional, vetted labeled data. More classes would also increase the chance for misclassifications by providing other class options that may not be significantly different. Additional classes would also add further challenges to balancing the test and validation sets (Section *2.2.1*). One way to mitigate the occasionally poor transient classification performance would be to merge the Explosions, Earthquake, and LP classes into a single "Transient" class. The transient classifications could then be passed to different models or methods for discrimination, such as P/S ratio for earthquakes versus explosions [Wang et al. 2020] and Frequency Index for earthquakes versus LPs [Buurman and West 2006].

Adding in additional spectrogram labels from other volcanoes and time periods would likely further increase `VOISS-Net`'s performance and generalization by providing a greater diversity of signals for the model to learn. However, as noted in this and previous work, hand-labeling spectrograms can be a time-consuming process and is subjective and based on the analyst's interpretation. Synthetic data could also be used to expand the training dataset, with care needed to ensure realistic signal and noise. Transfer Learning [Lapins et al. 2021] may also be effective in applying `VOISS-Net`, but we leave that for future testing and evaluation.

4.3 Comparing `VOISS-Net` models

We believe our generalized `VOISS-Net` seismic model is an improvement upon previous models by T24 and Lyons et al. [2025], and should be more effective at classifying signals from volcanoes worldwide. The generalized model test set accuracy is 87 % compared to 81 % for both T24 and Lyons et al. [2025]. Examination of the classifications in the individual case studies in Figures 6–10 and the comparison with T24 in Figure 11 shows that the generalized model typically agrees with our interpretations of the spectrograms. We acknowledge that validation of the classifications is challenging and subjective based on the analyst's and our interpretation. The switch from 4 to 2 minutes also provides additional granularity and more of a focus on transient events, while still capturing some of the longer duration features of tremor. The addition of the LP class is important to capture this frequent event type, although we find that it is occasionally misclassified as regional and teleseismic earthquakes and adds additional challenges for classifying transients. The T24 model also appears to be biased in classifying most impulsive signals as earthquakes. Our vastly larger training dataset adds diversity and numbers that aids in generalization and reducing bias. Lastly, we implement thresholding in our network-wide classification using P_{norm}. P_{norm} was introduced in T24 but only as a visual metric for comparison. Here we use it to discard low probability classifications and find it to be effective, similar to Lyons et al. [2025].

5 Conclusions

We present an updated ML model to detect and classify the main types of volcano seismicity. Our model is an attempt to generalize the model of Tan et al. [2024], which focused primarily on tremor. We utilize a broad, diverse training dataset of over 270,000 spectrograms that are split into seven seismicity classes: Broadband Tremor, Harmonic Tremor, Monochromatic Tremor, Explosions, Earthquakes, LPs, and Noise. We utilize the training data of Tan et al. [2024] and Lyons et al. [2025] and supplement it with additional expert-labeled data and catalog-derived earthquakes from multiple other volcanoes. This diverse training dataset is then used to train an image classification model that successfully classifies 87 % of the test set. The model can then be applied to continuous seismic data for a variable number of stations at any volcano. `VOISS-Net` returns both station-specific classifications as well as a network-averaged majority vote. We find the network vote and associated classification probability threshold is an effective way to mitigate inconsistent classifications between stations.

We apply the generalized `VOISS-Net` model to six volcanic unrest case studies from a diverse set of volcanoes and durations. These case studies illustrate how `VOISS-Net` can effectively classify the spectral content of continuous seismic data. It performs well for a diverse sequence of tremor, LPs, and explosions at Semisopochnoi, Alaska, similar to the results of Lyons et al. [2025]. `VOISS-Net` classifies some unique, mixed frequency and monochromatic events at Trident Volcano, Alaska, which are traditionally difficult to detect and catalog. We find that `VOISS-Net` successfully classifies the main features of the run-up and eruption signals from recent eruptions at Kīlauea (Hawai'i, USA), Great Sitkin (Alaska, USA), and Mt. Etna (Italy). In these cases it is able to successfully detect and classify precursory VT earthquakes, LP events, and tremor. Our model occasionally struggles with periods of simultaneous seismic events (i.e. earthquake and tremor occur concurrently) and performs poorly on signals it is not trained on (i.e. regional and teleseismic seismicity). The model also occasionally has challenges differentiating transient event types.

`VOISS-Net` provides a fast, interpretable means of characterizing volcano seismicity based on its spectral characteristics. It augments the role of a seismic analyst by providing seismic summaries in the form of timelines. It can be used to construct catalogs of volcano seismicity that are often difficult to catalog and provides a systematic and reproducible means of identifying volcano seismicity, ensuring consistency in long-term seismic databases. Users can easily apply the open-source model to data from an FDSN web service client. Alternatively, they can build their own labeled dataset and craft their own volcano-specific model. `VOISS-Net` is not intended to supplant existing techniques to detect individual types of seismicity (i.e. earthquake detection and location), but rather can be used in conjunction with traditional techniques to form more complete seismic catalogs. We envision and encourage future work that will examine the utility and limitations of `VOISS-Net`, as well as additional improvements to the training data and overall implementation. A similar framework could be used to create a generalized volcano infrasound model as well.

Author Contributions

DF and DT conceptualized the project and performed the analysis with feedback from all coauthors. JL, MS, and AC contributed data labels. DF led the writing of the manuscript. All authors read the manuscript and provided valuable comments.

Acknowledgements

We acknowledge helpful comments from the NSF PREEVENTS Volcano Forecasting team and members of AVO. Insightful reviews were provided by the Editor (Chiara Montagna), Matteo Bagagli, John Power, Michelle Coombs, and an anonymous reviewer. We used `ObsPy` [Beyreuther et al. 2010] extensively, as well as `PyGMT` [Uieda et al. 2021] and `Tensorflow` [Abadi et al. 2016]. DF, DT, and TG acknowledge funding by the NSF PREEVENTS Grant 1855126 and support by the U.S. Geological Survey under Cooperative Agreement No. G24AC00240. M.S. acknowledges the IMPACT PROJECT ("A multidisciplinary Insight on the kinematics and dynamics of Magmatic Processes at Mt. Etna Aimed at identifying preCursor phenomena and developing early warning sysTems"), INGV Department strategic Projects - 2019. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government. Observations of volcanic activity from Kīlauea are taken from the Hawaiian Volcano Observatory website (`https://www.usgs.gov/observatories/hvo`).

Data Availability

The `VOISS-Net` code is available on github at `https://github.com/darren-tpk/voiss-net` and Zenodo [Tan and Fee 2025]. A list of the stations and times for each of the spectrogram labels is available as a text file in the supplemental material. The Mt. Etna seismic data are available from the European Integrated Data Archive (EIDA) Italia (`https://eida.ingv.it/en`). All other seismic data are available through the EarthScope Consortium Web Services (`https://service.iris.edu/`), including the following seismic networks: (1) AV [Alaska Volcano Observatory/USGS 1988] (2) HV [USGS Hawaiian Volcano Observatory (HVO) 1956], and (3) 7G [Roman 2014].

Copyright Notice

References

Abadi, M., P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D. G. Murray, B. Steiner,

P. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu, and X. Zheng (2016). "TensorFlow: A System for Large-Scale Machine Learning". *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*. Savannah, GA: USENIX Association, pages 265–283. ISBN: 978-1-931971-33-1.

Aki, K. and R. Koyanagi (1981). "Deep volcanic tremor and magma ascent mechanism under Kilauea, Hawaii". *Journal of Geophysical Research: Solid Earth* 86 (B8), pages 7095–7109. DOI: 10.1029/JB086iB08p07095.

Alaska Volcano Observatory/USGS (1988). *Alaska Volcano Observatory*. DOI: 10.7914/SN/AV.

Andronico, D., A. Cannata, G. Di Grazia, and F. Ferrari (2021). "The 1986–2021 paroxysmal episodes at the summit craters of Mt. Etna: Insights into volcano dynamics and hazard". *Earth-Science Reviews* 220, page 103686. DOI: 10.1016/j.earscirev.2021.103686.

Ardid, A., D. Dempsey, C. Caudron, and S. Cronin (2022). "Seismic precursors to the Whakaari 2019 phreatic eruption are transferable to other eruptions and volcanoes". *Nature Communications* 13(1), page 2002. DOI: 10.1038/s41467-022-29681-y.

Bean, C., I. Lokmer, and G. O'Brien (2008). "Influence of near-surface volcanic structure on long-period seismic signals and on moment tensor inversions: Simulated examples from Mount Etna". *Journal of Geophysical Research: Solid Earth* 113 (B8), 2007JB005468. DOI: 10.1029/2007JB005468.

Beyreuther, M., R. Barsch, L. Krischer, T. Megies, Y. Behr, and J. Wassermann (2010). "ObsPy: A Python Toolbox for Seismology". *Seismological Research Letters* 81(3), pages 530–533. DOI: 10.1785/gssrl.81.3.530.

Buurman, H. and M. E. West (2006). *Seismic Precursors to Volcanic Explosions During the 2006 Eruption of Augustine Volcano*. U.S. Geological Survey Professional Paper 1769, pages 41–57.

Cameron, C., T. Orr T.R., J. P. Dixon, H. Dietterich, C. Waythomas, A. Iezzi, J. Power, C. Searcy, R. Grapenthin, G. Tepp, T. Lopez, K. DeGrandpre, and J. Perreault (2023). *2018 Volcanic Activity in Alaska—Summary of Events and Response of the Alaska Volcano Observatory*. Scientific Investigations Report 2023–5029. U.S. Geological Survey.

Cannata, A., G. Di Grazia, M. Aliotta, C. Cassisi, P. Montalto, and D. Patanè (2013). "Monitoring Seismo-volcanic and Infrasonic Signals at Volcanoes: Mt. Etna Case Study". *Pure and Applied Geophysics* 170(11), pages 1751–1771. DOI: 10.1007/s00024-012-0634-x.

Cannavo', F., M. Sciotto, A. Cannata, and G. Di Grazia (2019). "An Integrated Geophysical Approach to Track Magma Intrusion: The 2018 Christmas Eve Eruption at Mount Etna". *Geophysical Research Letters* 46(14), pages 8009–8017. DOI: 10.1029/2019GL083120.

De Angelis, S. (2006). "Analyses of unusual long-period earthquakes with extended coda recorded at Katmai National Park, Alaska, USA". *Geophysical Research Letters* 33(7), 2005GL025581. DOI: 10.1029/2005GL025581.

De Angelis, S. and S. R. McNutt (2007). "Observations of volcanic tremor during the January–February 2005 eruption of Mt. Veniaminof, Alaska". *Bulletin of Volcanology* 69(8), pages 927–940. DOI: 10.1007/s00445-007-0119-4.

Endo, E. T. and T. Murray (1991). "Real-time Seismic Amplitude Measurement (RSAM): a volcano monitoring and prediction tool". *Bulletin of Volcanology* 53(7), pages 533–545. DOI: 10.1007/BF00298154.

Fee, D., M. Haney, R. Matoza, C. Szuberla, J. Lyons, and C. Waythomas (2016). "Seismic Envelope-Based Detection and Location of Ground-Coupled Airwaves from Volcanoes in Alaska". *Bulletin of the Seismological Society of America* 106(3), pages 1024–1035. DOI: 10.1785/0120150244.

Fee, D., M. M. Haney, R. S. Matoza, A. R. Van Eaton, P. Cervelli, D. J. Schneider, and A. M. Iezzi (2017). "Volcanic tremor and plume height hysteresis from Pavlof Volcano, Alaska". *Science* 355(6320), pages 45–48. DOI: 10.1126/science.aah6108.

Fee, D., L. Toney, K. Kim, R. W. Sanderson, A. M. Iezzi, R. S. Matoza, S. De Angelis, A. D. Jolly, J. J. Lyons, and M. M. Haney (2021). "Local Explosion Detection and Infrasound Localization by Reverse Time Migration Using 3-D Finite-Difference Wave Propagation". *Frontiers in Earth Science* 9, page 620813. DOI: 10.3389/feart.2021.620813.

Goodfellow, I., Y. Bengio, and A. Courville (2016). *Deep Learning*. http://www.deeplearningbook.org. MIT Press.

Haney, M. M., D. Fee, K. F. McKee, J. J. Lyons, R. S. Matoza, A. G. Wech, G. Tepp, C. Searcy, and T. D. Mikesell (2020). "Co-eruptive tremor from Bogoslof volcano: seismic wavefield composition at regional distances". *Bulletin of Volcanology* 82(2), page 18. DOI: 10.1007/s00445-019-1347-0.

Hildreth, W., J. Fierstein, M. A. Lanphere, and D. F. Siems (2001). *Trident Volcano: Four Contiguous Stratocones Adjacent to Katmai Pass, Alaska Peninsula*. U.S. Geological Survey Professional Paper 1678, pages 153–180.

Hotovec-Ellis, A. J. and C. Jeffries (2016). "Near real-time detection, clustering, and analysis of repeating earthquakes: Application to Mount St. Helens and Redoubt volcanoes". Seismological Society of America Annual Meeting.

Hotovec-Ellis, A. J., D. R. Shelly, D. P. Hill, A. M. Pitt, P. B. Dawson, and B. A. Chouet (2018). "Deep fluid pathways beneath Mammoth Mountain, California, illuminated by migrating earthquake swarms". *Science Advances* 4(8), eaat5258. DOI: 10.1126/sciadv.aat5258.

Ioffe, S. and C. Szegedy (2015). "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift". *Proceedings of the 32nd International Conference on Machine Learning*, pages 448–456.

Ketner, D. and J. Power (2013). "Characterization of seismic events during the 2009 eruption of Redoubt Volcano, Alaska". *Journal of Volcanology and Geothermal Research* 259, pages 45–62. DOI: 10.1016/j.jvolgeores.2012.10.007.

Kingma, D. P. and J. Ba (2017). "Adam: A Method for Stochastic Optimization". *3rd International Conference for Learning Representation*. San Diego.

Klein, F. W. (2002). *User's guide to HYPOINVERSE-2000, a Fortran program to solve for earthquake locations and magnitudes*. Open-File Report. U.S. Geological Survey.

Kong, Q., R. Wang, W. R. Walter, M. Pyle, K. Koper, and B. Schmandt (2022). "Combining Deep Learning With Physics Based Features in Explosion-Earthquake Discrimination". *Geophysical Research Letters* 49(13), e2022GL098645. DOI: 10.1029/2022GL098645.

Lahr, J. C. (1989). *HYPOELLIPSE/version 2.0*: A computer program for determining local earthquake hypocentral parameters, magnitude, and first motion pattern*. Open-File Report. U.S. Geological Survey.

Lahr, J. C., B. A. Chouet, C. D. Stephens, J. A. Power, and R. A. Page (1994). "Earthquake classification, location, and error analysis in a volcanic environment: implications for the magmatic system of the 1989–1990 eruptions at redoubt volcano, Alaska". *Journal of Volcanology and Geothermal Research* 62(1), pages 137–151. DOI: 10.1016/0377-0273(94)90031-0.

Lapins, S., B. Goitom, J.-M. Kendall, M. J. Werner, K. V. Cashman, and J. O. S. Hammond (2021). "A Little Data Goes a Long Way: Automating Seismic Phase Arrival Picking at Nabro Volcano With Transfer Learning". *Journal of Geophysical Research: Solid Earth* 126(7), e2021JB021910. DOI: 10.1029/2021JB021910.

Linville, L., K. Pankow, and T. Draelos (2019). "Deep Learning Models Augment Analyst Decisions for Event Discrimination". *Geophysical Research Letters* 46(7), pages 3643–3651. DOI: 10.1029/2018GL081119.

Lomax, A., J. Virieux, P. Volant, and C. Berge-Thierry (2000). "Probabilistic Earthquake Location in 3D and Layered Models". *Advances in Seismic Event Location*. Edited by C. H. Thurber and N. Rabinowitz. Redacted by G. Nolet. Volume 18. Series Title: Modern Approaches in Geophysics. Dordrecht: Springer Netherlands, pages 101–134. ISBN: 978-90-481-5498-2 978-94-015-9536-0. DOI: 10.1007/978-94-015-9536-0_5.

Lyons, J., D. Tan, M. Angarita, M. Loewen, T. Lopez, R. Grapenthin, A. Hotovec-Ellis, D. Fee, and M. Haney (2025). "Identifying precursors and tracking pulses of magma ascent in multidisciplinary data during the 2018–2023 phreatomagmatic eruption at Semisopochnoi Island, Alaska". *Journal of Volcanology and Geothermal Research*, page 108329. DOI: 10.1016/j.jvolgeores.2025.108329.

Madshus, C., F. Løvholt, A. Kaynia, L. R. Hole, K. Attenborough, and S. Taherzadeh (2005). "Air–ground interaction in long range propagation of low frequency sound and vibration—field tests and model verification". *Applied Acoustics* 66(5), pages 553–578. DOI: 10.1016/j.apacoust.2004.09.006.

Maguire, R., B. Schmandt, R. Wang, Q. Kong, and P. Sanchez (2024). "Generalization of Deep-Learning Models for Classification of Local Distance Earthquakes and Explosions across Various Geologic Settings". *Seismological Research Letters* 95(4), pages 2229–2238. DOI: 10.1785/0220230267.

Malfante, M., M. Dalla Mura, J. I. Mars, J.-P. Métaxian, O. Macedo, and A. Inza (2018). "Automatic Classification of Volcano Seismic Signatures". *Journal of Geophysical Research: Solid Earth* 123(12). DOI: 10.1029/2018JB015470.

Matoza, R. S. and D. C. Roman (2022). "One hundred years of advances in volcano seismology and acoustics". *Bulletin of Volcanology* 84(9), page 86. DOI: 10.1007/s00445-022-01586-0.

McNutt, S. R. (1996). "Seismic Monitoring and Eruption Forecasting of Volcanoes: A Review of the State-of-the-Art and Case Histories". Scarpa, R. and R. I. Tilling. *Monitoring and Mitigation of Volcano Hazards*. Berlin, Heidelberg: Springer Berlin Heidelberg, pages 99–146. ISBN: 978-3-642-80089-4 978-3-642-80087-0. DOI: 10.1007/978-3-642-80087-0_3.

McNutt, S. R. and T. Nishimura (2008). "Volcanic tremor during eruptions: Temporal characteristics, scaling and constraints on conduit size and processes". *Journal of Volcanology and Geothermal Research* 178(1), pages 10–18. DOI: 10.1016/j.jvolgeores.2008.03.010.

McNutt, S. R. and D. C. Roman (2015). "Volcanic Seismicity". *The Encyclopedia of Volcanoes*. Elsevier, pages 1011–1034. ISBN: 978-0-12-385938-9. DOI: 10.1016/B978-0-12-385938-9.00059-6.

Minaee, S., Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos (2021). "Image segmentation using deep learning: A survey". *IEEE transactions on pattern analysis and machine intelligence* 44(7). Publisher: IEEE, pages 3523–3542.

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

Nishimura, T. (1998). "Source mechanisms of volcanic explosion earthquakes: single force and implosive sources". *Journal of Volcanology and Geothermal Research* 86(1), pages 97–106. DOI: 10.1016/S0377-0273(98)00088-2.

Orr, T. R., H. R. Dietterich, D. Fee, T. Girona, R. Grapenthin, M. M. Haney, M. Loewen, J. J. Lyons, J. A. Power, H. F. Schwaiger, D. J. Schneider, D. Tan, L. Toney, V. K. Wasser, and C. F. Waythomas (2024). *2021 Volcanic Activity in Alaska and the Commonwealth of the Northern Mariana Islands—Summary of Events and Response of the Alaska Volcano Observatory*. Scientific Investigations Report 2024–5014. U.S. Geological Survey, page 64. DOI: 10.3133/sir20245014.

Orr, T. R., H. R. Dietterich, R. Grapenthin, M. M. Haney, M. Loewen, P. Saunders-Shultz, D. Tan, C. F. Waythomas, and A. G. Wech (2025). *2022 Volcanic Activity in Alaska and the Northern Mariana Islands—Summary of Events and Response of the Alaska Volcano Observatory*. Scientific Investigations Report. U.S. Geological Survey. DOI: 10.3133/sir20245014.

Patanè, D., A. Aiuppa, M. Aloisi, B. Behncke, A. Cannata, M. Coltelli, G. Di Grazia, S. Gambino, S. Gurrieri, M. Mattia, and G. Salerno (2013). "Insights into magma and fluid transfer at Mount Etna by a multiparametric approach: A model of the events leading to the 2011 eruptive cycle". *Journal of Geophysical Research: Solid Earth* 118(7), pages 3519–3539. DOI: 10.1002/jgrb.50248.

Power, J. A., P. A. Friberg, M. M. Haney, T. Parker, S. D. Stihler, and J. P. Dixon (2019). *A unified catalog of earthquake hypocenters and magnitudes at volcanoes in Alaska—1989 to 2018*. Scientific Investigations Report 2019–5037. U.S. Geological Survey, page 17.

Power, J. A. and D. C. Roman (2024). "Event classification, seismicity, and eruption forecasting at Great Sitkin Volcano, Alaska: 1999–2023". *Journal of Volcanology and Geothermal Research* 454, page 108182. DOI: 10.1016/j.jvolgeores.2024.108182.

Roman, D. C. (2014). *Popocatepetl Tremor Experiment*. DOI: 10.7914/SN/7G_2014.

– (2017). "Automated detection and characterization of harmonic tremor in continuous seismic data". *Geophysical Research Letters* 44(12), pages 6065–6073. DOI: 10.1002/2017GL073715.

Sigmundsson, F., A. Hooper, S. Hreinsdóttir, K. S. Vogfjörd, B. G. Ófeigsson, E. R. Heimisson, S. Dumont, M. Parks, K. Spaans, G. B. Gudmundsson, V. Drouin, T. Árnadóttir, K. Jónsdóttir, M. T. Gudmundsson, T. Högnadóttir, H. M. Fridriksdóttir, M. Hensch, P. Einarsson, E. Magnússon, S. Samsonov, B. Brandsdóttir, R. S. White, T. Ágústsdóttir, T. Greenfield, R. G. Green, Á. R. Hjartardóttir, R. Pedersen, R. A. Bennett, H. Geirsson, P. C. La Femina, H. Björnsson, F. Pálsson, E. Sturkell, C. J. Bean, M. Möllhoff, A. K. Braiden, and E. P. S. Eibl (2015). "Segmented lateral dyke growth in a rifting event at Bárðarbunga volcanic system, Iceland". *Nature* 517(7533), pages 191–195. DOI: 10.1038/nature14111.

Soubestre, J., N. M. Shapiro, L. Seydoux, J. de Rosny, D. V. Droznin, S. Y. Droznina, S. L. Senyukov, and E. I. Gordeev (2018). "Network-Based Detection and Classification of Seismovolcanic Tremors: Example From the Klyuchevskoy Volcanic Group in Kamchatka". *Journal of Geophysical Research: Solid Earth* 123(1), pages 564–582. DOI: 10.1002/2017JB014726.

Srivastava, N., G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov (2014). "Dropout: A Simple Way to Prevent Neural Networks from Overfitting". *Journal of Machine Learning Research* 15(56).

Steinke, B., A. D. Jolly, R. Carniel, D. E. Dempsey, and S. J. Cronin (2023). "Identification of Seismo-Volcanic Regimes at Whakaari/White Island (New Zealand) Via Systematic Tuning of an Unsupervised Classifier". *Journal of Geophysical Research: Solid Earth* 128(3), e2022JB026221. DOI: 10.1029/2022JB026221.

Tan, D. and D. Fee (2025). "darren-tpk/voiss-net (v1.0.4)". *Zenodo*. [Software]. DOI: 10.5281/zenodo.15277852.

Tan, D., D. Fee, A. Witsil, T. Girona, M. Haney, A. Wech, C. Waythomas, and T. Lopez (2024). "Detection and Characterization of Seismic and Acoustic Signals at Pavlof Volcano, Alaska, Using Deep Learning". *Journal of Geophysical Research: Solid Earth* 129(6), e2024JB029194. DOI: 10.1029/2024JB029194.

Titos, M., A. Bueno, L. Garcia, and C. Benitez (2018). "A Deep Neural Networks Approach to Automatic Recognition Systems for Volcano-Seismic Events". *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 11(5), pages 1533–1544. DOI: 10.1109/JSTARS.2018.2803198.

Titos, M., A. Bueno, L. Garcia, M. C. Benitez, and J. Ibanez (2019). "Detection and Classification of Continuous Volcano-Seismic Signals With Recurrent Neural Networks". *IEEE Transactions on Geoscience and Remote Sensing* 57(4), pages 1936–1948. DOI: 10.1109/TGRS.2018.2870202.

Titos, M., L. Gutiérrez, C. Benítez, P. Rey Devesa, I. Koulakov, and J. M. Ibáñez (2023). "Multi-station volcano tectonic earthquake monitoring based on transfer learning". *Frontiers in Earth Science* 11, page 1204832. DOI: 10.3389/feart.2023.1204832.

U.S. Geological Survey (2017). "Advanced national seismic system (ANSS) comprehensive catalog of earthquake events and products". DOI: https://doi.org/10.5066/F7MS3QZH.

Uieda, L., D. Tian, W. J. Leong, L. Toney, W. Schlitzer, M. Grund, D. Newton, M. Ziebarth, M. Jones, and P. Wessel (2021). "PyGMT: A Python interface for the generic mapping tools". *Zenodo*. [Software]. DOI: 10.5281/ZENODO.4522136.

USGS Hawaiian Volcano Observatory (HVO) (1956). "Hawaiian Volcano Observatory Network". *International Federation of Digital Seismograph Networks*. DOI: 10.7914/SN/HV. [dataset].

Wang, R., B. Schmandt, and E. Kiser (2020). "Seismic Discrimination of Controlled Explosions and Earthquakes Near Mount St. Helens Using *P* / *S* Ratios". *Journal of Geophysical Research: Solid Earth* 125(10), e2020JB020338. DOI: 10.1029/2020JB020338.

Waythomas, C. F., M. M. Haney, K. L. Wallace, C. E. Cameron, and D. J. Schneider (2017). *The 2014 eruptions of Pavlof Volcano, Alaska*. Scientific Investigations Report No. 2017-5129. U.S. Geological Survey.

Waythomas, C. F., M. M. Haney, D. Fee, D. J. Schneider, and A. Wech (2014). "The 2013 eruption of Pavlof Volcano, Alaska: a spatter eruption at an ice- and snow-clad volcano". *Bulletin of Volcanology* 76(10), page 862. DOI: 10.1007/s00445-014-0862-2.

Wills, G., A. Nippress, D. N. Green, and P. J. Spence (2022). "Site-specific variations in air-to-ground coupled seismic arrivals from the 2012 October 16 explosion at Camp Minden, Louisiana, United States". *Geophysical Journal International* 231(1), pages 243–255. DOI: 10.1093/gji/ggac184.

Zali, Z., S. M. Mousavi, M. Ohrnberger, E. P. S. Eibl, and F. Cotton (2024). "Tremor clustering reveals pre-eruptive signals and evolution of the 2021 Geldingadalir eruption of the Fagradalsfjall Fires, Iceland". *Communications Earth & Environment* 5(1), page 1. DOI: 10.1038/s43247-023-01166-w.

Zhong, Y. and Y. J. Tan (2024). "Deep-Learning-Based Phase Picking for Volcano-Tectonic and Long-Period Earthquakes". *Geophysical Research Letters* 51(12), e2024GL108438. DOI: 10.1029/2024GL108438.