

Waveform Features Strongly Control Subcrater Classification Performance for a Large, Labeled Volcano Infrasound Dataset

Liam Toney^{*1}, David Fee¹, Alex Witsil², and Robin S. Matoza³

Abstract

Volcano infrasound data contain a wealth of information about eruptive patterns, for which machine learning (ML) is an emerging analysis tool. Although global catalogs of labeled infrasound events exist, the application of supervised ML to local (<15 km) volcano infrasound signals has been limited by a lack of robust labeled datasets. Here, we automatically generate a labeled dataset of >7500 explosions recorded by a five-station infrasound network at the highly active Yasur Volcano, Vanuatu. Explosions are located via backprojection and associated with one of Yasur's two summit subcraters. We then apply a supervised ML approach to classify the subcrater of origin. When trained and tested on data from the same station, our chosen algorithm is >95% accurate; when training and testing on different stations, accuracy drops to about 75%. The choice of waveform features provided to the algorithm strongly influences classification performance.

Cite this article as Toney, L., D. Fee, A. Witsil, and R. S. Matoza (2022). Waveform Features Strongly Control Subcrater Classification Performance for a Large, Labeled Volcano Infrasound Dataset, *The Seismic Record*, **2**(3), 167–175, doi: [10.1785/0320220019](https://doi.org/10.1785/0320220019).

Supplemental Material

Introduction

Machine learning (ML) has shown promise within seismology and infrasound studies for finding patterns in waveforms that are not easily perceived by humans, and for automating large-scale data analysis (Kong *et al.*, 2019). In a volcanic context, ML can identify eruptive trends and help reduce analyst workload (Carniel and Raquel Guzmán, 2020). However, the application of supervised ML to local volcano infrasound signals has to date been limited by a lack of large labeled datasets. Therefore, the previous research has focused on unsupervised techniques, with a limited number of supervised approaches using relatively small labeled datasets. Feature extraction—the process of summarizing data examples by calculating a collection of waveform attributes—is a key step in ML pipelines (Christ *et al.*, 2018). These studies have employed diverse approaches for feature extraction with variable results.

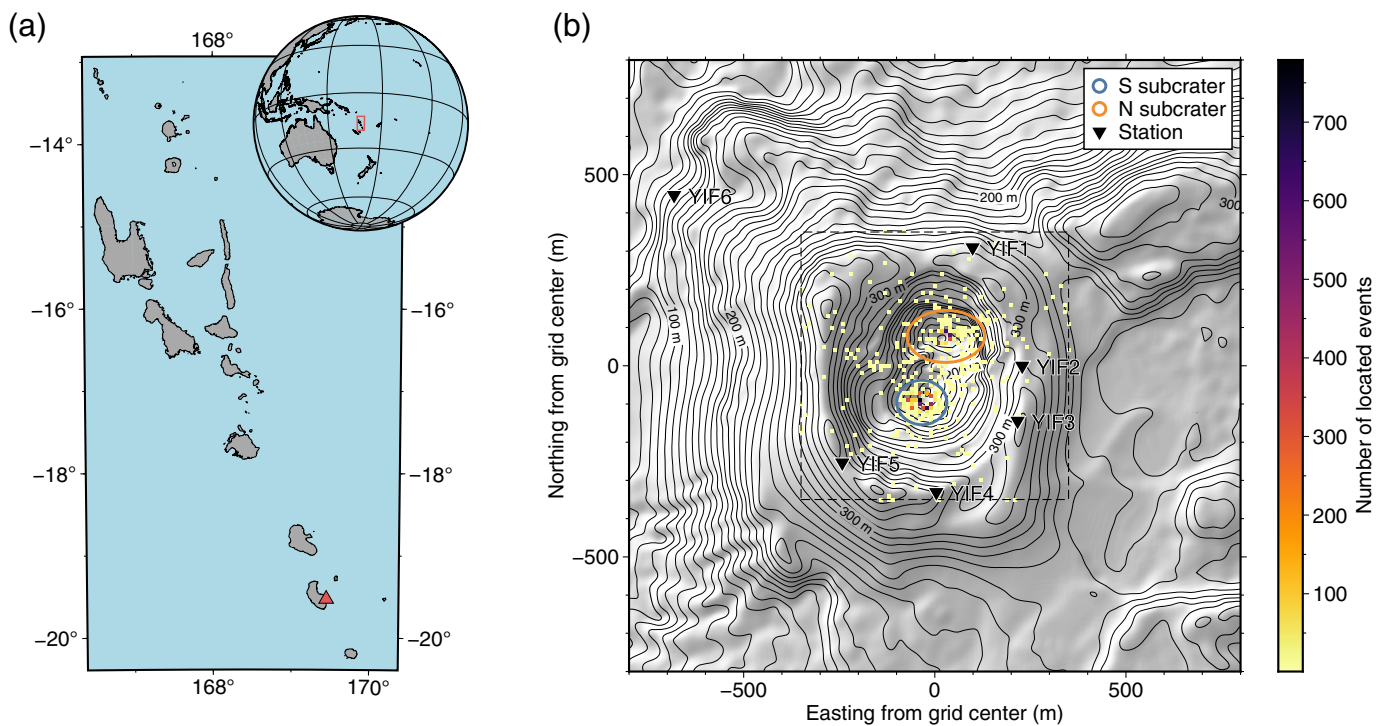
At Etna (Italy), Cannata *et al.* (2011) used an unsupervised algorithm to cluster infrasound signals into three groups and then used those groups to train a support vector machine (SVM) to recognize the active vent with implications for monitoring.

Witsil and Johnson (2020) and Watson (2020) used unsupervised learning to cluster recorded infrasound signals—at Stromboli (Italy) and Etna, respectively—into groups representing different modes of activity, and then analyzed the contributions of these groups over time to characterize changing eruptive modes. Liu *et al.* (2014) achieved an accuracy of 98% using SVMs to classify volcano, tsunami, and earthquake classes with features derived from the Hilbert–Huang transform, whereas Li *et al.* (2016) obtained an accuracy of 86% using a different feature extraction technique (spectral entropy). Ortiz *et al.* (2020) applied unsupervised clustering to infrasound array detections

1. Alaska Volcano Observatory and Wilson Alaska Technical Center, Geophysical Institute, University of Alaska Fairbanks, Fairbanks, Alaska, U.S.A., <https://orcid.org/0000-0003-0167-9433> (LT); <https://orcid.org/0000-0002-0936-9977> (DF); 2. Wilson Alaska Technical Center, Geophysical Institute, University of Alaska Fairbanks, Fairbanks, Alaska, U.S.A., <https://orcid.org/0000-0003-0200-0074> (AW); 3. Department of Earth Science and Earth Research Institute, University of California, Santa Barbara, California, U.S.A., <https://orcid.org/0000-0003-4001-061X> (RSM)

*Corresponding author: ldtoney@alaska.edu

© 2022. The Authors. This is an open access article distributed under the terms of the CC-BY license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



for three volcanoes in Ecuador, finding good agreement with analyst-derived explosion catalogs.

The aforementioned supervised approaches (Liu *et al.*, 2014; Li *et al.*, 2016) were successful but limited by their small, manually created training datasets. This motivates the development of methods that produce large volumes of training data automatically. Here, we algorithmically generate a labeled dataset and then apply a supervised learning approach to classify the subcrater of origin for Strombolian explosion signals recorded by a local infrasound network at Yasur Volcano, Vanuatu. Yasur is a basaltic-andesitic scoria cone located on Tanna Island in the island nation of Vanuatu (Fig. 1a; Iezzi *et al.*, 2019). The volcano is continuously active, producing thousands of explosions per day from various vents, and its summit consists of a bifurcated crater containing two subcraters, termed “S” and “N” for “north” and “south,” respectively (Jolly *et al.*, 2017; Simons *et al.*, 2020). There were multiple vents active in the two subcraters during the deployment whose data we use here (Jolly *et al.*, 2017; Matoza *et al.*, 2022). In addition to classifying the labeled explosion waveforms, we explore the effects of different extracted waveform features on classification performance.

Data

We use data from a six-day-long deployment at Yasur that took place in July–August 2016 (Fee *et al.*, 2016). The infrasound

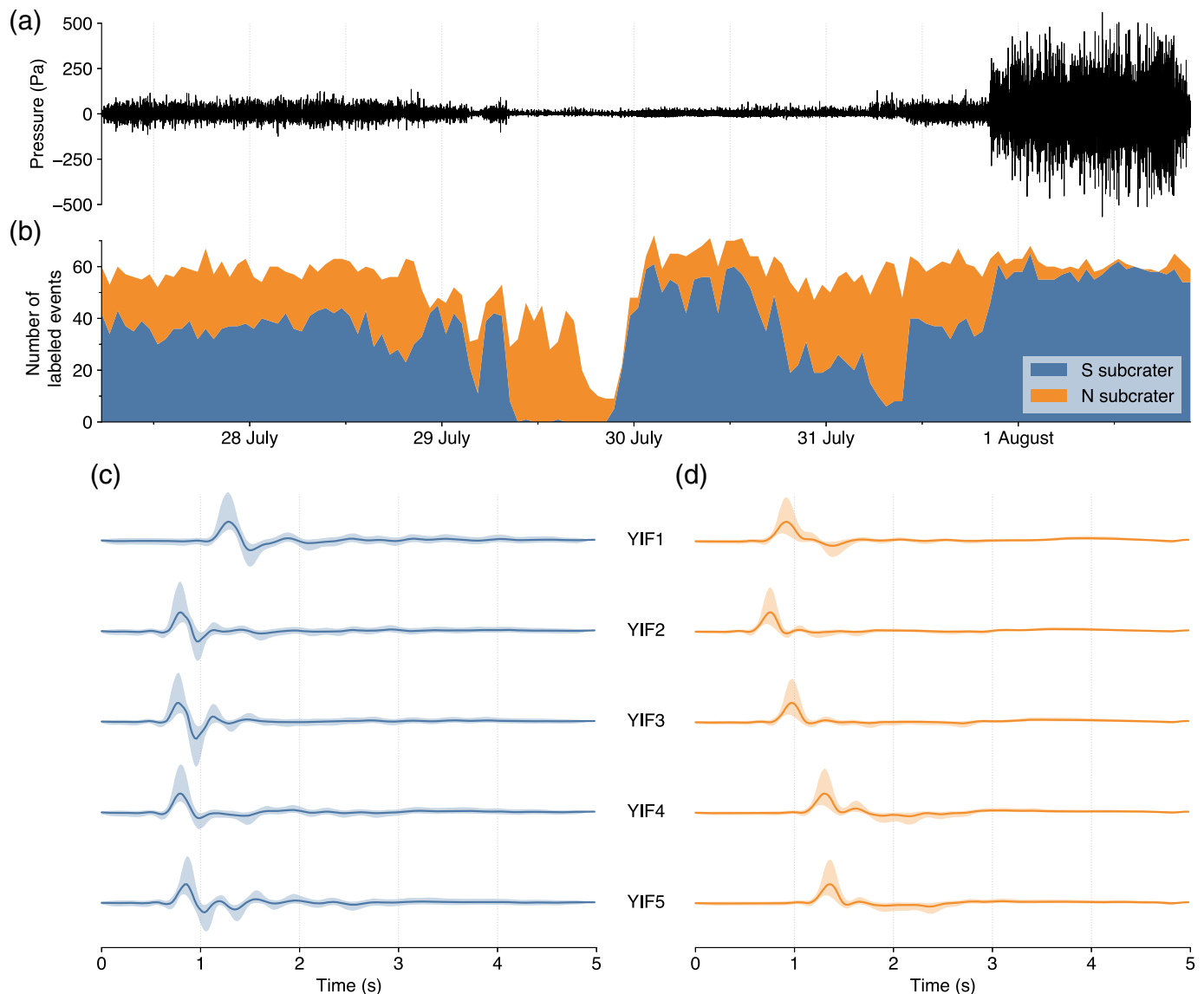
Figure 1. Map of the study region. (a) Location of the nation of Vanuatu. The red box in the globe inset has the same extent as the main map. Yasur Volcano is indicated with a red triangle. (b) Spatial histogram of 7877 RTM-FDTD locations generated over the six-day-long deployment overlain on a shaded relief map of Yasur Volcano. Ellipses denote $2\text{-}\sigma$ regions for the south and north subcraters. Dashed box delineates grid search bounds for RTM-FDTD. Infrasound stations are denoted by black inverted triangles.

component of the deployment comprised a ground-based network of sensors surrounding the summit crater of Yasur (Fig. 1b) as well as sensors connected to a tethered aerostat. For more deployment information, see Jolly *et al.* (2017), Iezzi *et al.* (2019), and Matoza *et al.* (2022).

In this work, we use data from ground-based stations YIF1–YIF5. Station YIF6 was not online for the entire deployment. These five Chaparral Model 60 sensors recorded data at 400 Hz sampling rate for approximately 138 hr from 04:00 on 27 July to 22:00 on 1 August (these and all following times in UTC; Fig. 2a). We downsample all data to 50 Hz and apply a 0.2–4 Hz band-pass filter before processing. This filter band proved effective for locating explosions in a previous study using this dataset (Fee *et al.*, 2021).

Methods

Our classification goal is to determine whether a given explosion waveform originates from a vent in the south subcrater or



a vent in the north subcrater. This is therefore a binary classification problem. To create our labeled dataset, we use a network-based algorithm to generate a catalog of event locations and associate these locations to a subcrater. We extract vectors of features from the waveforms in this dataset. Finally, we use the labeled feature vectors to train and test an ML classifier.

Creation of labeled dataset

Following [Fee et al. \(2021\)](#), we use the Python package *rtm* (see [Data and Resources](#)) to locate acoustic sources via reverse time migration with finite-difference time-domain travel-time removal (RTM-FDTD). RTM-FDTD computes travel times for acoustic waves propagating over a digital elevation model

Figure 2. (a) Pressure waveform for station YIF3, filtered as described in the [Data](#) section. (b) Number of labeled S and N subcrater events computed in hour-long rolling windows. Medians (lines) and 25th and 75th percentiles (shaded regions) of (c) 25,180 S subcrater waveforms and (d) 12,675 N subcrater waveforms, plotted by station. Traces are filtered as described in the [Data](#) section and individually normalized to their median.

(DEM) of the complex topography of Yasur, which improves location accuracy and precision. The source search grid is 350×350 m with a grid spacing of 10 m and is centered on the midpoint between the two DEM minima corresponding to the lowest points in each subcrater. The grid spacing, chosen primarily for speed, does not allow us to reliably differentiate

vents within each subcrater. Hence, we assume that inter-subcrater waveform variability is more significant than intra-subcrater waveform variability.

For RTM-FDTD, we process the waveforms from stations YIF1 to YIF5 by decimating to 20 Hz for computational speed and applying a 60 s long adaptive gain control, which reduces the dominance of large-amplitude signals on the stack function (Walker *et al.*, 2010). The stack function comprises the sums of the amplitude envelopes of the processed, normalized, and time-shifted waveforms of all five stations. To pick peaks in the stack function, we set a threshold of 4 (i.e., 80% of total possible stack value) and require a 30 s gap between adjacent peaks. This produces a catalog of 7877 events over six days, which is plotted as a spatial histogram in Figure 1b. Multiple explosions may occur within a single 30 s time window.

We fit a regularized two-component Gaussian mixture model to the collection of catalog locations to define a 2- σ confidence ellipse around each subcrater (see colored ellipses in Fig. 1b). Located waveforms are associated with either the S or N subcrater if they fall within that subcrater distribution's ellipse. After this association step, we arrive at a labeled catalog of 7571 events: 5036 S subcrater events and 2535 N subcrater events. The temporal evolution of the labeled catalog is plotted in Figure 2b. We window waveforms from the continuous time-series data using the catalog time t_0 (from RTM-FDTD) as the start time and $t_0 + 5$ s as the end time. Because each labeled event in the catalog is recorded on five stations, we have a total of 37,855 labeled waveforms (of which 25,180 are S subcrater and 12,675 are N subcrater). These labeled waveforms, summarized in Figure 2c,d, show impulsive initial transients followed by 1–2 s of coda.

Feature extraction

We extract features from the 250-sample-long labeled waveforms to input to our ML classifier. We remove the instrument response and normalize waveforms prior to feature extraction, and for two of the three extracted feature sets we apply a random time shift. These steps minimize the effect of distance from the source on the feature signature of the waveforms by removing amplitude and travel time information, respectively.

A common issue with feature engineering is that one does not usually know which features are optimal for a given classification problem. A more rigorous approach is to use a large number of features and let an algorithm determine which are the most relevant. We use the TSFRESH Python package (Christ *et al.*, 2018) to automatically extract over 700 time-

and frequency-domain features from the labeled waveforms. The full list of features is outlined on the TSFRESH website (see [Data and Resources](#)). TSFRESH has been used in a volcanic context on seismic data from Whaakari/White Island, New Zealand (Dempsey *et al.*, 2020).

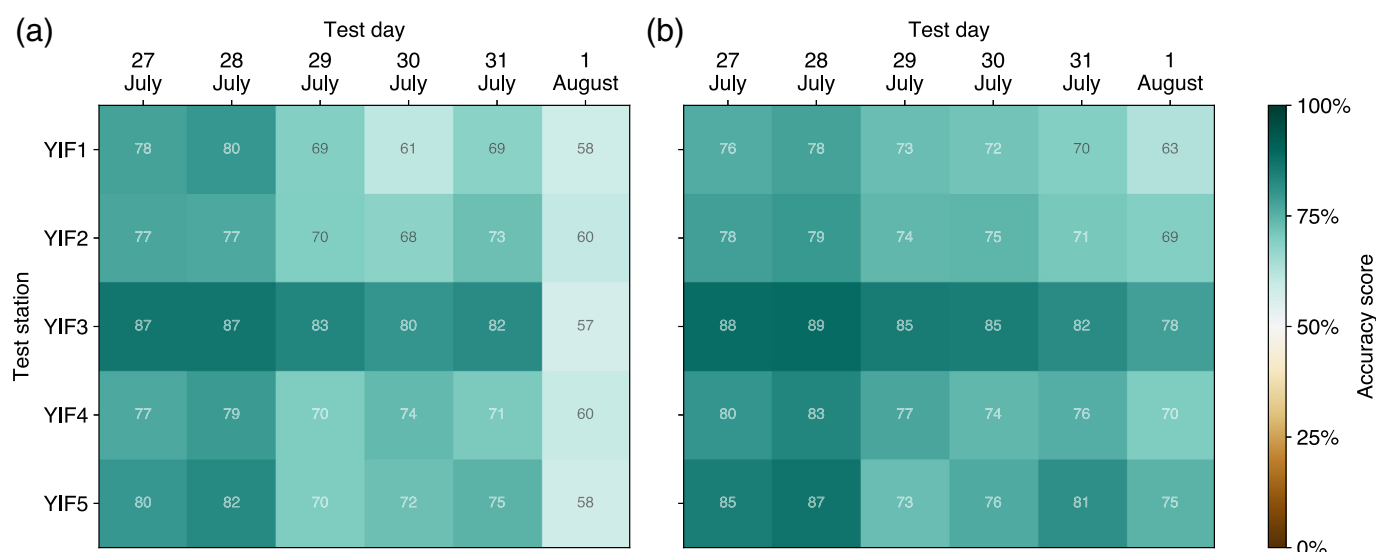
The collection of hundreds of features likely includes many that are irrelevant for the classification task. To improve classification results, as well as make the algorithm more efficient, we select an optimal subset of 10 TSFRESH features: We apply sequential feature selection (SFS; Ferri *et al.*, 1994) on data from the entire deployment across all stations. SFS is a greedy algorithm that iteratively picks the best-performing feature (evaluated using classification accuracy scores) from the available pool of features. These 10 features are listed in Table S1, available in the supplemental material to this article. They are specific to the context of Yasur Volcano and the timeframe of the 2016 deployment, and most do not have a clear physical meaning.

We additionally produce a full set of TSFRESH features from non-time-shifted waveforms (i.e., travel time is preserved), and, for comparison to the 10 SFS-selected TSFRESH features—a set of 10 features which have been employed for unsupervised ML classification on volcano infrasound (see, e.g., Watson, 2020; Witsil and Johnson, 2020). These previously used features comprise statistical moments and measures of the time- and frequency-domain amplitude distributions, as well as four additional frequency-domain features (for more information, see table 1 in Witsil and Johnson, 2020). We refer to these features as the “manual” features, and note that they have been successful in previous unsupervised clustering analyses (Watson, 2020; Witsil and Johnson, 2020).

Support vector classification

We train a linear SVM classifier, as implemented in scikit-learn (Pedregosa *et al.*, 2011), for each of the three extracted feature sets mentioned earlier. SVMs find the optimal hyperplane in feature space that maximizes the margin between the two classes of the training dataset. We initially chose SVMs over the myriad other options available, because (1) they have been applied to infrasound data before (Cannata *et al.*, 2011; Albert and Linville, 2020), and (2) they are a simple and easily interpretable classifier.

Temporal and station-wise subsets of the labeled catalog form our training and validation datasets. We formulate two classification problems to evaluate the performance of the SVM. We refer to these as the “generalization case” and



the “same-station” case. The generalization case trains an SVM on a subset comprised of all but one day and all but one station of the total labeled dataset. The validation subset consists of data from the excluded station on the excluded day. The same-station case uses the same temporal subsetting as the generalization case, but for station-wise subsetting we train and test on one station at a time, including training and testing on the same station. We only show results from a temporal subset of 27–31 July (train) and 1 August (test). For each classification problem, we balance classes, after temporally subsetting, by downsampling the majority class.

Results

Trends in labeled catalog

The labeled catalog of events shows clear trends that correlate with waveform characteristics (Fig. 2a,b). The relative contribution of events from the S and N subcraters is balanced, and the total number of events per hour is steady, until just before 29 July. The contribution from the S subcrater diminishes, and the total rate of events and average explosion amplitudes decline, until 30 July. At that time, the event rate returns to baseline, and the contribution from the S subcrater varies from 20% to 80%. Finally, just before 1 August, as explosion amplitudes increase dramatically, the contribution from the N subcrater shrinks to near zero.

Classification accuracy

We evaluate all classification problems using the classification accuracy score, which is the fraction of correct classifications out of the total number of classifications. For the generalization

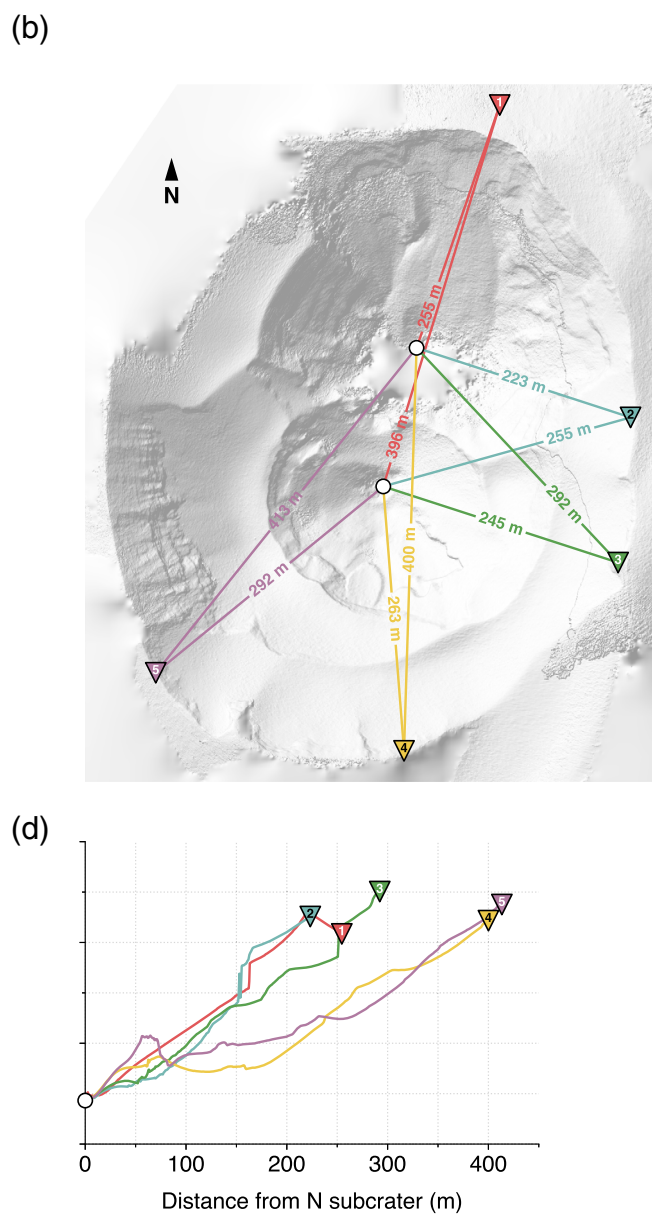
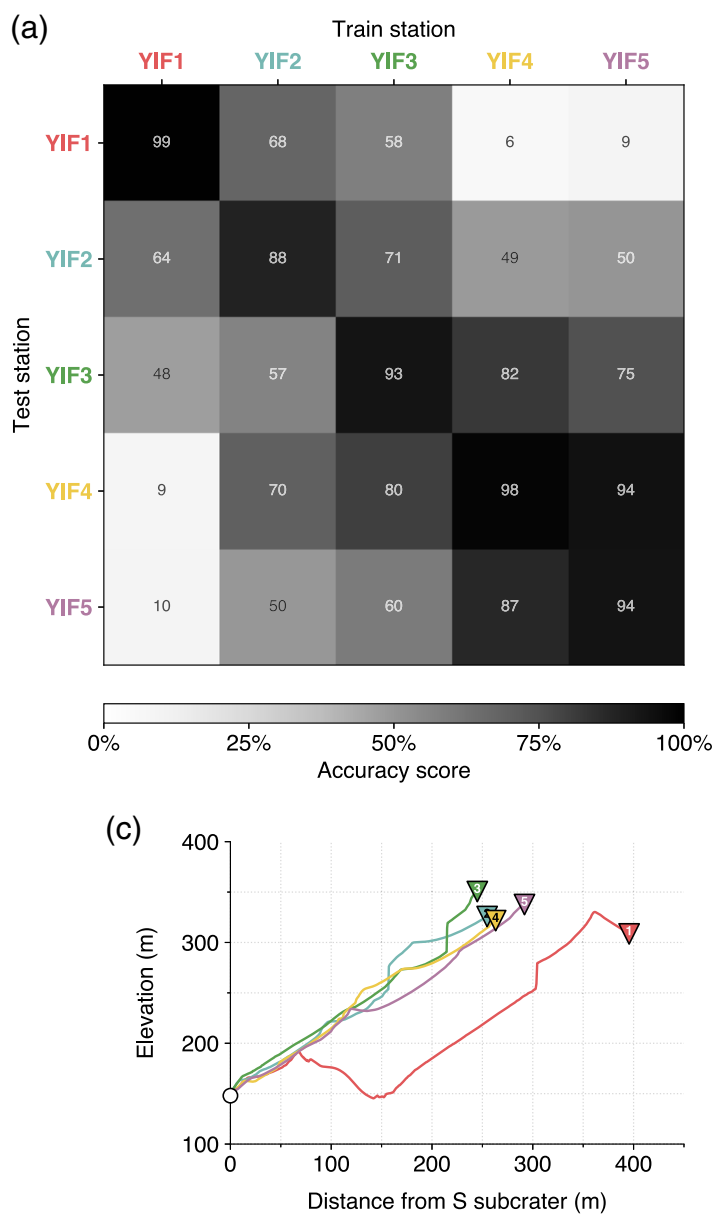
Figure 3. Accuracy scores for “generalization case” train–test subsets using (a) 10 manual features and (b) 10 sequential-feature-selected (SFS) TSFRESH features. Each cell’s score is obtained by training on all available data excluding the given day (column) and station (row) of the cell. For example, in panel (b), training on waveforms from stations YIF1 to YIF4 between 28 July and 1 August results in 85% accuracy on validation waveforms from YIF5 on 27 July.

case, we achieve mean classification accuracies of 73% (maximum 87%) for time-shifted manual features (Fig. 3a) and 77% (maximum 89%) for SFS-selected time-shifted TSFRESH features (Fig. 3b). The SFS-selected TSFRESH features have more uniform accuracy scores across the matrix, but especially during 1 August, when compared to the manual features.

Classification accuracies for the same-station case are shown in Figure 4a. We obtain a mean same-station (diagonal in Fig. 4a) classification accuracy of 96% using all non-time-shifted TSFRESH features, averaging over all six temporal train–test partitions. The maximum single-station score, 99%, occurs for station YIF1 when testing on 1 August, the final day of the deployment (top-left entry in Fig. 4a). Figure 4a shows the accuracies for a problem where we test on 1 August. The off-diagonal entries in Figure 4a, which represent the classifier’s generalization to different stations, generally show progressively poorer performance for stations that are farther away from each other. However, a correlation between the accuracies for stations YIF4 and YIF5 is evident.

Discussion

We show here that starting with a large collection of features and employing an automated feature selection scheme can



yield superior classification results when compared to using domain knowledge-derived (“manual”) features (Fig. 3). This automated feature extraction and selection scheme is superior, because it makes fewer assumptions about the system. However, we are still limited by the features built into TSFRESH. There are perhaps additional features that are relevant for a classification problem not contained within the TSFRESH feature set. This is where a deep learning approach may deliver a superior classification results, because such an approach avoids explicit feature specification altogether. One important trade-off, however, is interpretability; though outside the scope of this article, with linear SVMs we can

Figure 4. Accuracy scores and subcrater–station geometry and profiles. (a) Accuracy scores for various “same-station case” train–test pairings using non-time-shifted TSFRESH features. The training time window is 27–31 July; we test on 1 August. (b) Map view showing horizontal distances from each station to the bottom of each subcrater (subcrater locations are estimated from digital elevation model [DEM] minima). Vertical profiles from (c) the south subcrater and (d) the north subcrater to each station. Spatial scale is identical to (b). For station legend, see colors in (a).

examine the weights (hyperplane coefficients) of each feature to understand how the classifier is behaving. Furthermore, one can explore eruptive source processes and propagation effects

through the feature importances. This is more challenging, though not impossible, with deep learning algorithms.

Yasur Volcano is a dynamic system, and notable changes in the intensity of activity are observed even on the six-day-long timescale of this study (Fig. 2a,b). These are consistent with seismoacoustic analyses of Matoza *et al.* (2022). Furthermore, Yasur's bifurcated summit crater and its associated topographic features have a large impact on the character of recorded infrasound (Iezzi *et al.*, 2019; Fee *et al.*, 2021). Both of these volcano-specific factors pose challenges for classification, because a useful classification workflow inherently involves generalization either to new time periods or new station locations at the same volcano (or at a different volcano altogether). Furthermore, our catalog is automatically labeled, so there are likely misclassified events or noise (however, manual inspection of labeled waveforms did not indicate that such misclassifications were common). It is critical that these challenges can be addressed through clever training, design, and application of the classification algorithm. For example, training on a diverse collection of waveforms or leveraging data augmentation (Witsil *et al.*, 2022) can improve generalization.

The same-station classification results shown in Figure 4a demonstrate excellent generalization in time and poor generalization to other stations. Plots like Figure 4a produced for different test days show similar or superior results; this is because, by testing on the final day of the deployment when the eruptive pattern was different (Fig. 2a,b), we are challenging the algorithm to work on an eruptive period unlike what it encountered during training. Generalization to other stations is poor due to vast differences in topographic path between source and receiver. Figure 4b–d shows map and profile views of the paths from each subcrater DEM minimum to each station. Source–receiver distances and topographic complexity vary considerably. This has profound effects on generalization, which we illustrate with two examples:

1. Consider training on YIF1 and testing on YIF5 (Fig. 4a). The model learns to associate a smaller travel time with the N subcrater, as that subcrater is closer to YIF1. When we test this model using data from YIF5, the classification accuracy is 10%, that is, the algorithm picks the wrong subcrater 90% of the time. Examining Figure 4b we see that the smaller travel time for YIF5 is associated with the S subcrater, not the N subcrater.
2. There is a strong correlation between YIF4 and YIF5 accuracies (Fig. 4a). This is explained by the relative similarity of the path lengths between each subcrater and YIF4 and YIF5,

and the similar topographic profiles seen in Figure 4c,d. This similarity allows a model trained using YIF4 data to perform well on YIF5 data and vice versa; the topographic effect on the waveforms is similar for these similar paths.

Both of these examples suggest that the features used for this problem—non-time-shifted TSFRESH features—are strongly path dependent. (See Fig. S1 for a version of Fig. 4a made using time-shifted TSFRESH features.) These examples showcase the importance of feature extraction on classification results. This has implications for future ML studies in which feature engineering is necessary. Randomly time-shifting waveforms is a basic step toward removing the effect of path; a more rigorous approach could involve deconvolving the full numerically computed Green's function from each labeled explosion waveform. Such an approach would allow a classifier to train directly on features more closely linked to the explosion source, as opposed to path, and would help the classifier generalize to new network geometries. However, we note that to maximize classifier accuracy for a fixed path, waveform path information should be retained, because it provides additional information useful for determining the correct location (e.g., modification of waveform spectra by vent-proximal topography, Johnson *et al.*, 2018).

Using an SVM, we achieve classification accuracies on par with Cannata *et al.* (2011), who used a single station to achieve 95% accuracy classifying explosive events to either of Etna's southeast and northeast craters. Cannata *et al.* (2011) used a 3D feature space composed of frequency and quality factor and peak-to-peak amplitude. We discard amplitude information by normalizing each labeled waveform, but for same-station classification we retain travel-time information. Albert and Linville (2020) used SVMs to obtain 75% accuracy for a binary source-type classification (volcano vs. earthquake) problem. Large path differences between the globally recorded waveforms used in their study reduced the generalization performance of their algorithm. Although our study is not directly comparable due to its local scale and classification type, we also see a significant reduction in accuracy when we force the algorithm to generalize to new paths (e.g., compare Fig. 3 to Fig. 4a diagonal). Future improvements in ML-based volcano infrasound research could account for path effects through waveform modeling, as has been done in seismology (Kuang *et al.*, 2021) or through synthetic training data such as in Witsil *et al.* (2022).

To compare our ML-based classification to a more traditional method, we use the stacked waveforms shown in Figure 2c,d as templates for a correlation analysis (e.g., Green and Neuberg,

2006). Details on this method and results are available in the supplemental material. We obtain a mean same-station classification accuracy of 91%, which is lower than the accuracy of 96% obtained using ML. The key benefit of the ML application arises in situations in which generalization to other stations is desired, because this is not possible using waveform correlation methods.

Conclusions

We assemble a large, multistation, labeled dataset for a binary classification problem tasked with locating the source of explosive volcano infrasound signals within multiple subcraters using a crater rim infrasound network at Yasur Volcano. We experiment with three different strategies for extracting features from the labeled waveforms: 10 “manual” features, 10 automatically selected TSFRESH features, and a full set of non-time-shifted TSFRESH features. For each of these strategies, we explore SVM classification performance, and evaluate generalization potential for different time periods and different station combinations.

“Shallow learners” such as SVMs are readily interpretable, but they require explicitly defined features that are often obtained from domain knowledge; this can be arbitrary and can introduce bias. The choice of features depends on the classification goal (generalization, Fig. 3; or single-station performance, Fig. 4). The 10 SFS-selected, time-shifted TSFRESH features are better suited for studying source processes and their variation over time (Witsil and Johnson, 2020), whereas the full set of non-time-shifted TSFRESH features are more useful for reliable location of explosive activity using a single station. Feature engineering is the most important aspect of classification workflow design for these problems.

Our very high single-station performance could be exploited to locate explosions to a subcrater or vent using data from a single permanent infrasound station (once an SVM was trained with data from, e.g., a larger temporary network). The addition of a noise class would allow the algorithm, if running on rolling windows, to detect explosions as well as locate them. These applications are limited, however, in that any substantial change in crater morphology or atmosphere may violate the assumptions of the trained model.

This entire workflow (labeled dataset creation, training, and classification on new data) could be applied to other volcanoes or sources producing frequent explosions. Workflows such as this will only become more feasible and relevant, as infrasound data volumes increase and more readily labeled events are

observed. A more ambitious goal would involve the application of this automatic labeled dataset generation technique to a source type classification problem; for example, image processing on video data to create a catalog of event source types. This approach would then be feasible even for single-vent systems. Future work should additionally assess the performance of deep learners such as convolutional neural networks on large, labeled infrasound datasets such as the one we introduce here.

Data and Resources

The data from this deployment are available through the Incorporated Research Institutions for Seismology Data Management Center (IRIS DMC) as network code 3E and station codes YBAL, YIB2, YIF1, YIF2, YIF3, YIF4, YIF5, and YIF6 (Fee *et al.*, 2016). The Python code written to perform this work, including labeled dataset creation, is publicly available on GitHub at https://github.com/liamtoney/yasur_ml; the code relies heavily upon the Python seismological framework ObsPy (Beyreuther *et al.*, 2010). Figure 1 was made using PyGMT (v0.6.0; Wessel *et al.*, 2019; Uieda *et al.*, 2021). Figures 2–4 and Fig. S1 were made using Matplotlib (Hunter, 2007). The rtm Python package is available at <https://github.com/uafgeotools/rtm>. An outline of the features calculated by the TSFRESH Python package is available at https://tsfresh.readthedocs.io/en/latest/text/list_of_features.html. The supplemental material contains the methods for our cross-correlation analysis, supporting Table S1, and supporting Figure S1. All websites were last accessed in July 2022.

Declaration of Competing Interests

The authors declare that there are no conflicts of interest recorded.

Acknowledgments

The authors thank Leighton Watson and another anonymous reviewer for their helpful comments on this article. The authors thank the Vanuatu Meteorology and Geohazards Department (Esline Garaebiti, Sandrine Cevuard, Janvion Cevuard, Athanase Worwor, and Julius Mala) and the entire field team (David Fee, Robin Matoza, Alex Iezzi, Arthur Jolly, Richard Johnson, Bruce Christenson, Ben Kennedy, Geoff Kilgour, Nick Key, Rebecca Fitzgerald, Adrian Tessier, Allison Austin, and Di Christenson) for their efforts in obtaining the data used in this study. Robin S. Matoza acknowledges National Science Foundation (NSF) Grant Number EAR-1847736. This work was supported by the Nuclear Arms Control Technology (NACT) program at

the Defense Threat Reduction Agency (DTRA) through Contract Number HDTRA121C0030. Cleared for release.

References

- Albert, S., and L. Linville (2020). Benchmarking current and emerging approaches to infrasound signal classification, *Seismol. Res. Lett.* **91**, no. 2A, 921–929.
- Beyreuther, M., R. Barsch, L. Krischer, T. Megies, Y. Behr, and J. Wassermann (2010). ObsPy: A Python toolbox for seismology, *Seismol. Res. Lett.* **81**, no. 3, 530–533.
- Cannata, A., P. Montalto, M. Aliotta, C. Cassisi, A. Pulvirenti, E. Privitera, and D. Patanè (2011). Clustering and classification of infrasonic events at Mount Etna using pattern recognition techniques, *Geophys. J. Int.* **185**, no. 1, 253–264.
- Carniel, R., and S. Raquel Guzmán (2020). Machine learning in volcanology: A review, in *Volcanoes—Updates in Volcanology*, K. Nemeth (Editor), IntechOpen, London, United Kingdom.
- Christ, M., N. Braun, J. Neuffer, and A. W. Kempa-Liehr (2018). Time series Feature extraction on basis of scalable hypothesis tests (tsfresh—A Python package), *Neurocomputing* **307**, 72–77.
- Dempsey, D. E., S. J. Cronin, S. Mei, and A. W. Kempa-Liehr (2020). Automatic precursor recognition and real-time forecasting of sudden explosive volcanic eruptions at Whakaari, New Zealand, *Nat. Commun.* **11**, no. 1, doi: [10.1038/s41467-020-17375-2](https://doi.org/10.1038/s41467-020-17375-2).
- Fee, D., R. S. Matoza, and A. Jolly (2016). Yasur Volcano temporary deployment, summer 2016, doi: [10.7914/SN/3E_2016](https://doi.org/10.7914/SN/3E_2016).
- 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, *Front. Earth Sci.* **9**, doi: [10.3389/feart.2021.620813](https://doi.org/10.3389/feart.2021.620813).
- Ferri, F., P. Pudil, M. Hatef, and J. Kittler (1994). Comparative study of techniques for large-scale feature selection, in *Machine Intelligence and Pattern Recognition*, Vol. 16, Elsevier, Amsterdam, Netherlands, 403–413.
- Green, D. N., and J. Neuberg (2006). Waveform classification of volcanic low-frequency earthquake swarms and its implication at Soufrière Hills Volcano, Montserrat, *J. Volcanol. Geotherm. Res.* **153**, nos. 1/2, 51–63.
- Hunter, J. D. (2007). Matplotlib: A 2D graphics environment, *Comput. Sci. Eng.* **9**, no. 3, 90–95.
- Iezzi, A. M., D. Fee, K. Kim, A. D. Jolly, and R. S. Matoza (2019). Three-dimensional acoustic multipole waveform inversion at Yasur Volcano, Vanuatu, *J. Geophys. Res.* **124**, no. 8, 8679–8703.
- Johnson, J. B., L. M. Watson, J. L. Palma, E. M. Dunham, and J. F. Anderson (2018). Forecasting the eruption of an open-vent volcano using resonant infrasound tones, *Geophys. Res. Lett.* **45**, no. 5, 2213–2220.
- Jolly, A. D., R. S. Matoza, D. Fee, B. M. Kennedy, A. M. Iezzi, R. H. Fitzgerald, A. C. Austin, and R. Johnson (2017). Capturing the acoustic radiation pattern of strombolian eruptions using infrasound sensors aboard a tethered aerostat, Yasur Volcano, Vanuatu, *Geophys. Res. Lett.* **44**, no. 19, 9672–9680.
- Kong, Q., D. T. Trugman, Z. E. Ross, M. J. Bianco, B. J. Meade, and P. Gerstoft (2019). Machine learning in seismology: Turning data into insights, *Seismol. Res. Lett.* **90**, no. 1, 3–14.
- Kuang, W., C. Yuan, and J. Zhang (2021). Real-time determination of earthquake focal mechanism via deep learning, *Nat. Commun.* **12**, no. 1, doi: [10.1038/s41467-021-21670-x](https://doi.org/10.1038/s41467-021-21670-x).
- Li, M., X. Liu, and X. Liu (2016). Infrasound signal classification based on spectral entropy and support vector machine, *Appl. Acoust.* **113**, 116–120.
- Liu, X., M. Li, W. Tang, S. Wang, and X. Wu (2014). A New classification method of infrasound events using Hilbert-Huang transform and support vector machine, *Math. Probl. Eng.* **2014**, doi: [10.1155/2014/456818](https://doi.org/10.1155/2014/456818).
- Matoza, R. S., B. A. Chouet, A. D. Jolly, P. B. Dawson, R. H. Fitzgerald, B. M. Kennedy, D. Fee, A. M. Iezzi, G. N. Kilgour, E. Garaebiti, and S. Cevuard (2022). High-rate very-long-period seismicity at Yasur volcano, Vanuatu: Source mechanism and decoupling from surficial explosions and infrasound, *Geophys. J. Int.* **187**, no. 4, 717–725.
- Ortiz, H. D., R. S. Matoza, C. Garapaty, K. Rose, P. Ramón, and M. C. Ruiz (2020). Multi-year regional infrasound detection of Tungurahua, El Reventador, and Sangay volcanoes in Ecuador from 2006 to 2013, *Proc. Mtgs. Acoust.* **41**, no. 1, doi: [10.1121/2.0001362](https://doi.org/10.1121/2.0001362).
- Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, et al. (2011). Scikit-learn: Machine learning in Python, *J. Mach. Learn. Res.* **12**, 2825–2830.
- Simons, B., A. Jolly, J. Eccles, and S. Cronin (2020). Spatiotemporal relationships between two closely-spaced strombolian-style vents, Yasur, Vanuatu, *Geophys. Res. Lett.* **47**, no. 5, 1–10.
- Uieda, L., D. Tian, W. J. Leong, M. Jones, W. Schlitzer, M. Grund, L. Toney, J. Yao, Y. Magen, K. Materna, et al. (2021). PyGMT: A Python interface for the Generic Mapping Tools, *Zenodo*, doi: [10.5281/zenodo.6349217](https://doi.org/10.5281/zenodo.6349217).
- Walker, K. T., M. A. H. Hedlin, C. de Groot-Hedlin, J. Vergoz, A. Le Pichon, and D. P. Drob (2010). Source location of the 19 February 2008 Oregon bolide using seismic networks and infrasound arrays, *J. Geophys. Res.* **115**, no. B12, doi: [10.1029/2010JB007863](https://doi.org/10.1029/2010JB007863).
- Watson, L. M. (2020). Using unsupervised machine learning to identify changes in eruptive behavior at Mount Etna, Italy, *J. Volcanol. Geotherm. Res.* **405**, 107042, doi: [10.1016/j.jvolgeores.2020.107042](https://doi.org/10.1016/j.jvolgeores.2020.107042).
- Wessel, P., J. F. Luis, L. Uieda, R. Scharroo, F. Wobbe, W. H. F. Smith, and D. Tian (2019). The Generic Mapping Tools Version 6, *Geochem. Geophys. Geosys.* **20**, no. 11, 5556–5564.
- Witsil, A., D. Fee, J. Dickey, R. Pena, R. Waxler, and P. Blom (2022). Detecting large explosions with machine learning models trained on synthetic infrasound data, *Geophys. Res. Lett.* **49**, no. 11, doi: [10.1029/2022GL097785](https://doi.org/10.1029/2022GL097785).
- Witsil, A., and J. B. Johnson (2020). Analyzing continuous infrasound from Stromboli volcano, Italy using unsupervised machine learning, *Comput. Geosci.* **140**, 104494, doi: [10.1016/j.cageo.2020.104494](https://doi.org/10.1016/j.cageo.2020.104494).

Manuscript received 16 May 2022
Published online 28 July 2022