Examining the Evolution of Seismicity in Volcanic Eruptions

CSE 6242: Final Report

Alex Dreo, Bear Jordan, Min Gu Kang, David Koch,
Juanwen Lu, and Jeanette Schofield

April 22, 2022

1 Introduction

Of the thousands of volcanoes on Earth, only a handful are actively monitored. Manual data analysis using highly-trained researchers requires a large time investment and prevents more volcanoes from being monitored.

The present project seeks to classify seismic events generated by a volcano using machine learning (ML) models to streamline the data pipeline and allow more volcanoes to be monitored. This study will compare the efficacy of several data treatments and classification models on a single dataset to select the highest performer [HQ1].

1.1 Literature Survey

Currently, volcanoes are monitored using a mix of seismicity, ground deformation, volcanic gases, and water level measurements. This data allows researchers to predict volcanic events [1]. Of these, geophysical techniques provide the most information on the evolution of a volcanic system [2, 3].

Specifically, there are four types of volcanic earthquakes of interest: high-frequency or volcano-tectonic (VT) events, low-frequency or long-period events (LP), tremors, and explosion earthquakes [4]. Classifying the type of earthquake provides information about what is happening within the volcanic system [5, 6]. High-frequency earthquakes occur in swarms as fluid filled fractures propagate in the volcanic system [7]. Long-period earthquakes are related to changes in fluid pressure within the magma and are a strong indicator of magma being enriched in volatiles—a main driver of eruptions [8]. Exsolution of these volatiles and bubbling cause the random turbulence found in tremor waveforms [9]. Finally, explosion earthquakes occur as acoustic waves generated by the rapid expansion of volatiles during an eruptive event [10].

Volcanoes progress in a sequence of various events: high-frequency events where magma moves in and forms the magma chamber, long-period events where volatiles exsolve in magma, tremor events where the magma moves to shallow depths, and explosive earthquake events when eruption occurs. As expected, the frequency of these events occurring is not uniform. Magma at shallow depths constantly generate bubbles as gasses exsolve. Long-period earthquakes are therefor common. In contrast, it is more rare for magma to generate enough internal pressure to rupture the surrounding rock. High-frequency events are more rare as a result [4, 11].

Researchers historically took advantage of this progression by manually classifying volcanic earthquakes. For example, Jiménez and others manually examined 1,700 earthquakes leading to the 1982 El Chichón Volcano eruption to generate an understanding of the seismic fingerprint leading up to an eruption at the volcano [12] [HQ2].

With the time-cost of manually classifying an earthquake, several researchers have turned to ML methodologies [13]. Dr. Lara-Cueva and others have produced one of the largest ML training datasets of volcanic earthquakes by hand labeling seismograms over the period of nearly 10 years from the Cotopaxi volcano in Ecuador. The group has performed foundational work in extracting features from raw waveforms, and implemented ML algorithms like kNN and decision tree classifiers to identify long-period earthquakes [14, 15]. When contacted about the present project, Dr. Lara-Cueva was generous enough to share his training data.

Several other researchers have investigated this approach with other ML models, new feature extraction techniques, and their applications to different volcanoes [16, 17, 18]. Researchers have looked towards artificial neural networks, DT, multi-class SVM, linear SVM, GMM, Hidden Markov Models to identify high performers [19, 20, 21]. Since waveforms are noisy and unstructured, empirical mode decomposition (EMD) has emerged as a key feature extraction technique that works by producing several characteristic intrinsic mode frequencies that when added together produce a signal like the input [22, 23, 24, 25].

Volcanic eruptions occur infrequently over the course of many thousands of years. Data collection related to these events has occurred only in very recent history. While the dataset we have is quite large, it only represents a small fraction of all volcanic eruptions. Thus, it is potentially-biased towards the recent state of Cotopaxi. Another potential drawback is that many physically-based models assume underlying mechanics are the same for all eruptions. This may not always be the case [26].

2 Method

There are several primary innovations this project seeks to implement: the use of EMD to transform seismic data, building a classification model on the transformed data to compare and contrast with literature survey models, adding a user interface for visualizing data, and investigating other noise reduction and signal reduction methods.

Current papers in this field put more emphasis on finding well-fitted classification models while the discussion on removing noise in the signal is limited. EMD is an effective technique widely used for decomposing signals into critical components in the analysis of non-linear and non-stationary signals. Such signal properties are also found in seismic signals which are the focus area of this project. Through the process of EMD and the well-designed process of Clear Iterative Interval Thresholding (CIIT) [27, 28], the noise embedded in the signals could be reduced using a soft threshold created for each EMD component while the important information from the signals, such as the trend, is retained. The advantage of using this EMD-CIIT method is extraction of features using less noisy signals and then building a classification model to identify only the most significant differences among types.

In the design of this project, there are two phases. In Phase 1, EMD-CIIT method is introduced to process the raw signals and to filter the noise of signal. This EMD-CIIT process includes several steps.

1. EMD is implemented and the signal with noise is decomposed into its subsequent components or Intrinsic Mode Functions (IMF).
2. Each IMF is denoised using soft thresholding. The soft thresholding is relevant to median IMF value and sample size. Any value within the range of +/- soft threshold is set to be zero and any value outside the range is reduced by the proportion of the difference between absolute value of the IMF extrema and the threshold.
3. The denoised version of the 1st IMF is subtracted from the original 1st IMF to obtain its noise-only version.
4. The signal is reconstructed using all IMFs by adding them except the 1st IMF.
5. The signals from Steps 2 and 4 are combined.
6. The positions of data from the noise-only part of 1st IMF are changed randomly. This restructured noise-only data set is combined with the output from Step 5.
7. Again, the newly reconstructed signal is decomposed from the output in Step 6 using EMD.
8. Repeat Steps 2 to 7 until M sets of reconstructed signal are obtained. M signals are averaged to create the denoised version of the signal.

In Phase 2, the filtered data is then processed to extract three key features proposed in the paper [14] as the attributes to train the models. To compare the performance among different models, supervised models (KNN, random forest, SVM, and logistic regression) and unsupervised model (k-means) are implemented using training set and evaluated using testing set with confusion matrix and accuracy.

One challenge we faced while creating our classification model is that our data set is extremely imbalanced. In our data set, the most frequent type of earthquake is LP, which accounts for 87.95% of the data. Second most frequent type is VT, which accounts for 8.51% of the data. The rest of the types combined accounts for less than 4% of the entire dataset. This imbalance in data could cause our model to have a poor performance when classifying the classes that are in the minority. Since it is critical for our model to correctly classify even the minority classes, we needed a method to re-balance the classes.

For Phase 3, the technique we applied to our model is Synthetic Minority Oversampling Technique (SMOTE). SMOTE creates synthetic data points randomly along the lines that connects two nearest neighbor data points in same minority class. The process is repeated until desired balance of the majority and minority class is reached. This technique allows us to simulate more training points without collecting additional data for the under-represented classes.

2.1 Data Engineering and Transformation

Due to the nature of our complex data preprocessing and our team size of six, splitting our code into a preprocessing notebook and a modeling/evaluation notebooks was crucial. It allowed us to divide tasks, avoid source control merge conflicts, and avoid running the lengthy transformation code multiple times. This separation of concerns, as opposed to a monolithic notebook, is a tenant of modern software engineering and would be essential to productionalizing our code if that became necessary.

2.2 Visualization

Educational outreach is intrinsic to volcanology because it directly impacts public response to hazards. The better these phenomena are understood, the more likely the public is to react appropriately. While there are many good educational resources for volcanoes, there are not many data visualizations about them — let alone about volcanic seismicity. Our visualization will be geared towards educating a general audience, with the goal of helping them understand the evolution of volcanic seismicity.

Specifically, the expected results from the above procedure will be a dataset for each classification scheme that includes the date, duration, and type of earthquakes. Earthquake type will be visually encoded to the hue of the circle, duration will be encoded to the radius of the circle, and the date will be encoded to the x-position. For the visualization to be interactive, this project adapt D3's force graph library to produce a beeswarm plot. The main innovation here is the ability to encode data to the radius of the circle (something not allowed in D3's standard beeswarm).

3 Experiments/Evaluations

3.1 Analysis of EMD-CIIT

With the goal of evaluating the performance of the EMD-CIIT procedure to remove noise before creating classification models, the process design includes filtering noise in raw data using EMD-CIIT, separating the data set into a training set (80%) and test set (20%), training the classification models, and calculating the accuracy using testing set. The raw seismic signal data are from Cotopaxi volcano. Each signal sample contains 2,000-5,000 data points and its event type. Most of the types are long-period and volcano-tectonic event classes. Regional, hybrid, and icequakes types have a minor presence. In order to evaluate the performance difference due to the number of the iteration, four scenarios are implemented and compared: using the raw signal, and the signal after two, four, and six iterations (Fig. 1). With signals output from these scenarios, the value of three important features are extracted using the definition in [13, 14]: the maximum threshold frequency, maximum peak in the frequency domain, and the entropy.

3.2 Analysis of Model Performance

Supervised and non-supervised models are then trained using the normalized training data set from raw signals and the signal data after two, four, and six iterations. Accuracy is evaluated using normalized testing data set.

Accuracy is not significantly improved when the filtering process is considered (Fig. 2). From an overall perspective, most event types have a similar accuracy value when rounded to two decimals. The number of samples predicted in certain types of events changes slightly. One of the reasons might be the sample size of LP and VT (8%, 100 samples) event type in the original data set and the model might not get enough data points to get

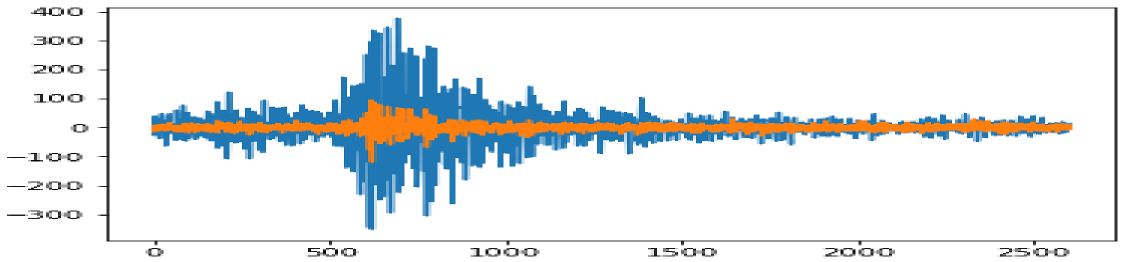


Figure 1: Raw signal of one sample in blue and its average denoised version after six iterations of EMD-CIIT procedure

Scenario	KNN	SVM	RF	Logistics Regression	K-means
Original Data	84.87%	84.87%	86.13%	85.29%	23.84%
EMD-2 iterations	84.87%	84.87%	84.87%	85.29%	6.32%
EMD-4 iterations	84.87%	84.87%	84.87%	85.29%	6.32%
EMD-6 iterations	84.87%	84.87%	86.55%	85.29%	29.74%

Figure 2: Classification accuracy comparison among 4 scenarios

trained. The second reason might be the randomized locations of the points (the variance between the original IMF1 and the denoised version of IMF1) in each iterations and such design introduces uncertainty into the system. Another reason is that in the design of EMD-CIIT procedure: in each iteration, it adds random noise from 1st IMF to represent the changes in original signals. After extracting the feature value from the average denoised version after two, four, and six iterations, a large amount of data points assumed to be noise remained. Output from 6-iteration scenario obtains accuracy improvements compared with the 2-iteration scenario. This is because the value of noise part becomes smaller as more the number of iterations increase. At the same time, the average value becomes smaller.

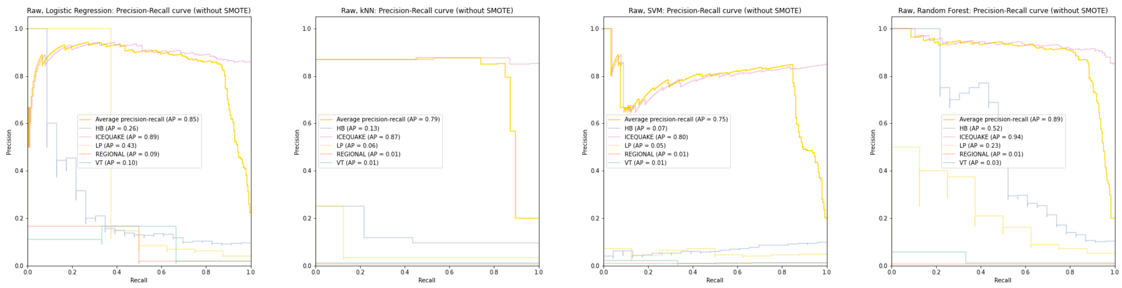
For classification models, relying solely on accuracy as a performance metric can give misleading results. This is especially true if a training set is unbalanced, as ours is. For this reason, precision and recall were calculated for our one-hot-encoded classification data. For the raw data (i.e. without any decomposition), the precision-recall curves were then plotted below without and with SMOTE (Figs. 3 and 4). This process was then repeated on the EMD-CIIT data with six iterations (Fig. 5). This seemed to have very little effect on models' precision/recall.

Differences among the classifiers are also apparent in their results (Fig. 2). Random Forest (RF) achieves better and more stable performance than other classifiers. k-Means performed far worse than the other classifiers used. Such output might result from the insignificant point-to-point distance in the hyperspace due to extracting the features from normalized data. The actual data label in the classifier enables the models to find direction in improving accuracy.

3.3 Analysis of SMOTE

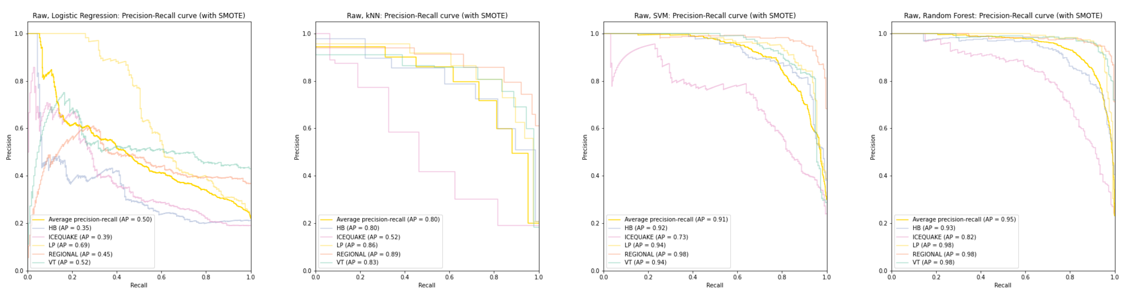
We can clearly see the benefit of correcting our data set for class imbalances, using SMOTE (Figs. 3 and 4). This is apparent from the increase in precision and recall of under-represented classes, i.e. all classes except LP. Logistic regression (LR) clearly performed the worst, with low recall and precision across most classes. KNN also seemed to perform worse than both SVM and RF models (Figs. 4 and 5).

Another clear trend is the below average performance of ICEQUAKE class across all models. Poor performance for these models on the ICEQUAKE class could be an sign that this generalized dataset is missing a key feature specific to icequakes. Creating an icequake specific model, with additional feature(s) that can explain seismic characteristics specific to an icequake could improve performance.



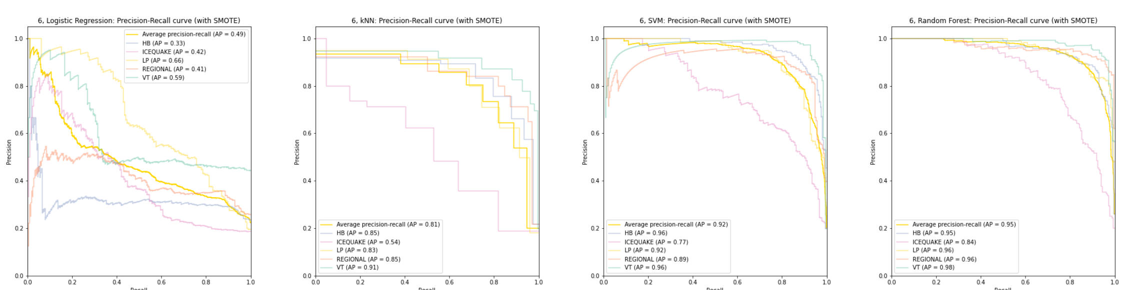
Logistic Regression kNN SVM Random Forest

Figure 3: Precision-Recall Curves for Raw Data (without SMOTE)



Logistic Regression kNN SVM Random Forest

Figure 4: Precision-Recall Curves for Raw Data (with SMOTE)



Logistic Regression kNN SVM Random Forest

Figure 5: Precision-Recall Curves after 6 EMD Iterations (with SMOTE)

3.4 Visualization Results

The visualization produced by this study adapted D3's force-layout in order to produce small multiples of a beeswarm plot (Fig. 6). This allows for quick comparison between how different classification models predicted the class of seismic waveforms. By encoding the color to the type of earthquake, viewers are able to see the distribution and imbalance within the seismicity of volcanoes.

Several options for exploration and analysis are offered in the visualisation. The drop-down menu in the lower left offers a selection of the data treatment options—Raw data, EMD-CIIT with two, four, and six iterations. This allows for the user to directly see how different treatment methods effected the results of the study.

The visualization also considers the performance of the output. In the lower left, users can select a toggle that dynamically transitions the data to cluster into two columns. One representing cases where the models predicted the true class, and one where the models did not predict the true class. This allows users to quickly view the relative performance of each model.

A few interesting trends can be observed in the visualization. While only looking at the classification data, the majority of models predict LP events as the majority class. This is consistent with real world data. K-Means, however, shows a wide diversity of earthquake predictions that differ from the other models. From the visualization alone, it is clear that

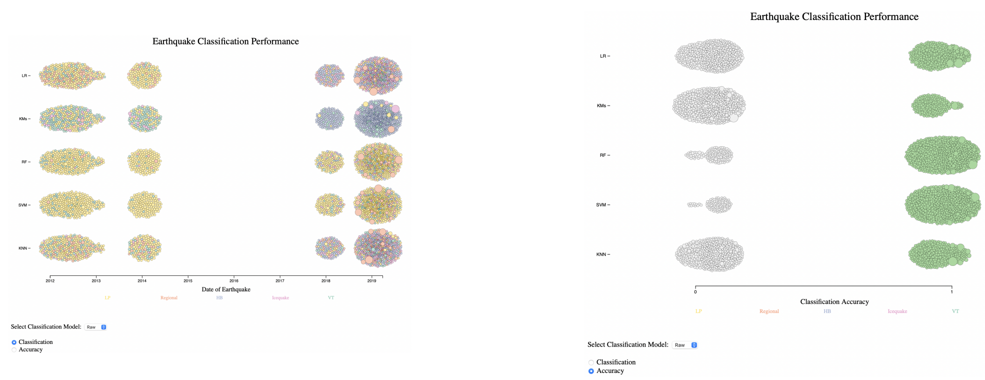


Figure 6: (a) This visualization shows the predicted classification of earthquakes for several different ML models allowing for easy comparison. (b) When the accuracy toggle is selected, the visualization updates to show the number of correctly identified earthquakes.

k-Means performs worse than other models.

Another trend observed is an omission of data around 2015. Unfortunately for the people living close to Cotopaxi, the volcano entered a period of extreme unrest, resulting in an eruption that started in August and lasted into 2016 [25]. As volcanoes enter the pre-eruptive stage, volcanic earthquakes regularly occur simultaneously increasing the complexity of the denoising and classification problem. As a result, these data were omitted in the Lara-Cueva dataset.

4 Conclusion

The precison-recall curves (Figs. 3-5) clearly show that SMOTE greatly improves performance on the imbalanced training set. Meanwhile, while EMD-CIIT shows negligible improvement on the same dataset. Since actual earthquake classification distribution may not be strictly balanced, working with imbalanced data is a problem worth spending time on.

Our results can also be viewed through our d3 visual, which is intended as a tool to help both explore the data, as well as analyze model performance. These results show that using machine learning (ML) models to classify volcanic earthquakes could be critical to the development of comprehensive volcanic hazard assessments. This ML approach would alleviate current pain points, such as the need to manually analyze the data, as well as reduce amount of domain specific expertise to perform analysis.

4.1 Team Effort

All group members contributed to this project equally.

4.2 Future Work

Given more time and resources, numerous areas that would benefit from further exploration. One such area is infrasound data. Acoustic signals in this spectrum provide insight as to when degassing events are occurring, which can in turn provide reliable, cheap, and near real-time indicators of volcanic activity [29]. By applying the same data treatment and classification techniques, researchers can develop comprehensive datasets of regional volcanic activity to monitor states of unrest.

Streaming data is another area worth pursuing. Steaming data can significantly reduce the lag time between a seismic event occurring and data being collected, leading to faster classification and prediction of seismic events. In the days leading up to an eruption, real-time monitoring is critical to enable local authorities to keep their communities safe.

References

[1] Sparks R and Steve J. "Forecasting volcanic eruptions". In: *Earth and Planetary Science Letters* 210.1-2 (2003), pp. 1–15.

[2] RI Tilling. "The critical role of volcano monitoring in risk reduction". In: *Advances in Geosciences* 14 (2008), pp. 3–11.

[3] Bruno Massa et al. "Determining the stress field in active volcanoes using focal mechanisms". In: *Frontiers in Earth Science* 4 (2016), p. 103.

[4] Stephen McNutt. "Volcanic seismology". In: *Annu. Rev. Earth Planet. Sci.* 32 (2005), pp. 461–491.

[5] John C Lahr et al. "Earthquake classification, location, and error analysis in a volcanic environment: Implications for the magmatic system of the 1989–1990 eruptions at Redoubt Volcano, Alaska". In: *Journal of Volcanology and Geothermal Research* 62.1-4 (1994), pp. 137–151.

[6] David N Green and Jürgen Neuberg. "Waveform classification of volcanic low-frequency earthquake swarms and its implication at Soufrière Hills Volcano, Montserrat". In: *Journal of Volcanology and Geothermal Research* 153.1-2 (2006), pp. 51–63.

[7] Laura Becerril et al. "Depth of origin of magma in eruptions". In: *Scientific reports* 3.1 (2013), pp. 1–6.

[8] Oleg Melnik et al. "Deep long period volcanic earthquakes generated by degassing of volatile-rich basaltic magmas". In: *Nature communications* 11.1 (2020), pp. 1–7.

[9] D Seidl, R Schick, and M Riuscetti. "Volcanic tremors at Etna: a model for hydraulic origin". In: *Bulletin volcanologique* 44.1 (1981), pp. 43–56.

[10] MT Hagerty et al. "Analysis of seismic and acoustic observations at Arenal Volcano, Costa Rica, 1995–1997". In: *Journal of Volcanology and Geothermal Research* 101.1-2 (2000), pp. 27–65.

[11] Thomas R Walter. "How a tectonic earthquake may wake up volcanoes: Stress transfer during the 1996 earthquake–eruption sequence at the Karymsky Volcanic Group, Kamchatka". In: *Earth and Planetary Science Letters* 264.3-4 (2007), pp. 347–359.

[12] Z Jiménez, VH Espindola, and JM Espindola. "Evolution of the seismic activity from the 1982 eruption of El Chichón Volcano, Chiapas, Mexico". In: *Bulletin of Volcanology* 61.6 (1999), pp. 411–422.

[13] Noel Pérez et al. "ESeismic: Towards an Ecuadorian volcano seismic repository". In: *Journal of Volcanology and Geothermal Research* 396 (2020), p. 106855.

[14] Roman Lara-Cueva et al. "Comparative analysis of automated classifiers applied to volcano event identification". In: *2016 ieee Colombian conference on communications and computing (colcom)*. IEEE. 2016, pp. 1–6.

[15] RA Lara-Cueva et al. "Feature selection of seismic waveforms for long period event detection at Cotopaxi Volcano". In: *Journal of Volcanology and Geothermal Research* 316 (2016), pp. 34–49.

[16] Marielle Malfante et al. "Machine learning for volcano-seismic signals: Challenges and perspectives". In: *IEEE Signal Processing Magazine* 35.2 (2018), pp. 20–30.

[17] Román Lara-Cueva et al. "On the use of multi-class support vector machines for classification of seismic signals at Cotopaxi volcano". In: *2017 IEEE International Autumn Meeting on Power, Electronics and Computing (ROPEC)*. IEEE. 2017, pp. 1–6.

[18] Román Lara-Cueva et al. "On finding possible frequencies for recognizing microearthquakes at Cotopaxi volcano: A machine learning based approach". In: *Journal of Volcanology and Geothermal Research* 407 (2020), p. 107092.

[19] Pablo Venegas et al. "Building machine learning models for long-period and volcano-tectonic event classification". In: *2019 IEEE CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON)*. IEEE. 2019, pp. 1–6.

[20] Pablo Venegas et al. "Combining filter-based feature selection methods and gaussian mixture model for the classification of seismic events from cotopaxi volcano". In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 12.6 (2019), pp. 1991–2003.

[21] Gloria Curilem et al. "Classification of seismic signals at Villarrica volcano (Chile) using neural networks and genetic algorithms". In: *Journal of volcanology and geothermal research* 180.1 (2009), pp. 1–8.

[22] Ya-juan Xue et al. "Application of local wave decomposition in seismic signal processing". In: *Earthquakes-Tectonics, Hazard and Risk Mitigation*. IntechOpen, 2017.

[23] AO Boudraa, JC Cexus, and Z Saidi. "EMD-based signal noise reduction". In: *International Journal of Signal Processing* 1.1 (2004), pp. 33–37.

[24] Guillermo Cortés et al. "A comparative study of dimensionality reduction algorithms applied to volcano-seismic signals". In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 9.1 (2015), pp. 253–263.

[25] Noel Perez et al. "A new volcanic seismic signal descriptor and its application to a data set from the cotopaxi volcano". In: *IEEE Transactions on Geoscience and Remote Sensing* 58.9 (2020), pp. 6493–6503.

[26] Grace Manley et al. "Machine Learning Approaches to Identifying Changes in Eruptive State Using Multi-Parameter Datasets From the 2006 Eruption of Augustine Volcano, Alaska". In: *JGR Solid Earth* 126 (2021).

[27] Yannis Kopsinis and Stephen McLaughlin. "Development of EMD-based denoising methods inspired by wavelet thresholding". In: *IEEE Transactions on signal Processing* 57.4 (2009), pp. 1351–1362.

[28] Hassan Ashraf et al. "Threshold parameters selection for empirical mode decomposition-based EMG signal denoising". In: *INTELLIGENT AUTOMATION AND SOFT COMPUTING* 27.3 (2021), pp. 799–815.

[29] Silvio De Angelis et al. "Uncertainty in Detection of Volcanic Activity Using Infrasound Arrays: Examples From Mt. Etna, Italy". In: *Frontiers in Earth Science* 8 (2020). ISSN: 2296-6463. DOI: 10.3389/feart.2020.00169. URL: https://www.frontiersin.org/article/10.3389/feart.2020.00169.

[30] Michael Manga and Emily Brodsky. "Seismic triggering of eruptions in the far field: Volcanoes and geysers". In: *Annu. Rev. Earth Planet. Sci* 34 (2006), pp. 263–291.

Appendix A: Heilmeier Question Locations

1. What are you trying to do? Articulate your objectives using absolutely no jargon.
   Problem Definition.

2. How is it done today; what are the limits of current practice?
   Literature Survey.

3. What's new in your approach? Why will it be successful?
   Plan of Activities.

4. Who cares?
   Plan of Activities.

5. If you're successful, what difference and impact will it make, and how do you measure them (e.g., via user studies, experiments, ground truth data, etc.)?
   Plan of Activities.

6. What are the risks and payoffs?
   Plan of Activities.

7. How much will it cost?
   Plan of Activities.

8. How long will it take?
   Plan of Activities.

9. What are the midterm and final "exams" to check for success? How will progress be measured?
   Plan of Activities.