Heartbeat of a Volcano: Detecting Seismicity in the Valles Caldera, New Mexico, Using Machine Learning

Yair Franco[1], Luke Frazeur[2], Nadine Igonin[2], Akram Mostafanejad[3] and John F. Ferguson[2]
(1)University of California, Berkeley, (2)University of Texas at Dallas, (3)EarthScope

Poster #1292954
AGU 2023

Abstract

The Valles Caldera is a dormant volcanic complex in northern New Mexico which began forming 1.6 Ma, and last erupted 52,000 years ago. Despite its very recent eruptive history, the current state of volcanic and seismic activity within the Valles Caldera is not well understood. Recent studies aiming to gain a better understanding of the current inner structure of the caldera have provided some evidence of seismic activity in the system, but results so far are only preliminary. This study presents observations made with data collected from Fairfield nodes loaned from the PASSCAL Instrument Center, installed in the Valles Caldera in 2019 during the Summer of Applied Geophysical Experience (SAGE), which made measurements for four weeks. Data was processed with several different computational methods: Obspy's STA/LTA coincidence trigger function, the Seisbench library of machine learning methods, and EQCCT, a newly developed machine learning method. **Our findings indicate that the Valles Caldera is still seismically active.** Furthermore, we show how to overcome false detections from machine learning detection by using a combined time and frequency-domain approach.

Background and Data

Area of study: **Valles Caldera**, New Mexico
- Roughly 20 x 18 km area

Data:
- Four arrays of Fairfield ZLAND 3C nodes (58 stations)
- Collected June 1-29, 2019[2]

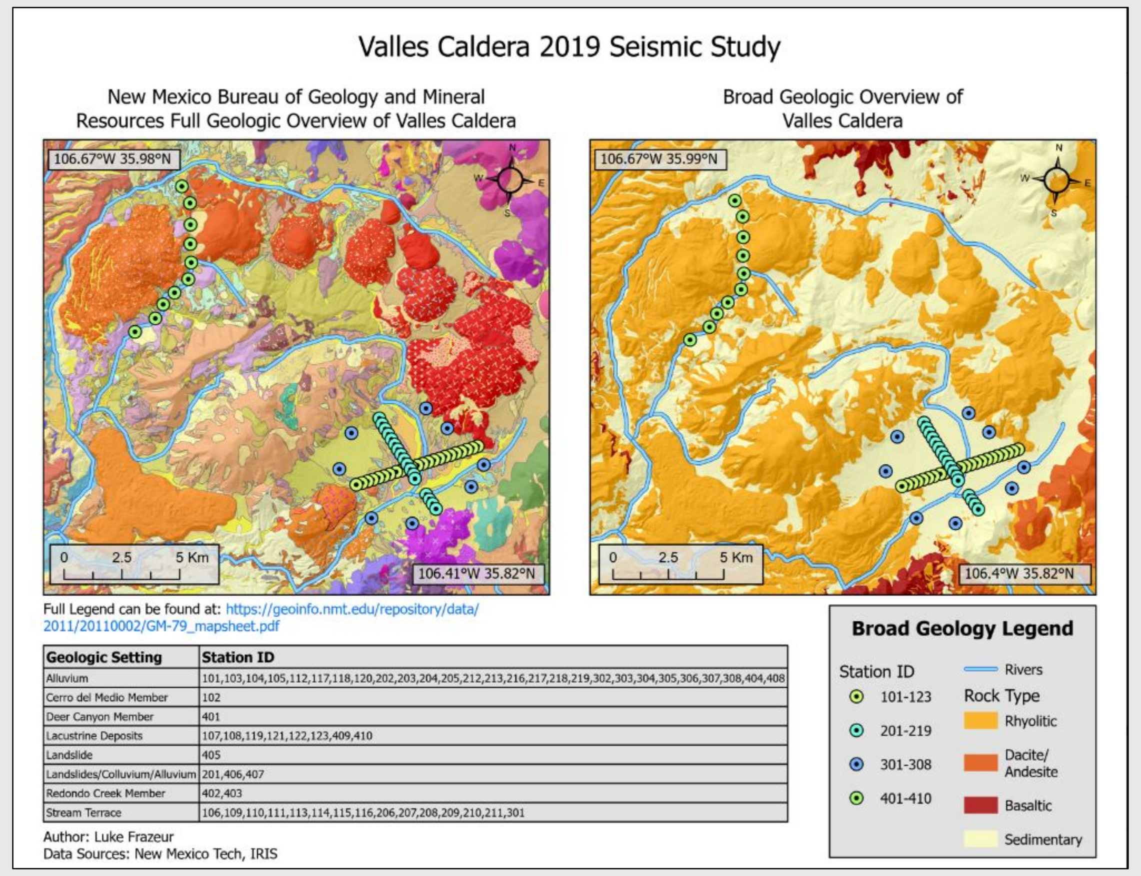


Figure 1: Map figure depicting geophone locations and geology of the region. The map on the left shows the detailed geologic map developed by the NM Bureau of Geology and Mineral Resources. The map on the right is a simplified derivation developed by the author. Geophones are mostly set in sedimentary rock which can account for slightly slower wave speeds, however the subsurface is still mostly igneous. Stations 401-410 are separated from the others by over 10 kilometers.

Method & Challenges

Three earthquake detection methods:
1. **Obspy STA/LTA**; baseline for detections.
2. **Seisbench EQTransformer**[3]; machine learning method for P and S-wave picks.
   - Pretrained models compared: ETHZ, SCEDC, LenDB, and STEAD (Figure 2). STEAD chosen as best model.
3. **EQCCT**[4]: New method developed at TexNet. Compared with STA/LTA and EQT.

Detection workflow:
a. Run all three models on geophone data.
b. Compare **coincident** phase detections.
   - *High coincidence between STA/LTA and correctly ordered phase picks (i.e. P-wave picks before S picks) corresponds to higher probability for finding an event.*

c. Confirm event nature **empirically**: earthquake vs. lightning vs. random noise, etc.

Frequency analysis:
- **Frequency analysis** was used to distinguish between earthquakes and false detections, with two approaches: 1) clustering, 2) cross-correlation
- Cross correlation-based approach: a Master Spectra Template (MST) was made for the confirmed lightning and earthquake events. Remaining events were correlated to each template.
- Clustering approach: Using Python's Scikit-Learn package. Events were clustered between 3 groups and 4 groups.

Monsoon detections:
- All models had many false detections due to **lightning** (Figure 3).
- Lightning confirmed empirically from waveforms and radar archive[5].
- Direction of origin estimated from station position.

All event origins estimated from station position and velocity estimates (P ~ 3 km/s, S ~ 5.5 km/s). Velocities estimated based on arrival times for USGS-catalogued earthquakes (Figure 7), and station position.

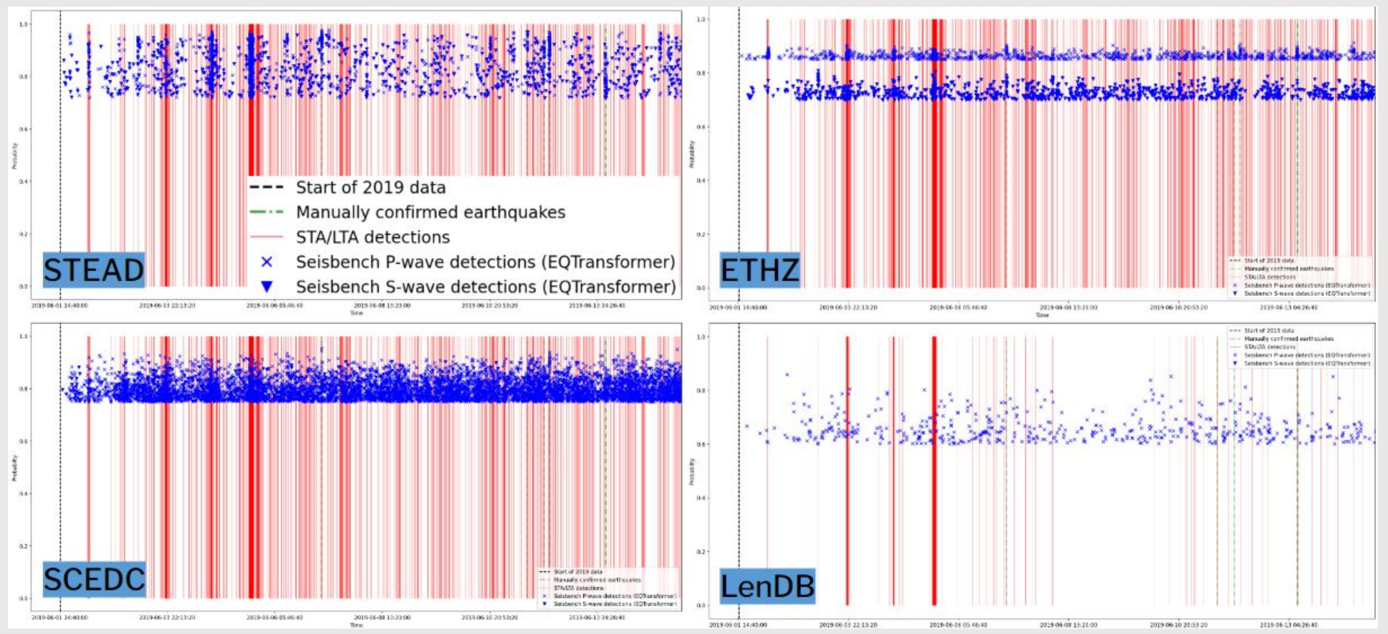


Figure 2: Comparison of returned picks from the four different EQTransformer models tested.

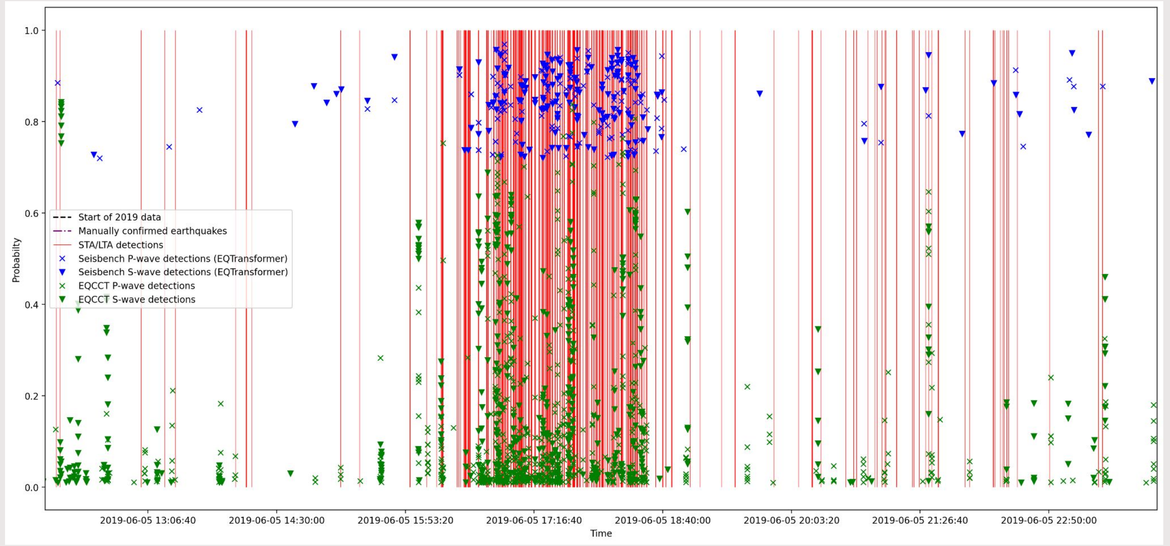


Figure 4: One of several detection "clusters" returned in unison by all three models. These clusters correspond to thunderstorms.

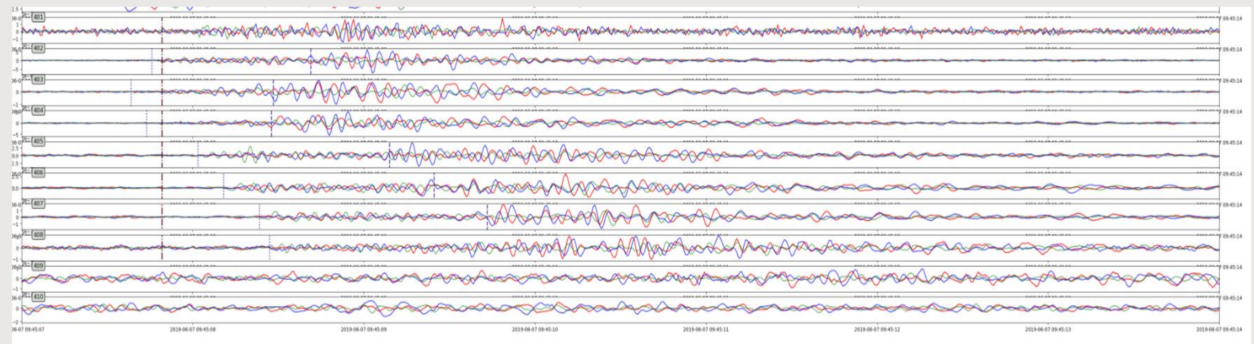


Figure 5: Waveform of a possible local earthquake. Wave arrivals suggest very high hypocentral proximity.

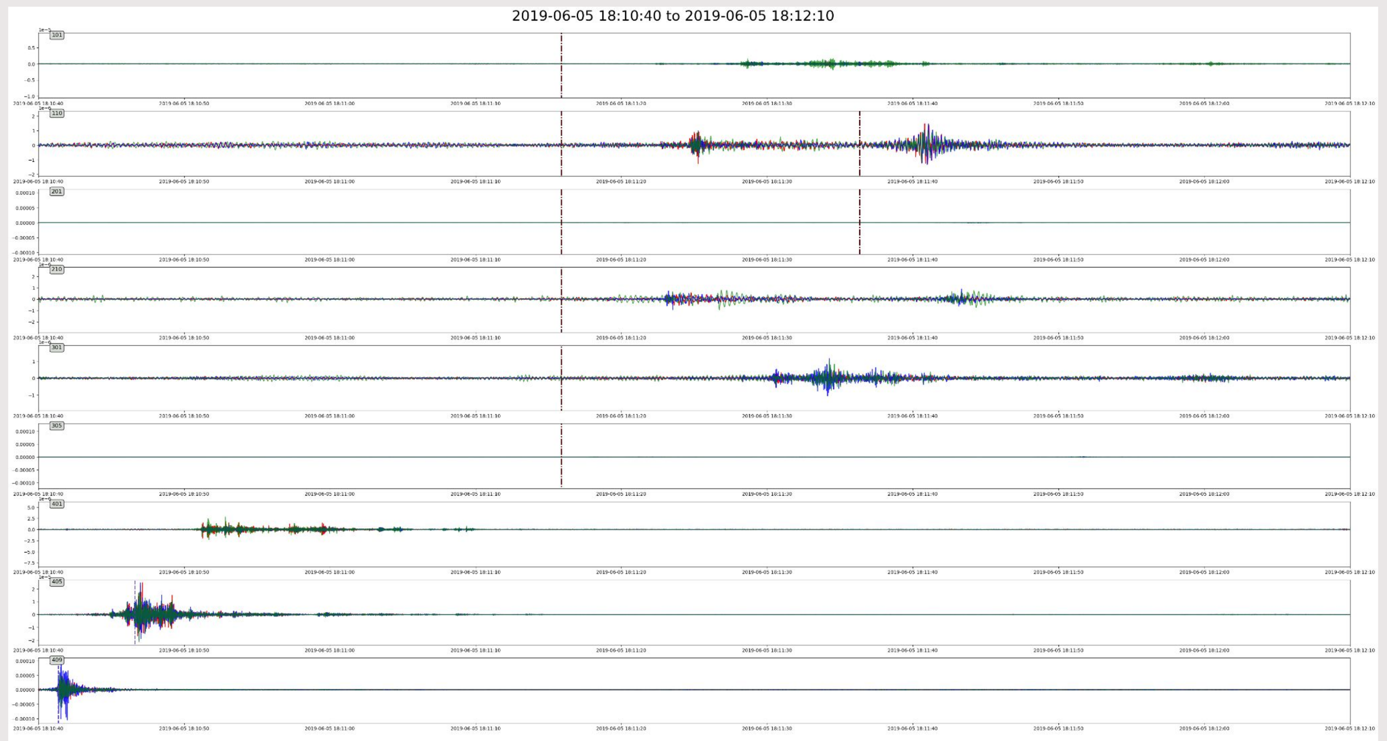


Figure 3: Data from June 5, 2019, 18:10 UTC, with an event marked as an earthquake by EQT (blue vertical dashed lines) and STA/LTA (red vertical dashed lines). Based on station positioning, waves from these event are estimated to have traveled at around 300 m/s, close to the speed of sound.

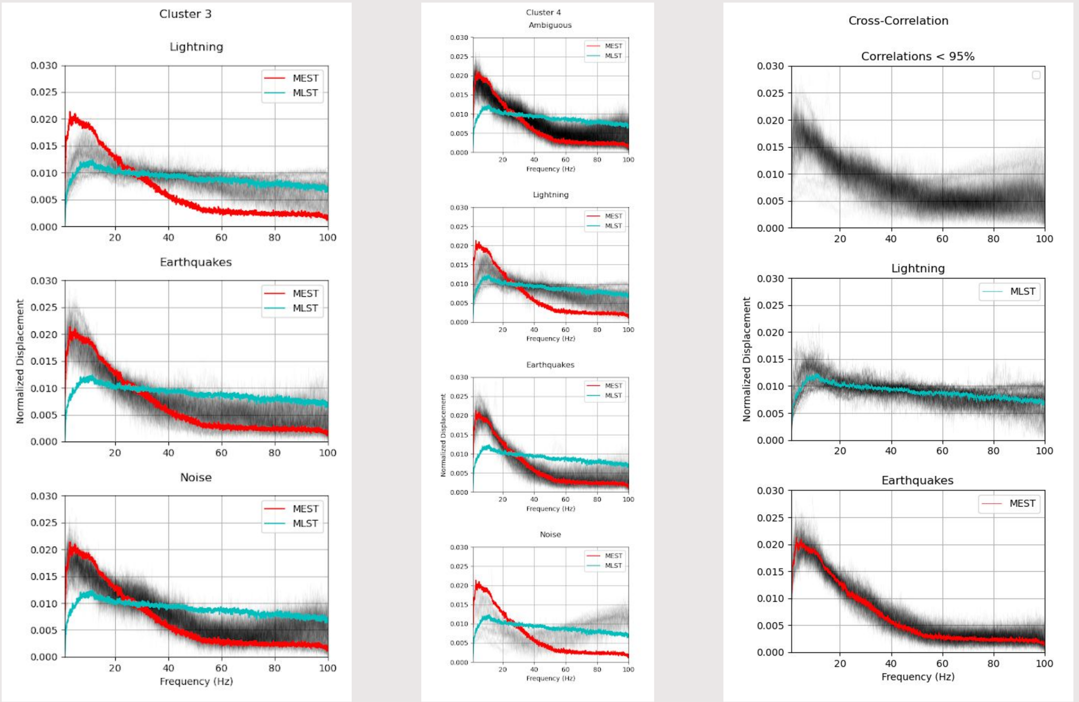


Figure 6: 336 events classified by clustering and cross correlation. Master Earthquake/Lightning Spectra Templates (MEST/MLST) were generated from empirically confirmed events and then compared to cluster groups and correlated to each event. Events had to be at least 95% correlated to be classified.

Results

15 confirmed earthquakes (Figure 7):
- 12 regional, uncatalogued
- 2 regional, catalogued on USGS
- 1 caldera earthquake (Figure 5)
  - Wave arrival times consistent with estimated wave velocities, radar, and time of day rule out lightning strike
  - Only seen on '400' stations; other stations saw too much noise/too low amplitudes
  - Estimated origin: Sulphur Springs hydrothermal area

Frequency analysis (Figure 8):
- Of 336 recordings, 207 are likely to be earthquakes, 107 are likely to originate from lightning strikes, and 22 are random noise or ambiguous.
- Moment magnitudes from the likely earthquakes, using estimated variables, ranged from M 0.68-2.40 with a median of M 1.51.

Figure 7: Classification of all manually confirmed earthquakes.

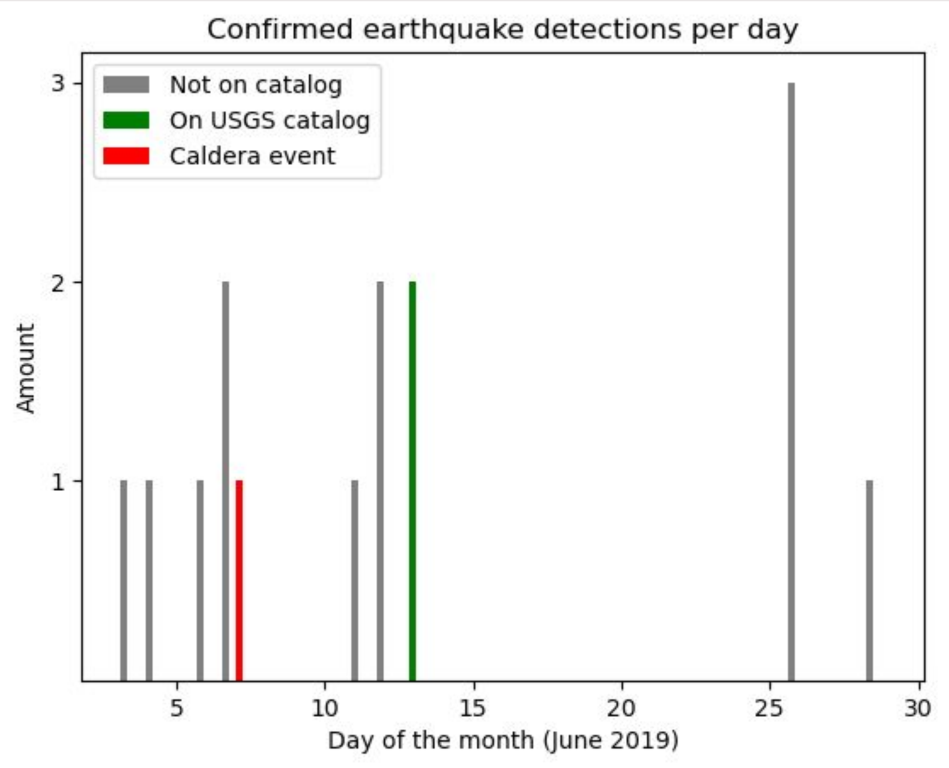
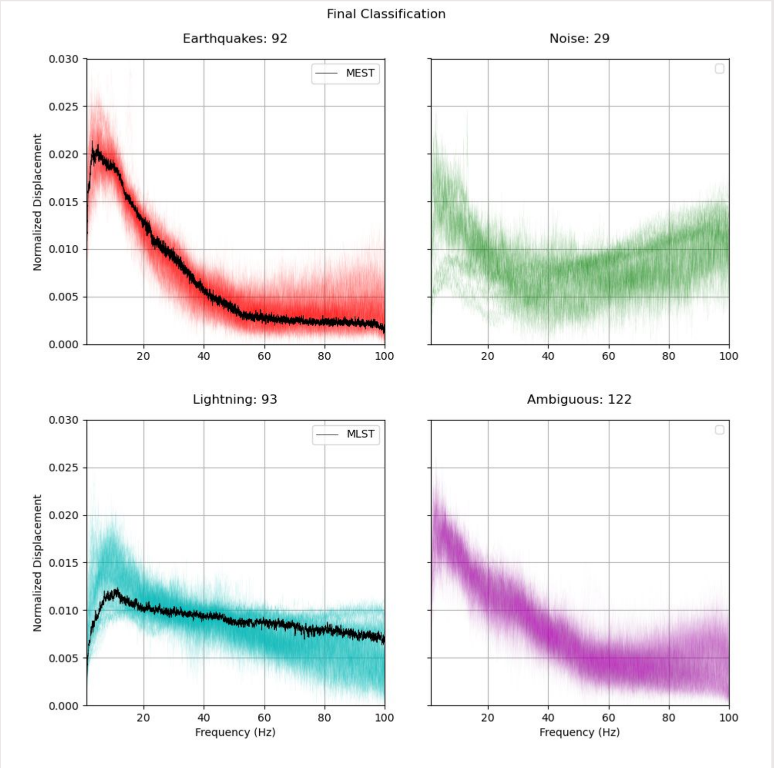


Figure 8: Cross-referencing the analysis from figure 6, events were appropriately classified. Results are confident that 92 of the events are earthquakes and should be further studied. Another 122 events are still unclassified and can be evaluated using more complex methods, if necessary.



Conclusion and Future Work

Given the limited recording time, only a few earthquakes were detected in this region, and we can only confidently label one of these as a caldera event. More data is being collected at the Valles Caldera and the next step is to start locating the local events relative to the complex internal structure of the caldera.

Another interesting implication was presented by the challenges in relying on machine learning for earthquake detection. We deduce that a lot of our initial challenges were a result of assuming machine learning would yield the desired results without much need for adjustment, and conclude that machine learning still has progress to be made in event discrimination. Ideally, future methods will be able to differentiate earthquakes from other events that excite seismic waves.

Our frequency domain analysis yields promise for the prospect of improving machine learning for earthquake detection. We find that analyzing the frequency content of waveforms can help differentiate an event from a lightning strike and an earthquake, a distinction that can be used in future machine learning methods. Further research into lightning detection with seismic instruments, their frequency content, and their comparison to earthquakes can open the doors to further improvements to earthquake detection algorithms.

References

1. Goff et al. "Geologic map of the Valles Caldera, Jemez Mountains, New Mexico." 2011.
2. Mostafanejad et al. *SAGE investigations of the Valles Caldera* [Data set]. 2019.
3. Woollam et al. "SeisBench—A Toolbox for Machine Learning in Seismology." 2022.
4. Saad et al. "EQCCT: A Production-Ready Earthquake Detection and Phase-Picking Method Using the Compact Convolutional Transformer." 2022.
5. NOAA National Weather Service (NWS) Radar Operations Center: *NOAA Next Generation Radar (NEXRAD) Level 3 Products*. Accessed July 2023.

This material is based upon work supported by the National Science Foundation under Grant No. EAR-1852339. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.