Geophysical Research Letters

RESEARCH LETTER
10.1029/2022GL101528

Global Nuclear Explosion Discrimination Using a Convolutional Neural Network

Louisa Barama[1,2], Jesse Williams[3], Andrew V. Newman[1], and Zhigang Peng[1]

[1]School of Earth and Atmospheric Sciences, Georgia Institute of Technology, Atlanta, GA, USA, [2]Lawrence Livermore National Laboratory, Livermore, CA, USA, [3]Global Technology Connection, Inc., Atlanta, GA, USA

Key Points:

- We successfully discriminate underground nuclear explosions with a Convolutional Neural network (CNN) trained on P-wave seismograms
- Robust global seismic event discrimination is possible with machine learning trained on regional and teleseismic data
- A CNN trained with historical nuclear explosion data can be applied with high accuracy to other regions, like the six Democratic People's Republic of Korea's test explosions

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

L. Barama,
barama1@llnl.gov

Citation:

Barama, L., Williams, J., Newman, A. V., & Peng, Z. (2023). Global nuclear explosion discrimination using a convolutional neural network. *Geophysical Research Letters*, *50*, e2022GL101528. https://doi.org/10.1029/2022GL101528

Received 4 OCT 2022
Accepted 6 AUG 2023

Author Contributions:

Conceptualization: Louisa Barama, Andrew V. Newman, Zhigang Peng
Data curation: Louisa Barama
Formal analysis: Louisa Barama, Jesse Williams, Andrew V. Newman
Funding acquisition: Zhigang Peng
Investigation: Louisa Barama
Methodology: Louisa Barama, Jesse Williams
Project Administration: Andrew V. Newman, Zhigang Peng
Software: Jesse Williams
Supervision: Jesse Williams, Andrew V. Newman, Zhigang Peng

Abstract

Using P-wave seismograms, we trained a seismic source classifier using a Convolutional Neural Network. We trained for three classes: earthquake P-wave, underground nuclear explosion (UNE) P-wave, and noise. With the current absence of nuclear testing by countries that have signed the Comprehensive Test Ban Treaty, high quality seismic data from UNEs is limited. Even with limited training data, our model can accurately characterize most events recorded at regional and teleseismic distances, finding over 95% signals in the validation set. We applied the model on holdout datasets of the North Korean test explosions to evaluate the performance on unique region and station-source pairs, with promising results. Additionally, we tested on the Source Physics Experiment events to investigate the potential for chemical explosions to act as a surrogate for nuclear explosions. We anticipate that machine-learning models like our classifier system can have broad application for other seismic signals including volcanic and non-volcanic tremor, anomalous earthquakes, ice-quakes or landslide-quakes.

Plain Language Summary

We train a global seismic event classifier using machine learning on underground nuclear test explosion seismic data. Our classifier model can successfully discriminate (with over 95% accuracy) between underground nuclear explosion, earthquake, and noise signals from stations both regionally and far-field. Since this model was trained on a relatively small data set (for machine learning applications) we expect that similar methods can be applied to event or discrimination of other unique seismic sources like those from volcanoes, landslides, or glaciers.

1. Introduction

Nuclear weapons can have disastrous effects on human life, ecological environments, and public health; ramifications that can last for generations (Wu et al., 2020). According to the 2020 Global Nuclear Power Report (Yearbook, 2021) the number of nuclear weapons is generally reduced in 2020 but states with nuclear weapons continue to modernize nuclear arsenals (Yearbook, 2021). Robust global underground nuclear explosion (UNE) detection is essential for international regulatory bodies such as the Comprehensive Nuclear Test Ban Treaty Organization (CTBTO) and the Air Force Technical Applications Center (AFTAC) to monitor the near-worldwide cessation of nuclear explosive testing and eventually to verify the compliance of the Comprehensive Nuclear-Test-Ban Treaty (CTBT). Although the CTBTO uses a variety of data including continuous seismic, acoustic, hydro-acoustic signals, and radionuclide data, seismic methods have proved to be the most robust and rapid for the purpose of detecting, locating and discriminating UNEs from natural earthquakes and other near surface seismic signals (Maceira et al., 2017). Improved and continued development of underground explosion detection using seismic methods is imperative given that UNE testing is the most likely scenario for all future testing as it allows for control of radioactive explosive products (Maceira et al., 2017).

Existing methods for seismic detection and discrimination include beamforming, template matching, waveform autocorrelation, ratio of body to surface wave magnitudes, and P to S wave amplitude ratio calculations (Bowers & Selby, 2009; Maceira et al., 2017; Schaff et al., 2018; Tibi, 2021; Walter et al., 2018). Although these tools are robust for large yield (>0.5 kt) UNE monitoring (Koper, 2020; Maceira et al., 2017), the ever increasing amounts of seismic data, and lower detection thresholds drives the need for rapid and automated single-station detection algorithms.

Machine learning (ML) methods are widely spreading as a tool in seismology and real-time monitoring (Kong et al., 2019; Yeck et al., 2020) with encouraging potential. ML methods can allow generalizable models that can identify events outside of those used in initial training. This is essential, especially for nuclear explosion

10.1029/2022GL101528

Validation: Louisa Barama, Jesse Williams
Visualization: Louisa Barama
Writing – original draft: Louisa Barama
Writing – review & editing: Louisa Barama, Jesse Williams

monitoring where nuclear weapon test explosions are rare after almost all the world's countries signed the CTBT, meaning that all current and future monitoring research and development is dependent on historical nuclear explosion data, with the exception of the Democratic People's Republic of Korea's (DPRK) tests.

The first applications of ML in seismology focused on discriminating the amplitude spectra of seismic waveforms of natural earthquakes, and nuclear and chemical explosions (Dowla et al., 1990; Dysart & Pulli, 1990; Ren et al., 2020). ML has successfully been applied to discriminate between natural earthquakes and other types of seismic events (UNE's, underwater explosions, volcano-tectonic events) at local or regional distances (Romeo, 1994). These earlier studies typically use limited training labels (e.g., on the order of a few tens or hundreds) and Artificial Neural Networks (ANNs) with shallow fully-connected feed-forward neural networks and simple recurrent networks (Del Pezzo et al., 2003; Dowla et al., 1990). These networks tend to limit performances (e.g., up to 90% of classification accuracy), limit spatial application, and can be computationally intensive for standard CPU-based computations. More recent developments in deep learning (LeCun et al., 2015; Rouet-Leduc et al., 2017) opened doors for scientists in many fields to be able to utilize historical and relatively small datasets for classifying and discriminating seismic event types or phase determinations (Bergen et al., 2019; Kong et al., 2019; Nakano et al., 2019) with encouraging results.

Evolved from traditional ML techniques (G. E. Hinton et al., 2006), deep Convolutional Neural Networks (CNNs) have been shown to be highly successful in both image (Krizhevsky et al., 2012) and speech (G. Hinton et al., 2012) processing. In the past few years, seismologists started applying CNNs on seismic waveform data for phase determinations and associations (W. Zhu & Beroza, 2019; Ross et al., 2018, 2019; L. Zhu et al., 2019), event detection and discrimination (Mousavi & Beroza, 2019; Mousavi et al., 2019; Nakano et al., 2019) and source location (Mousavi & Beroza, 2019; Perol et al., 2018; Zhang et al., 2020), and magnitude determination (Mousavi & Beroza, 2020). Other machine learning methods such as Support Vector Machines (SVM), random forests, self-organizing maps, and naive bayes classification have also been used for natural earthquake and explosion discrimination (Pu et al., 2019; Sick et al., 2015; Kim et al., 2020; L. Dong et al., 2014). Studies have found improved recognition ability with CNNs over SVMs for detecting nuclear and chemical explosions with infrasound data. CNNs have been used successfully for non-nuclear explosion and earthquake discrimination in numerous studies, however always in a regional or local setting and not always on the seismic waveforms (Tian et al., 2022; L.-j. Dong et al., 2020; L. Linville et al., 2019; Trani et al., 2022; Huang et al., 2018; Song et al., 2020). Mousavi et al. (2019)'s earthquake detector “CRED” didn't detect more events than conventional template matching when applied to a region outside of its training, it did perform faster without the need region-specific waveform templates, which highlights the computational efficiency of CNN's and potential for use in data-limited regions. Generalized Global ML models have been successfully applied to real-time magnitude estimation (Chakraborty et al., 2021; Mousavi & Beroza, 2020) and regional focal mechanism predictions (Kuang et al., 2021), as well as for global operational processing systems at monitoring institutions like the USGS NEIC (Yeck et al., 2020). Both the ML architecture and training data choice are important for developing a robust model. Inconsistency of neural networks for seismic phase prediction are strongly influenced by the training data, and biases there-within that influence a model's generalizability and performance (Park et al., 2023).

With the increasing availability of seismic data recently (up to ten of thousands sensors recording in near real-time), it is virtually impossible to manually flag suspicious events (e.g., conventional or nuclear explosions) for further analysis. Barama et al. (2022) compiled the Georgia Tech Underground Nuclear Explosions (GTUNE) data set which is a comprehensive seismic waveform and event catalog from a wide range of sources for declassified underground nuclear explosions. With this, we explore the utility of older and lower quality historical data from UNEs in combination with high quality data from the more recent events for developing more robust ML-based global UNE detectors. Here we show that this data set can be used to train a CNN model to automatically discriminate seismic waves generated by earthquakes and nuclear explosions.

2. Data

The labels for the three training classes of: earthquake P-waves, UNE P-waves and noise are sourced from Barama et al. (2022), including global seismic data stored at the Incorporated Research Institutions in Seismology (IRIS) and Japanese national waveforms from their National Research Institute for Earth Science and Disaster Resilience (NIED), and other networks. The NIED Hi-Net array of data are particularly useful for dense observations of the recent DPRK nuclear tests at local to regional distances. The UNE data include 600 underground nuclear tests from the year 1961–2017 (Figure S1 in Supporting Information S1). Earthquakes used range in magnitude from 4.5 to 6.5 (comparable wave amplitude to used nuclear tests), are shallower than 50 km, and recorded at stations less than 90° away. Figure S1 in Supporting Information S1 shows the distribution of UNEs and stations. Since most historical data was recorded on single channel seismographs, we use only vertical component waveforms for training. L. Linville et al. (2019) showed successful discrimination between surface mine explosions and seismic events at local distances and found that although 3-component data made the CNN predictions more accurate, it was not required for prediction. The training labels are formatted in 20 s windows with the earthquake and UNE P-arrivals fixed at 5 s, which means we are training on 15 s of earthquake and UNE signal. We choose this window length in order to capture the more impulsive nature of the explosion waveforms when compared to earthquakes recorded at teleseismic distances. Window lengths of 30 and 60 s and windows with random P-arrival positioning were also tested but we found that waveform duration after the P-wave does not significantly affect the final accuracy. The 20 s windows with fixed P-arrival resulted in the best performing model, but additionally enabled us to have smaller training data, a shallower CNN thus faster training.

A data processing pipeline was created that ingests raw seismic signals and produced a 3-class probability (earthquake, UNE, and noise). We did not remove the instrument response, because such information was not readily available for many historic seismic stations. All methods and pipelines are organized in a Python library we call *PyWave*. The pipeline was designed to produce a consistent data structure regardless of the recording stations' characteristics to create a user-friendly system, and speed-up processing time. Pre-processing functions included are for demeaning, resampling, applying a cosine envelope function and a four-corner bandpass filter. Additionally we developed and integrated a method into the *PyWave* library that measures the energy (the sum of the square of the amplitudes) of pre-event noise and event signal, and calculates the signal-to-noise ratio (SNR). This energy filter method was used to further refine the data set by excluding those waveforms with a very low SNR.

The windowed data is then centered by subtracting the windowed mean value from the signal vector. The resulting time series is then re-sampled to 20 samples per second (sps) using a Fourier method to produce consistent results between stations with different sampling rates. Any signals sampled at greater than 20 sps are down-sampled, while signals sampled at less than 20 sps are discarded. The waveform is demeaned and then a cosine envelope function is applied at the beginning and end 4% of the re-sampled window. This minimizes edge effects that may otherwise be created in subsequent processing. Because data cover a wide range of distances, a number of filter ranges were tested that could accommodate valuable information for all data. We found we retained the most data of energy filter threshold greater than 5 as well as model performance with data filtered between 1 and 5 Hz using a digital Butterworth filter.

Maceira et al. (2017) showed that seismo-acoustic wave amplitude—yield relationship can vary and depends on many factors such as emplacement geologic conditions and depth. To address such amplitude variations, we applied a signal normalization function to the data before feeding them into the CNN classifier. Signal normalization scales all signals so they are roughly the same order of magnitude, ranging from −1 to 1, while retaining all salient features. When training an algorithm, feature importance can be associated with feature amplitude. Thus an algorithm generally considers waveforms with a strong feature as more important and focus on these features during the training process. The normalization process is used to prevent this deleterious behavior.

3. CNN Model Architecture and Training

The deep learning classifier we used is composed of many fully-connected one-dimensional convolutional layer bundles, following an architecture used by L. Zhu et al. (2019). In total there are 6 CNN layers and one dense output layer as shown in Figure 1. The number of inputs is dependent on the initial window size, which has an initial input of 400 data points, corresponding to a 20 s window sampled at 20 sps. The output size is always 3: probabilities of earthquake, UNE, and noise, respectively. The number of convolution filters in each layer is designed to map a few low complexity features, many mid complexity features and then a few compound, top-level features.

Each layer in the CNN bundle is composed of multiple sub-layers as shown in Figure 1. These layers are a 1D convolution, batch normalization, activation, dropout, and a 1D max pooling layer. The batch normalization layer scales the inputs for better generalization and also improves computation speed. The activation layer applies the "ReLU" (rectified linear activation unit) function to each input. The dropout layer randomly re-initializes a

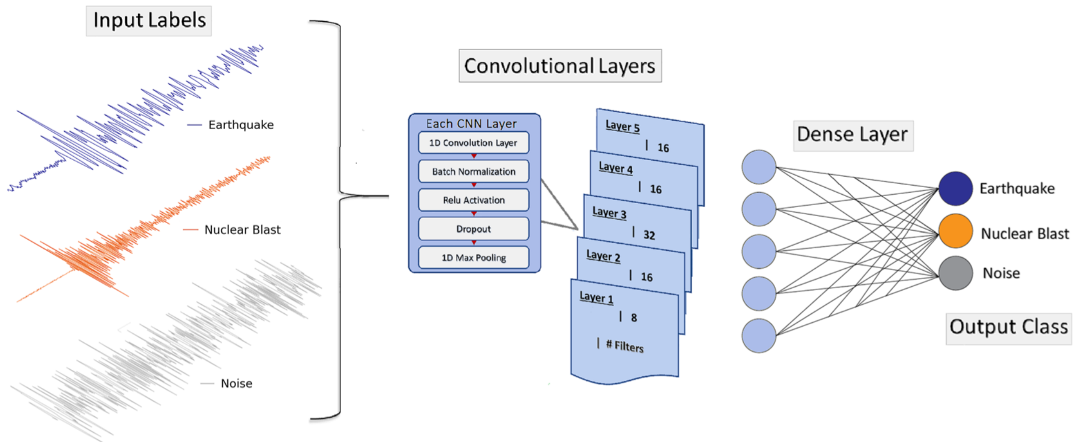

Figure 1. Convolutional Neural Network architecture: the input layering design, and the number of convolutional filters per layer. The number of inputs is dependent on the initial window size. The shown example has an initial input of 400 data points, corresponding to a 20 s window sampled at 20 sps. The output size is always 3: probability of earthquake, UNE, and noise. The number of filters is designed to map a few low complexity features, many mid complexity features and then a few compound, top-level features. This is represented in the number of filters present in the respective layers: 8 → 16 → 32 → 16 → 16.

fraction of the inputs to prevent over-fitting. Dropout fractions were tested up to 0.1. The 1D max pooling layer combines all kerneled layers back into one layer. Pooling operations allow the model to learn multiple templates at different resolutions by changing the resolution of the filtered input waveform (Ren et al., 2020). Lastly, a soft-max layer is included, so that the output predicted probability distribution total sums up to 1.

The CNN classifier was trained and tested on a subset of the cumulative data set discussed above. Interestingly, the best performing models were those with an energy threshold of ~5; higher signal-to-noise ration (SNR) thresholds would substantially reduce the training data used. After applying the energy filter, the nuclear data set retained 14,302 observations, 44.7% of the original data set. One thousand traces of this data set were randomly removed as a holdout set for final testing for nuclear, earthquake and noise labels. An additional 10% was randomly removed from the remaining data set for training and testing. Since it is important to keep a balanced training and testing data set to equally represented classes and give similar internal weighting to results (Megahed et al., 2021), the earthquake and noise datasets were randomly reduced to match the nuclear data set training and test size. In Table S1 in Supporting Information S1 and Table S2 detail number of waveforms used in training and testing and the CNN classifier structure as adjusted for a 20 s window sampled at 20 sps (input size of 400) with 1 channel vertical component data, respectively.

4. Results and Testing Model Performance

Using the training and test data presented in Table S1 in Supporting Information S1, a single-channel CNN model was trained on the vertical component, pre-processed seismograms. After tuning the hyper-parameters, a dropout of 8% and a batch size of 500 were selected. We trained the model for 100 epochs. The trained CNN was used to predict the classes within the holdout sets. All datasets tested above have very near or above 95% accuracy (Figure 2). The holdout datasets include earthquakes, noise, UNE, source physics experiment (SPE) chemical explosions and DPRK holdout data set. For the holdout sets, the class-specific noise set has the highest accuracy at 99.6%, while the nuclear set was 98.4% accurate and the earthquake set was the lowest with a 95.6% accuracy. From our confusion matrix in Figure 2b we calculate class-specific F1-scores to be 0.976, 0.999, and 0.979 for earthquake, noise and UNE, respectively. This shows that the trained CNN was highly capable of classifying signals curated in manner similar to this study, that is, window size and positioned event arrival. Figure 3 shows the accuracy of the model with the source-receiver distance.

4.1. Source Physics Experiments

It is important to have confidence that current empirical and limited test-site based methods will work in all regions of the world, even in under-tested emplacement conditions. This drives physics-based simulation and

10.1029/2022GL101528

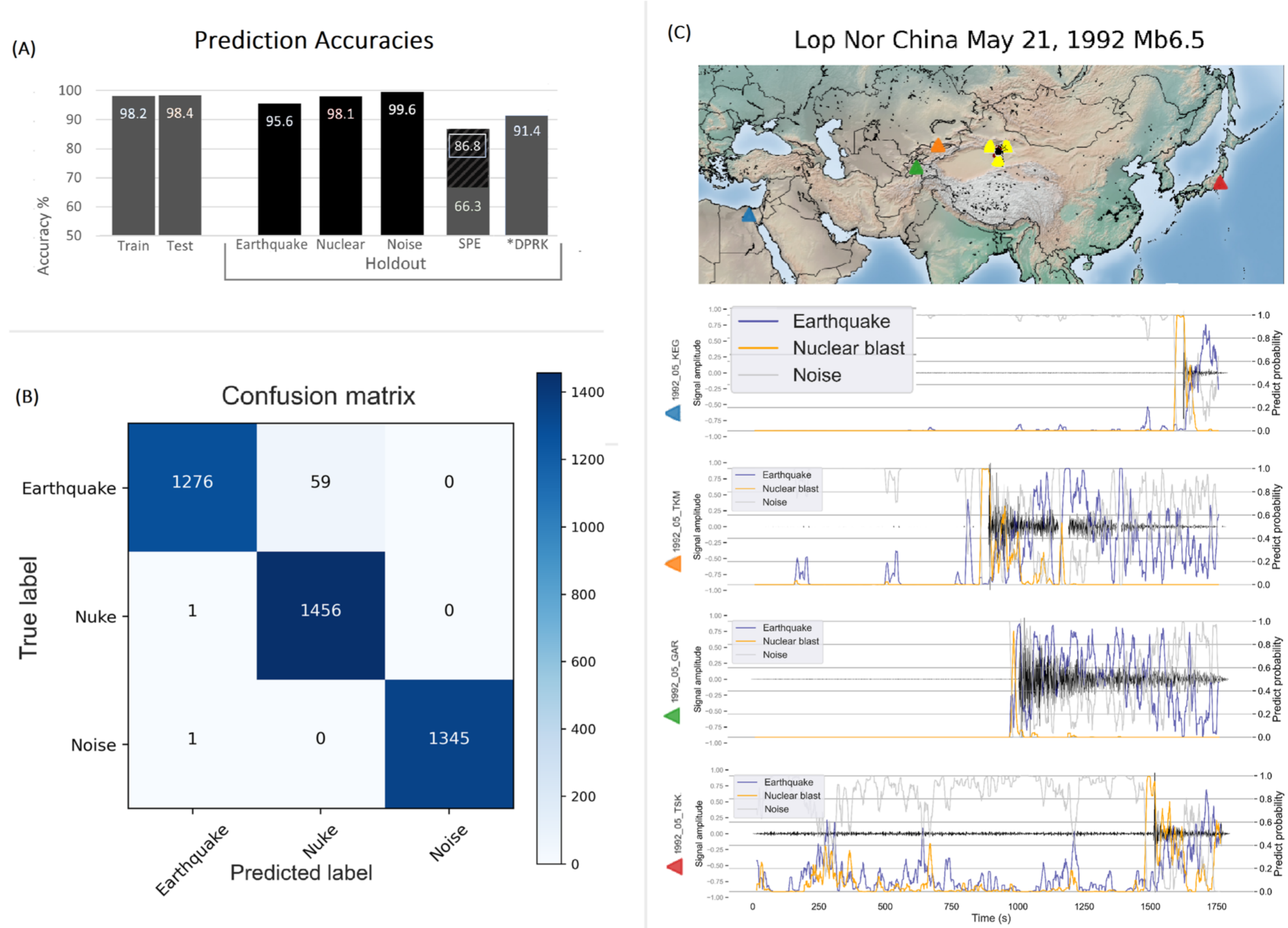

Figure 2. (a) CNN accuracy during training and on earthquake, UNE, noise, SPE, and DPRK holdout sets. The SPE results include results for energy threshold of 5 and 10. The confusion Matrix (b) shows number of events the model predicted incorrectly in the holdout sets. All testing was done with the CNN model not trained with SPE as UNE with the exception of the SPE holdout set. (c) Example showing the model predictions on continuous 20 min seismograms from the 21 May 1992 Mb6.5 Lopnor nuclear test in China.

modeling supported by chemical explosion data like the Source Physics Experiments (SPE). Waveforms from the SPE tests Snelson et al. (2013) were used to test how well chemical explosions can emulate low-yield nuclear explosions for seismic detection. When the CNN classifier was applied to the SPE data, the model performed poorly at classifying SPE explosions as UNEs (66.3%). When the energy threshold was set higher to 10, the model identified SPE events accurately for 86.0% of SPE events.

When the SPE waveforms were included as UNE labels in the CNN training data, the CNN showed only a slightly reduced performance of 89% testing accuracy. This seems to suggest that even though the chemical SPE explosions are tested as a surrogate for small nuclear explosion data at local and regional distances, their resulting regional waveforms do not quite emulate those of UNEs. This could also be due to their relatively high SNRs due the smaller yields of chemical explosions (Mellors et al., 2018; Stump et al., 1999). Even though the waveforms are normalized, the SPE explosions have much smaller yield than UNEs we trained the CNN model with. This test highlights the limitation of our global CNN classifier to apply to chemical explosions. To detect lower yield explosions with a CNN likely requires a model trained with more local/regional data.

Previous results from non-Proliferation Experiments (NPE) (Denny, 1994) showed that the amplitudes for chemical explosions were larger than those of UNEs for similar depths. Additionally, the NPE findings concluded that chemical explosions couple more energy into the ground than UNEs do of the same yield and total energy (Denny, 1994). Our results here emphasize the distinction of source-physics of the explosion to the propagation and how that affects the resulting waveforms.

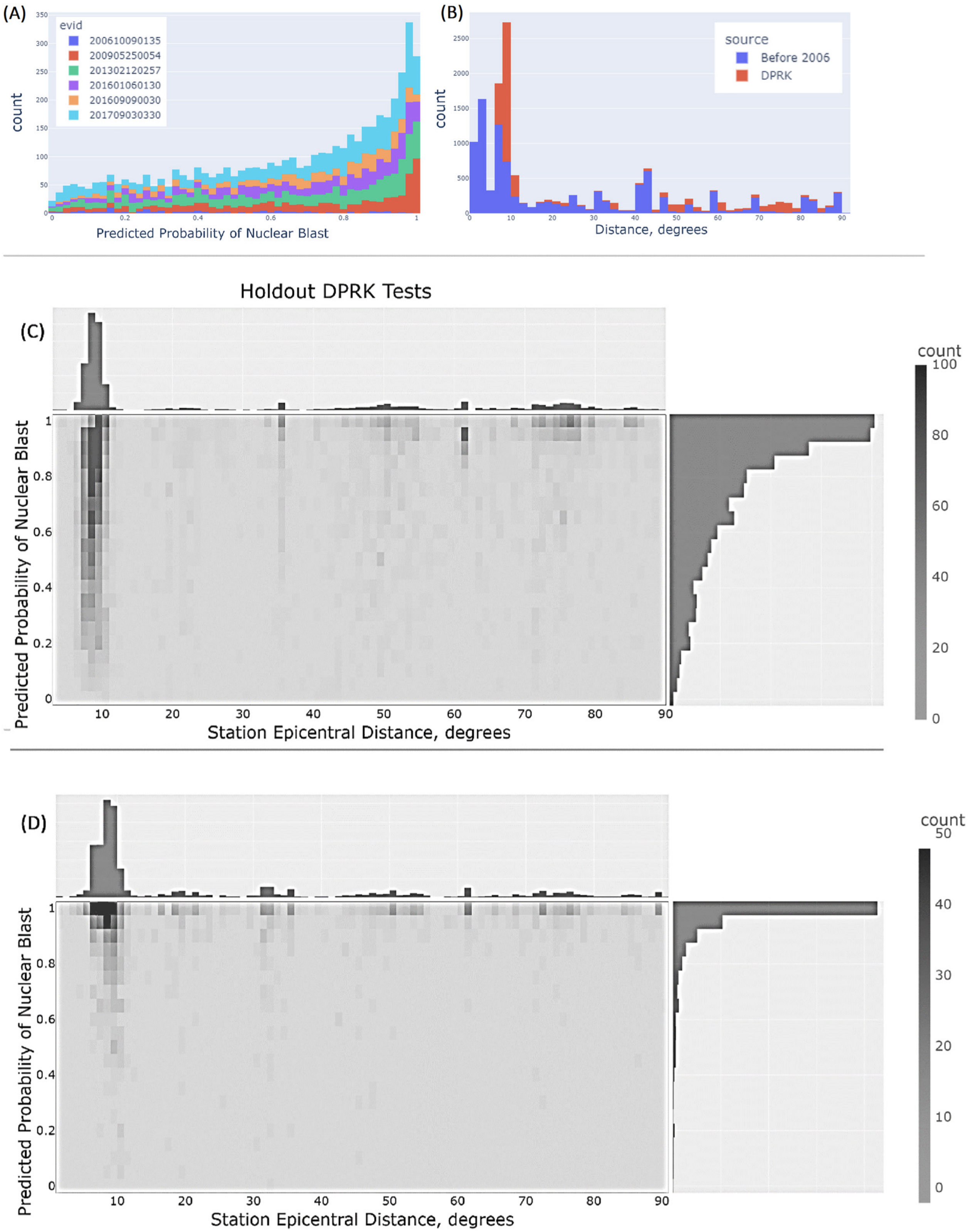

Figure 3. Testing on the DPRK holdout data set model and the global data set (a) Predicted probabilities of UNE for each of the 6 DPRK nuclear test explosions (events 2006 to 2017, with Mb 4.3 to 6.3). (b) Count of label waveform station distance in training (blue) and DPRK (red) datasets. (c) Distribution of predicted UNE probability and station distance is shown for the DPRK holdout model and (d) the global model. Histogram distribution of waveform station distance and model prediction probability are top and bottom bar graphs.

4.2. Holdout Testing on the North Korea Nuclear Test Explosions

Up to now we randomly split the data set into training, holdout, and test sets. However because nuclear waveforms from any specific source-receiver pair are likely to be highly similar (e.g., the 6 DPRK nuclear tests) (Lay, 2021; Voytan et al., 2019), it is possible that our algorithm has already seen the waveform from that particular source-receiver path during the training process, resulting in a high success rate (i.e., data leakage). This likely occurs because the algorithm is learning the Green's function that represents the path-effect rather than training on the fundamental differences in the waveform and source characteristics. To test the model's performance on unique source-receiver pairs as well as the usefulness of the historical seismograms, we trained an additional model excluding all 6 DPRK tests. Figure S1 in Supporting Information S1 shows the distribution of stations used for DPRK events for testing and global historical data. Following the same procedure for data preparation and model training as using the full data set, the model accuracy for nuclear, noise, and earthquake holdout sets was 90.1%, 99.7%, and 95.8%, respectively. When applying the model to the 4871 DPRK traces, the model prediction accuracy for UNE is 91.4%. Figure 3 displays the details of the difference in training data for the global model and the DPRK holdout data, as well as difference in model performance with station distance. Our results suggest that our CNN model is capable of identifying UNEs from a region excluded from the training data set as well as a range of magnitudes from Mb 4.3 to 6.3 (Figure 3a). Although the historical waveforms are from more diverse set of stations and have source-receiver pairs all over the world, the model does generalize to be applicable to other regions.

4.3. Application to Continuous Data Sets

The results so far have shown that our CNN classifier was effective on a curated data set. Next we perform an additional test for its performance on continuous waveform data where we apply a moving average to a three-class predictor (Figure 2). The algorithm class prediction is assigned to the class with the highest probability. Note that for a three-class algorithm, the highest probability may not be a majority. In the example of the 12 May 1992 Lopnor China nuclear test (Figure 2), as the P-arrival appears into a sliding window, a nuclear event is briefly predicted with a confidence higher and over a longer period for the first large event. In these instances, the algorithm tends to predict an earthquake when it is exposed to a limited P-arrival signal. When it is exposed to a larger portion of the signal, it predicts a UNE with much higher confidence. Because the three-class classifier does not recognize other signals, any nuisance event is necessarily categorized as either an earthquake, nuclear test, or noise.

5. Discussion

Given the modest size of our training set, we have higher prediction variance and more false positives (predicting earthquakes as UNE) than we would expect from a deep-learning model with deeper layers. This is suggested by the results of L. Zhu et al. (2019) for earthquake phase detection, in which the authors did not have such issues while using a similar approach but with a training set that was an order of magnitude larger than ours. Without substantially more nuclear event waveforms this will remain a limitation of our work, however, event-based training could prove useful to combat these limitations (L. M. Linville, 2022). The issue is still quite small, affecting only 2% of the signals. Too, as this method is testing solely on individual waveforms, network-based solutions should quickly correct for most, if not all of these false positives.

In a comparative study, Chen et al. (2018) used a small data set of 72 earthquakes and 101 man-made explosions around Beijing, China to train a CNN to discriminate explosions from earthquakes. Unlike this study, Chen et al. (2018) trained their CNN on the extracted Mel Frequency Cepstrum Coefficient (MFCC) map, not seismic waveforms, and found an average recognition rate of 95.78%. Though there are limitations, our model result of holdout accuracy of over 98% for global UNEs is a promising result.

Koper (2020) stressed the importance of lowering the detection threshold of explosions using the International Monitoring System (IMS) global seismic network at local and regional distances. Kong et al. (2022) used local augmented waveforms and P-S ratios from 90 explosions in Northwest United States to train a deep learning model as an explosion-earthquake discriminator with high rates of success. However, their model did not perform well when applied to other regions. Our preliminary tests on the DPRK data already demonstrated that our model successfully generalizes to global (both regional and teleseismic) sources and stations. However, there are opportunities to test how to create improved performance of the algorithm such as including synthetic tests or augmented waveforms to improve the diversity and size of the training datasets. Kong et al. (2022) increased their explosion training explosion data from 8,502 to 178,059 traces by using data augmentation. For our case, data augmenting to create a larger data set would allow us to train station-distance dependent models to test performance compared to the global model. Modifying the training method to require unique source-receiver pairs between training and test sets could also improve performance, however since many locations were re-used for nuclear tests, this would exclude a lot of data, and especially high-quality digital data from the recent DPRK events. There remains potential for our current algorithm to be fine-tuned toward specific regions or station distances. Furthermore, developing these algorithms have far-reaching potential beyond automatic nuclear detection as well, and may be adapted to automatic detection of seismic signals associated with deep earthquakes, glacier or ice-quakes, magmatic movements at depth, dangerous slow-source tsunami-generating earthquakes, or other features that may rapidly and automatically illuminate a host of geophysical hazards (e.g., initiation phase of seismogenic landslides).

6. Conclusions

In this study, we successfully built a robust global earthquake, UNE, and noise discriminator with available limited and balanced data. The prediction accuracy on the class-specific holdout sets were 99.6% for noise, 98.4% for UNE, and 95.6% for earthquake. This work shows promising applications in single-station discrimination earthquakes and UNEs recorded at a global network of stations. Furthermore, our results suggest that ML classifier systems could have broad application for other global and "small data" seismic signals without necessarily relying on the use of synthetic or augmented data.

Data Availability Statement

Most of the waveforms and metadata used in this work are sourced from the GTUNE repository (Barama et al., 2022), the facilities of EarthScope Consortium and Japan's Data Management Center of the National Research Institute for Earth Science and Disaster Resilience (NIED DMC) Hi-Net array data (National Research Institute for Earth Science and Disaster Resilience, 2019). EarthScope Consortium services are funded through the Seismological Facility for the Advancement of Geoscience (SAGE) Award of the National Science Foundation under Cooperative Support Agreement EAR-1851048. Python was used for seismic data processing, specifically using the ObsPy package (Beyreuther et al., 2010) and Tensorflow (Abadi et al., 2016) for machine learning. The pre-processing library and CNN model structure *PyWave* is available on Zenodo (https://doi.org/10.5281/zenodo.7809356).

Acknowledgments

This project was completed as part of an Air Force Research Laboratory (AFRL) Small Business Innovation Research (SBIR) grant funded collaboration between Global Technology Connection Inc. and the Georgia Institute of Technology, (Grant AF19A-T012, and contract number FA9453-19-P-0684). The authors thank Eli Baker for guidance, questions, and comments throughout this work, as well as Ben Baker and an anonymous reviewer for their comments and suggestions on the article. Revision of this article was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., & others (2016). Tensorflow: A system for large-scale machine learning. In *12th {USENIX}* symposium on operating systems design and implementation ({*OSDI*} 16) (pp. 265–283). Retrieved from https://www.usenix.org/system/files/conference/osdi16/osdi16-abadi.pdf

Barama, L., Peng, Z., Newman, A. V., & Williams, J. (2022). Gtune: An assembled global seismic dataset of underground nuclear test blasts. *Seismological Research Letters*, *93*(6), 3514–3523. https://doi.org/10.1785/0220220036

Bergen, K. J., Johnson, P. A., de Hoop, M. V., & Beroza, G. C. (2019). Machine learning for data-driven discovery in solid Earth geoscience. *Science*, *363*, eaau0323. https://doi.org/10.1126/science.aau0323

Beyreuther, M., Barsch, R., Krischer, L., Megies, T., Behr, Y., & Wassermann, J. (2010). ObsPy: A python toolbox for seismology. *Seismological Research Letters*, *81*(3), 530–533. https://doi.org/10.1785/gssrl.81.3.530

Bowers, D., & Selby, N. D. (2009). Forensic seismology and the comprehensive nuclear-test-ban treaty. *Annual Review of Earth and Planetary Sciences*, *37*(1), 209–236. https://doi.org/10.1146/annurev.earth.36.031207.124143

Chakraborty, M., Li, W., Faber, J., Ruempker, G., Stoecker, H., & Srivastava, N. (2021). A study on the effect of input data length on deep learning based magnitude classifier. arXiv preprint arXiv:2112.07551.

Chen, R.-H., Huang, H.-M., & Chai, H.-M. (2018). Study on the discrimination of seismic waveform signals between earthquake and explosion events by convolutional neural network. *Progress in Geophysics*, *33*(4), 1331–1338.

Del Pezzo, E., Esposito, A., Giudicepietro, F., Marinaro, M., Martini, M., & Scarpetta, S. (2003). Discrimination of earthquakes and underwater explosions using neural networks. *Bulletin of the Seismological Society of America*, *93*(1), 215–223. https://doi.org/10.1785/0120020005

Denny, M. D. (1994). In *Proceedings of the symposium on the non-proliferation experiment: Results and implications for test ban treaties, Rockville, Maryland, (Tech. Rep.)*. Lawrence Livermore National Lab (LLNL).

Dong, L., Li, X., & Xie, G. (2014). Nonlinear methodologies for identifying seismic event and nuclear explosion using random forest, support vector machine, and naive bayes classification. In *Abstract and applied analysis* (Vol. 2014, pp. 1–8).

Dong, L.-J., Tang, Z., Li, X.-B., Chen, Y.-C., & Xue, J.-C. (2020). Discrimination of mining microseismic events and blasts using convolutional neural networks and original waveform. *Journal of Central South University*, *27*(10), 3078–3089. https://doi.org/10.1007/s11771-020-4530-8

Dowla, F. U., Taylor, S. R., & Anderson, R. W. (1990). Seismic discrimination with artificial neural networks: Preliminary results with regional spectral data. *Bulletin of the Seismological Society of America*, *80*(5), 1346–1373.

Dysart, P. S., & Pulli, J. J. (1990). Regional seismic event classification at the NORESS array: Seismological measurements and the use of trained neural networks. *Bulletin of the Seismological Society of America*, *80*(6B), 1910–1933.

Hinton, G., Deng, L., Yu, D., Dahl, G., Mohamed, A.-R., Jaitly, N., et al. (2012). Deep neural networks for acoustic modeling in speech recognition. *IEEE Signal Processing Magazine*, *29*(6), 82–97. https://doi.org/10.1109/msp.2012.2205597

Hinton, G. E., Osindero, S., & Teh, Y.-W. (2006). A fast learning algorithm for deep belief nets. *Neural Computation*, *18*(7), 1527–1554. https://doi.org/10.1162/neco.2006.18.7.1527

Huang, L., Li, J., Hao, H., & Li, X. (2018). Micro-seismic event detection and location in underground mines by using convolutional neural networks (CNN) and deep learning. *Tunnelling and Underground Space Technology*, *81*, 265–276. https://doi.org/10.1016/j.tust.2018.07.006

Kim, S., Lee, K., & You, K. (2020). Discrimination of imbalanced seismic events between earthquakes and artificial explosions using support vector machine (Vol. *7*).

Kong, Q., Trugman, D. T., Ross, Z. E., Bianco, M. J., Meade, B. J., & Gerstoft, P. (2019). Machine learning in seismology: Turning data into insights. *Seismological Research Letters*, *90*(1), 3–14. https://doi.org/10.1785/0220180259

Kong, Q., Wang, R., Walter, W. R., Pyle, M., Koper, K., & Schmandt, B. (2022). Combining deep learning with physics based features in explosion-earthquake discrimination. *Geophysical Research Letters*, *49*, e2022GL098645. https://doi.org/10.1029/2022GL098645

Koper, K. D. (2020). The importance of regional seismic networks in monitoring nuclear test-ban treaties. *Seismological Research Letters*, *91*(2A), 573–580. https://doi.org/10.1785/0220190160

Krizhevsky, A., Sutskever, I., & Hinton, G. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097–1105).

Kuang, W., Yuan, C., & Zhang, J. (2021). Real-time determination of earthquake focal mechanism via deep learning. *Nature Communications*, *12*(1), 1432. https://doi.org/10.1038/s41467-021-21670-x

Lay, T. (2021). *DPRK nuclear explosion yields from teleseismic modeling (Tech. Rep.)*. University of California.

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

Linville, L., Pankow, K., & Draelos, T. (2019). Deep learning models augment analyst decisions for event discrimination. *Geophysical Research Letters*, *46*(7), 3643–3651. https://doi.org/10.1029/2018GL081119

Linville, L. M. (2022). Event-based training in label-limited regimes. *Journal of Geophysical Research: Solid Earth*, *127*(1), e2021JB022820. https://doi.org/10.1029/2021jb022820

Maceira, M., Blom, P. S., Maccarthy, J. K., Marcillo, O. E., Euler, G. G., Begnaud, M. L., & Slinkard, M. E. (2017). Trends in nuclear explosion monitoring research & development - A physics perspective. 17-21274.

Megahed, F. M., Chen, Y.-J., Megahed, A., Ong, Y., Altman, N., & Krzywinski, M. (2021). The class imbalance problem. *Nature Methods*, *18*(11), 1270–1272. https://doi.org/10.1038/s41592-021-01302-4

Mellors, R. J., Pitarka, A., Matzel, E., Magana-Zook, S., Knapp, D., Walter, W. R., et al. (2018). The source physics experiments large n array. *Seismological Research Letters*, *89*(5), 1618–1628. https://doi.org/10.1785/0220180072

Mousavi, S. M., & Beroza, G. C. (2019). A machine-learning approach for earthquake magnitude estimation. *Geophysical Research Letters*, *47*(1), e2019GL085976. https://doi.org/10.1029/2019GL085976

Mousavi, S. M., & Beroza, G. C. (2020). A machine-learning approach for earthquake magnitude estimation. *Geophysical Research Letters*, *47*(1), e2019GL085976. https://doi.org/10.1029/2019gl085976

Mousavi, S. M., Zhu, W., Sheng, Y., & Beroza, G. C. (2019). CRED: A deep residual network of convolutional and recurrent units for earthquake signal detection. *Scientific Reports*, *9*(1), 10267. https://doi.org/10.1038/s41598-019-45748-1

Nakano, M., Sugiyama, D., Hori, T., Kuwatani, T., & Tsuboi, S. (2019). Discrimination of seismic signals from earthquakes and tectonic tremor by applying a convolutional neural network to running spectral images. *Seismological Research Letters*, *90*(2), 530–538. https://doi.org/10.1785/0220180279

National Research Institute for Earth Science and Disaster Resilience. (2019). Nied hi-net. National research institute for earth science and disaster resilience. https://doi.org/10.17598/NIED.0003

Park, Y., Beroza, G. C., & Ellsworth, W. L. (2023). A mitigation strategy for the prediction inconsistency of neural phase pickers. *Seismological Society of America*, *94*(3), 1603–1612. https://doi.org/10.1785/0220230003

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

Pu, Y., Apel, D. B., Liu, V., & Mitri, H. (2019). Machine learning methods for rockburst prediction-state-of-the-art review. *International Journal of Mining Science and Technology*, *29*(4), 565–570. https://doi.org/10.1016/j.ijmst.2019.06.009

Ren, C. X., Hulbert, C., Johnson, P. A., & Rouet-Leduc, B. (2020). Chapter two - Machine learning and fault rupture: A review. In B. Moseley & L. Krischer (Eds.), *Machine learning in geosciences* (Vol. 61, pp. 57–107). Elsevier. https://doi.org/10.1016/bs.agph.2020.08.003

Romeo, G. (1994). Seismic signals detection and classification using artificial neural networks. *Annals of Geophysics*, *37*(3).

Ross, Z. E., Meier, M.-A., Hauksson, E., & Heaton, T. H. (2018). Generalized seismic phase detection with deep learning. *Bulletin of the Seismological Society of America*, *108*(5), 2894–2901. https://doi.org/10.1785/0120180080

Ross, Z. E., Yue, Y., Meier, M.-A., Hauksson, E., & Heaton, T. H. (2019). PhaseLink: A deep learning approach to seismic phase association. *Journal of Geophysical Research: Solid Earth*, *124*(1), 856–869. https://doi.org/10.1029/2018jb016674

Rouet-Leduc, B., Hulbert, C., Lubbers, N., Barros, K., Humphreys, C. J., & Johnson, P. A. (2017). Machine learning predicts laboratory earthquakes. *Geophysical Research Letters*, *44*(18), 9276–9282. https://doi.org/10.1002/2017gl074677

Schaff, D. P., Kim, W.-Y., Richards, P. G., Jo, E., & Ryoo, Y. (2018). Using waveform cross correlation for detection, location, and identification of aftershocks of the 2017 nuclear explosion at the North Korea test site. *Seismological Research Letters*, *89*(6), 2113–2119. https://doi.org/10.1785/0220180132

Sick, B., Guggenmos, M., & Joswig, M. (2015). Chances and limits of single-station seismic event clustering by unsupervised pattern recognition. *Geophysical Journal International*, *201*(3), 1801–1813. https://doi.org/10.1093/gji/ggv126

Snelson, C. M., Abbott, R. E., Broome, S. T., Mellors, R. J., Patton, H. J., Sussman, A. J., & Walter, W. R. (2013). Chemical explosion experiments to improve nuclear test monitoring. *Eos, Transactions American Geophysical Union*, *94*(27), 237–239. https://doi.org/10.1002/2013eo270002

Song, G., Cheng, J., & Grattan, K. T. V. (2020). Recognition of microseismic and blasting signals in mines based on convolutional neural network and Stockwell transform. *IEEE Access*, *8*, 45523–45530. https://doi.org/10.1109/ACCESS.2020.2978392

Stump, B. W., Pearson, D. C., & Reinke, R. E. (1999). Source comparisons between nuclear and chemical explosions detonated at Rainier Mesa, Nevada Test Site. *Bulletin of the Seismological Society of America*, *89*(2), 409–422. https://doi.org/10.1785/BSSA0890020409

10.1029/2022GL101528

Tian, X., Wang, M., Zhang, X., Wang, X., Sheng, S., & Lü, J. (2022). Discrimination of earthquake and quarry blast based on multi-input convolutional neural network. *Chinese Journal of Geophysics*, *65*(5), 1802–1812.

Tibi, R. (2021). Discrimination of seismic events (2006–2020) in North Korea using P/Lg amplitude ratios from regional stations and a bivariate discriminant function. *Seismological Research Letters*, *92*(4), 2399–2409. https://doi.org/10.1785/0220200432

Trani, L., Pagani, G. A., Zanetti, J. P. P., Chapeland, C., & Evers, L. (2022). Deepquake — An application of CNN for seismo-acoustic event classification in The Netherlands. *Computers & Geosciences*, *159*, 104980. https://doi.org/10.1016/j.cageo.2021.104980

Voytan, D. P., Lay, T., Chaves, E. J., & Ohman, J. T. (2019). Yield estimates for the six North Korean nuclear tests from teleseismic *P* wave modeling and intercorrelation of *P* and *Pn* recordings. *Journal of Geophysical Research: Solid Earth*, *124*(5), 4916–4939. https://doi.org/10.1029/2019jb017418

Walter, W. R., Dodge, D. A., Ichinose, G., Myers, S. C., Pasyanos, M. E., & Ford, S. R. (2018). Body-wave methods of distinguishing between explosions, collapses, and earthquakes: Application to recent events in North Korea. *Seismological Research Letters*, *89*(6), 2131–2138. https://doi.org/10.1785/0220180128

Wu, Y., Zhang, J., Chen, X., Zou, S., & Yang, M. (2020). Research on the recognition of infrasound signal of nuclear explosion by SVM and CNN. In *IOP conference series: Earth and environmental science* (Vol. 610).012010

Yearbook, S. (2021). *Armaments, disarmament and international security summary*. Stockholm International Peace Research Institute.

Yeck, W. L., Patton, J. M., Ross, Z. E., Hayes, G. P., Guy, M. R., Ambruz, N. B., et al. (2020). Leveraging deep learning in global 24/7 real-time earthquake monitoring at the national earthquake information center. *Seismological Research Letters*, *92*(1), 469–480. https://doi.org/10.1785/0220200178

Zhang, X., Yuan, C., Zhang, J., Liu, S., Chen, Z., & Li, W. (2020). Locating induced earthquakes with a network of seismic stations in Oklahoma via a deep learning method (Vol. *10*, p. 1941).

Zhu, L., Peng, Z., McClellan, J., Li, C., Yao, D., Li, Z., & Fang, L. (2019). Deep learning for seismic phase detection and picking in the aftershock zone of 2008 M_w7.9 Wenchuan earthquake. *Physics of the Earth and Planetary Interiors*, *293*, 106291.

Zhu, W., & Beroza, G. C. (2019). PhaseNet: A deep-neural-network-based seismic arrival-time picking method. *Geophysical Journal International*, *216*(1), 261–273.