Global Nuclear Blast Discrimination using a Convolutional Neural Network

Barama Louisa¹, Williams Jesse², Newman Andrew V¹, and Peng Zhigang¹

¹Georgia Institute of Technology
²Global Technology Connection

November 16, 2022

Abstract

Using P-wave seismograms by Barama et al. (2022), we trained a seismic source classifier using a Convolutional Neural Network. We trained for three classes: earthquake P-wave, nuclear P-wave, and noise. Seismograms with low signal to noise ratios (SNR) adversely affect the model performance, thus a threshold was applied, limiting the training set size. Our method can accurately characterize most events, finding over 95% signals in the validation set, even with the SNR-limited training data. We applied the model on holdout datasets of the North Korean test blasts to evaluate the performance on unique region and station-source pairs. Additionally, we tested on the Source Physics Experiment events to investigate the potential for chemical blasts to act as a surrogate for nuclear blasts. 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.
Global Nuclear Blast Discrimination using a Convolutional Neural Network

Louisa Barama1*, Jesse Williams2, Andrew V. Newman1, and Zhigang Peng1

1School of Earth and Atmospheric Sciences, Georgia Institute of Technology, 311 Ferst Drive, Atlanta GA, 30332
2Global Technology Connection, Inc., 2839 Paces Ferry Rd, Atlanta GA, 30339

Key Points:

• We successfully discriminate underground nuclear blasts with a Convolutional Neural network (CNN) trained on seismic data.
• Robust global seismic event discrimination is possible with machine learning trained on regional and teleseismic data.
• A CNN trained with historical nuclear blast data can be applied with high accuracy to other regions, like the 6 DPRK test blasts.

*Ford ES&T Building, Room 2106, School of Earth and Atmospheric Sciences, Georgia Institute of Technology, 311 Ferst Drive, Atlanta GA, 30332

Corresponding author: Louisa Barama, lbarama@gatech.edu
Abstract

Using P-wave seismograms by Barama et al. (2022), we trained a seismic source classifier using a Convolutional Neural Network. We trained for three classes: earthquake P-wave, nuclear P-wave, and noise. Seismograms with low signal to noise ratios (SNR) adversely affect the model performance, thus a threshold was applied, limiting the training set size. Our method can accurately characterize most events, finding over 95% signals in the validation set, even with the SNR-limited training data. We applied the model on holdout datasets of the North Korean test blasts to evaluate the performance on unique region and station-source pairs. Additionally, we tested on the Source Physics Experiment events to investigate the potential for chemical blasts to act as a surrogate for nuclear blasts. 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.

1 Introduction

Nuclear weapons can have disastrous effect 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 nuclear blast 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 moratorium on 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 underground nuclear blasts from natural earthquakes and other near surface seismic signals (Maceira et al., 2017). Improved and continued development of nuclear blast detection using seismic methods is imperative given that underground nuclear blast 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, and P to S wave ratio calculations. Machine learning (ML) methods are widely spreading as a tool in seismology (Kong et al., 2019) and detection methods with encouraging potential such as with the USGS National Earthquake Information Center (NEIC) that is using ML for simultaneously local, regional, and global real-time monitoring processing systems (Yeck et al., 2020). ML methods can allow generalizable models that can identify events outside of those used in initial training. This is essential, especially for nuclear explosion monitoring where nuclear weapon test blasts are completely banned according to the CTBT, meaning that all current and future monitoring research and development is dependent on historical nuclear explosion data.

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 (underground nuclear explosions, underwater explosions, volcano-tectonic events) for events recorded 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 (Dowla et al., 1990; Del Pezzo et al., 2003), which tend to limit
their 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 dataset for classifying and discriminating seismic event types or phase determinations (Nakano et al., 2019; Bergen et al., 2019; Kong et al., 2019) with encouraging results. Mousavi and Beroza (2020) showed that ML methods can outperform more standard approaches for global earthquake detection, including those that compare short- to long-term waveform amplitudes.

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 (W. Zhu & Beroza, 2019; Ross et al., 2018, 2019; L. Zhu et al., 2019), event detection and discrimination (Mousavi et al., 2019; Nakano et al., 2019; Mousavi & Beroza, 2019) and source location (Perol et al., 2018; Zhang et al., 2020), and magnitude determination (Mousavi & Beroza, 2020). Generalized Global ML models have been successfully applied to real-time magnitude classification (Mousavi & Beroza, 2019; Chakraborty et al., 2021) and regional focal mechanism predictions (Kuang et al., 2021), as well as for global operational processing systems at monitoring institutions like the USGS NEIC (Yech et al., 2020). Other machine learning methods such as Support Vector Machines (SVM) and self-organizing map (SOM) have also been used for natural earthquake and mining blast discrimination (Pu et al., 2019; Sick et al., 2015; Kim et al., n.d.). Studies have found improved recognition ability with CNNs over Support Vector Machines (SVM) 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; Dong et al., 2020; Linville et al., 2019; Trani et al., 2022; Huang et al., 2018; Song et al., 2020). However, models like Mousavi et al. (2019)’s earthquake detector "CRED" fail to successfully apply to other regions which suggests that complex models built upon large databases can have difficulties generalizing to other areas.

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 blasts) for further analysis. Recently, Barama et al. (2022) compiled a comprehensive seismic waveform and event catalog termed GTUNE (Georgia Tech Underground Nuclear Explosions) from a wide range of sources for declassified underground nuclear explosions. With this development, we see a potential to use the older and lower quality historical data from nuclear blasts in combination with high quality data from the more recent events for developing more robust ML-based global nuclear blast detectors. Here we show that this dataset can be used to train a CNN model to automatically discriminate seismic waves generated by earthquakes and nuclear explosions. Developing these algorithms have far-reaching potential beyond automatic nuclear detection as well, and may be adapted to automatic detection of seismic signals associated magmatic movements at depth, dangerous slow-source or giant tsunami-generating earthquakes, or other features that may rapidly and automatically illuminate a host of geophysical hazards (e.g. initiation phase of seismogenic landslides).

2 Data

The labels for the three training classes of earthquake $P$-waves, underground nuclear blast (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 Sci-
ence and Disaster Resilience (NIED), as well as other networks. The NIED Hi-Net ar-
array of data are particularly useful for dense observations of recent North Korean (DPRK) nuclear tests at regional to teleseismic distances. Overall, the nuclear blast data are from 600 underground nuclear tests from the year 1961 to 2017 (Figure S1). Earthquake 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 ?? shows the distribution of underground nuclear blasts and stations. Since most historical data was recorded on single channel seismographs, we use only vertical component waveforms for training. Linville et al. (2019) showed successful discrimination between surface mine blasts 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 nuclear blast P-arrivals fixed at 5 seconds, which means we are training on 15 seconds of earthquake and nuclear blast signal. We choose this window length in order to capture the more impulsive nature of the blast waveforms when compared to earthquakes recorded at teleseismic dis-
tances.

A data processing pipeline was created that ingests raw seismic signals and produced a 3-class probability (earthquake, nuclear blast, 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. Functions for demeaning, resampling, applying a cosine enve-
lope function and four-corner bandpass filter. Additionally we developed and integrated a method into the PyWave library that measures the energy (square of the amplitudes) of pre-event noise and event signal, and calculates the ratio. This energy filter method was used to further refine the dataset by excluding those waveforms with a very low sig-
nal to noise ratio (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 differ-
ent sampling rates. Only down-sampling is performed with all original signals below the 20 sps discarded. The waveform is demeaned and then a cosine envelope function is ap-
plicated at the beginning and end 4% of the re-sampled window. This minimizes edge ef-
fects 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 valu-
able information for all data. We found we retained the most data of energy filter thresh-
old greater than 5 as well as model performance with data filtered between 1 Hz and 5 Hz using a Butterworth filter.

Maceira et al. (2017) showed that seismo-acoustic wave amplitude - yield relation-
ship 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 re-
taining all salient features. When training an algorithm, feature importance can be as-
sociated with feature amplitude. Thus an algorithm generally considers waveforms with a strong feature as more important and focus on these features during the training pro-
cess. 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
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, nuclear blast, 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: $8 \rightarrow 16 \rightarrow 32 \rightarrow 32 \rightarrow 16 \rightarrow 16 \rightarrow 8 \rightarrow 8$.

In the supplementary material we detail the CNN classifier structure as adjusted for a 60 second window sampled at 20 sps (input size of 1200) with 1 channel vertical component data. The CNN classifier was trained and tested on a subset of the cumulative dataset discussed above. Interestingly, the best performing models were those with an energy threshold of $\sim 5$; higher signal to noise ratio (SNR) thresholds would substantially reduce the training data used. After applying the energy filter, the nuclear dataset retained 14,841 observations. 1000 traces of this dataset 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 dataset for training and testing. Since it is important to keep a balanced training and testing dataset to equally represented classes and give similar internal weighting to results, the earthquake and noise datasets were
randomly reduced to match the nuclear dataset training and test size. For additional information, see Table 1.

### 4 Results and Testing Model Performance

Using the training and test data presented in Table 1, a single-channel CNN model was trained on the vertical component, pre-processed seismograms. After tuning the hyperparameters, 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 (Fig. 2). The holdout datasets include earthquakes, noise, nuclear blast, source physics experiment (SPE) chemical blasts and DPRK holdout dataset. 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. This shows that the trained CNN was highly capable of classifying signals curated in manner similar to this study, i.e. 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 modeling supported by chemical blast data like the Source Physics Experiments (SPE). Waveforms from the SPE tests Snelson et al. (2013) were used to test how well chemical blasts can emulate low-yield nuclear blasts for seismic detection. When the CNN classifier was applied to the SPE data, the model performed poorly at classifying SPE blasts as nuclear blasts (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 nuclear blast labels in the CNN training data, the CNN showed only a slightly reduced performance of 89% testing accuracy. This seems to supports the assumption that such chemical blasts may act as a surrogate for small nuclear blast data at local and regional distances, where their SNRs are relatively high (Mellors et al., 2018; Stump et al., 1999). Even though the waveforms are normalized, the SPE blasts have much smaller yield than nuclear test blasts we trained the CNN model with. This test highlights the limitation of our global CNN classifier and suggests that lower yield blasts may be detected when trained with more local/regional data.

---

**Table 1.** Number of observations in training, testing, and holdout datasets.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Earthquake</th>
<th>Nuclear</th>
<th>Noise</th>
</tr>
</thead>
<tbody>
<tr>
<td>Train</td>
<td>10349</td>
<td>10349</td>
<td>10349</td>
</tr>
<tr>
<td>Test</td>
<td>1068</td>
<td>1068</td>
<td>1069</td>
</tr>
<tr>
<td>Earthquake holdout</td>
<td>1000</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Nuclear holdout</td>
<td>0</td>
<td>1000</td>
<td>0</td>
</tr>
<tr>
<td>Noise holdout</td>
<td>0</td>
<td>0</td>
<td>1000</td>
</tr>
<tr>
<td>SPE holdout</td>
<td>0</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td><em>DPRK holdout</em></td>
<td>0</td>
<td>4871</td>
<td>0</td>
</tr>
</tbody>
</table>

* DPRK holdout is total waveforms from DPRK events not included in training of the CNN model.
Figure 2. (A) CNN accuracy during training and on earthquake, nuclear blast, 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 nuclear blasts with the exception of the SPE holdout set. (C) Example showing the model predictions on continuous 20 minute seismograms from the May 21, 1992 Mb6.5 Lopnor nuclear test in China.

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

4.2 Holdout Testing on the North Korea Nuclear Test Blasts

Up to now we randomly split the dataset 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), 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. This reduced the training dataset from 11,349 to 9,338 traces per label. Figure S1 in the supplementary information shows the distribution of stations used for DPRK events for test-
Figure 3. Testing on the DPRK holdout dataset model and the global dataset (A) Predicted probabilities of nuclear blast for each of the 6 DPRK nuclear test blasts (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 nuclear blast probability and station distance for 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.
quake 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 nuclear blast 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 result suggest that our CNN model is capable of identifying nuclear blasts from a region excluded from the training dataset 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 datasets

The results so far have shown that our CNN classifier was effective on a curated dataset. Next we perform additionally test for its performance on continuous waveform data where we apply a moving average to a three-class predictor (Fig. 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 May 12, 1992 Lopnor China nuclear test (Fig. 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 nuclear blast 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

This work shows promising applications in discrimination of both regular earthquake and nuclear events recorded at a global network of stations. Given the modest size of our training set, we have higher prediction variance and more false positives (predicting earthquakes as nuclear blast) 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. The issue is still quite small, affecting only 2% of the signals. Too, as this method is testing solely on individual waveforms, network base solutions should quickly correct for most, if not all of these false positives.

Chakraborty et al. (2021) found that length of waveform used for classification did not affect earthquake magnitude predictions. Here, we found that waveform length (15, 30, and 60 s) also doesn’t significantly affect the final accuracy, resulting in close to equal accuracy but smaller training dataset. However, shorter wavelengths converge in training sooner than longer waveforms.

In a comparative study, Chen et al. (2018) used a small dataset 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 nuclear blasts is a promising result. 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 both global seismic sources and stations. How-
ever, there are opportunities to explore on how to create improved performance tests on
our 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. With
a larger training dataset we could test how a regional model, trained based on seismic
stations within selected source-receiver distances in a region, applying more specific pre-
processing and training of the data based on station distance, compares to the global model.
Modifications in the training method to require unique source-receiver pairs between train-
ing and test sets could be helpful, 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.

6 Conclusions

In this study, we successfully built a robust global earthquake, nuclear blast, 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 underground nuclear
blast, and 95.6% for earthquake. This study demonstrate that ML classifier systems can
have broad application for other global and “small data” seismic signals including deep
earthquakes, volcanic and non-volcanic tremor, anomalous earthquakes, including slow-
sourced tsunami earthquakes, glacier/ice-quakes or landslides, without necessarily rely-
ning on the use of synthetic or augmented data.

7 Open Research

AGU requires an Availability Statement for the underlying data needed to under-
stand, evaluate, and build upon the reported research at the time of peer review and pub-
lication.

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 number AF19A-
T012, and contract number FA9453-19-P-0684). This study was performed using python
and TensorFlow.

References

bled global seismic dataset of underground nuclear test blasts. Seismological
Research Letters.
learning for data-driven discovery in solid earth geoscience., 363(eaauf0323).
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
Chen, R.-h., HUANG, H.-m., & CHAI, H.-m. (2018). Study on the discrimina-
tion of seismic waveform signals between earthquake and explosion events by
Del Pezzo, E., Esposito, A., Giudicepietro, F., Marinaro, M., Martini, M., & Scar-
Denny, M. D. (1994). Proceedings of the symposium on the non-proliferation ex-


Supporting Information for “Global Nuclear Blast Discrimination using a Convolutional Neural Network”

Louisa Barama¹, Jesse Williams², Andrew V. Newman¹, and Zhigang Peng¹

¹School of Earth and Atmospheric Sciences, Georgia Institute of Technology, 311 Ferst Drive, Atlanta GA, 30332
²Global Technology Connection, Inc., 2839 Paces Ferry Rd, Atlanta GA, 30339

Contents of this file

1. Figures S1
2. Table S1 Description

Additional Supporting Information (Files uploaded separately as excel file)

1. Captions for large Figure S1: Global map of all Underground Nuclear Explosions (UNE) (red stars) and stations (triangles) used in this study. The dark blue triangles are stations that recorded UNE’s from 1961 to 1998, and the purple triangles are the stations that recorded the DPRK test blasts 2006 to 2017. NIED Hi-Net stations, which are a rich regional dataset to evaluated the DPRK tests, are represented with light blue triangles.

2. Captions for large Table S1: Details of the CNN architecture and layers used for the model in this study. Example shown and number of possible training parameters is for a one minute long training waveform sampled at 20 sps (1200 samples long).
Introduction

The supplementary information includes one figure that shows a global map of all blast sources and data station locations and one table that details the architecture and training parameters of the Convolutional Neural Network trained as our classifier. The architecture shown is for the example case of a one minute long input waveform (1200 samples).
Figure S1. Global map of all Underground Nuclear Explosions (UNE) (red stars) and stations (triangles) used in this study. The dark blue triangles are stations that recorded UNE’s from 1961 to 1998, and the purple triangles are the stations that recorded the DPRK test blasts 2006 to 2017. NIED Hi-Net stations, which are a rich regional dataset to evaluated the DPRK tests, are represented with light blue triangles.

Table S1. Details of the Convolutional Neural Network (CNN) layers and training parameters used for our classifier. Shown, is the example case for a one-minute long input waveform (1200 samples).

\(^{a}\) Uploaded separately as an excel .csv file.