Deep learning and transfer learning of earthquake and quarry-blast discrimination: Applications to southern California and eastern Kentucky

Jun Zhu[1], Lihua Fang[2], Fajun Miao[3], Liping Fan[4], Ji Zhang[1], and Zefeng Li[1]

[1]University of Science and Technology of China
[2]Institute of Geophysics, China Earthquake Administration
[3]Jiangsu Earthquake Administration, China Earthquake Administration
[4]China Earthquake Administration

November 24, 2022

Abstract

Discrimination between tectonic earthquakes and quarry blasts is important for accurate earthquake cataloging and seismic hazard analysis. However, reliable classification is challenging with raw waveforms and no prior knowledge of source parameters. Here we apply deep learning to perform this task in southern California and eastern Kentucky, which differ significantly in available labelled data, class imbalance and waveform characteristics. Accordingly, we adopt different strategies for the two regions. First, we directly train a convolutional neural network (CNN) for southern California due to its data abundancy. To alleviate class imbalance, the blast data are augmented by randomly shifting waveform windows. The model for California yields an accuracy of 91.97% for single-station classification and 97.54% for network-averaged classification. Second, as eastern Kentucky has a much smaller data size, we fine-tune the pretrained California model to fit the Kentucky data. The fine-tuned model yields an accuracy of 97.35% for single-station classification and 99.46% for network-averaged classification. The fine-tuned model outperforms the model trained from scratch. Finally, we use occlusion test and gradient-weighted class activation mapping to illuminate which parts of waveforms are important for model prediction. Our results demonstrate that deep learning can achieve high accuracy in seismic event discrimination with raw waveforms and that transfer learning is effective and efficient to generalize deep learning models across different regions.

Deep learning and transfer learning of earthquake and quarry-blast discrimination: Applications to southern California and eastern Kentucky

Jun Zhu[1], Lihua Fang[2], Fajun Miao[3], Liping Fan[2], Ji Zhang[1], and Zefeng Li[1,4*]

1. *Laboratory of Seismology and Physics of Earth's Interior, School of Earth and Space Sciences, University of Science and Technology of China, Hefei, 230026, China*
2. *Institute of Geophysics, China Earthquake Administration, Beijing, 100081, China*
3. *Jiangsu Earthquake Administration, China Earthquake Administration, Nanjing, 210014, China*
4. *Mengcheng National Geophysical Observatory, University of Science and Technology of China, Mengcheng, 233500, China*

*Corresponding author: Zefeng Li (zefengli@ustc.edu.cn)

Submitted to Journal of Geophysical Research: Solid Earth

Date: 26 April 2022

Key Points:

1. We train deep learning and transferred models to classify local earthquakes and quarry blasts in southern California and eastern Kentucky.
2. These models directly take inputs of minimally processed waveforms and have potential to operate in real time.
3. We show that transfer learning is effective and efficient to generalize deep learning models across different regions.

Abstract

Discrimination between tectonic earthquakes and quarry blasts is important for accurate earthquake cataloging and seismic hazard analysis. However, reliable classification is challenging with raw waveforms and no prior knowledge of source parameters. Here we apply deep learning to perform this task in southern California and eastern Kentucky, which differ significantly in available labelled data, class imbalance and waveform characteristics. Accordingly, we adopt different strategies for the two regions. First, we directly train a convolutional neural network (CNN) for southern California due to its data abundancy. To alleviate class imbalance, the blast data are augmented by randomly shifting waveform windows. The model for California yields an accuracy of 91.97% for single-station classification and 97.54% for network-averaged classification. Second, as eastern Kentucky has a much smaller data size, we fine-tune the pretrained California model to fit the Kentucky data. The fine-tuned model yields an accuracy of 97.35% for single-station classification and 99.46% for network-averaged classification. The fine-tuned model outperforms the model trained from scratch. Finally, we use occlusion test and gradient-weighted class activation mapping to illuminate which parts of waveforms are important for model prediction. Our results demonstrate that deep learning can achieve high accuracy in seismic event discrimination with raw waveforms and that transfer learning is effective and efficient to generalize deep learning models across different regions.

Plain Language Summary

Discrimination between tectonic earthquakes and quarry blasts is needed to properly evaluate seismic hazards. In this study, we build two deep learning models to automatically discriminate them in southern California and eastern Kentucky. As California has more seismic data than Kentucky, we first train a deep learning model for California and then fine tune it for Kentucky. Using minimally processed seismic waveforms as input, both the California and the Kentucky models achieve classification accuracy of 91.97% and 97.35%, respectively. In Kentucky, we compare the performances of the fine-tuned California model and the one trained with the Kentucky data from scratch, and find that the fine-tuned model outperforms the other by 2.99%. This demonstrates that transfer learning is an economic way to build a high-quality deep learning model with small data sets. Our results show that deep learning can achieve a high accuracy in seismic event discrimination with raw waveforms.

1 Introduction

Modern seismic networks have improved considerably over the past decades and increasingly recorded diverse seismic signals other than earthquakes. These signals have natural origins such as landslides and debris flow, or anthropogenic origins such as industry exploitation blasts and traffic flow. On one hand, discrimination between local tectonic earthquakes and other seismic signals is important for seismic hazard analysis. Recent studies show that seismic catalogs contaminated by quarry blasts can result in an overestimated b-value and an underestimated probability of large earthquakes (Gulia & Gasperini, 2021; Gulia & Wiemer, 2019; Tang et al., 2020). On the other hand, classification of these signals can provide valuable resources in environmental seismology (Larose et al., 2015), urban seismology (Díaz et al., 2017) and forensic seismology (Douglas, 2013). For example, in mining areas, monitoring blasting events with small magnitudes has practical uses for supervising safe production and detecting illegal mining activities (Banchirigah, 2008). Either to clean up earthquake catalogs or to construct catalogs of non-earthquake events, interests to classify different events have grown rapidly over the past years. However, visual identification is laborious and subjective. Reliable automated classification methods are necessary, especially in real-time processing and earthquake early warning (Li et al., 2018).

Discrimination between earthquakes and quarry blasts has been a challenging task in seismic network operation for a long time owing to their apparent similarity in waveform characteristics (Astiz et al., 2014). Seismologists have proposed a variety of automated approaches, which can be roughly divided into two categories: source-parameter-based and waveform-based. The former relies on source information like location and/or origin time (Fereidoni & Atkinson, 2017; Renouard et al., 2021; Wiemer, 2000). For example, blasts tend to occur in known quarry sites and in the daytime. However, source information requires accurate preceding analyses, which limits its application in real-time processing. Comparatively, waveform-based approaches use seismic waveform features only, which can be manually defined by experienced experts or automatically extracted from data (Allmann et al., 2008; Hartse et al., 1997; Koper et al., 2016, 2021; Korrat et al., 2022; L. Linville et al., 2019; Miao et al., 2020; Rodriguez Asihama, 2016; Su et al., 1991; Tibi et al., 2018, 2019; Wang et al., 2020, 2021).

Manually defined features include spectral ratios of Lg (Bennett & Murphy, 1986) and Rg (Tibi et al., 2018), P-to-S phase ratios (Hartse et al., 1997; Wang et al., 2020, 2021), $M_L - M_C$ (M_L, local magnitude; M_C, coda magnitude. Koper et al., 2016, 2021; Wang et al., 2021), and the misfit of P-wave spectra to an ω^{-2} source model (Allmann et al., 2008). Particularly, blast signals tend to have emergent P waves and weak S waves, upward P wave polarity and Gaussian-like envelopes, whereas earthquake signals often have clear P and S waves (Miao et al., 2020; Stump et al., 2002; Tang et al., 2020). Generally, the performance of waveform-based approaches depends on the generalizability of the extracted features. Features crafted for one region may be unsuitable for another so that the classification performance could drop significantly. Moreover, waveform features cannot be reliably measured in presence of high noise and thus small events could be difficult to classify (L. Linville et al., 2019; Tibi et al., 2019).

Instead of using manually defined features, deep learning enables direct extraction of implicit features from data (L. Linville et al., 2019; Liu et al., 2020; Ross, Meier, & Hauksson, 2018; Ross, Meier, Hauksson, et al., 2018; L. Zhu et al., 2019; W. Zhu & Beroza, 2018). The features are sought greedily during the training process that maps the input waveforms to the output class. Thereby, they are likely more representative than manually-defined features. However, deep learning models contain a large number of parameters and thus require large sets of high-quality labelled data to avoid overfitting (i.e., the model memorizes the data rather than learns the generalizable rules). Such data sets may not be always available, especially in areas with a short monitoring history, a low activity level, or lack of manual labels. In this case, transfer learning could be helpful. It leverages the knowledge learned from a rich dataset and applies to another dataset with minor modifications (Chai et al., 2020; Do & Ng, 2005; Ismail Fawaz et al., 2018; Pan & Yang, 2010; Yosinski et al., 2014; Y. Zhu et al., 2011). With a well-trained base model, transfer learning only takes a small number of labels to adequately fine-tune the model.

In this study, we build a deep learning model for southern California using a large number of labelled earthquakes and blasts, and on this basis apply transfer learning to build another model for eastern Kentucky. We use different strategies to the two regions because they have distinct amount of data. Moreover, we implement data augmentation to mitigate the effect of imbalanced data in the two regions. The southern California model achieves an accuracy of 91.97%, an F1-score of 95.00% for earthquakes and 79.62% for blasts. The transfer-learned model to eastern Kentucky achieves an accuracy of 97.35%, an F1-score of 74.59% for earthquakes and 98.60% for blasts. In both cases, the smaller class (blasts in southern California, earthquakes in eastern California) tends to have a lower precision and F1-score.We implement occlusion (Zeiler & Fergus, 2014) test and gradient-weighted class activation mapping (Grad-CAM) (Selvaraju et al., 2020) to investigate which parts of the waveforms are important for decision making. Finally, we discuss the advantages and practical considerations of our models and offer suggestions for further use of them.

2 Data

2.1 Southern California

Southern California is instrumented by the Southern California Seismic Network (SCSN) with a long earthquake monitoring history since 1932. Presently it provides not only waveform data archives but also high-quality earthquake catalogs with information of origin time, location, magnitudes, phase picks, and event types (Hutton et al., 2010). The event types cataloged in SCSN include local, regional and remote earthquakes, as well as quarry blasts and sonic events. In this study we keep one of every six earthquakes in the SCSN catalogs from 2011 to 2020 to mitigate data imbalance, and finally obtain 37,702 local earthquake events and 6,690 blast events recorded by 329 seismic stations with a maximal epicentral distance of 100 km (Figures 1a and b). This results in a total of 1,721,092 three-component earthquake recordings and 312,911 blast recordings, a class ratio approximately 5.5:1. Data augmentation for blasts is still needed and will be described in the method section.

The waveforms are preprocessed as follows. First, the waveforms are linearly detrended, tapered with a Hanning window, and normalized by the maximal standard deviation of three components. Second, we pick the P arrivals by applying a short-term-average/long-term-average picker (Allen, 1978) within 3 s around predicted arrivals on the 1-15 Hz bandpass filtered vertical component (Figure 2a). The theoretical prediction uses the Hadley-Kanamori 1D model (Hadley & Kanamori, 1977; Kanamori & Hadley, 1975). The short-term and long-term windows are 0.4 s and 4 s, respectively. Third, we only keep the recordings with signal-to-noise ratio (SNR) >3 dB for further analysis. We use a relatively low SNR threshold to include many noisy waveforms so that the model performance could be improved in low SNRs. Finally, the raw waveforms are cut 5 s before and 50 s after the first arrival (Figure 2b). The resulting data set contains 17,336 M_L −0.48 ~ 5.51 earthquake events (319,576 recordings) and 3,671 M_L −0.17 ~ 2.56 blast events (64,655 recordings).

2.2 Eastern Kentucky

Another dataset used in this study is from Miao et al. (2020) who compiled waveforms of earthquakes and blasts in the Rome trough of eastern Kentucky. It contains 148 natural earthquakes (1,198 recordings) and 3,542 quarry blasts (27,854 recordings) from June 2015 to March 2019, recorded by 20 regional seismic stations and a temporary network EKMMP of 6 seismic stations (Figures 1c and d). The P arrivals are picked by generalized phase detection (Ross, Meier, Hauksson, et al., 2018) and associated as seismic events using PhasePAPy (Chen & Holland, 2016). Other preprocessing steps are similar to those for the California data set.

The eastern Kentucky data set differs from the California one in terms of data size and sense of data imbalance. Southern California has active seismicity and a long history of earthquake monitoring. There are abundant manual labels of local earthquakes and quarry blasts. Eastern Kentucky is relatively quiet in seismicity and has frequent quarry blasts due to prosperous mining industry in the region (Carpenter et al., 2020). As a result, earthquakes outnumber the quarry blasts in southern California, whereas it is the opposite in eastern Kentucky. In addition, most events in Kentucky have an epicentral distance longer than 100 km due to relatively sparse seismic networks, compared to the shorter epicentral distance and denser seismic networks in southern California. These differences are test stones to evaluate the efficacy of transfer learning between two distinct regions.

3 Methods

3.1 Data division and augmentation

We divide the southern California dataset into three subsets, training (70%), validation (10%) and test set (20%, Table 1). To avoid data leakage (i.e., the model training uses information that is unavailable in real-world applications), we split the dataset at the event level rather than the station level. Specifically, all recordings from the same event are assigned to the same subset. To mimic the scenario of near real-time classification, the events in the training, validation and test sets are split chronologically as recommended by Linville et al. (2019) (Table 1). This practice allows assessing the model performance under more realistic monitoring conditions.

As local earthquakes are predominant in the southern California dataset, the trained model tends to predict a recording as an earthquake owing to the Bayesian nature of deep learning. Hence, data augmentation is needed for quarry blasts. We randomly shift (< 5 s) the waveform window to make multiple copies of quarry blasts (Figure 2b). The random shifts also avoid the model's sensitivity to phase picking timing (W. Zhu & Beroza, 2018). We take 5 random shifts for recordings of quarry blasts (hence augment quarry blasts for 5 times) and only 1 random shift for recordings of local earthquakes. Conversely, to balance the two classes in the eastern Kentucky dataset, we augment the number of earthquakes by 23 random shifts and keep 1 random shift for quarry blasts.

3.2 Deep learning: the CNN architecture

CNNs essentially learn a nonlinear function that maps 2-D image or 1-D time-series data to class labels (Lecun et al., 1998). Compared with a multilayer perceptron, a CNN introduces a feature extractor which consists of several convolutional layers (the blue rectangles in Figure 2c). Each layer receives the locally connected input from its previous layer. After convolving the input with a filter (convolutional kernel), the layer subsamples the output and sends it to the subsequent layer at the locally connected position. The convolutional kernels are determined during the training process to parameterize the general object features. As these kernels are shared across a layer and convolutional downsampling is done through the model, the feature extractor is generally insensitive to specific locations and scales of the salient features. Because seismic signals vary widely owing to diverse path and source effects, the invariance in feature shift, scale and distortion makes CNNs suitable for seismic waveforms. Since 2018, CNNs have been widely applied to seismic data processing, such as earthquake detection (Perol et al., 2018), phase picking (Ross, Meier, Hauksson, et al., 2018; L. Zhu et al., 2019; W. Zhu & Beroza, 2018), first-motion polarity identification (Ross, Meier, & Hauksson, 2018) and seismic source discrimination (L. Linville et al., 2019; Liu et al., 2020; Tibi et al., 2019). Especially, CNNs are proposed to distinguish between earthquakes and quarry blasts in Utah with spectrograms as input (Linville et al., 2019; Tibi et al., 2019).

We design a new CNN model to discriminate earthquake and blasts in southern California using minimally processed 50 s three-component waveforms (dimension [5000, 1, 3]). The 50 s window is sufficiently long to include the major phases as well as the coda in both California and Kentucky. The model contains seven convolutional layers and one fully-connected layer (Figure 2c). A Rectified Linear Unit (ReLU) nonlinear activation function is used in each convolutional layer. Each convolutional layer is followed with a batch normalization and a max pooling layer. In the bottom, a fully-connected layer, together with a dropout layer, serves as a downstream classifier. The final output is a two-node probability vector representing the likelihoods of an earthquake and a blast, respectively. We add a softmax activation function in the last layer so that the values of the two nodes are non-negative and their sum is 1.

The CNN model has a total of 150,154 trainable parameters and is built on the Keras framework (Chollet & others, 2015). To avoid overfitting, we terminate training when the loss on validation set fails to decrease over five epochs and save the model with the lowest validation loss, a common training strategy called early stopping in machine learning. The training is run on a work station with Nvidia Graph Process Unit GeForce RTX 2070 for approximately 1 hour.

3.3 Transfer learning: fine-tuning the pretrained model

A deep learning model typically contains tens of thousands to millions of parameters to be determined during the training process. Sufficient data are needed to constrain them without overfitting. As aforementioned, the California model contains 150,154 parameters and is trained on 268,958 recordings. In comparison, the Kentucky data have only 29,052 recordings, an order of magnitude less. Moreover, the Kentucky data have more quarry blasts than earthquakes, different from those in southern California. Besides, the epicentral distance is generally longer in Kentucky. Finally, the differences in velocity and attenuation structures of southern California and eastern Kentucky could lead to differences in the recorded earthquake and blast waveforms. Therefore, direct application of the California model to Kentucky could be unsatisfactory and model modification is required.

Transfer learning provides a convenient tool to modify deep learning models. Typically, deep learning models hypothesize that the training data are independent and identically distributed with the test data; transfer learning relaxes this hypothesis to extend the applications to similar tasks (Tan et al., 2018). Transfer learning proves to be efficient and effective in many other fields (Long et al., 2013, 2015; Pan et al., 2011), whereas its applications in seismology are rather limited. As an early example, Zhu et al. (2019) trained a phase picker on the 2008 Mw 7.9 Wenchuan earthquake sequence and fine-tuned it for an Oklahoma data set. By modifying the PhaseNet (W. Zhu & Beroza, 2018) model which is learned from regional seismic networks, Chai et al. (2020) made it work well for microseismic data with a much higher sample rate and a smaller data size than those for training the original PhaseNet. They reported that the transfer-learned model outperforms both the original PhaseNet and even human analysts.

In practice, fine-tuning a deep learning model is straightforward. Instead of re-training the model from scratch where the model parameters are randomly initialized, transfer learning uses the original model as a starting point and continues to train the model with new data. Alternatively, one can choose to freeze most parameters of the original model and only allow modifying the rest. Here we adopt the former strategy which continues to train the California model with the eastern Kentucky data. This allows tuning the parameters at the widest range while using the knowledge from the California model.

4 Results

4.1 Deep learning for southern California

For each seismogram, the model outputs two probabilities corresponding to the likelihood of each class (earthquake or blast), which sums to be 1. If the probability for natural earthquakes is above/below a threshold level (here set as 0.5), the seismogram is assigned to the earthquake/blast class. As an event is often recorded by multiple stations, more reliable classification can be achieved by averaging over all the available stations. We evaluate the classification performance quantitatively with confusion matrices (Table 2) at both the station level (individual recordings) and the network level (averaged over all stations). To describe the performance in the framework of typical classification tasks, we define blasts as positive predictions and earthquakes as negative predictions. Following this definition, a confusion matrix consists of four elements: (1) True positive (TP): Correctly classified blasts; (2) True negative (TN): Correctly classified earthquakes; (3) False positive (FP): Earthquakes misclassified as blasts; (4) False negative (FN): Blasts misclassified as earthquakes. We further define recall as fraction of true samples that are correctly classified, and precision as fraction of classified samples that are true samples.

The results show that most events are correctly classified (Figure 3), with overall 91.97% accuracy at the station level and 97.54% accuracy at the network level. Figure 3b shows that events with mean probability around 0.5 (uncertain predictions) also have large inter-station standard deviations, indicating discrepancy among individual stations. The recall is 91.73% for earthquakes and 93.18% for quarry blasts at the station level (Figure 4). The precision is 98.52% for earthquakes and 69.50% for blasts. Earthquakes have higher precision because of more earthquakes in the data set. We use the area under the precision-recall curve (AUC) to evaluate our model. AUC measures overall performance by taking both precision and recall into account, which is particularly useful for imbalanced data classification. Figure 4c shows that the AUC for earthquakes is 0.996 and the AUC for blasts is 0.921, confirming a generally good classification performance.

We investigate the performance dependency on SNR, epicentral distance, magnitude and focal depth (Figure 5). The results show that recalls of both earthquakes and blasts increase with SNR and magnitude, and decrease with epicentral distance, suggesting that signal quality is a major impact factor on classification accuracy. Our model shows a slightly increasing recall with focal depth (Figure 5d). This is consistent with previous findings that focal depth is an important discriminant (Koper et al., 2016, 2021) given that most natural earthquakes are deeper than quarry blasts. Finally, we plot the misclassified earthquakes and blasts (FNs and FPs) but find no clear spatial patterns (Figure 3d).

4.2 Transfer learning to eastern Kentucky

As eastern Kentucky has a much smaller data set and different characteristics from California, we compare the performance of different strategies in eastern Kentucky: (1) Directly apply the California model to the Kentucky test set; (2) Re-train the model with the Kentucky data from scratch; (3) Fine-tune the California model with the Kentucky data. Directly applying the California model to the Kentucky test set yields the lowest accuracy (45.38%). The re-trained model performs significantly better with an accuracy of 94.36%. The transfer-learned model achieves the highest 97.35% accuracy which outperforms the other two (Table 3 and Figure 6). The better performance of the transfer-learned model demonstrates that the California model does provide additional useful knowledge to discriminate events in Kentucky.

4.3 Model interpretation: occlusion and Grad-CAM tests

While deep learning proves to be powerful in a wide range of applications, its decision-making process has been known for poor interpretability. As an end-to-end approach, CNNs do not provide direct clues whether it truly identifies major seismic phases (e.g., P, S and coda) like human experts do. However, computer scientists have developed various techniques to shed light on their decision-making process. We use two auxiliary interpretation techniques, namely the occlusion test and the Grad-CAM test, to illustrate the primary characteristics that the models utilize to distinguish two signal types. Both techniques evaluate the importance of given waveform sections for classification. Specifically, occlusion test evaluates the average importance using the entire data set (L. Linville et al., 2019; Zeiler & Fergus, 2014), whereas Grad-CAM evaluates on individual recordings (Selvaraju et al., 2020).

Specifically, occlusion iteratively masks each waveform section (i.e., replaced by zeros) and then monitors the change in classification performance. If a given waveform section is important for decision-making but is masked, the classification accuracy is expected to drop significantly. In this study, the mask window is 1-s long and nonoverlapped. Figures 7a and e show that when the waveforms following the first P arrival are masked, the number of TNs decreases. This indicates that P waves and coda are important signatures for earthquakes on average. The local minima in TN curves are likely caused by the temporally separated P and S arrivals, which are also diagnostic features of natural earthquakes compared to the Gaussian-like envelopes of blasts. As for quarry blasts, the most important waveform windows are around the P wave arrival (Figures 7b and f). This is consistent with previous results that the most obvious characteristic of natural earthquakes is the well-developed S waves due to shear rupture, whereas quarry blasts, often have smaller S wave energy due to volumetric change. We notice that the F1-scores of both classes slightly increase when masking the noise before P arrival or after coda, likely because the removal of irrelevant noise reduces interference in model prediction.

Compared to the occlusion test that evaluates the average contribution, Grad-CAM visualizes the model sensitivity for individual samples (Selvaraju et al., 2020). It first computes the gradients of either class's score with respect to the last convolutional layer's output, i.e., the 256-feature maps in Figure 2c. These gradients are related to the importance of each feature map for the class of interest. Grad-CAM sums the absolute feature maps with the gradients as weights. This produces a weighted forward activation heatmap which represents the significance of the waveforms. The heatmap has the same size as but a different resolution from the input seismogram.

Figure 8 shows example waveforms and the heatmaps for earthquakes and blasts in southern California. From top to bottom, the California model predicts their earthquake probabilities as approximately 1, 0.5 and 0, respectively. Both the correctly and the incorrectly predicted earthquakes have an impulsive P wave followed by a high-frequency S wave (Figures 8a and b). Comparatively, both the incorrectly and the correctly predicted quarry blasts have the relatively long-duration and low-frequency S waves (Figures 8e and f). The events with earthquake probability near 0.5 exhibit weak high-frequency S waves contaminated by P wave coda, possibly resulting in uncertain predictions (Figures 8c and d). In general, these results agree with previous observations that quarry blasts have lower frequency content than earthquakes (Allmann et al., 2008; Korrat et al., 2022; Kortström et al., 2016; Su et al., 1991). Particularly, quarry blasts in southern California are often characterized by long-duration and low-frequency S-waves, a feature that the SCSN analysts often use to distinguish from local earthquakes (personal communication with Shang-Lin Chen).

5 Discussion and conclusions

We have developed two deep learning models to distinguish between local earthquakes and quarry blasts in southern California and in eastern Kentucky. The CNN model takes waveforms as direct input and automatically extracts implicit features that are optimized for classification. These models show generally high classification accuracy (>90%) for natural earthquakes and quarry blasts. Our results demonstrate that reliable classification can be achieved using raw waveforms without source information, which has potential to speed up real-time applications. We find that although Kentucky has distinct data characteristics from California, transfer learning of the California model to eastern Kentucky still outperforms the retrained model. This demonstrates the efficacy of transfer learning in generalizing deep learning across different regions.

5.1 Comparison with previous methods

Allmann et al. (2008) proposed to use the misfit of P wave spectra to an ω^{-2} source model as a discriminant between local earthquakes and blasts in southern California. They examined the events recorded by at least three stations with SNR >3 dB on three frequency bands, and reported a 90% classification accuracy. Comparatively, our California model achieves overall 91.97% accuracy on a much larger data set, with a looser data selection criterion and much simpler preprocessing. In eastern Kentucky, Miao et al. (2020) carefully designed multiwindow spectral features and trained a two-layer artificial neural network that achieved an accuracy of 97%. Here we use a much deeper model to automatically extract the features and achieve a comparable accuracy. Although the simpler model by Miao et al. (2020) indicates that the carefully handcrafted features may help reduce model complexity, our approach with transfer learning is potentially more generalizable to other regions. Because the model can be automatically fine-tuned and requires little expert knowledge. Our results show that transfer learning is a viable approach to obtain a state-of-the-art deep classification model for regions like Kentucky where data are relatively scarce.

Without handcrafted features, the deep learning models proposed by Linville et al. (2019) also automatically distinguish between mining blasts and tectonic earthquakes in Utah. Tibi et al. (2019) compared the models to amplitude-ratio-based methods and concluded that deep learning methods are generally more robust for low-SNR events. Our CNN models differ from Linville et al.'s in input data formats and data balancing. First, they use the spectrograms of three-component waveforms which are rotated to radial, transverse and vertical directions. In comparison, our model uses unfiltered raw waveforms as input. Also, we do not rotate the waveforms so that our model can work without knowing the event locations. Compared to the fixed P wave arrivals used in Linville's model, the random window shifts also make our model insensitive to P wave onsets. Second, the number of earthquakes and blasts are fairly balanced in the Utah dataset (L. Linville et al., 2019; Tibi et al., 2019), compared to the extreme class imbalance in southern California and eastern Kentucky. The data imbalance inevitably biases the model training and results in overall performance drop. The opposite sense of class imbalance in the two study regions could also have decreased the effectiveness of transfer learning, even though we mitigate the data imbalance by taking random shifted copies of the waveforms.

5.2 Practical considerations in deep learning strategies

We design our models to take input with minimal preprocessing for more generalizability and ease of use. For this consideration, conventional preprocessing operations, such as filtering, spectrogram calculation and component rotation, are all skipped. Possibly, the model performance could be further enhanced if some of preprocessing operations are applied to enhance signal quality. However, we argue that minimal human interference of the data not only reduces on-line runtime, but also offers flexibility to transfer the models to other regions. Our model can classify 76,848 samples within 88 seconds during on-line test, and the model file is as small as 1.8 Mb. The fast-processing speed and small file size enable easy integration of the models into real-time seismic monitoring systems.

Class imbalance is a major problem in the classification for both California and Kentucky, where different treatments can lead to different results. In the California case, training without augmentation produces a lower recall of 63.05% for the minority class. Besides data augmentation, we test two other strategies i.e., downsampling the majority class and cost-sensitive learning, to improve the performance. Compared with data augmentation, downsampling the majority class reduces both quantity and diversity of the training data and the accuracy drops by 5.99% (the recall drops by 6.32%/4.38% for earthquakes/blasts, respectively). As for cost-sensitive learning, we increase the weight of the loss term for the minority class i.e., misclassifying blasts as earthquakes is penalized more than misclassifying earthquakes as blasts. However, cost-sensitive learning yields only a recall of 82.81% for blasts, as compared to a recall of 93.18% for blasts with data augmentation. Thereby, data augmentation appears more effective to mitigate the imbalance problem, likely because various versions of time shifts increase the overall data diversity. Nonetheless, even with data augmentation, the smaller class (i.e., blasts in southern California, earthquakes in eastern Kentucky) inevitably tends to have a lower precision and F1-score.

Finally, the input window length also affects the performance. We have tested a window length of 30 s. Compared to the 50-s window, the 30-s window yields a better performance in southern California (92.59% accuracy) but a worse performance in eastern Kentucky (94.58% accuracy) than the window length of 50 s. This is likely associated with the longer source-receiver distance (mostly >100 km) and longer P/S separation and coda waves in Kentucky. A longer window at the face value can incorporate more coda waves. However, the occlusion tests suggest that it might not improve the performance, because a longer window tends to include more irrelevant noise, especially in southern California where the seismic networks are denser and epicentral distances are generally shorter.

5.3 Suggestions for further use

Although our deep learning models produce generally reliable results, there remains room for improvement given that some waveforms are misclassified. We could not rule out the possibility of false manual labels, as some waveforms are difficult to label even for experienced experts (Allmann et al., 2008). Figure 9 shows an earthquake and a blast misclassified on almost all stations, which are likely false labelled. Despite this, path and site effects can significantly modify the waveforms and results in potential misclassification. Averaging the probabilities of multi-station recordings can help mitigate the misclassification caused by the variations of path and site effects. Compared to single-station classification, network-averaged classification improves accuracy by 5.57% in the California case, 2.11% in the Kentucky case (Table 4). Similarly, majority voting, which simply counts the output classes of all stations, may improve performance at the network level (Liu et al., 2020). Besides, L. M. Linville (2022) proposed an event-based training strategy to promote consistency across different stations, which could be also useful. Finally, in real-time operation, there might be events neither natural earthquakes nor quarry blasts. For example, mining-induced earthquakes might be similar to both earthquakes and blasts as their source mechanisms can be double-couple, isotropic, or a combination of both (Koper et al., 2016). Including other classes of events to further reduce ambiguity remains a subject of future work.

As different areas seem to have different characteristics of blasts and earthquakes, when adopting our classification models to other areas, we strongly recommend using the transfer learning strategy. We have shown that transfer learning is the best solution among three strategies for the eastern Kentucky case. Although deep learning models are increasingly used in seismology and many efforts have been made to pursue a universal model that works best for all data (Ross et al., 2019; Zhu et al., 2020; Mousavi et al., 2020), we argue that rather than pursuing an optimal universal model, customizing models that work optimally for specific regions is probably a more viable pathway. With the help of transfer learning, regions with short seismic monitoring histories could have high-quality deep learning models comparable to those in well-instrumented regions at an economical price.

Data Availability Statement

Seismic waveforms and catalogs used in the California study are from the Southern California Earthquake Data Center (doi: 10.7909/C3WD3XH1). The temporary network EKMMP used in Kentucky is a part of the Kentucky Seismic and Strong Motion Network (doi: http://dx.doi.org/10.7914/SN/KY). The regional network N4 (doi: 10.7914/SN/N4), ET (https://www.fdsn.org/networks/detail/ET/) and NM (https://www.fdsn.org/networks/detail/NM/), USArray Transportable Array TA (doi: 10.7914/SN/TA) data are available at IRIS DMC (https://ds.iris.edu/mda/N4/, https://ds.iris.edu/mda/ET/, https://ds.iris.edu/mda/NM/, https://ds.iris.edu/mda/TA/).

Acknowledgements

We thank the Southern California Seismic Network and the Kentucky Seismic and Strong-Motion Network for providing seismic data. Jun Zhu thanks Shang-Lin Chen for email communication on how seismic events are classified by SCSN analysts. This research was supported by the National Key R&D Program of China (No. 2021YFC3000700) and the Special Fund of the Institute of Geophysics, China Earthquake Administration (No. DQJB21Z05).

References

Allen, R. V. (1978). Automatic earthquake recognition and timing from single traces. *Bulletin of the Seismological Society of America*, *68*(5), 1521–1532. https://doi.org/10.1785/BSSA0680051521

Allmann, B. P., Shearer, P. M., & Hauksson, E. (2008). Spectral Discrimination between Quarry Blasts and Earthquakes in Southern California. *Bulletin of the Seismological Society of America*, *98*(4), 2073–2079. https://doi.org/10.1785/0120070215

Astiz, L., Eakins, J. A., Martynov, V. G., Cox, T. A., Tytell, J., Reyes, J. C., et al. (2014). The Array Network Facility Seismic Bulletin: Products and an Unbiased View of United States Seismicity. *Seismological Research Letters*, *85*(3), 576–593. https://doi.org/10.1785/0220130141

Banchirigah, S. M. (2008). Challenges with eradicating illegal mining in Ghana: A perspective from the grassroots. *Resources Policy*, *33*(1), 29–38. https://doi.org/10.1016/j.resourpol.2007.11.001

Bennett, T. J., & Murphy, J. R. (1986). Analysis of seismic discrimination capabilities using regional data from western United States events. *Bulletin of the Seismological Society of America*, *76*(4), 1069–1086. https://doi.org/10.1785/BSSA0760041069

Carpenter, N. S., Holcomb, A. S., Woolery, E. W., Wang, Z., Hickman, J. B., & Roche, S. L. (2020). Natural Seismicity in and around the Rome Trough, Eastern Kentucky, from a Temporary

Seismic Network. *Seismological Research Letters*, *91*(3), 1831–1845. https://doi.org/10.1785/0220190015

Chai, C., Maceira, M., Santos-Villalobos, H. J., Venkatakrishnan, S. V., Schoenball, M., Zhu, W., et al. (2020). Using a Deep Neural Network and Transfer Learning to Bridge Scales for Seismic Phase Picking. *Geophysical Research Letters*, *47*(16). https://doi.org/10.1029/2020GL088651

Chen, C., & Holland, A. A. (2016). PhasePApy: A Robust Pure Python Package for Automatic Identification of Seismic Phases. *Seismological Research Letters*, *87*(6), 1384–1396. https://doi.org/10.1785/0220160019

Chollet, F. & others. (2015). Keras. Retrieved from https://keras.io

Díaz, J., Ruiz, M., Sánchez-Pastor, P. S., & Romero, P. (2017). Urban Seismology: on the origin of earth vibrations within a city. *Scientific Reports*, *7*(1), 15296. https://doi.org/10.1038/s41598-017-15499-y

Do, C. B., & Ng, A. Y. (2005). Transfer learning for text classification. In Y. Weiss, B. Schölkopf, & J. Platt (Eds.), *Advances in Neural Information Processing Systems* (Vol. 18). MIT Press. Retrieved from https://proceedings.neurips.cc/paper/2005/file/bf2fb7d1825a1df3ca308ad0bf48591e-Paper.pdf

Douglas, A. (2013). *Forensic Seismology and Nuclear Test Bans*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139524001

Fereidoni, A., & Atkinson, G. M. (2017). Discriminating Earthquakes from Quarry Blasts Based on ShakeMap Ground-Motion Parameters. *Bulletin of the Seismological Society of America*, ssabull;0120160308v1. https://doi.org/10.1785/0120160308

Gulia, L., & Gasperini, P. (2021). Contamination of Frequency–Magnitude Slope (b-Value) by Quarry Blasts: An Example for Italy. *Seismological Research Letters*. https://doi.org/10.1785/0220210080

Gulia, L., & Wiemer, S. (2019). Real-time discrimination of earthquake foreshocks and aftershocks. *Nature*, *574*(7777), 193–199. https://doi.org/10.1038/s41586-019-1606-4

Hadley, D., & Kanamori, H. (1977). Seismic structure of the Transverse Ranges, California. *Geological Society of America Bulletin*, *88*(10), 1469. https://doi.org/10.1130/0016-7606(1977)88<1469:SSOTTR>2.0.CO;2

Hartse, H. E., Taylor, S. R., Phillips, W. S., & Randall, G. E. (1997). A preliminary study of regional seismic discrimination in central Asia with emphasis on western China. *Bulletin of the Seismological Society of America*, *87*(3), 551–568. https://doi.org/10.1785/BSSA0870030551

Hutton, K., Woessner, J., & Hauksson, E. (2010). Earthquake Monitoring in Southern California for Seventy-Seven Years (1932-2008). *Bulletin of the Seismological Society of America*, *100*(2), 423–446. https://doi.org/10.1785/0120090130

Ismail Fawaz, H., Forestier, G., Weber, J., Idoumghar, L., & Muller, P.-A. (2018). Transfer learning for time series classification. In *2018 IEEE International Conference on Big Data (Big Data)* (pp. 1367–1376). Seattle, WA, USA: IEEE. https://doi.org/10.1109/BigData.2018.8621990

Kanamori, H., & Hadley, D. (1975). Crustal structure and temporal velocity change in southern California, *113*, 24.

Koper, K. D., Pechmann, J. C., Burlacu, R., Pankow, K. L., Stein, J., Hale, J. M., et al. (2016). Magnitude-based discrimination of man-made seismic events from naturally occurring earthquakes in Utah, USA. *Geophysical Research Letters*, *43*(20), 10,638-10,645. https://doi.org/10.1002/2016GL070742

Koper, K. D., Holt, M. M., Voyles, J. R., Burlacu, R., Pyle, M. L., Wang, R., & Schmandt, B. (2021). Discrimination of Small Earthquakes and Buried Single-Fired Chemical Explosions at Local Distances (<150 km) in the Western United States from Comparison of Local Magnitude (ML) and Coda Duration Magnitude (MC). *Bulletin of the Seismological Society of America*, *111*(1), 558–570. https://doi.org/10.1785/0120200188

Korrat, I. M., Lethy, A., ElGabry, M. N., Hussein, H. M., & Othman, A. S. (2022). Discrimination Between Small Earthquakes and Quarry Blasts in Egypt Using Spectral Source Characteristics. *Pure and Applied Geophysics*. https://doi.org/10.1007/s00024-022-02953-w

Kortström, J., Uski, M., & Tiira, T. (2016). Automatic classification of seismic events within a regional seismograph network. *Computers & Geosciences*, *87*, 22–30. https://doi.org/10.1016/j.cageo.2015.11.006

Larose, E., Carrière, S., Voisin, C., Bottelin, P., Baillet, L., Guéguen, P., et al. (2015). Environmental seismology: What can we learn on earth surface processes with ambient noise? *Journal of Applied Geophysics*, *116*, 62–74. https://doi.org/10.1016/j.jappgeo.2015.02.001

Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278–2324. https://doi.org/10.1109/5.726791

Li, Z., Meier, M.-A., Hauksson, E., Zhan, Z., & Andrews, J. (2018). Machine Learning Seismic Wave Discrimination: Application to Earthquake Early Warning. *Geophysical Research Letters*, *45*(10), 4773–4779. https://doi.org/10.1029/2018GL077870

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). https://doi.org/10.1029/2021JB022820

Liu, X., Ren, T., Chen, H., & Chen, Y. (2020). Classification of tectonic and non-tectonic seismicity based on convolutional neural network. *Geophysical Journal International*, *224*(1), 191–198. https://doi.org/10.1093/gji/ggaa444

Long, M., Wang, J., Ding, G., Sun, J., & Yu, P. S. (2013). Transfer Feature Learning with Joint Distribution Adaptation. In *2013 IEEE International Conference on Computer Vision* (pp. 2200–2207). Sydney, Australia: IEEE. https://doi.org/10.1109/ICCV.2013.274

Long, M., Cao, Y., Wang, J., & Jordan, M. I. (2015). Learning Transferable Features with Deep Adaptation Networks. *ArXiv:1502.02791 [Cs]*. Retrieved from http://arxiv.org/abs/1502.02791

Miao, F., Carpenter, N. S., Wang, Z., Holcomb, A. S., & Woolery, E. W. (2020). High-Accuracy Discrimination of Blasts and Earthquakes Using Neural Networks With Multiwindow Spectral Data. *Seismological Research Letters*, *91*(3), 1646–1659. https://doi.org/10.1785/0220190084

Pan, S. J., & Yang, Q. (2010). A Survey on Transfer Learning. *IEEE Transactions on Knowledge* *and Data Engineering*, *22*(10), 1345–1359. https://doi.org/10.1109/TKDE.2009.191

Pan, S. J., Tsang, I. W., Kwok, J. T., & Yang, Q. (2011). Domain Adaptation via Transfer Component Analysis. *IEEE Transactions on Neural Networks*, *22*(2), 199–210. https://doi.org/10.1109/TNN.2010.2091281

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

Renouard, A., Maggi, A., Grunberg, M., Doubre, C., & Hibert, C. (2021). Toward False Event Detection and Quarry Blast versus Earthquake Discrimination in an Operational Setting Using Semiautomated Machine Learning. *Seismological Research Letters*. https://doi.org/10.1785/0220200305

Rodriguez Asihama, P. (2016). Development of a Semi-Automated Methodology for Discriminating Between Natural and Manmade Seismic Events Using the OIINK Seismic Array [PDF]. https://doi.org/10.13023/ETD.2016.526

Ross, Z. E., Meier, M., Hauksson, E., & Heaton, T. H. (2018). Generalized Seismic Phase Detection with Deep Learning. *Bulletin of the Seismological Society of America*, *108*(5A), 2894–2901. https://doi.org/10.1785/0120180080

Ross, Z. E., Meier, M.-A., & Hauksson, E. (2018). P Wave Arrival Picking and First-Motion Polarity Determination With Deep Learning. *Journal of Geophysical Research: Solid Earth*, *123*(6), 5120–5129. https://doi.org/10.1029/2017JB015251

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2020). Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. *International Journal of Computer Vision*, *128*(2), 336–359. https://doi.org/10.1007/s11263-019-01228-

7

Stump, B. W., Hedlin, M. A. H., Pearson, D. C., & Hsu, V. (2002). CHARACTERIZATION OF MINING EXPLOSIONS AT REGIONAL DISTANCES: IMPLICATIONS WITH THE INTERNATIONAL MONITORING SYSTEM: REGIONAL SEISMOGRAMS FROM MINES. *Reviews of Geophysics*, *40*(4), 2-1-2–21. https://doi.org/10.1029/1998RG000048

Su, F., Aki, K., & Biswas, N. N. (1991). Discriminating quarry blasts from earthquakes using coda waves. *Bulletin of the Seismological Society of America*, *81*(1), 162–178. https://doi.org/10.1785/BSSA0810010162

Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C., & Liu, C. (2018). A Survey on Deep Transfer Learning. In V. Kůrková, Y. Manolopoulos, B. Hammer, L. Iliadis, & I. Maglogiannis (Eds.), *Artificial Neural Networks and Machine Learning – ICANN 2018* (Vol. 11141, pp. 270–279). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-01424-7_27

Tang, L., Zhang, M., & Wen, L. (2020). Support Vector Machine Classification of Seismic Events in the Tianshan Orogenic Belt. *Journal of Geophysical Research: Solid Earth*, *125*(1). https://doi.org/10.1029/2019JB018132

Tibi, R., Koper, K. D., Pankow, K. L., & Young, C. J. (2018). Depth Discrimination Using Rg-to-Sg Spectral Amplitude Ratios for Seismic Events in Utah Recorded at Local Distances. *Bulletin of the Seismological Society of America*, *108*(3A), 1355–1368. https://doi.org/10.1785/0120170257

Tibi, R., Linville, L., Young, C., & Brogan, R. (2019). Classification of Local Seismic Events in the Utah Region: A Comparison of Amplitude Ratio Methods with a Spectrogram-Based

Machine Learning Approach. *Bulletin of the Seismological Society of America*, *109*(6), 2532–2544. https://doi.org/10.1785/0120190150

Wang, R., Schmandt, B., & Kiser, E. (2020). Seismic Discrimination of Controlled Explosions and Earthquakes Near Mount St. Helens Using *P* / *S* Ratios. *Journal of Geophysical Research: Solid Earth*, *125*(10). https://doi.org/10.1029/2020JB020338

Wang, R., Schmandt, B., Holt, M., & Koper, K. (2021). Advancing Local Distance Discrimination of Explosions and Earthquakes with Joint P/S and M_L-M_C Classification. *Geophysical Research Letters*. https://doi.org/10.1029/2021GL095721

Wiemer, S. (2000). Mapping and Removing Quarry Blast Events from Seismicity Catalogs. *Bulletin of the Seismological Society of America*, *90*(2), 525–530. https://doi.org/10.1785/0119990104

Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). How transferable are features in deep neural networks? *ArXiv:1411.1792 [Cs]*. Retrieved from http://arxiv.org/abs/1411.1792

Zeiler, M. D., & Fergus, R. (2014). Visualizing and Understanding Convolutional Networks. In D. Fleet, T. Pajdla, B. Schiele, & T. Tuytelaars (Eds.), *Computer Vision – ECCV 2014* (Vol. 8689, pp. 818–833). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-10590-1_53

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 Mw7.9 Wenchuan Earthquake. *Physics of the Earth and Planetary Interiors*, *293*, 106261. https://doi.org/10.1016/j.pepi.2019.05.004

Zhu, W., & Beroza, G. C. (2018). PhaseNet: A Deep-Neural-Network-Based Seismic Arrival Time

Picking Method. *Geophysical Journal International*. https://doi.org/10.1093/gji/ggy423

Zhu, Y., Chen, Y., Lu, Z., Pan, S. J., Xue, G.-R., Yu, Y., & Yang, Q. (2011). Heterogeneous Transfer Learning for Image Classification. In *Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence* (pp. 1304–1309). San Francisco, California: AAAI Press.

Tables

Table 1. Southern California data division

	Recordings	Events	Time span	Split ratio
Quake	223,701	12,941	Jan 2011 - Dec 2017	70%
Blast	45,257	2,542	Jan 2011 - Dec 2015	(training)
Quake	31,956	1,481	Dec 2017 - July 2019	10%
Blast	6,469	219	Dec 2015 - May 2016	(validation)
Quake	63,919	2,914	July 2019 - Dec 2020	20%
Blast	12,929	910	May 2016 - Dec 2020	(test)

Table 2. Performance of the California model

	Station		Network	
	Predicted EQ	Predicted QB	Predicted EQ	Predicted QB
True EQ	58,632	5,287	2,842	72
True QB	882	12,047	22	888
Recall	91.73%	93.18%	97.53%	97.58%
Precision	98.52%	69.50%	99.23%	92.50%
F1-score	95.00%	79.62%	98.37%	94.97%
Accuracy	91.97%		97.54%	

Note. EQ, natural earthquakes are defined as "negative"; QB, quarry blasts;

Recall of earthquakes: TN/(TN+FP); Recall of blasts: TP/(TP+FN);

Precision of earthquakes: TN/(TN+FN); Precision of blasts: TP/(TP+FP);

F1-score: $2 \times Precison \times Recall/(Precision + Recall)$;

Accuracy: (TP+TN)/(TP+FP+TN+FN)

Table 3. Performance comparison of three models in the eastern Kentucky dataset

Model	Original California		Retrained		Transfer-learned	
Accuracy	45.38%		94.36%		97.35%	
	Predicted EQ	Predicted QB	Predicted EQ	Predicted QB	Predicted EQ	Predicted QB
True EQ	220	24	222	22	226	18
True QB	3,154	2,420	306	5,268	136	5,438
Recall	90.16%	43.42%	90.98%	94.51%	92.62%	97.56%
Precision	6.52%	99.02%	42.05%	99.58%	62.43%	99.67%
F1-score	12.16%	60.36%	57.51%	96.98%	74.59%	98.60%

Table 4. Performance of the transfer-learned model in eastern Kentucky

	Station		Network	
	Predicted EQ	Predicted QB	Predicted EQ	Predicted QB
True EQ	226	18	27	2
True QB	136	5,438	2	711
Recall	92.62%	97.56%	93.10%	99.72%
Precision	62.43%	99.67%	93.10%	99.72%
F1-score	74.59%	98.60%	93.10%	99.72%
Accuracy	97.35%		99.46%	

Figures

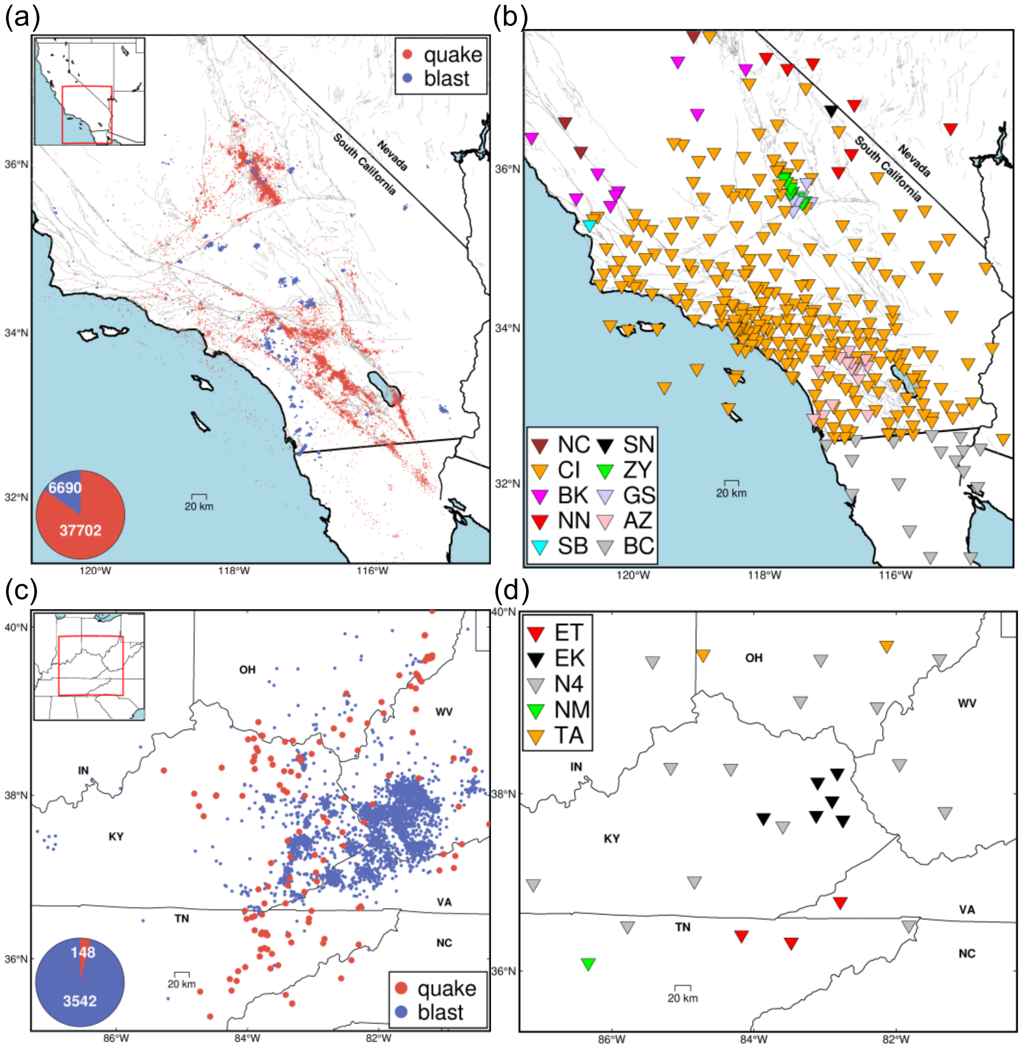

Figure 1. Distribution of local earthquakes, quarry blasts and seismic stations. (a) Earthquakes (red dots) and blasts (blue dots) in southern California. The bottom left inset shows the statistics of 6,690 quarry blasts and 37,702 natural earthquakes. The top left inset marks the study area. (b) Seismic stations used in this study, colored by different networks. (c) and (d): symbols are the same as (a) and (b) respectively but for eastern Kentucky. Note that EKMMP, abbreviated as EK in (d), is a part of the temporary Eastern Kentucky Microseismic Monitoring Network.

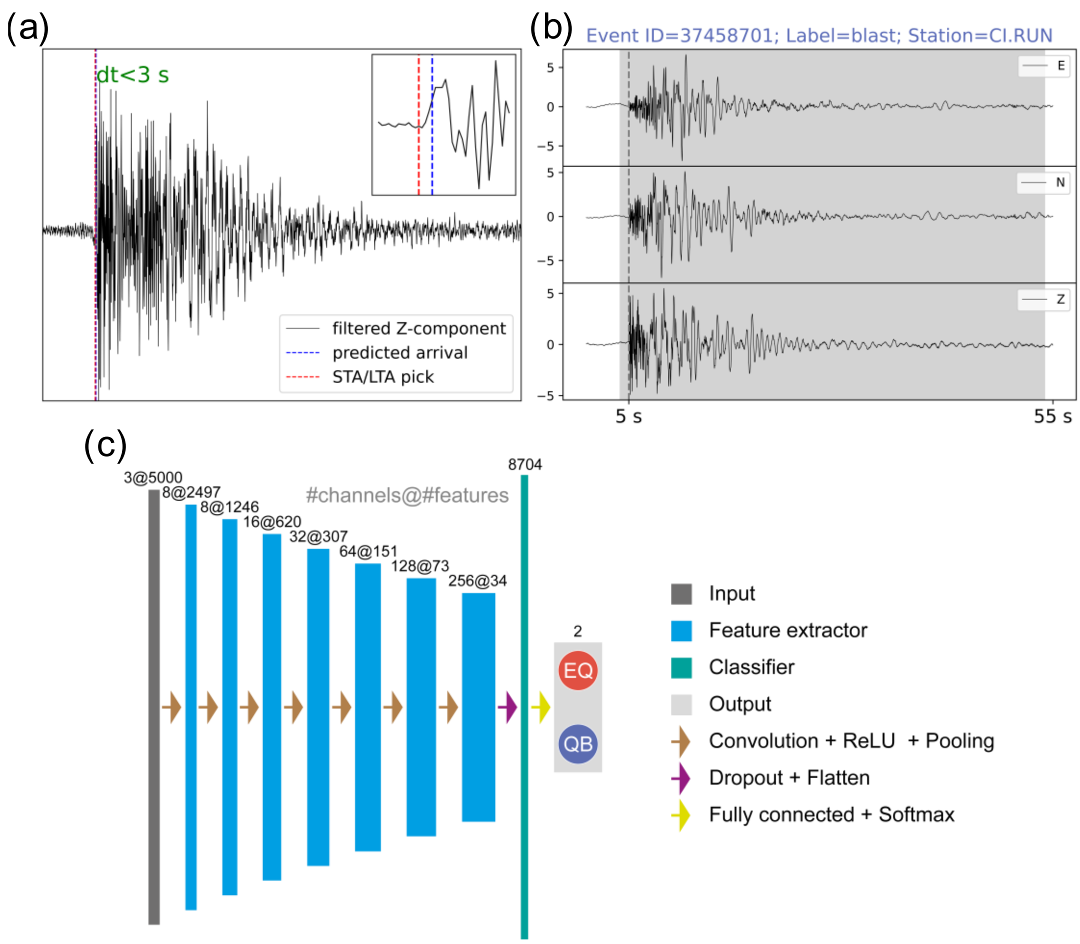

Figure 2. Preprocessing workflow and the convolutional neural network (CNN) architecture. (a) A phase picking example for a 55 s quarry blast waveform. The blue dashed line marks the predicted arrival time; the red dashed line marks the STA/LTA pick. The top right inset shows the zoom-in plot around the P arrival. (b) A window of 50 s waveforms is used for model input (shaded area). Waveforms are augmented via a sliding window whose onset is random within 5 s before the arrival. (c) The CNN model consists of seven convolutional layers (each including three basic operations i.e., convolution, the ReLU activation function and max pooling), one dropout and one fully connected layer. The dimensions of the model input and each layer's output are annotated on top of the layers.

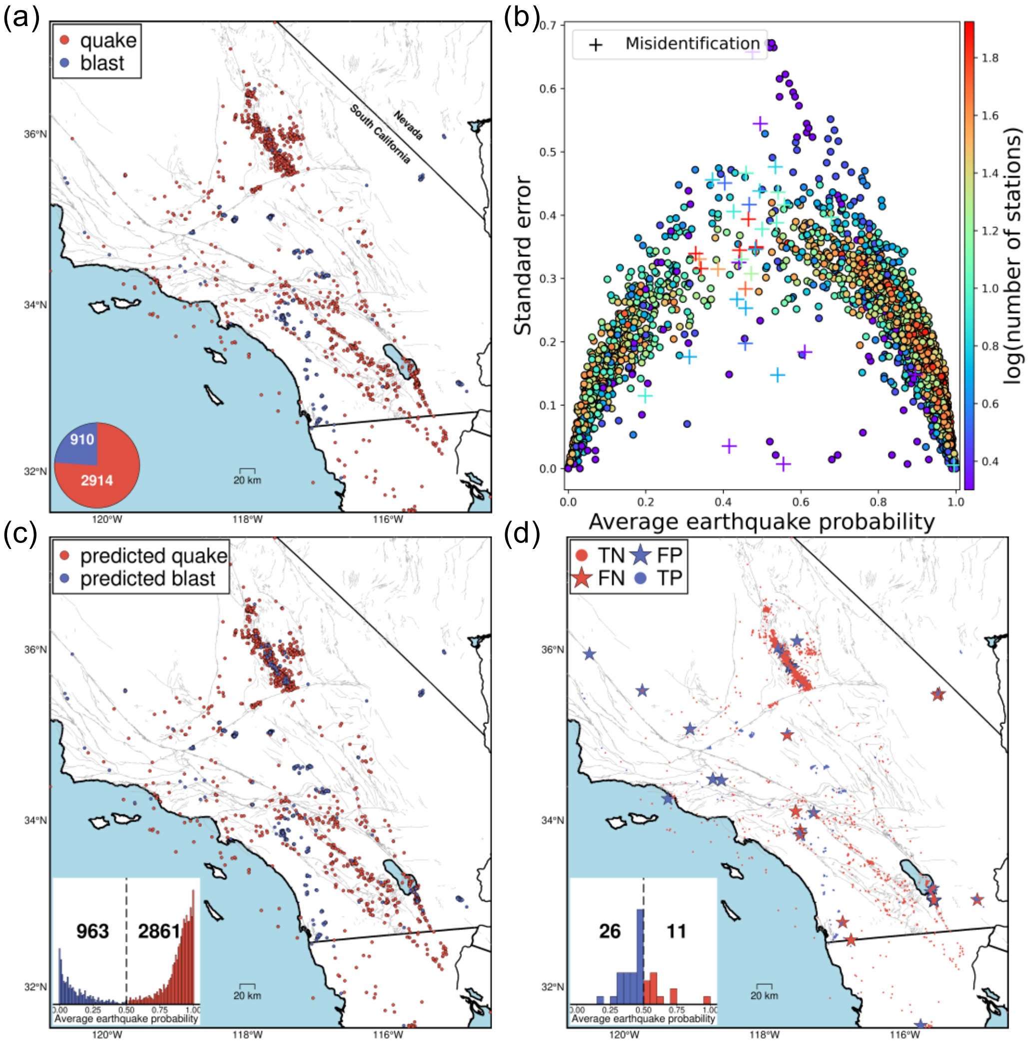

Figure 3. Classification results of the CNN California model on the test set in southern California. (a) Distribution of the test set including 910 blasts and 2,914 earthquakes. (b) Consistency across multiple stations. Each circle (correctly classified) or cross (misclassified) represents the standard deviation versus the average of output earthquake probability on multiple stations for a given event. Symbols are colored by the log-scaled number of available stations for that event. (c) Network-averaged predictions by the California model. The inset shows the histogram of predicted earthquake probabilities. (d) Symbols are similar to (c), but only for events misclassified by at least two stations. Blue/red stars mark the FPs/FNs, respectively. Blue/red circles mark the TPs/TNs, respectively. The inset is the histogram of network-averaged earthquake probabilities for misclassified events.

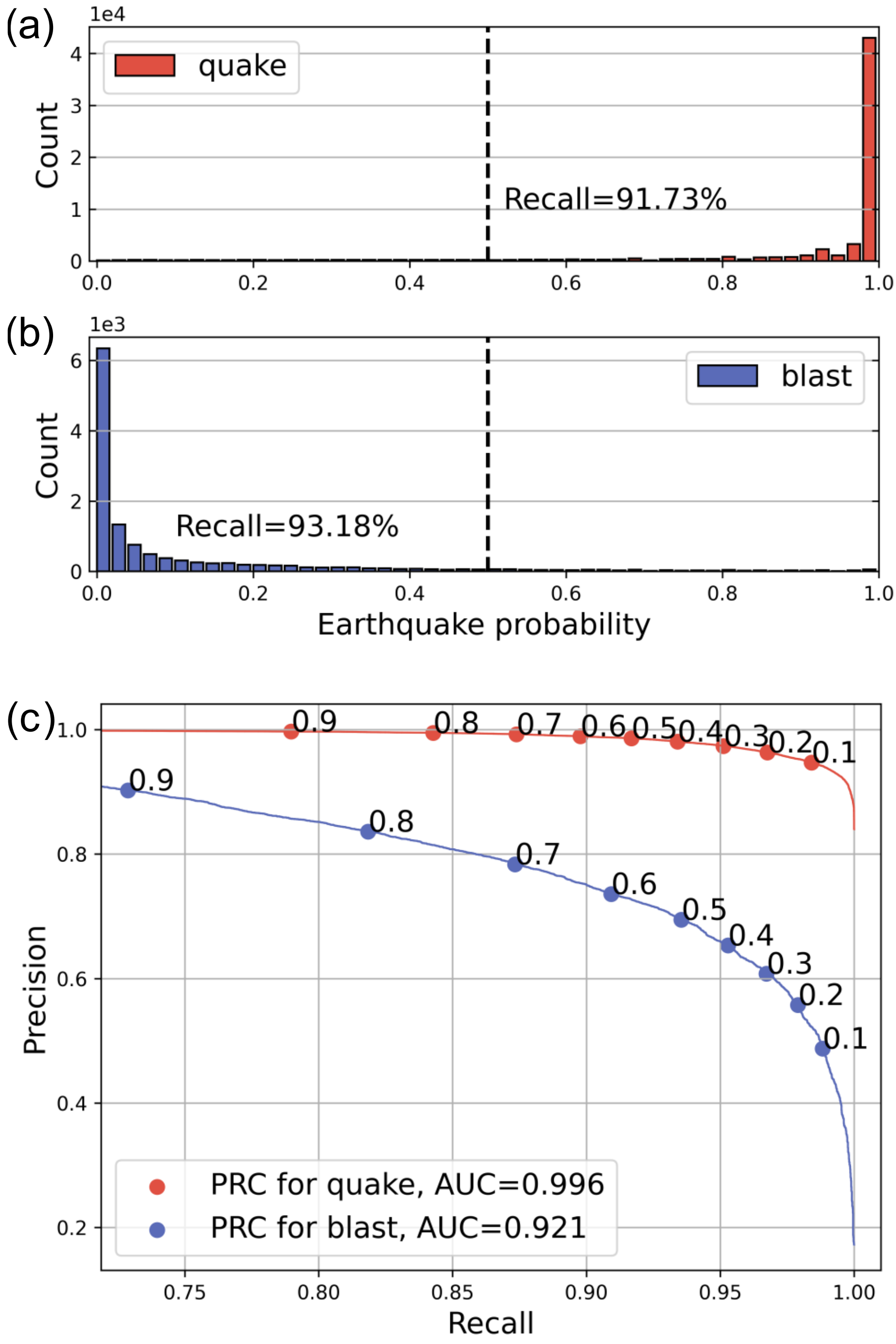

Figure 4. Performance of the California model on the test set in southern California. (a) Histogram of predicted earthquake probability for earthquakes. The black dashed line marks the threshold of 0.5. (b) Histogram of predicted earthquake probabilities for blasts. (c) Precision-recall curves for earthquakes (red) and blasts (blue).

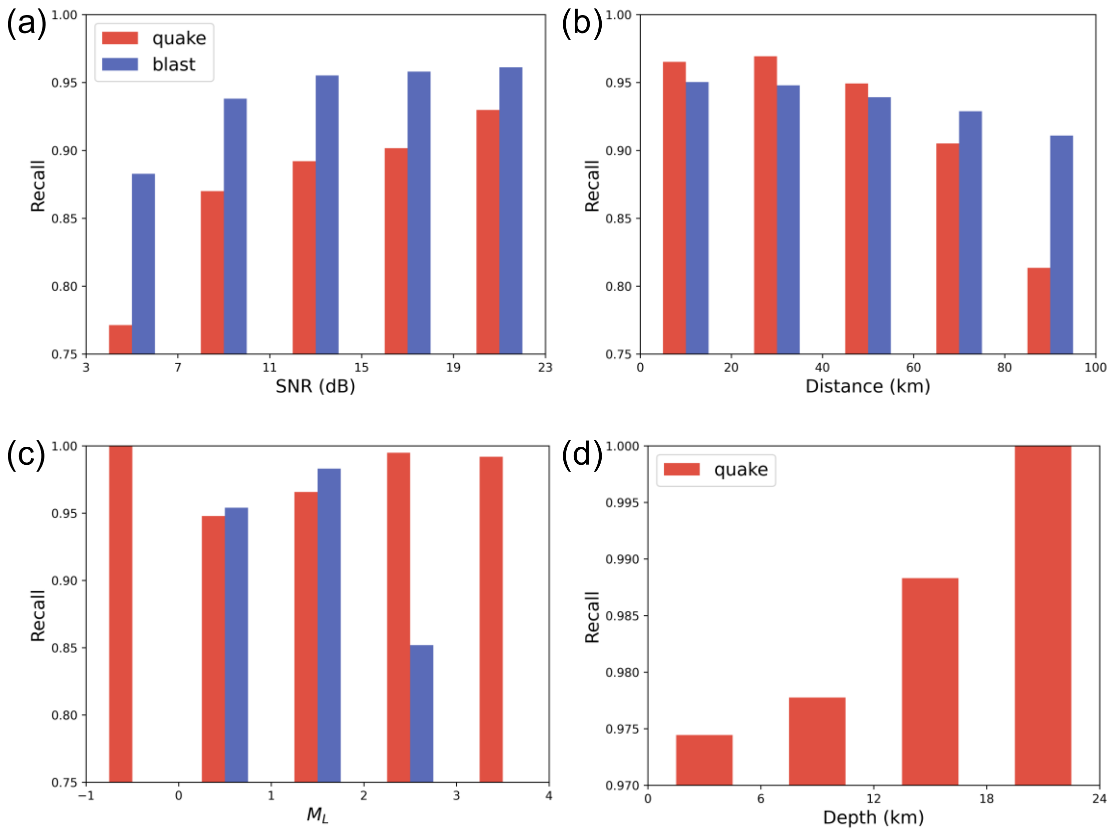

Figure 5. Variations of model recall on SNR (a), epicentral distance (b), magnitude (c) and earthquake depth (d).

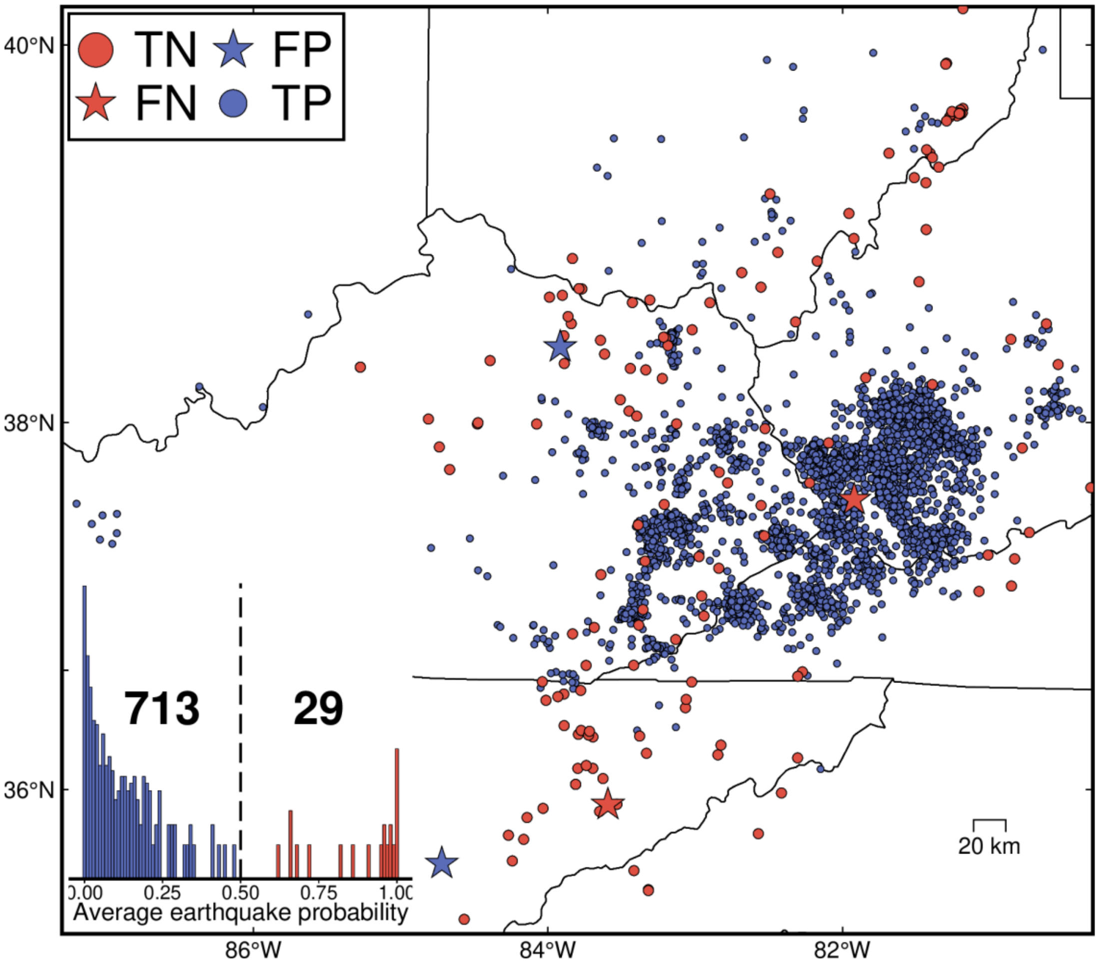

Figure 6. Classification results of the transferred CNN model for the test set in eastern Kentucky: TNs (red circles), FPs (blue stars), FNs (red stars) and TPs (blue circles). The inset is the histogram of network-averaged earthquake probabilities. Both the number of predicted blasts and earthquakes are annotated.

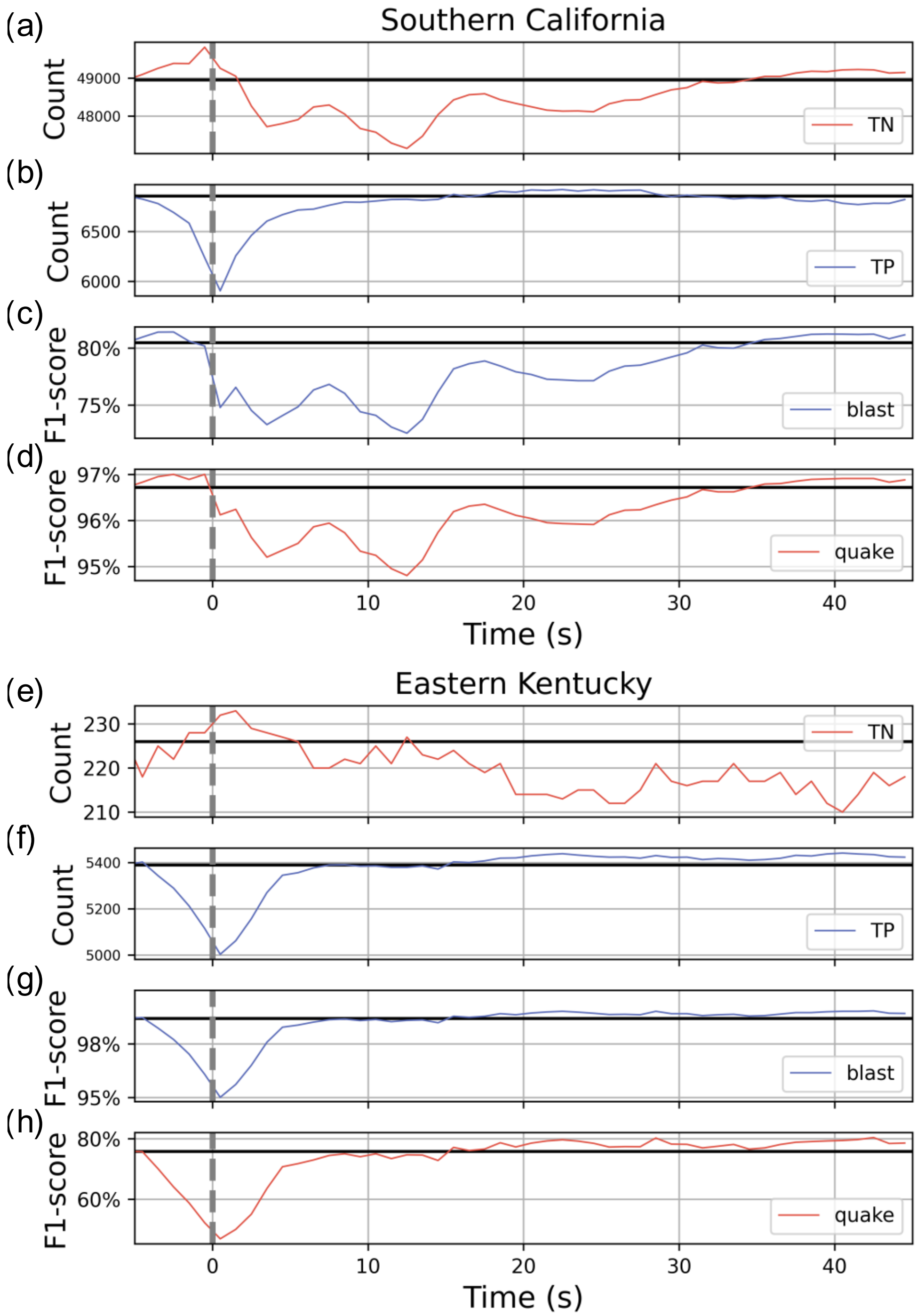

Figure 7. Occlusion tests for model performance dependence with masked waveform sections (nonoverlapping three-component signals of 1 s) in southern California and eastern Kentucky. From top to bottom are (a) the number of TNs and (b) TPs, (c) the F1-score of blasts and (d) earthquakes for the California model. The horizontal black line marks the baseline performance without occlusion. (e)-(h) Same as (a)-(d), but for the transfer-learned model in eastern Kentucky.

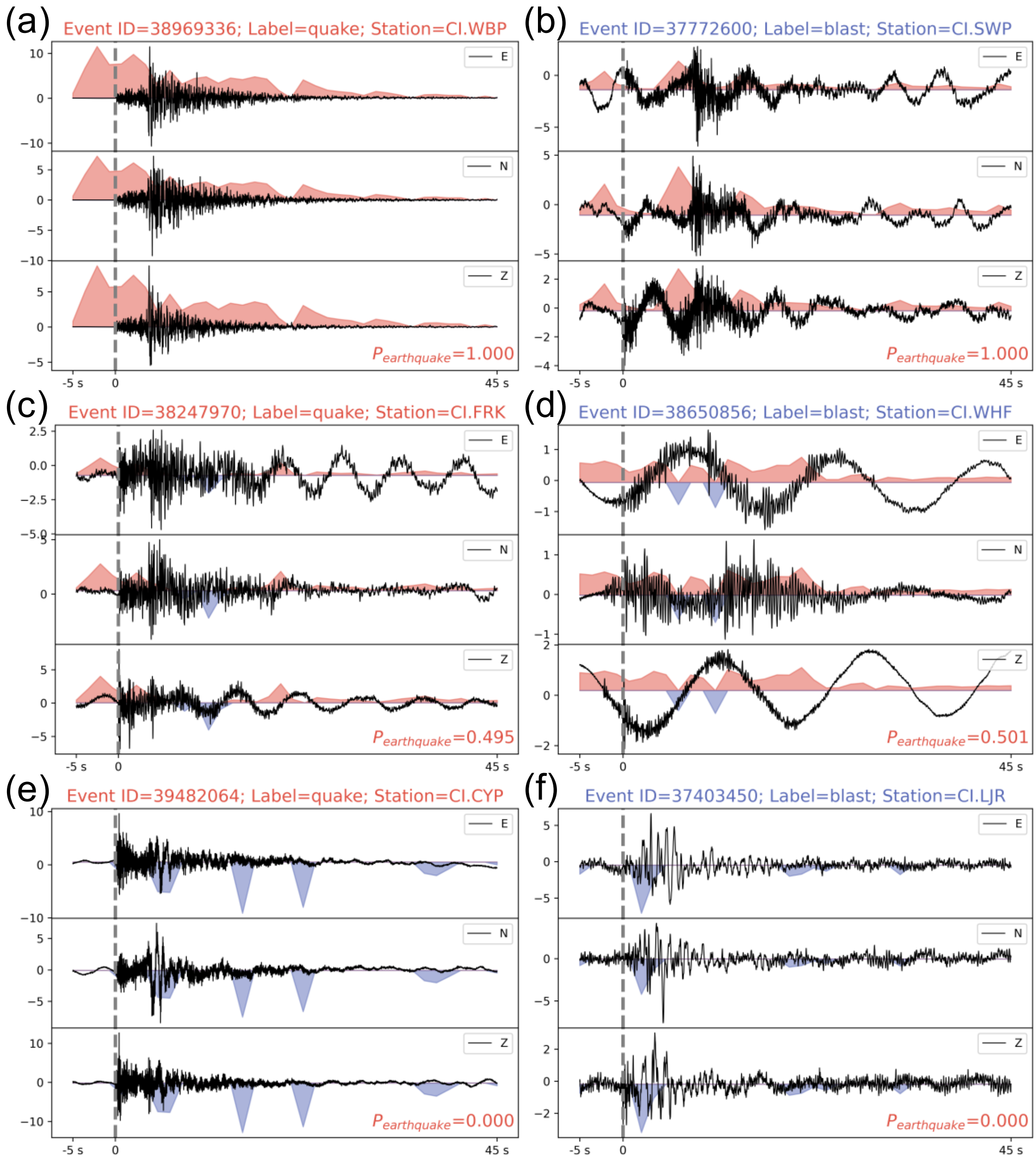

Figure 8. Gradient-weighted class activation mapping (Grad-CAM) for example earthquakes (left) and blasts (right) in the southern California test set. From top to bottom are from most earthquake-like (a)-(b) to most blast-like (e)-(f), with earthquake probabilities annotated in the Z-component of each example. The red/blue shade is the heatmap representing the relative contribution of the waveforms towards an earthquake/blast prediction, respectively.

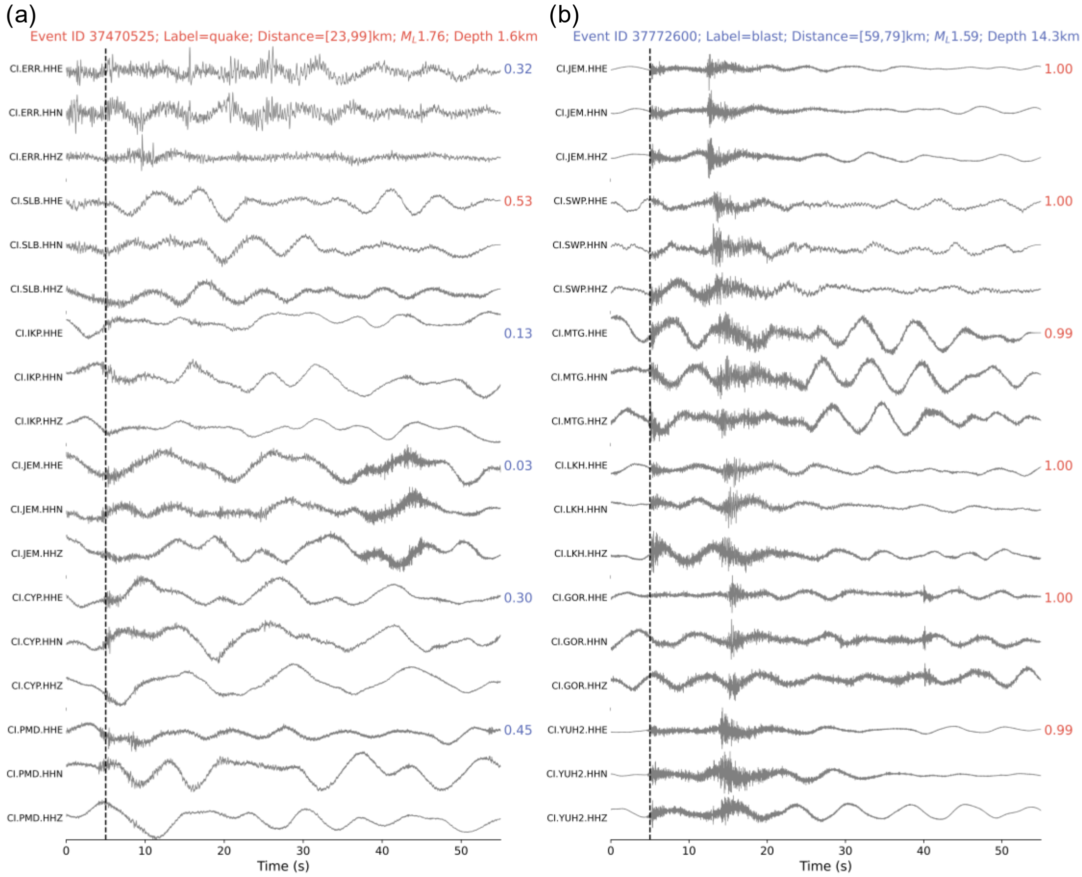

Figure 9. Examples of misclassified events. (a) An earthquake (SCEDC event ID 37470525) classified as a blast by almost all stations except CI.SLB. Earthquake probabilities are annotated on the right. (b) A quarry blast (SCEDC event ID 37772600) classified as an earthquake by all stations. Notice that the reported depth is 14.3 km, indicating that it is likely an earthquake but falsely labelled.