Application of Convolutional Neural Network for Seismic Event Classification: Impact of Dataset Quality, Distribution, and Human-in-the-Loop Feedback

Yanwei Zhang[1], Stephen S. Gao[1], Xiaoyi Fan[2], Shaohui Zhou[3], and Junhao Qu[*1,3]

ABSTRACT

Convolutional neural networks (CNNs) have gained popularity in geophysical research due to their exceptional performance in various areas. However, achieving reliable results with CNNs typically requires a significant amount of high-quality data for training. In this study, we develop a CNN to classify natural earthquakes, mine collapses, and explosions using seismic waveforms from 287 stations in Shandong Province, China. The dataset comprises 1035 earthquakes, 159 mine collapses, and 586 explosions. To address the impact of unreliable measurements, we employ cross validation to screen, manually correct, or discard measurements with inconsistent labels assigned by human experts and CNN. By refining the dataset through these methods, classification accuracies for the three event types improved substantially, reaching over 95%. Notably, CNN outperforms human classification in this task, with its performance heavily influenced by the quality and distribution of the training dataset. Our research demonstrates the potential of CNNs for classifying seismic events while emphasizing the crucial role of human-in-the-loop feedback and the significance of cross-validation techniques in ensuring the accuracy and reliability of the CNN model.

KEY POINTS

- Convolutional neural network (CNN) is powerful for classification of natural earthquakes, mine collapses, and explosions.
- Dataset quality and distribution strongly impact CNN model accuracy and reliability.
- Cross validation and human-in-the-loop techniques enhance accuracy and reliability of training dataset.

INTRODUCTION

The ever-increasing demand for mineral products in contemporary society has led to a marked rise in nontectonic seismic events, such as mine collapses and explosions. Consequently, the accurate classification of natural earthquakes, collapses, and explosions has emerged as a formidable challenge in the field of applied seismological research (Astiz *et al.*, 2014). The task of distinguishing between these event types is particularly challenging, as their waveforms can exhibit significant overlap, making it difficult to differentiate between them using traditional signal processing techniques. Furthermore, the classification process requires a high level of accuracy to help government departments respond differently to different events.

Over the years, a range of different automated and semiautomated techniques have been developed to classify seismic sources based on the characteristics of their waveforms. Fäh and Koch (2002) compared the ratios between *P* and *S* phases with different time windows and frequency bands to differentiate between earthquakes and chemical explosions. In addition, the *Lg* and *Rg* phases have been analyzed to classify various source types (Baumgardt and Young, 1990; O'Rourke and Baker, 2017). These studies have demonstrated that seismic waveforms can contain significant information about their sources, but the effectiveness of waveform-based methods is strongly influenced by a variety of factors, including the depth and distance of the event, the environment in which it occurred, and other potential sources of noise. Given the lack of clear and

1. Department of Earth Sciences and Engineering, Missouri University of Science and Technology, Rolla, Missouri, U.S.A., https://orcid.org/0000-0003-1809-0174 (YZ); https://orcid.org/0000-0001-7530-7128 (SSG); https://orcid.org/0000-0002-9108-5112 (JQ); 2. Jiangsu Earthquake Agency, Nanjing, Jiangsu, China; 3. Shandong Earthquake Agency, Jinan, Shandong, China

*Corresponding author: gisqjh@126.com

Cite this article as Zhang, Y., S. S. Gao, X. Fan, S. Zhou, and J. Qu (2024). Application of Convolutional Neural Network for Seismic Event Classification: Impact of Dataset Quality, Distribution, and Human-in-the-Loop Feedback, *Bull. Seismol. Soc. Am.* **115**, 106–114, doi: 10.1785/0120240179

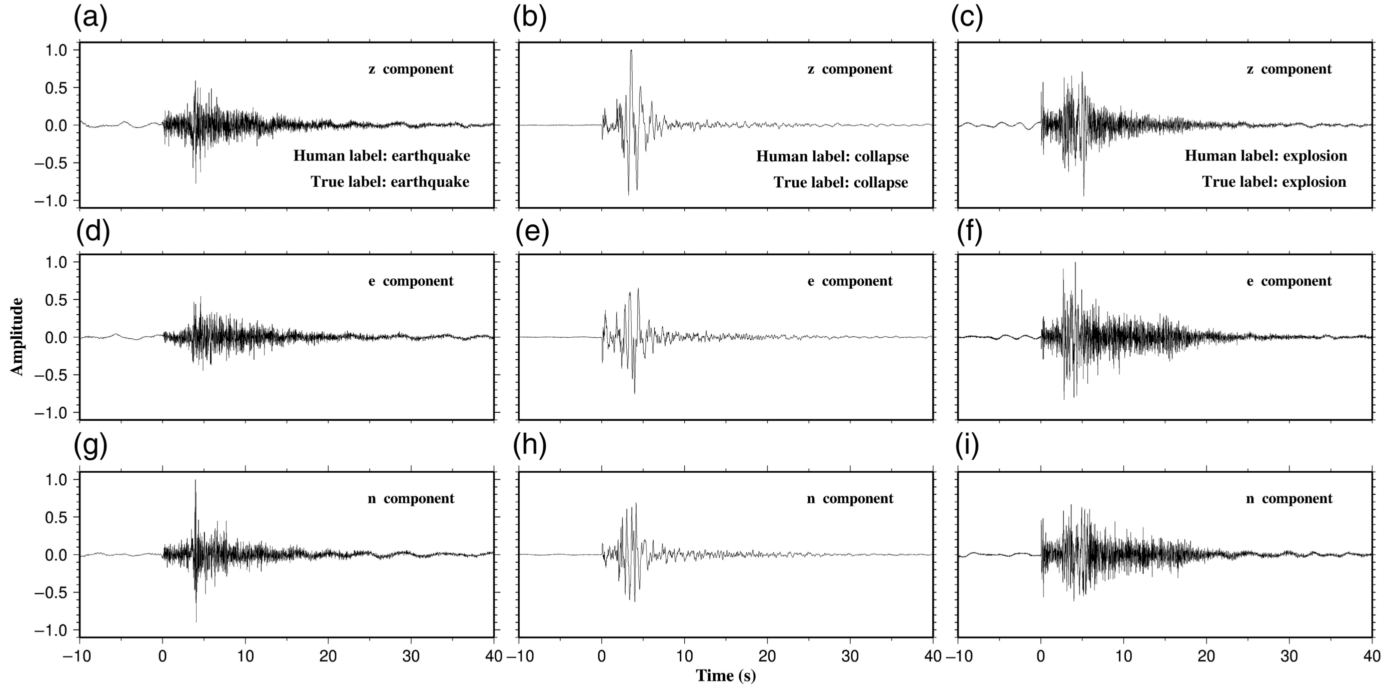

Figure 1. Three types of seismic events are demonstrated by the following examples: (a,d,g) three-component traces of an earthquake recorded at station PENL during event 20170922055. (b,e,h) A collapse captured at station ZCH during event 201710272144. (c,f,i) An explosion recorded at station LZH during event 201805031120.

consistent characteristics that can differentiate between the waveforms produced by earthquakes, mine collapses, and explosions, the classification of these types of events has traditionally been performed by experienced human analysts who can use their knowledge and expertise to identify key features and make accurate determinations about the source of each seismic signal.

Based on previous studies and empirical experience, seismologists commonly categorize seismic events such as earthquakes, explosions, and collapses using several key parameters. These include dominant frequency (normally lower frequency for collapses), the ratio of *P*-wave to *S*-wave energy (normally greater *P*-wave energy for explosions), the duration of the event (normally longer duration for earthquakes), the polarization direction of the initial arrival (depending on event type and back azimuth), and characteristics of surface waves (normally more clear surface waves for earthquakes and explosions) (Fig. 1) (Baumgardt and Young, 1990; Kim *et al.*, 1997; Fäh and Koch, 2002; O'Rourke and Baker, 2017). However, it is crucial to recognize that in different geological settings, some events may display overlapping features, which can make classification more challenging.

In recent years, the use of high-performance machine learning (ML) techniques in geophysics has gained significant attention from researchers. Among the ML techniques, the convolutional neural network (CNN) has emerged as a widely used tool in solving classification problems based on seismic waveforms, owing to its remarkable feature extraction ability. The potential of CNNs in seismic event classification has been demonstrated in numerous studies, including those by Perol *et al.* (2018), Linville *et al.* (2019), Zhang and Gao (2022), and Barama *et al.* (2023). However, it is essential to have a large quantity of high-quality data for training the CNNs to obtain reliable results. Some earlier studies have pointed out that signals with mislabeled and ambiguous measurements are often present in human-labeled datasets (Zhu and Beroza, 2018; Garcia *et al.*, 2021; Zhang and Gao, 2022). These contaminated datasets can negatively impact the results of CNN-based classifications. Linville *et al.* (2019) reported that over 50% of events with different labels between CNNs and humans were mislabeled by human analysts. This situation could potentially be attributed to various factors, including typing errors and/or the limited expertise of a beginning operator. As a matter of fact, it is commonly recognized in the ML community that the quality of the utilized dataset is as critical as the model (Chen *et al.*, 2021; Gong *et al.*, 2023). The inclusion of low-quality measurements in the training dataset can compromise the training process and adversely affect the performance of CNNs. In addition, previous studies revealed that the performance of ML methods exhibited minimal discrepancies across different methods and models in seismic classification projects (Linville *et al.*, 2019; Wang *et al.*, 2023). Consequently, we aim to address the issue at the dataset level.

In this study, we have aimed to develop a more accurate and reliable CNN model for the classification of three types of seismic events: tectonic earthquakes, mine collapses, and explosions. To achieve this, we acquired a human-labeled dataset from the Shangdong Seismic Network Center (SSNC) in

China. However, we have found that such datasets may contain unreliable or ambiguous measurements, which can negatively impact the performance of CNN. Therefore, we have employed a tenfold cross-validation method to scan the entire dataset and minimize the influence of these measurements. The measurements with different labels, provided by both humans and CNN, have been carefully evaluated, and a decision has been made to either keep, correct, or abandon them in the dataset. The resulting CNN model with eleven layers has shown significant improvements in classification accuracies for all three types of events. Specifically, the classification accuracy increased from 97.3% to 99.2% for earthquakes, from 84.9% to 95.8% for mine collapses, and from 93.7% to 98.1% for explosions. Our study emphasizes the importance of reliable datasets and demonstrates that cross validation can effectively evaluate and improve the performance of ML methods.

DATA AND METHODS

The seismic waveform dataset used in this study is provided by the SSNC and covers a period of over four years, from August 2017 to January 2022. The dataset comprises a total of 31,754 three-component waveforms that were recorded by 287 seismic stations including 4,410 events, as illustrated in Figure 2. To categorize each waveform correctly, human experts manually picked the *P*-wave arrival time and assigned it to one of the three categories: earthquake, collapse, or explosion. The earthquake frequency–magnitude distribution is estimated by applying the magnitude of completeness (M_c) determination method developed by Cao and Gao (2002). The dataset from the SSNC has an M_c-value of 1.5. This value suggests that the SSNC is capable of detecting all seismic events with magnitudes equal to or exceeding 1.5. Thus, we have filtered out all events with magnitudes lower than 1.5 to ensure the completeness and reliability of our dataset. After filtering, 17,557 seismic velocity waveforms are retained for further analysis, including 11,901 waveforms from 1,035 earthquakes, 1,511 waveforms from 159 collapses, and 4,145 waveforms from 586 explosions. To standardize the data, detrending and normalizing are performed, and the time window of 10 s before and 40 s following each *P*-wave arrival time is used. In addition, the sampling interval is set to 0.01 s, and zeros are appended to waveforms that do not have sufficient length. Because CNNs cannot be fully understood, we chose not to apply filters to the waveforms, preserving all the original information to allow CNN to learn more features for this task.

In our dataset, there is a significant imbalance in the numbers of waveforms among the different events, which could lead to bias toward the majority class during the training process (Goodfellow *et al.*, 2016). To address this issue, we employ a common approach of setting different class weights for different event types (He and Garcia, 2009; Zhang and Gao, 2022). Specifically, we set the class weights as 1 for earthquakes, 8 for collapses, and 2.5 for explosions based on the ratio between numbers of waveforms corresponding to each event type. One-hot encoding is used to label each type of event, where each event is represented by an array with three elements, corresponding to earthquakes, collapses, and explosions, respectively (Goodfellow *et al.*, 2016). The earthquakes are labeled as [1, 0, 0], the collapses are labeled as [0, 1, 0], and the explosions are labeled as [0, 0, 1].

Based on practices from previous studies, we established a CNN with 10 convolutional layers and 2 fully connected layers using Keras (Gulli and Pal, 2017) (Fig. 3). The inputs are 5000-node three-component seismic waveforms, and after convolution, the outputs are the probabilities of the three types of events. The activation function between each layer is LeakyReLU with a 0.05 negative slope (Maas *et al.*, 2013). During the convolutional process, 3-by-1 filters are applied to extract features from the seismic waveforms. The stride is set to 2, so the length of each layer is down-sampled to half of the previous length. Softmax is used as the activation function to classify each seismic waveform in the output layer (Goodfellow *et al.*, 2016):

$$p(x)_i = \frac{e^{x_i}}{\sum_{j=1}^{3} e^{x_j}}, \tag{1}$$

in which x represents each node in the final layer of CNN, $j = 1, 2, 3$ and $i = 1, 2, 3$ represent the three nodes of the last layer, and $p(x)_1$, $p(x)_2$, and $p(x)_3$ represent the probabilities of earthquakes, collapses, and explosions, respectively. The result of the CNN follows from the highest value of $p(x)$.

The cross entropy between the human-labeled and CNN-predicted results is defined as the loss function of our CNN (Goodfellow *et al.*, 2016):

$$L = -\sum_{i=1}^{n} p_i \log(q_i), \tag{2}$$

in which n is 3, representing the three types of events (earthquakes, collapses, and explosions), p is the probability of the CNN-predicted result given by the softmax function, and q is the human-labeled result. The weights of the filters and fully connected layer are automatically optimized by minimizing the loss function during the training process. The optimizer is Adam with a learning rate of 0.001 (Kingma and Ba, 2014), and the number of epochs is 32.

RESULTS

Results with original dataset

Initially, we conduct tenfold cross validation to explore the entire dataset. The complete dataset is uniformly divided into ten groups, where each group is used as the testing dataset in each of the ten iterations. The remaining nine groups serve as the training dataset in each iteration. In each iteration of cross validation, a new model is trained independently of the model trained on the previous iteration. Because the same event can

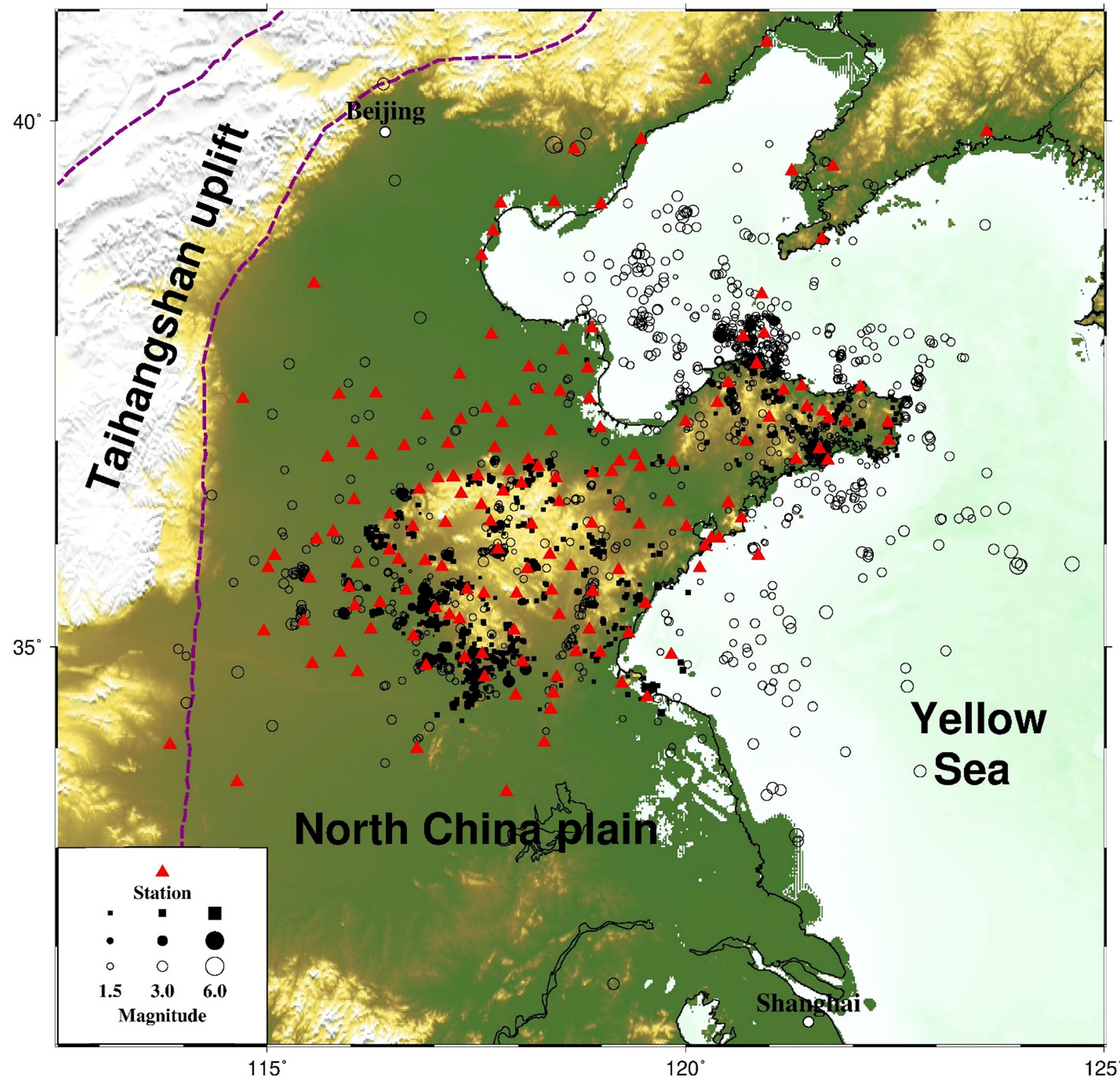

Figure 2. The distribution of monitoring stations and seismic events in the study region, with the dataset sourced from the Seismic Network of Shandong Province (SSNC). The locations of monitoring stations are denoted by triangles, and the natural earthquakes, mine collapses, and explosions are indicated by circles, dots, and squares, respectively, as initially labeled by human experts. The purple-dashed lines are major geological block boundaries. The white dots are major cities. The color version of this figure is available only in the electronic edition.

be recorded by multiple stations, we employ the majority voting algorithm to determine each event at the network level. The output of each event is obtained by summing the outputs of all the stations that record the corresponding event. Ultimately, our CNN model successfully identifies 97.3% (1007/1035) of the earthquakes, 84.9% (135/159) of the collapses, and 93.7% (549/586) of the explosions (as depicted in Fig. 4a,c,e, respectively). For the comparison at the station level, our CNN model successfully identifies 92.3% (10,982/11,901) of the earthquakes, 80.7% (1,219/1,511) collapses, and 83.1% (3,443/4,145) explosion.

The seismograms of all the measurements that exhibit different classifications assigned by the CNN and human experts are subjected to manual evaluation. Our findings indicate that 10 out of 28 earthquakes, 18 out of 24 collapses, and 18 out of 37 explosions were mislabeled by the previous human operators. Here, we selected three events to exemplify the

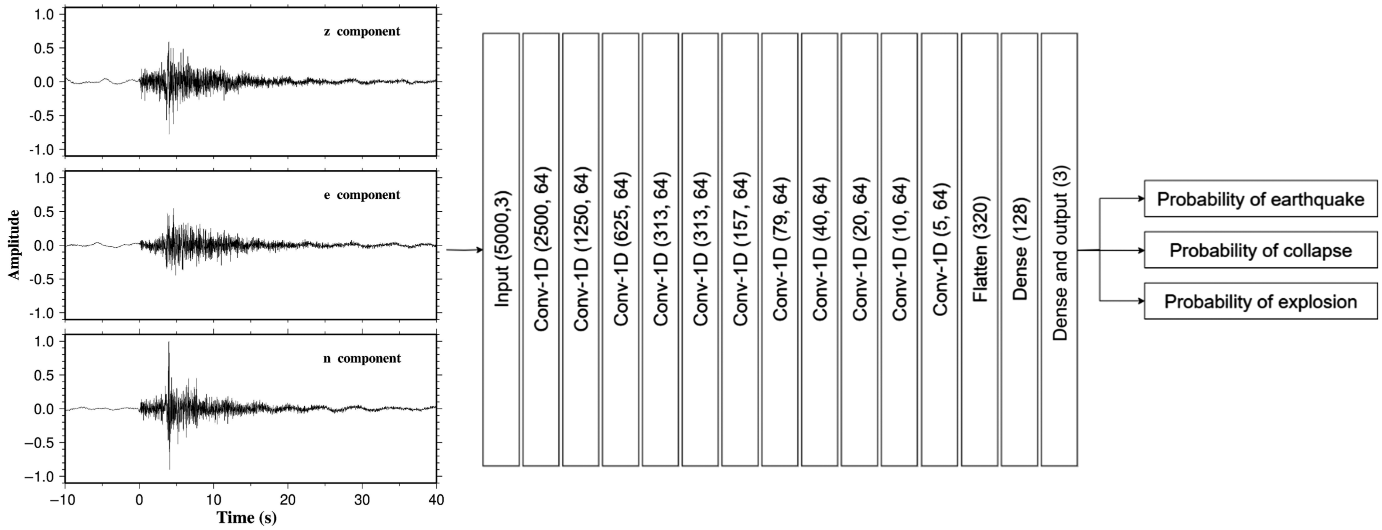

Figure 3. The operational workflow of the convolutional neural network (CNN) adopted in this study. The CNN receives three-component (N, E, and Z) seismic waveforms as inputs, and produces outputs in the form of probability estimates for natural earthquakes, mine collapses, and explosions, represented as three distinct probabilities.

characteristic waveforms associated with different seismic sources, as illustrated in Figure 5. Collapse events typically exhibit longer periods compared to earthquakes and explosions with more clear surface waves. In the z-component, the initial particle motion of vertical component generally shows downward displacement. Therefore, event 202002050239 has been reclassified as a collapse event, diverging from its initial categorization as an earthquake based on its long periods waveform (Fig. 5a,d,g). Explosion events, in contrast, manifest as a uniform spectrum, with an upward first motion in the vertical component. Furthermore, the amplitude of the *P* wave is generally larger than the *S* wave in the case of explosions. Earthquakes, on the other hand, display a more complex spectral pattern. The initial particle motion of vertical component varies depending on the coordinate of the back azimuth. In addition, earthquakes typically show a slower rate of signal attenuation than explosions or collapses. Consequently, events 201804182153 and 202003222154, originally labeled as a collapse and an explosion, respectively, have been reinterpreted as earthquakes based on these specific seismic characteristics based on their large *S*-wave amplitude and complex spectral pattern (Fig. 5b,c,e,f,h,i).

Moreover, due to the presence of indistinct features on the seismograms, 5 earthquakes, 1 collapse, and 12 explosions are regarded as ambiguous events. Consequently, it is difficult to determine whether these events were mislabeled by the CNN or previous human operators. For examples, Figure 6a,d,g is classified as an ambiguous earthquake due to the overlapping features from all three types of events in their traces (large *P*-wave amplitude, long-period waveform after *S* wave); Figure 6b,c,e,f,h,i is classified as an ambiguous collapse and explosion respectively due to low signal-to-noise ratio. Even if we assume that the previous human operators' labeling is correct for all the ambiguous events, more than 50% of the conflicting measurements are still misclassified by the human operators with varying event labels. These results suggest that the CNN outperforms human operators when working with the original dataset.

Results with corrected dataset

Mislabeled and ambiguous measurements commonly exist in various datasets and decrease the accuracy of ML models. To minimize the impact of unreliable measurements, we perform correction of all mislabeled measurements and removal of ambiguous ones with unclear features. We then conduct the same tenfold cross-validation process using the corrected dataset. After the same process as that used for the first scan is performed, the accuracies at the network level of the CNN increase from 97.3% to 99.2% (1040/1048) for natural earthquakes, from 84.9% to 95.9% (137/143) for collapses, and from 93.7% to 98.1% (560/571) for explosions (Fig. 4b,d,f, respectively). At the station level, CNN successfully classified 94.7% (11,409/12,053) earthquakes, 88.7% (1,228/1,385) collapses, and 88.1% (3,549/4,030) explosions. The overall accuracy at the network level of each type is over 95% with the corrected dataset. The CNN produces discrepant labels compared to previous human operators in 25 events, 10 of which are consistent with the initial training process utilizing the original dataset, whereas 15 are novel in the second training process that incorporated the corrected dataset. Furthermore, the CNN labels all corrected events identically to the rectified labels in the initial training process. Thus, the CNN agrees with all the manually corrected events and provides labels that align with those assigned by human experts for the 15 mislabeled events in the initial training process. We also manually assess the 15 unique events and find that 1 earthquake and 1 collapse are considered mislabeled by human, and the features of the 2 explosions and 1 earthquake are unclear on the seismograms.

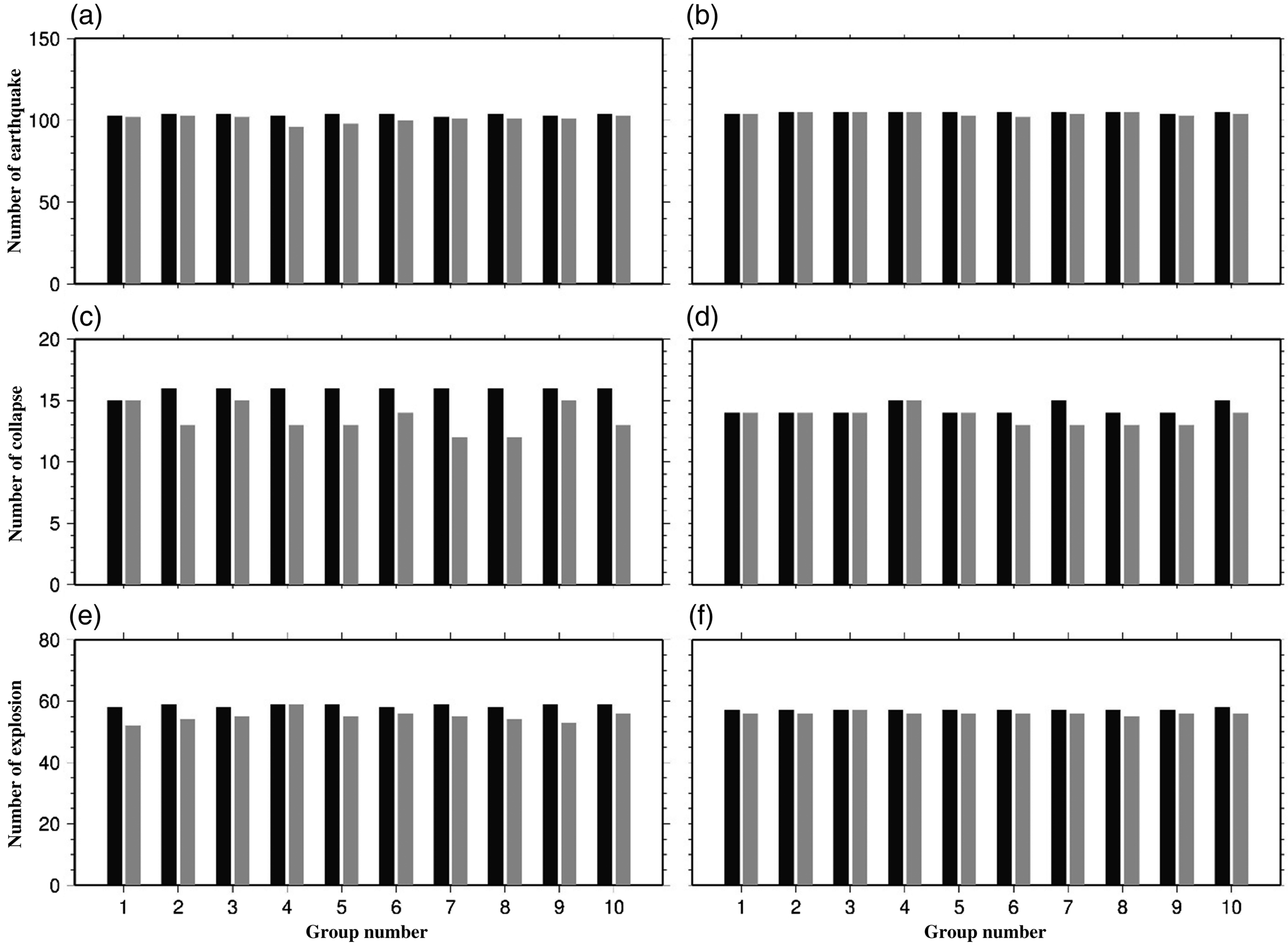

Figure 4. Performance evaluation of the CNN in the study, with the number of various types of events in each group denoted by left bars at each group number, and the number of events with matching labels assigned by both the CNN and human evaluators are shown by right bars at each group number. (a,c,e) The outcomes generated using the original dataset. (b,d,f) The results obtained with the corrected dataset.

DISCUSSION

Impact of data quality and human-in-the-loop feedback

After the correction of mislabeled and ambiguous events, CNN exhibits a noticeable improvement in accuracy. The true mislabeled events, which are correctly labeled by human experts, but misclassified by CNN, exhibit vague results, with the maximum value being close to the other two values. Cross validation is a powerful technique for evaluating the performance of ML models. When cross validation is combined with CNNs, the resulting methodology becomes a potent tool for identifying mislabeled data. By iteratively training and testing the model using different subsets of the data, researchers can gain insights into the quality of the labels assigned to each data point. Any instances of mislabeling can then be flagged and corrected, leading to improved accuracy and reliability of the resulting ML model.

Moreover, even with the most advanced ML techniques, the role of the human-in-the-loop feedback remains critical (Xin *et al.*, 2018; Monarch, 2021). This is particularly true when dealing with complex datasets, where the insights and experience of a human expert can help guide the evaluation process and ensure that the model is performing as expected. By combining the power of cross validation with the human expertise of researchers, it is possible to create ML models that are accurate, reliable, and effective in real-world applications.

Impact of data distribution

Even after balancing, CNN still exhibits superior accuracy in categorizing types with increasing event frequencies. This may be due to the fact that a large dataset includes various situations, and CNN attains a higher probability of avoiding overfitting. Overfitting occurs when a model is trained too well on a specific dataset, but cannot generalize to new, unseen data. Hence, it is recommended to utilize a balanced dataset for CNN training, whenever feasible, to ensure high classification accuracy, particularly for infrequent events that could be inadequately represented in the dataset. A balanced dataset can potentially enhance the model's ability to identify all classes equally well, which is crucial in applications such as event

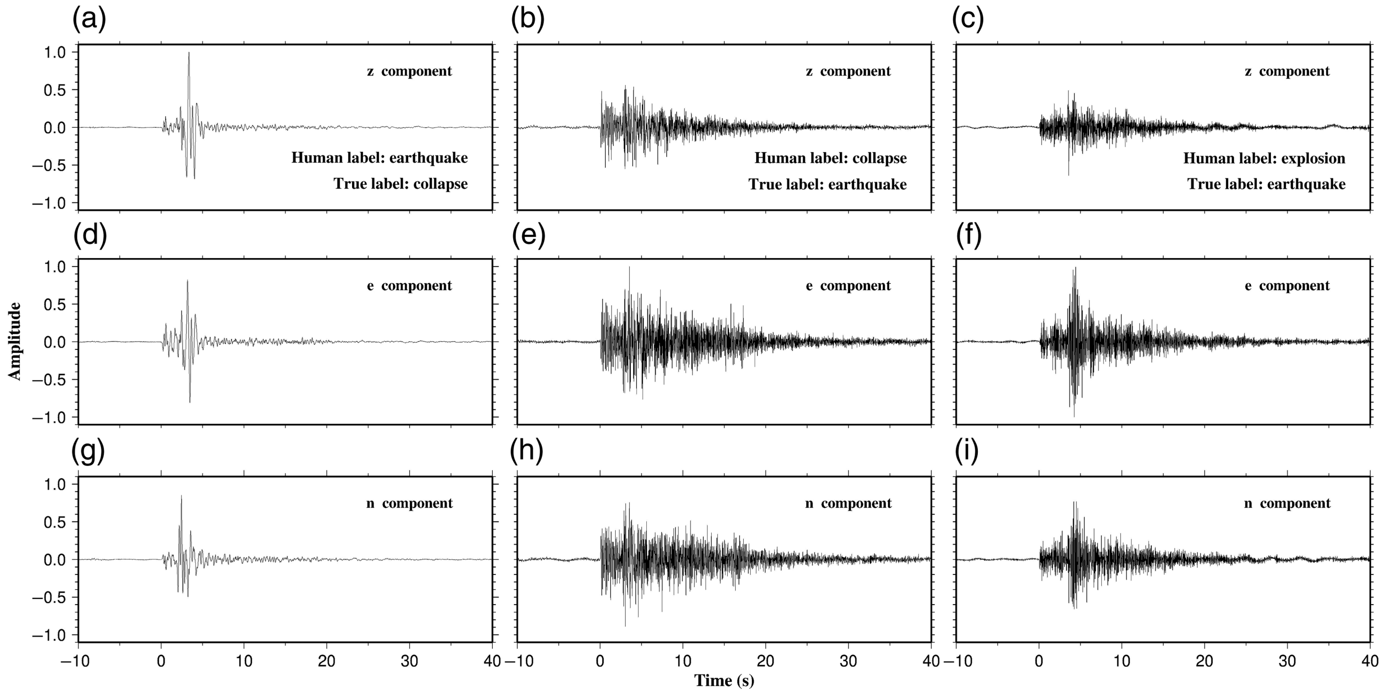

Figure 5. Three types of seismic events are demonstrated by the following examples: (a,d,g) three-component traces of event 202002050239 at station ZCH. (b,e,h) Event 201804182153 at station DSD. (c,f,i) Event 202003222154 at station XIT.

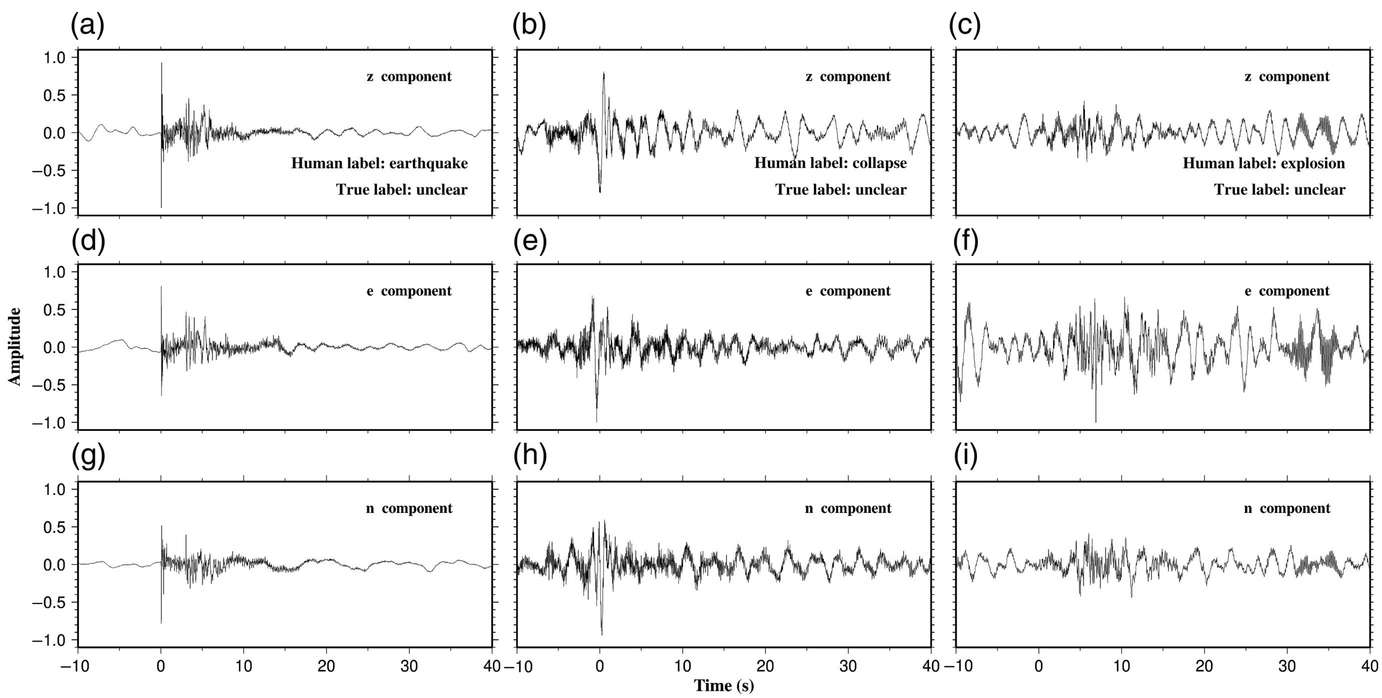

Figure 6. Three types of seismic events are demonstrated by the following examples: (a,d,g) three-component traces recorded at station PENL during event 201811091057, which is categorized as an ambiguous earthquake. (b,e,h) Traces at station RSH during event 201811201143, which is categorized as an ambiguous collapse. (c,f,i) Traces at station CHD during event 201810241237, which is categorized as an ambiguous explosion.

detection where each class is of equal importance. In addition, balanced datasets may assist in reducing the influence of any bias or noise that may be present in the training data.

To further investigate the performance of CNN, the distribution of the true mislabeled events in the results of CNN is evaluated. To accomplish this, we plot the accuracy of the CNN

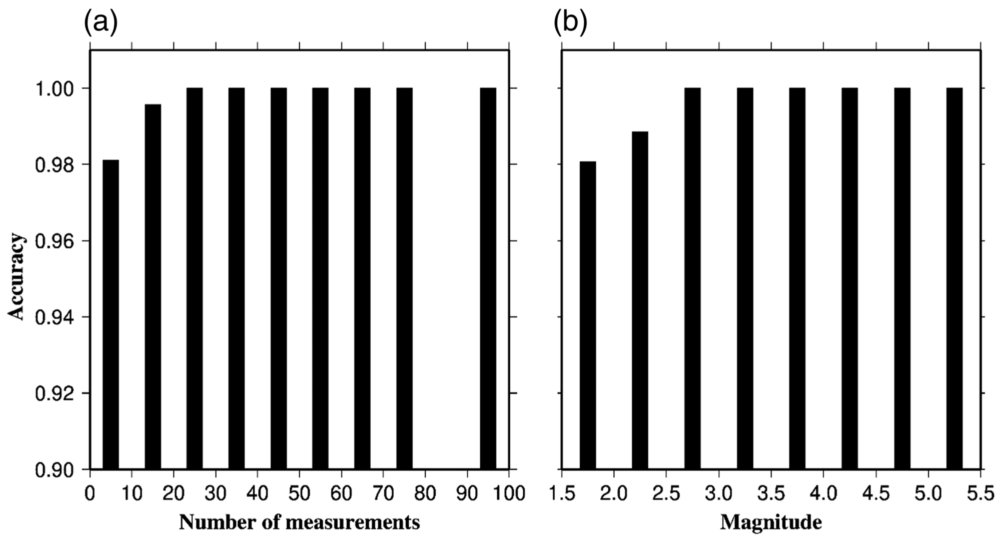

Figure 7. Two plots highlighting the influence of various factors on the performance of the CNN using the corrected dataset. (a) The accuracy of the CNN as a function of the number of measurements acquired from each event. (b) The accuracy of the CNN with respect to the magnitude of the events.

versus the magnitude and the number of measurements of all events, as shown in Figure 7. The plot demonstrates that the accuracy of the CNN increases with the magnitude and the number of measurements. When the magnitude is larger than 2.5 and the number of measurements is larger than 20, the CNN can successfully classify all the events. This is consistent with the commonsense idea that a larger magnitude would generate clearer features on the seismograms, and a larger number of measurements would reduce the influence of bias.

Data-centric artificial intelligence

In addition to the challenges of interpretability and calibration, CNNs can also be impacted by the quality and quantity of data available for training. The accuracy and reliability of the CNN model are highly dependent on the quality of the dataset used for training. Insufficient or poorly curated data can lead to underfitting or overfitting of the model, resulting in a less reliable and accurate classification. Ng *et al.* (2021) proposed a data-centric artificial intelligence strategy that emphasizes the importance of improving datasets instead of solely focusing on calibrating models. This approach involves developing better data collection strategies, incorporating human-in-the-loop feedback, and ensuring the quality and diversity of the data used for training. By improving the quality of the dataset and incorporating human-in-the-loop feedback, the resulting CNN models can be more accurate, reliable, and interpretable.

CONCLUSION

In this study, we establish an 11-layer CNN to classify natural earthquakes, mine collapses, and explosions in the Shandong Province, China. We perform cross validation to comprehensively scan the entire dataset, and the findings indicate that CNN outperforms human analysts. Furthermore, we manually evaluate all the measurements in the dataset with different labels assigned by both the CNN and human evaluators, retaining, correcting, or discarding them as appropriate. By training CNN on the corrected dataset, we observe a noticeable enhancement in accuracy, with all accuracy values of the three event types surpassing 95%. In addition, we find that the performance of CNN is strongly dependent on the number and magnitude of measurements from each event received by the monitoring stations.

These outcomes indicate that the quality and distribution of the dataset used to train and validate CNN play a critical role in its performance. Therefore, special attention must be paid to the dataset when using CNNs for seismic waveform classification.

DATA AND RESCOURCES

The code and trained models utilized in this study can be accessed on Github (https://github.com/YW-Zhang94/CNN_Classification_Earthquake_Collapse_Explosion.git, last accessed March 2023), whereas the dataset has been made available on Figshare (doi: 10.6084/m9.figshare.22207498). Waveform data has been made available with limitations due to confidentiality policies, consisting of 10 s before and 40 s after the *P*-wave arrival without location information. Furthermore, the authors would like to acknowledge the significant contribution of ChatGPT in refining and enhancing the text of our article.

DECLARATION OF COMPETING INTERESTS

The authors acknowledge that there are no conflicts of interest recorded.

ACKNOWLEDGMENTS

The authors express our gratitude to the SSNC for providing the valuable dataset. This study is supported by the Natural Science Foundation of Shandong Province (Grant ZR2020KF003), the US National Science Foundation (Grants 1919789 and 2305799 to S. G.) and Research Team on Monitoring and Activity Mechanisms of Unnatural Earthquakes of Shandong Earthquake Agency (TD202301).

REFERENCES

Astiz, L., J. A. Eakins, V. G. Martynov, T. A. Cox, J. Tytell, J. C. Reyes, R.L. Newman, G.H. Karasu, T. Mulder, M. White, *et al.* (2014).

The array network facility seismic bulletin: Products and an unbiased view of United States seismicity, *Seismol. Res. Lett.* **85,** no. 3, 576–593, doi: 10.1785/0220130141.

Barama, L., J. Williams, A. V. Newman, and Z. Peng (2023). Global nuclear explosion discrimination using a convolutional neural network, *Geophys. Res. Lett.* **50,** no. 17, e2022GL101528, doi: 10.1029/2022GL101528.

Baumgardt, D. R., and G. B. Young (1990). Regional seismic waveform discriminants and case-based event identification using regional arrays, *Bull. Seismol. Soc. Am.* **80,** no. 6B, 1874–1892.

Cao, A., and S. S. Gao (2002). Temporal variation of seismic b-values beneath northeastern Japan island arc, *Geophys. Res. Lett.* **29,** no. 9, 48-1–48-3, doi: 10.1029/2001GL013775.

Chen, H., J. Chen, and J. Ding (2021). Data evaluation and enhancement for quality improvement of machine learning, *IEEE Trans. Reliab.* **70,** no. 2, 831–847, doi: 10.1109/TR.2021.3070863.

Fäh, D., and K. Koch (2002). Discrimination between earthquakes and chemical explosions by multivariate statistical analysis: A case study for Switzerland, *Bull. Seismol. Soc. Am.* **92,** no. 5, 1795–1805, doi: 10.1785/0120010166.

Garcia, J. A., L. Waszek, B. Tauzin, and N. Schmerr (2021). Automatic identification of mantle seismic phases using a convolutional neural network, *Geophys. Res. Lett.* **48,** no. 18, e2020GL091658, doi: 10.1029/2020GL091658.

Gong, Y., G. Liu, Y. Xue, R. Li, and L. Meng (2023). A survey on dataset quality in machine learning, *Inf. Software Technol.* doi: 10.1016/j.infsof.2023.107268.

Goodfellow, I., Y. Bengio, and A. Courville (2016). Machine learning basics, *Deep Learn.* **1,** no. 7, 98–164.

Gulli, A., and S. Pal (2017). *Deep Learning with Keras*, Packt Publishing Ltd.

He, H., and E. A. Garcia (2009). Learning from imbalanced data, *IEEE Trans. Knowl. Data Eng.* **21,** no. 9, 1263–1284, doi: 10.1109/TKDE.2008.239.

Kim, W. Y., V. Aharonian, A. L. Lerner-Lam, and P. G. Richards (1997). Discrimination of earthquakes and explosions in southern Russia using regional high-frequency three-component data from the IRIS/JSP Caucasus network, *Bull. Seismol. Soc. Am.* **87,** no. 3, 569–588, doi: 10.1785/BSSA0870030569.

Kingma, D. P., and J. Ba (2014). Adam: A method for stochastic optimization, doi: 10.48550/arXiv.1412.6980.

Linville, L., K. Pankow, and T. Draelos (2019). Deep learning models augment analyst decisions for event discrimination, *Geophys. Res. Lett.* **46,** no. 7, 3643–3651, doi: 10.1029/2018GL081119.

Maas, A. L., A. Y. Hannun, and A. Y. Ng (2013). Rectifier nonlinearities improve neural network acoustic models, *Proc. of the 30th International Conf. on Machine Learning*, Atlanta, Georgia, 16 June 2013, Vol. 30, 3 pp.

Monarch, R. M. (2021). *Human-in-the-Loop Machine Learning: Active Learning and Annotation for Human-Centered AI*, Simon and Schuster, New York.

Ng, A., D. Laird, and L. He (2021). Datacentric AI competition, available at https://https-deeplearning-ai.github.io/data-centric-comp/ (last accessed June 2021).

O'Rourke, C. T., and G. E. Baker (2017). A spectrogram-based method of Rg detection for explosion monitoring, *Bull. Seismol. Soc. Am.* **107,** no. 1, 34–42, doi: 10.1785/0120160184.

Perol, T., M. Gharbi, and M. Denolle (2018). Convolutional neural network for earthquake detection and location, *Sci. Adv.* **4,** no. 2, e1700578, doi: 10.1126/sciadv.1700578.

Wang, T., Y. Bian, Y. Zhang, and X. Hou (2023). Using artificial intelligence methods to classify different seismic events, *Seismol. Res. Lett.* **94,** no. 1, 1–16, doi: 10.1785/0220220055.

Xin, D., L. Ma, J. Liu, S. Macke, S. Song, and A. Parameswaran (2018). Accelerating human-in-the-loop machine learning: Challenges and opportunities, *Proc. of the Second Workshop on Data Management for End-to-End Machine Learning*, 1–4, doi: 10.1145/3209889.3209897.

Zhang, Y., and S. S. Gao (2022). Classification of teleseismic shear wave splitting measurements: A convolutional neural network approach, *Geophys. Res. Lett.* e2021GL097101, doi: 10.1029/2021GL097101.

Zhu, W., and G. C. Beroza (2018). PhaseNet: A deep-neural-network-based seismic arrival-time picking method, *Geophys. J. Int.* **216,** no. 1, 261–273, doi: 10.1093/gji/ggy423.

Manuscript received 23 July 2024
Published online 31 December 2024