EQTypeNet: Deep Learning Tri-Branch Earthquake Automatic Classification Model and Its Application in China

Article *in* Geophysical Journal International · July 2025

DOI: 10.1093/gji/ggaf263

CITATIONS
0

READS
63

7 authors, including:

Lianqing Zhou
Institute of Earthquake Forecasting, China Earthquake Administration, China, Beij…
61 PUBLICATIONS **442** CITATIONS

SEE PROFILE

Ming Zhao
Chinese Academy of Sciences
15 PUBLICATIONS **281** CITATIONS

SEE PROFILE

Geophysical Journal International

Geophys. J. Int. (2025) **242,** 1–19
Advance Access publication 2025 July 08
Seismology

https://doi.org/10.1093/gji/ggaf263

EQTypeNet: deep learning tri-branch earthquake automatic classification model and its application in China

Tianran Lu,[1] Lianqing Zhou ,[1] Na Zhang,[2,3] Mengqiao Duan,[1] Ziyi Li,[2,4] Ming Zhao[2] and Li Sun[5]

[1] *Institute of Earthquake Forecasting, China Earthquake Administration, Beijing* 100036*, China. E-mail: zhoulq@ief.ac.cn, duanmq@ief.ac.cn*
[2] *Institute of Geophysics, China Earthquake Administration, Beijing* 100081*, China*
[3] *Shanxi Earthquake Bureau, Taiyuan* 030000*, China*
[4] *School of Earth and Space Sciences, Peking University, Beijing* 100871*, China*
[5] *China Earthquake Networks Center, Beijing* 100045*, China*

Accepted 2025 July 1. Received 2025 March 19; in original form 2024 October 25

SUMMARY

The rapid development of the global economy brings frequent artificial seismic events related to industrial activities. The identification of faults revealed by natural earthquakes, along with the temporal and spatial distribution of *b*-values derived from these seismic events, is influenced by their occurrence. Correcting identification of seismic events has significant influence on seismic risk assessment. In this paper, we present a tri-branch convolutional neural network model, known as EQTypeNet, to automatically classify three classes of events (natural earthquakes, explosions and collapses) in China, using waveform, spectrogram and amplitude ratio features as inputs. The EQTypeNet has developed in two steps: (1) Binary classification: mitigating data imbalance by supplementing non-natural events outside the test set area. In the classification of explosions and earthquakes, we obtained an individual station macroF1 of 0.99 and a classification accuracy of 98.7 per cent. In the classification of collapses and earthquakes, we obtained an individual station macroF1 of 0.99 and a classification accuracy of 99.5 per cent. (2) Ternary classification of natural earthquakes, explosions, and collapses: enhancement of the model's classification in small areas by transfer learning. In the transfer learning of Shanxi and Northeast China, the F1-score of individual stations reached 0.98 and 0.97, with corresponding accuracies of 98.6 and 96.8 per cent. The EQTypeNet applies to different magnitude ranges for seismic event discrimination and shows a satisfactory application performance in a large area. It will be expected to achieve higher performance in the ternary classification by further supplementing the training with suitable physical features.

Key words: Image processing; Machine learning; Earthquake hazards..

1 INTRODUCTION

The rapid development of the global economy brings frequent artificial seismic events related to industrial activities. Many artificial earthquakes have been classified as natural earthquakes in the network monitoring due to previous disagreements over the definition of seismic categories and a lack of knowledge in related studies (Zhao *et al.* 1995). Misidentified artificial earthquakes in seismic catalogues can cause significant problems when deducing underground fault tectonics from seismic events. Furthermore, they can also lead to anomalies in the temporal and spatial distribution of the regional *b*-value (Tang *et al.* 2020; Gulia & Gasperini 2021). These anomalies have implications for studies such as precursor discrimination of large earthquakes (Gulia & Wiemer 2019) and delineation of source regions for future strong earthquakes (Yi *et al.* 2008). Similarly, shallow natural earthquakes are often incorrectly labelled as artificial events, which may precipitate unwarranted public panic and affect the accuracy of earthquake catalogues. Correct identification of shallow natural earthquakes can provide abundant case data for related studies of the focal mechanism and seismogenic environment (Cornou *et al.* 2022). The accurate classification of natural and artificial earthquakes, along with the study of specific types of explosions and collapses within the latter, is important for the identification of regional underground faults, seismic hazard analysis, global nuclear explosion monitoring, risk assessment of artificial earthquakes and earthquake prediction.

Theoretically, artificial earthquakes, such as explosions and collapses, differ from natural earthquakes in terms of their focal mechanisms and waveforms due to the different mechanisms of generation. Traditional seismic classification studies start with waveforms or spectra and classify events by extracting a certain feature or combining discriminant functions of multiple features (Kim *et al.* 1994; Fereidoni & Atkinson 2017; Chen *et al.* 2021). These functions have a reasonable physical interpretation, but their generalization ability is limited by the extent of the study area and a single categorical feature. On the other hand, increasingly dense station networks and frequent industrial activities have led to an explosive growth in seismic data. Traditional classification has a relatively low-data utilization rate, making it difficult to meet the demand for automated classification driven by data growth and practical applications. Recently, the artificial intelligence (AI) seismic classification methods have obtained results rivaling the accuracy of traditional techniques (Hammer *et al.* 2012; Wei *et al.* 2019; Li *et al.* 2020; Tang *et al.* 2020; Wenner *et al.* 2021; Manley *et al.* 2022). Compared to the traditional method, the AI method can combine features more easily, handle large amounts of data, extract detailed information through automated clustering or by using high-quality data labels and enable automated classification where possible.

When labels are sufficient, most seismic classification studies tend to use supervised learning methods with high accuracy and intuitive results (Mousavi & Beroza 2023). An objective reality is that the size of the training set significantly impacts models' performance. However, the number of non-natural events is substantially smaller than that of natural earthquakes, resulting in a limited training set for classification tasks. Machine learning (ML) algorithms suitable for small data sets such as Support Vector Machines (SVM) are more accurate in seismic classification (Kortström *et al.* 2016; Tang *et al.* 2020; Liang *et al.* 2023). Building on this, the generalization performance of models has also garnered attention. Wang *et al.* (2023) noted that data sets incorporating multidimensional features benefit model generalization. Their data set includes 36-D features such as the Pg/Sg amplitude ratios and 201-D spectral amplitude features. The eXtreme Gradient Boosting (XGBoost) ensemble algorithm trained on this data set outperformed the standalone SVM classifier in distinguishing natural earthquakes, explosions and collapses. While ML training sets with manually selected feature parameters offer diversity, their construction is time-consuming and costly. Additionally, such data sets may introduce subjective biases that impact model performance. Koper *et al.* (2024) advocated selecting physical features less correlated with the study region to improve generalization. They observed that the performance of the Naive Bayes approach declined when more than six features were combined. Despite the effectiveness of ML in seismic classification, their reliance on human-selected features still constrains the performance of the models. Deep learning (DL) techniques provide a new direction to address this problem.

Mature data enhancement techniques and transfer learning methods have effectively mitigated the data imbalance problem between natural and non-natural earthquake events, providing data support for the application of DL in seismic classification (Titos *et al.* 2023). DL algorithms that automatically extract features from data reduce the potential for contamination of feature labels by human selection, allowing raw data containing multiple pieces of information to be used as input data (Jiang *et al.* 2023). Duan *et al.* (2021) employed ML and DL algorithms to classify underground mining-induced seismicity. Their results showed that the deep convolutional neural network, which used raw waveforms as input, demonstrated superior classification performance compared with SVM and Random Forest models trained on manually selected features. Different model architectures exhibit variations in classification results, while the composition of the training set may play a more critical role in determining model generalization capability. In seismic classification tasks, current methodologies to enhance model generalization through data set diversity primarily focus on two approaches.

(1)Expanding study area scope

Seismic signals from different regions carry diverse geological information, which enhances data set diversity to some extent. Taking composite small regions as an example. For data balancing, the study area for DL earthquake classification tends to be a small area with a concentration of artificial events (e.g. explosions, collapses, etc.). Maguire *et al.* (2024) constructed a data set comprising approximately 35 000 earthquake and explosion events from eight distinct U.S. regions, incorporating multiple blast types. They trained a convolutional neural network (CNN) model using scalograms, which are similar to spectrograms (calculated using the discrete Fourier transform). The model could be effectively applied to classifying earthquakes and explosions at local distances (within 200 km) in various geologic contexts, outperforming the uncorrected *P*/*S* ratios classification models. In contrast to localized studies, the classification performance of DL models deteriorates in spatially extensive regions, where data imbalance arises not only from event type disparities but predominantly stems from geological heterogeneity. Liu *et al.* (2020) applied waveforms as input to a CNN model to achieve binary classification of natural and non-natural earthquakes in China, excluding Yunnan, Sichuan and Inner Mongolia, but the accuracy was only 90 per cent. Due to the constraints of complicated geological characteristics and numerous types of non-natural events, DL seismic classification still faces many difficulties and challenges at larger regional scales.

(2)Enhancing data feature diversity

Raw seismic data inputs for DL predominantly consist of waveforms and spectrograms (Hourcade *et al.* 2022), whose rich seismological information enhances the upper-bound classification performance of deep learning models. Meanwhile, physics-informed features derived from empirical studies improve model convergence during training. Physics-based *P*/*S* amplitude ratio features have proven effective in discriminating small-magnitude earthquakes, explosions and collapse events (Kolaj 2018; Walter *et al.* 2018; Wang *et al.* 2023). As early as 1996, Musil & Plešinger (1996) demonstrated that the combination of spectra and *P*/*S* ratios was more effective than spectra alone as input to CNN models. Recent work by Koper *et al.* (2024) showed that models developed with physics-based waveform features achieve classification performance comparable to high-dimensional DL models in distinguishing small earthquakes, collapses and explosions. Integrating waveforms, spectrograms and generalizable physical features, while accepting modest reductions in model automation, might further optimize classification performance. Kong *et al.* (2022) presented a two-branch CNN model, trained on approximately 350 000 samples incorporating both waveform and *P*/*S* ratio features. Compared to waveforms or *P*/*S* ratios alone, this combination has better generalization in classifying small-magnitude earthquakes and underground chemical explosions.

In this study, the main problem is to develop automated DL earthquake classification models, to improve their performance on

a broader regional scale and to lower the magnitude threshold for classification. We employed the EQTypeNet that improved on the model of Kong *et al.* (2022), with waveform, spectrogram and *P*/*S* ratio features as inputs, to construct a classification model applicable for ternary classifications: earthquakes, explosions and collapses, in China. Based on the EQTypeNet, transfer learning was used to reduce the reliance of DL on data size: first pre-trained on ComCat, a Pacific Northwest data set with sufficient artificial events, then fine-tuned using the China DiTing 2.0 data set. In order to ensure that the EQTypeNet generalizes to different regions and achieves the best classification performance, we finally employed a EQTypeNet that had been fine-tuned or retrained on the DiTing 2.0 as our pre-trained model and further fine-tuned it with corresponding regional events when applying it to different regions in China.

2 DATA

2.1 ComCat data set

The PNW-ML data set (Ni *et al.* 2023) from the Pacific Northwest of the United States is a fundamental data set suitable for the study of seismological artificial intelligence techniques. The Pacific Northwest region is a plate boundary where the active margin is a subduction zone between the continental plate and the oceanic plate. And the region, seismically active, contains multiple active volcanoes. One of the subdatasets is ComCat, which consists of earthquakes and explosions. It contains 65 384 events of magnitude 0 or greater for the period 2002 January 1 to 2022 December 31, and the sources of explosions include quarry explosions, borehole explosions, etc. Each waveform recorded is at least 150 s long, with the epicentral distance mainly within 400 km, preserving the pre-seismic noise, and all waveforms are annotated with *P* wave and *S* wave arrivals. Most of the events in the ComCat data set are earthquakes. Considering the data balance, we selected 5 705 earthquakes with the same magnitude distribution ratio as the explosions, and together with 5 705 explosions, these events formed the data set used in this paper for the earthquakes and explosions classification study. Events in the study data set are concentrated at the edge of the plate boundary along the west coast of the United States, and the intraplate distribution is relatively sparse (Fig. 1a). There are 11 858 earthquake waveform records and 12 103 explosion waveform records. And small-magnitude events dominate the data set, with the largest magnitude being only M_D 3.27 (Fig. 2a).

2.2 DiTing 2.0 data set

The DiTing 2.0 data set includes multiple types of seismic events. It was updated by Zhao et al. at the Institute of Geophysics, China Earthquake Administration, based on DiTing 1.0 (Zhao *et al.* 2023). The data set contains a total of 264 298 events from March 2020 to February 2023 in China. Similar to the ComCat data set, the data set is primarily composed of earthquakes, with only 536 explosions and 381 collapses (Fig. 1b). Non-natural events exhibit a concentrated phenomenon in central China, possibly due to a concentrated distribution of mining or quarries in the area. Based on the event information of the DiTing 2.0 data set, Chinese non-natural events have a certain regional distribution. Most of the recorded waveforms are annotated with *P*-wave and *S*-wave arrivals, and some of them are shorter than 60 s in length. Taking into account the data balance, we selected 996 earthquakes. This selection was based on the magnitude distribution ratio of explosions and the number of three component records of selected earthquakes. Together with all 536 explosions and 381 collapses, these events formed the data set used in this study. There are 6213 earthquake waveform records, 5513 explosion waveform records and 5206 collapse waveform records. Compared to the ComCat data set, the events of DiTing 2.0 are concentrated around M_L 2, with a maximum magnitude of M_L 4.4 (Fig. 2b).

2.3 ShanXi and DongBei data sets

The ShanXi and DongBei data sets were provided by the Shanxi Earthquake Bureau and the China Earthquake Networks Centre, respectively, and include three types of events: earthquakes, explosions and collapses. The epicentral distribution region of the Shanxi or DongBei data sets partially overlaps with DiTing 2.0, but there are no common events. Both data sets were used for small region classification applications in China, so we did not balance the data.

The events in the ShanXi data set are located in central China, the region that has dense mining areas and where collapses are frequent and concentrated (Fig. 3b). Event waveforms in the data set are all 60 s with a sampling rate of 100 Hz. After removing anomalous data, such as missing data in some channels, a total of 41 earthquakes (1076 waveform records), 25 explosions (446 waveform records) and 119 collapses (1436 waveform records) with magnitudes ranging from M_L 2.5 to 4.0.

The events in the DongBei data set are located in northeastern China, the region with relatively weak seismicity (Fig. 3d). Event waveforms in the data set are relatively long with a sampling rate of 100 Hz. After removing anomalous data, a total of 188 earthquakes (1 347 waveform records), 188 explosions (656 waveform records) and 97 collapses (1 442 waveform records) with magnitudes ranging from M_L 2.2 to 3.0.

2.4 Data set production

In order to unify the event data from different regions, we performed some pre-processing steps to prepare waveforms, spectrograms and *P*/*S* ratios.

Waveform: After removing the mean and trend from waveforms, we applied a 5 per cent Hanning window taper followed by a four-corner high-pass filter at 2 Hz, and then performed normalization for amplitude (divide the amplitude of each point in the waveform by the maximum absolute amplitude of the waveform). The waveform sampling rate was unified to 100 Hz, and the length is intercepted to 6001 data points (60 s). The cutting time of each waveform started randomly 5–20 s before the *P*-wave arrival time. If the interval between the waveform start time and the *P*-wave arrival time was less than 5 s, the cutting time was adjusted to the waveform start time.

Spectrogram: Following the study of Hourcade *et al.* (2022), we used the discrete Fourier transform to calculate the spectrogram. Since explosions generally occur near the surface, *Rg* waves are an effective indicator of shallow source events of small magnitude at local distances (Koper *et al.* 2021). O'Rourke & Baker (2017) also found significant *Rg* waves in the spectrogram of explosions. Due to the greater diversity of non-natural events in the DiTing 2.0 and to account for the *Rg* waves in shallow source events, we increased the frequency resolution and computed discrete Fourier transforms

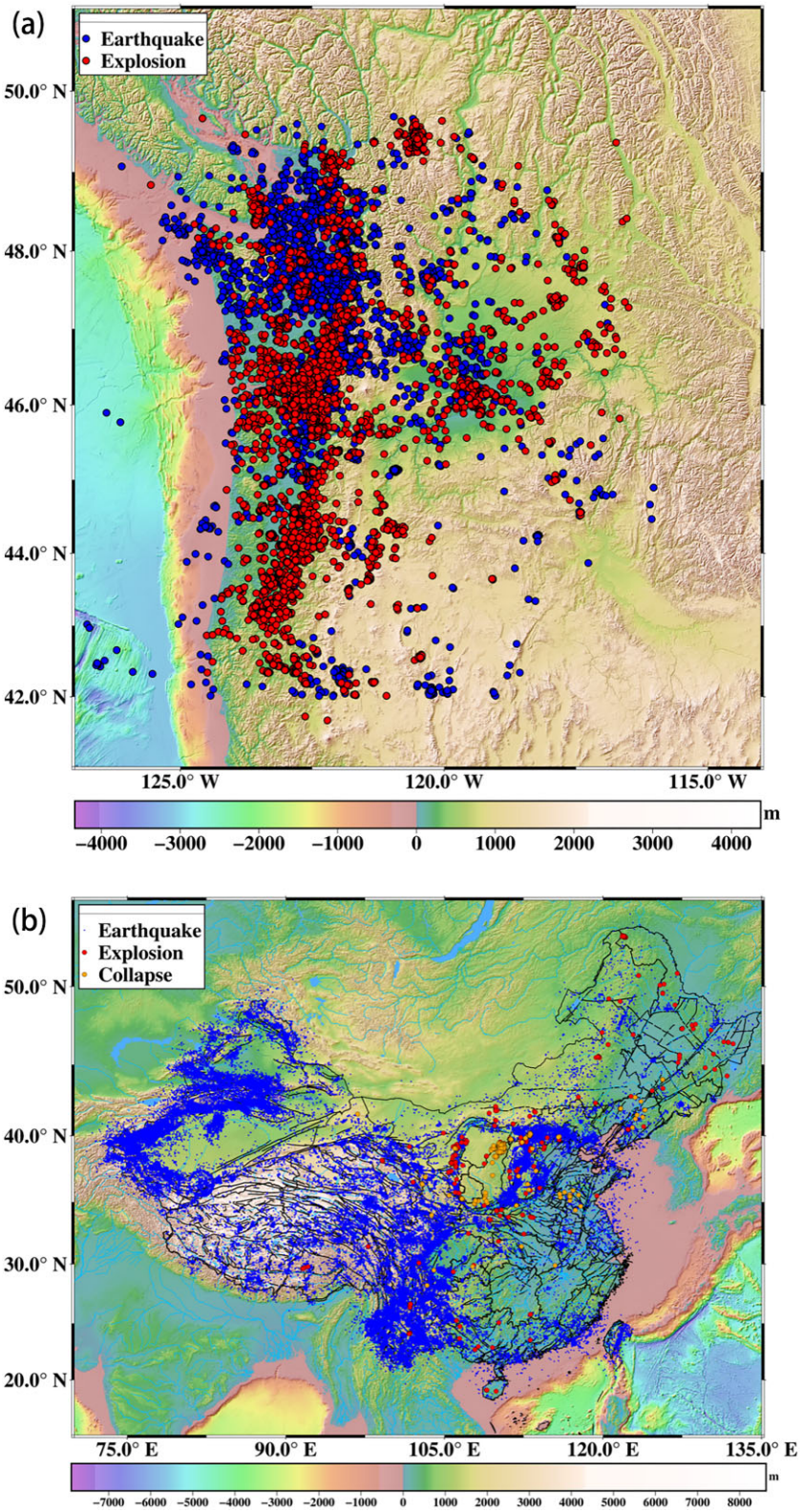

Figure 1. Distribution of event epicentres used in this study: (a) ComCat data set: distributions of earthquakes and explosions used in this study. (b) DiTing 2.0 data set: distributions of earthquakes, explosions and collapses.

of waveforms from 0.5 to 50 Hz using a 2 s sliding window with 75 per cent overlap.

P/*S* ratio: The selection of *P*/*S* ratio features is relatively complex compared to waveforms and spectrograms. They are affected by the calculation frequency band and exhibit varying classification performance in different regions and at different distances. The ComCat and DiTing 2.0 data sets encompass both local and regional events. *P*/*S* ratios obtained with a 2 Hz high-pass filter are effective in discriminating regional earthquakes, explosions and collapses (Walter *et al.* 2018), but they are less effective for local

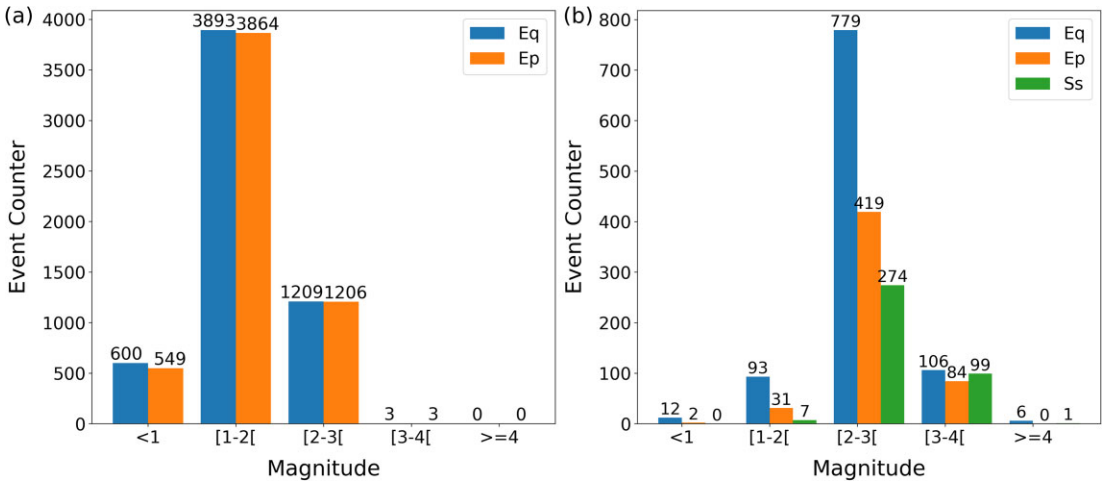

Figure 2. Distribution of event magnitude(M_L) for the data set: (a) ComCat; (b) DiTing 2.0. where Eq are earthquakes, Ep are explosions and Ss are collapses. The number of events here differs from the text due to the existence of 83 events with magnitude type M_D (duration magnitude) for the explosions in the ComCat data set.

events of small magnitude. *P*/*S* ratios calculated by a high-frequency bandpass filter (e.g. 10–18 Hz) can effectively discriminate small-magnitude local earthquakes from explosions (Wang *et al.* 2020). Since *Pg* and *Sg* waves of events are evident within 400 km, this study adopted the formula of Wang *et al.* (2021) to calculate the *P*/*S* ratios. The formula was somewhat simplified to reduce pre-processing steps for feature engineering (Zhu *et al.* 2023) and to ensure the final sufficient study data volume (1). The *P*/*S* ratios are no longer calculated by converting the ENZ channel to the RTZ channel. Rather, the *Z* component alone is calculated.

$$Pg/Sg = \sqrt{P_Z^2 - N_Z^2}/\sqrt{S_Z^2 - N_Z^2} \quad (1)$$

Referring to Kong *et al.* (2022), the *P*–*S* time difference is used as the basis for the selection of windows for calculating the *P*/*S* ratio. The windows for *Pg* and *Sg* start at 5 per cent of the *P*–*S* time difference before the corresponding seismic phase and end at 60 per cent of the *P*–*S* time difference after the corresponding seismic phase, with the windows length limited to 3 s (Fig. 4). The noise window was set before the arrival of the *Pg* to record the corresponding background noise. Since the event origin times of the DiTing 2.0 data set were updated during the study period, 60 per cent of the *P*–*S* time difference after waveform start time was taken as the noise window to harmonize the data (for all events, the time from the waveform start time to the *Pg* wave arrival time is greater than 60 per cent of the *P*–*S* time difference). When calculating the *P*/*S* ratios by high-frequency band filters according to formula (1), the number of usable *P*/*S* ratios was reduced. Therefore, we calculated the *P*/*S* ratios using a 2 Hz high-pass filter and used the average *P*/*S* ratio within the window as the *P*/*S* ratio feature. Given the large variability in *P*/*S* ratios of single waveform records (Pyle & Walter 2019), the final *P*/*S* ratio features for these records were calculated. These features were taken as the median of the *P*/*S* ratios of the *Z* component from all corresponding event records.

In order to alleviate the influence of the smaller training set size on the model performance after data balancing, we randomly selected the cutting time 5 times for each event record to generate different noise information for data enhancement (Table 1). The ShanXi training set was not processed for data enhancement due to waveform length limitations. And the DongBei training set, due to its fewer explosion records, we randomly selected the cutting time 10 times for each explosion record. To avoid potential seismic source information leakage during model training and to evaluate the model capabilities under more realistic monitoring conditions, we split the data set chronologically based on event number at a 9:1 training-to-test set ratio, considering the number of events rather than individual records. Meanwhile, 20 per cent of the training set was allocated as a validation set for parameter tuning. Because of the much smaller data size of the ShanXi data set, the training-test split ratio was adjusted to 5:5.

For simplicity, the data sets used in this study based on all the above pre-processing on the ComCat, DiTing, ShanXi and DongBei data sets will be referred to as the CC, DT, SX and DB data sets, respectively, in the following Sections.

3 METHODS

3.1 Tri-branch convolutional neural network

CNN is a deep neural network that can effectively learn spatial relationships in images. As the number of layers increases, the network is more sensitive to capturing localized information from large amounts of data and is able to handle complex abstract feature representations. However, deep CNN poses the risk of overfitting while improving accuracy. Adding dropout layers to the network framework can inhibit model overfitting and improve classification performances (Hinton *et al.* 2012; Krizhevsky *et al.* 2017). The lateral deepening of the two-branch CNN network proposed by Kong *et al.* (2022) allows for more diverse information in the representation of seismic data. And the dropout layers in the convolutional and fully connected layers suppress the risk of overfitting with lateral deepening. In our study, waveforms, spectrograms and *P*/*S* ratios are used as inputs to the EQTypeNet, a tri-branch CNN model. We fine-tuned the model parameters to achieve the best classification results.

We used the 6001 × 1 × 3 three component waveforms, the 117 × 100 × 3 three component spectrograms and the corresponding *P*/*S* ratios by a 2 Hz high-pass filter as inputs (Fig. 5). Next, we tested different architectures by varying the number of convolutional layers, the number of convolutional kernels and the parameters of the fully connected layers. Eventually, the EQTypeNet architecture is shown in Fig. 6, with waveform and *P*/*S* ratio branches referencing

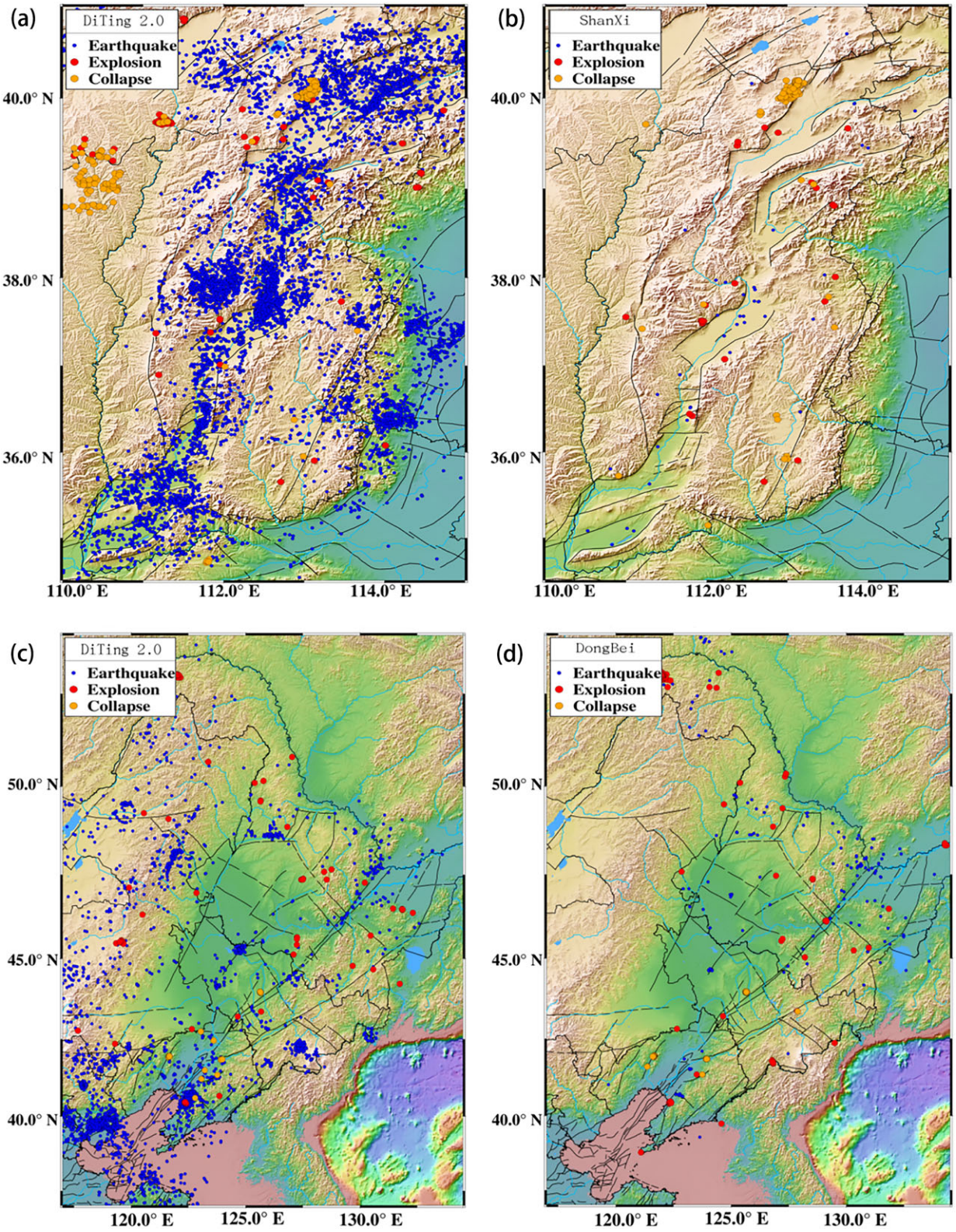

Figure 3. Distribution of events in Shanxi and Northeast China: (a) Distribution of events in Shanxi in the DiTing 2.0 data set; (b) distribution of events in the ShanXi data set; (c) distribution of events in Northeast China in the DiTing 2.0 data set; (d) distribution of events in the DongBei data set.

(Kong *et al.* 2022) and a spectrogram branch referencing (Hourcade *et al.* 2022). The waveform branch is composed of four convolutional layers with respectively 64, 128, 256 and 512 neurons for each layer. To simplify the computation and to highlight the most salient feature information, we apply a 4 × 1 max-pooling layer following each convolutional layer. To alleviate overfitting, dropout layers are introduced, beginning with the second convolutional layer. The convolutional part of the spectrogram branch is consistent with the

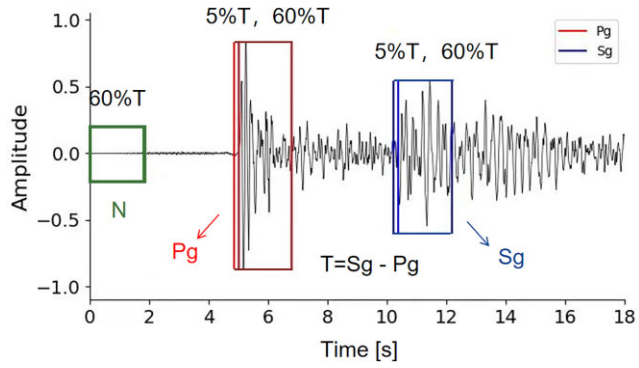

Figure 4. Schematic diagram of the *P*/*S* ratios calculation window: where *N* is the noise window and *T* is the *P*–*S* time difference, with the total window length limited to 3 s.

CNN model of Hourcade *et al.* (2022). After the flattening layer of the spectrogram branch, we included a dropout layer, followed by a fully connected layer of 420 neurons. The concatenation layer merges the outputs of the fully connected layer from the waveform branch, the spectrogram branch, and the *P*/*S* ratio branch along the feature dimension, forming a 1260-D fused feature vector. Then, after passing through three dropout layers and three fully connected layers, the input data is transformed into event category outputs. All layers use ReLU (Rectified Linear Unit) as the activation function except the last fully connected layer, which uses Softmax activation function.

3.2 Transfer learning

We further applied the transfer learning approach to improve DL models on small samples. Domain (source and target domains) and task (source and target tasks) are the two fundamental concepts of transfer learning. When transfer learning is used to improve the performance of a model on small samples, a larger data set is usually used as the source domain and a small data set corresponds to the target domain. If the distribution of data features in the source and target domains is similar, transfer learning can extract knowledge from the source task and apply it to the target task, further improving the model's result in the target domain. However, if the source and target domains exhibit low correlation, directly applying transfer learning may lead to negative transfer—a phenomenon where knowledge transferred from the source domain and task impairs the model's generalization capability in the target domain (Rosenstein *et al.* 2005). To avoid negative transfer, as Pan & Yang (2010) suggest, we should consider three key questions when applying transfer learning in practical research: 1 what to transfer, 2 how to transfer, and 3 when to transfer.

What to transfer: Current research on non-natural earthquakes shows that, although they are strongly influenced by path effect and the geological context, non-natural earthquakes in different regions still show partially uniform characteristics compared to natural earthquakes due to the different types of seismic sources: For example, the explosion source exhibits regular temporal and spatial features, it theoretically yields only *P* waves. Consequently, the *S* waves transformed during seismic propagation are significantly weaker than those from natural sources. Additionally, the first motion of the vertical component *P* wave typically points upward; Collapses have significant low-frequency components, and their surface waves with large periods can occur at local distances. The uniform characters of non-natural events across various regions answer what to transfer. These features may be common knowledge that can be learned as the model is applied to different regions, thus improving the classification performance of the target domain task.

How to transfer: When there are features that can be transferred between the source and target domains, transfer learning methods that utilize DL can be divided into four categories: instances transfer, mapping transfer, adversarial transfer and network transfer (Tan *et al.* 2018). The network transfer reduces a lot of tuning work and application deployment is relatively simple. In this study, we implemented transfer learning through a network transfer approach, which consists of two key steps: 1. Pre-trained Model Initialization: A model pre-trained on a larger-scale data set provides a foundational network architecture and optimized parameters, enabling high classification performance on the target task. 2. Task-Specific Fine-tuning: To address domain discrepancies between the source (pre-training) and target tasks, we replaced and retrained the top fully connected layer of the pre-trained network, adapting it to the target task's unique characteristics (Oquab *et al.* 2014). Finally, the model is fine-tuned during training on target domain data, freezing the convolution layers of the pre-trained model. In the fine-tuning process, we adjusted the initial learning rate from 0.01 based on the EQTypeNet classification results. When the initial learning rate was set to 0.001, the performance of the EQTypeNet was better after transfer learning. The final training settings for the transfer learning process are the same as in Section 4.1.1.

Table 1. Number of training and test set events used in this study: The number of waveforms (three components) in the training set is the total number after data enhancement. The *P*/*S* ratio calculations involve noise windows, and some of the waveforms are enhanced with a lower signal-to-noise ratio. After removing these data, the number of waveforms in the training set is not 5 or 10 times as strict as the waveforms in the data sets of Sections 2.1, 2.2 and 2.3. Eq, Ep and Ss denote earthquakes, explosions and collapses.

Data set		Number of events			Number of waveforms			Magnitude
		Eq	Ep	Ss	Eq	Ep	Ss	(M_L)
CC	Training Set	5135	5135	–	49 362	53 700	–	0–3.27
	Test Set	570	570	–	1982	1364	–	
DT	Training Set	896	482	343	27 632	24 927	23 528	0.6–4.4
	Test Set	100	54	38	686	527	500	
SX	Training Set	21	13	60	543	257	627	2.5–4.0
	Test Set	20	12	59	533	189	809	
DB	Training Set	169	169	87	6060	6205	6494	2.2–3.0
	Test Set	19	19	10	106	50	141	

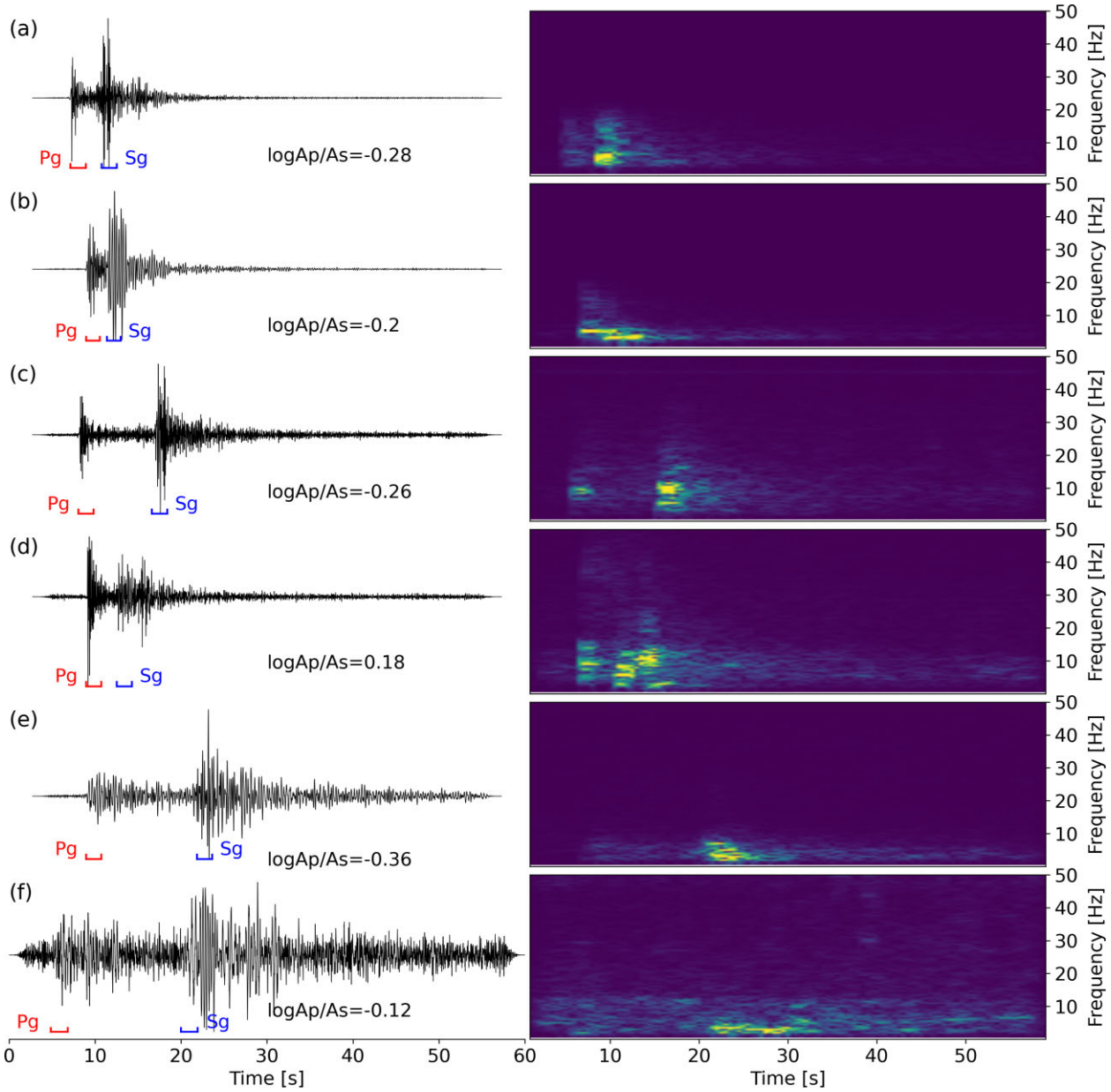

Figure 5. Waveform and spectrogram of the *Z*-component: (a) Example of a M_L 1.2 earthquake in the CC data set. (b) Example of a M_L 0.93 explosion in CC data set. (c) Example of a M_L 2.3 earthquake in the DT data set. (d) Example of a M_L 2.0 explosion in the DT data set. (e) Example of a M_L 3.1 collapse in the DT data set. (f) Example of a M_L 3.4 collapse in the DT data set. The waveforms correspond to the normalized raw data. The spectrograms are normalized by their maximum. Pg indicates the *P*-wave amplitude calculation window and Sg indicates the *S*-wave amplitude calculation window.

When to transfer: earthquake classification models should be fine-tuned using transfer learning when applied to other regions rather than pursuing a generalized model (Zhu *et al.* 2023). Taking the EQTypeNet used in this paper as an example, When applying a pre-trained EQTypeNet on the CC data set to fit the target domain DT data set, the convolutional layers should be partially frozen to achieve better classification after transfer learning. Thus, the question of when to transfer becomes a matter of selecting which convolutional layers to freeze. The distinct regional characteristics of seismic events make negative transfer likely in transfer learning across regions. Deep layers of a CNN learn more local features, while shallow layers extract features that tend to be global. Therefore, freezing only the shallow convolutional layers is more beneficial for the model's application on the target task. The results of our tests bear this out. We kept all other parameters consistent, unfreezing started from the fourth convolutional layer of the waveform or spectrogram branch. After multiple tests, the results show that the optimal transfer learning outcome was achieved by freezing the first two convolutional layers of the waveform and spectrogram branches (Fig. 6).

4 EQTypeNet TRAINING AND TESTING

4.1 Earthquakes and explosions binary classification model training and testing

4.1.1 Model training

In all training processes, we use the SparseCategoricalCross entropy function, which is commonly used in classification tasks, as the loss

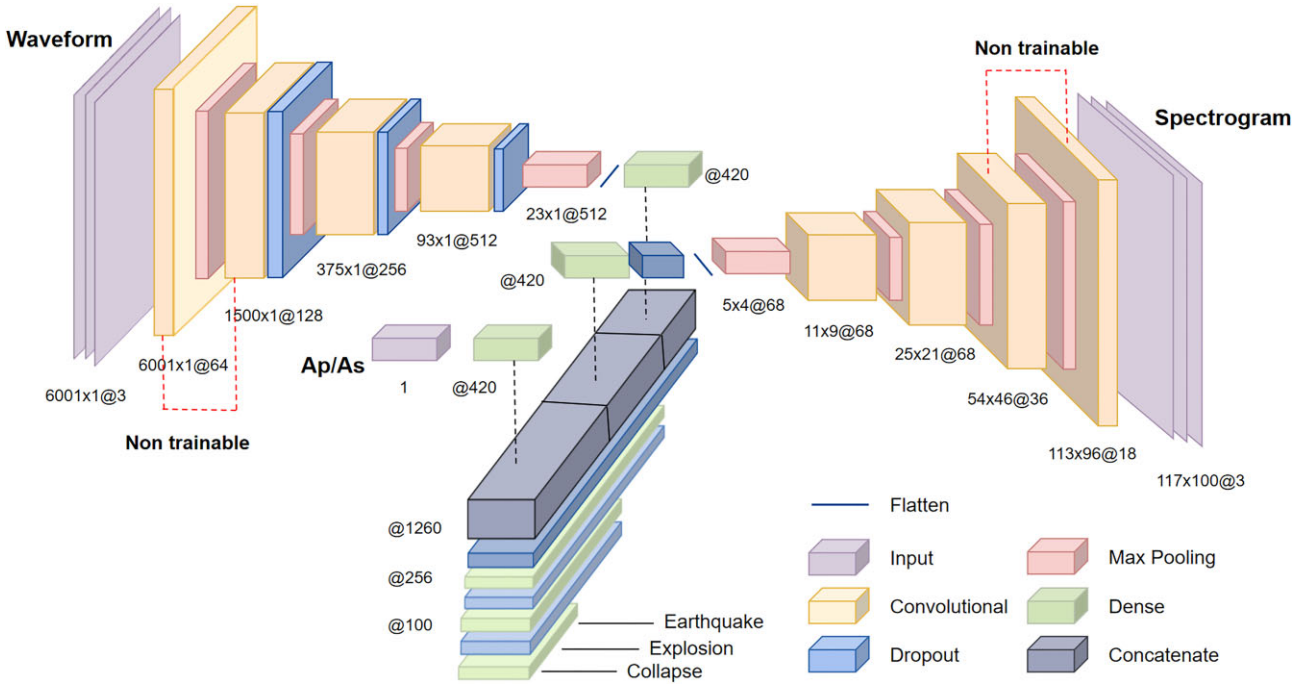

Figure 6. EQTypeNet architecture. The convolutional layers in the figure are all 2-D convolutional layers. The waveform branch employs convolution kernels sized (3,1) with corresponding pooling kernels of (4,1), with a dropout rate of 0.2. The spectrogram branch employs convolution kernels of sizes (5,5), (3,3), (3,3) and (2,2), along with a pooling kernel size of (2,2), and with a dropout rate of 0.2. In the dense layers, a higher dropout rate of 0.3 is applied.

Table 2. Test results of different branches of the CNN model on the CC data set: Where Ap/As denotes *P*/*S* ratio branch; waveform denotes waveform branch; spectrogram denotes spectrogram branch; and F1 denotes macroF1. Eq. and Ep. denote earthquakes and explosions. Pr and re denote precision and recall.

No. of data branches	Data input	Eq. pr	Ep. pr	Eq. re	Ep. re	F1
1	Waveform	97.2 per cent	91.2 per cent	93.6 per cent	96.0 per cent	0.94
2	Spectrogram	98.5 per cent	92.3 per cent	94.4 per cent	97.9 per cent	0.96
3	Waveform + Ap/As	97.6 per cent	91.1 per cent	93.5 per cent	96.6 per cent	0.95
4	Spectrogram + Ap/As	98.7 per cent	92.9 per cent	94.9 per cent	98.2 per cent	0.96
5	Waveform + Spectrogram	**99.2 per cent**	93.6 per cent	95.3%	**98.9 per cent**	**0.97**
6	Waveform + Spectrogram + Ap/As	99.0 per cent	**93.8 per cent**	**95.5%**	98.6 per cent	**0.97**

function. To reduce the model's sensitivity to changes in the learning rate, the Adam optimizer was used with the initial learning rate set to 0.001. To improve the model generalization ability (Hoffer *et al.* 2017), we monitored the accuracy of the validation set before and after each iteration during training and timely adjusted the learning rate. If accuracy no longer increases at the 30th iteration of training, halve the learning rate. At the same time, to alleviate overfitting, if accuracy no longer increases at the 50th iteration, training is stopped. The model with the highest accuracy in all iterations was used as the best model. For the stability of test results, we produced five test sets using the data enhancement scheme of Section 2.4 for all tests. The final metrics for evaluating the model are the average of the precision, recall, and F1, etc., of the five test sets. We trained and evaluated the classification effect of different branch combinations on the CC data set. Since the temporal division of the test set by events, the data were not balanced in terms of waveform records (Table 1). To simulate actual seismic classification, we did not balance the CC test set. Finally, average model performances are shown in Table 2.

The model classification results show that: (1) Spectrogram branch achieves better results compared to waveform branch. (2) The combination of *P*/*S* ratios with waveforms or spectrograms can both improve model result slightly. The extent of the improvement is smaller when the *P*/*S* ratio is combined with the spectrogram and waveform branches. (3) The combinations of branches are better at detecting earthquakes, and the identification of explosions exhibits low precision and high recall. When waveform, spectrogram and *P*/*S* ratio are inputs, the tri-branch CNN model achieves the best classification performance. And its F1-score of earthquakes is slightly higher than the waveform and spectrogram branch model. Therefore, we choose the EQTypeNet as the training model for earthquakes and explosions classification.

4.1.2 Generalizability test

Currently, different study areas do not have a uniform standard for seismic event classification. The differences in the criteria for selecting data labels across various regions may lead to corresponding impacts on the application of the EQTypeNet. In order to rationally transfer the pre-trained model on the United States West Coast CC data set to China, we compared the similarity of the data. The 50 iteration CC or DT spectrogram models were trained on spectrograms of the CC or DT training sets. We applied them and the France spectrogram model, which Hourcade *et al.* (2022) trained

Table 3. Generalizability of model trained on the French (Hourcade *et al.* 2022), CC or DT data sets: Since the French model has only spectrograms as input, all tests are results of the spectrogram branch (Table 2, Combination 2), epoch = 50. Eq.F1 and Ep.F1 denote earthquakes F1 and explosions F1. F1 denotes macroF1.

Training set	DT test set			CC test set		
	Eq.F1	Ep.F1	F1	Eq.F1	Ep.F1	F1
France	0.71	0.73	0.72	0.64	0.72	0.68
DT	0.90	0.87	0.89	0.65	0.68	0.67
CC	0.67	0.69	0.68	0.95	0.94	0.95

Table 4. Model generalization tests: 5 indicates that the data set is a waveform and spectrogram combination (Table 2, Combination 5), and 6 indicates that the data set is the addition of *P*/*S* ratio data to Combination 5 (Table 2, Combination 6). Eq. and Ep. denote earthquakes and explosions. Pr and re denote precision and recall. F1 denotes macroF1.

Training set/Test set	Eq. pr	Ep. pr	Eq. re	Ep. re	F1
CC (5)/DT	80.8 per cent	57.6 per cent	52.7 per cent	83.7 per cent	0.66
CC(6)/DT	85.5 per cent	60.9 per cent	57.0 per cent	87.4 per cent	0.70
DT(5)/CC	90.6 per cent	55.7 per cent	49.2 per cent	92.6 per cent	0.67
DT(6)/CC	93.8 per cent	69.1 per cent	71.4 per cent	93.1 per cent	0.80

with France data, to test their generalizability in the Pacific Northwest and China (Table 3). The France data is more similar to CC in terms of magnitude and epicentre distance distribution.

From the results of the generalization test, we can find: (1) When the French spectrogram model is generalized directly to the Pacific Northwest and China, compared to the CC test set, it achieves better earthquake identification in the DT test set, which generally has higher magnitudes. (2) The CC spectrogram model is slightly less effective than the French spectrogram model in detecting the DT test set. Active seismicity in the oceanic crust subduction zone of the northwestern Pacific Ocean may differ in terms of focal mechanisms and tectonic settings from that concentrated in the southern mountains of the French region (Marin 2004) and in the Chinese region. This difference may have a greater impact on event features than the magnitude. (3) Different regionally trained spectrogram models' generalization results of explosions are all higher than earthquakes. In addition, the French spectrogram model applied to the CC or DT test sets has almost the same explosion identification. Explosions may be less influenced by geological factors than earthquakes, which is consistent with the results of the Tang *et al.* (2020) model generalization. Differences in the data itself may have a greater impact on model performance than differences in the standard classification of different regions. Therefore, we believe that it is feasible to transfer the pre-trained models from other regions to the DT data set in China to fine-tune them for better classification.

This study aims to develop an earthquake classification model for China and realize its applications in small regions. Since the training set covers the event distribution region of the test set, there are no conditions for the generalization test in the strict sense. However, the fact that China is a vast area with a complex geological and tectonic background means that when good generalization features are input to the EQTypeNet, it will be beneficial for the classification of small areas within this study area. Although the *P*/*S* ratio features showed good generalizability in the test of Kong *et al.* (2022), their contribution to enhancing the classification performance of the tri-branch model was less significant on the CC test set. To explore the generalization of the *P*/*S* ratio features, we trained a two-branch (waveform and spectrogram) model and a EQTypeNet for the Chinese region in the DT training set. These models, along with those trained on the CC training set in Section 4.1.1, were tested for generalization on the respective CC or DT test sets. The test results show that the use of *P*/*S* ratio features significantly improves the generalization of the models (Table 4). The results of generalizing from the CC training set to the DT test set indicated the *P*/*S* ratio features improved the F1-score by 0.04. This increase was further extended in the generalization from the DT training set to the CC test set, with an F1-score increase of 0.13. Specifically analysing the identification of earthquakes and explosions using *P*/*S* ratios, we found that the *P*/*S* ratios improve the F1-score of the model's generalization test mainly by improving the model's identification of earthquakes. Whereas the difference between the *P*/*S* ratios of the CC and DT data sets on the performance improvement of the EQTypeNet may be due to the fact that the data of DT is more diverse, which potentially improves the generalizability of the EQTypeNet.

4.1.3 Transfer learning

Although the generalization of the French spectrogram model performs better in China, high-resolution features may be lost in its spectrogram inputs due to the spectral range of the French data from 1 to 50 Hz (with a resolution of 1 Hz). Therefore, we still choose the CC data set as the source domain in transfer learning, expecting to improve EQTypeNet's classification performance on the DT data set and its sensitivity to small magnitude events ($M_L \leq 2$) in China. The classification results of the non-transfer learning (EQTypeNet trained with DT in Section 4.1.2) and transfer learning EQTypeNets are shown in Tables 5 and 6. The four test sets used for testing are: DT, a DT earthquakes and explosions test set (Table 1); DT$*$, a DT test set in which human intervention has resulted in a large proportion of explosions (ratio of number of events of each type: Eq:Ep = 1:5); DT ($M_L \leq 2$), a test set consisting of events with magnitude less than or equal to 2 in DT. There are 28 earthquakes (127 waveform records) and 3 explosions (4 waveform records); DT ($M_L \leq 2$)$**$, a test set of 4 earthquake and explosion waveform records in DT ($M_L \leq 2$), constituted a balanced data set.

However, the transfer learning model, which has an F1-score of 0.94 on the DT test set (Table 6, Line 1), does not perform better than the model without transfer (Table 5, Line 1) and only accelerated the training speed. Combining the results of the poor generalization of the CC and DT spectrogram models to the DT or CC test sets in Table 3 with the curves of the model accuracies in the training phase (Fig. 7) and the performance of the models in the involved human intervention data set, we analyse that:

Table 5. Test results of the non-transfer learning EQTypeNet on the DT data set: ∗ indicates that this test set involved human intervention and ∗∗ indicates that this test set involved data balancing. Eq. and Ep. denote earthquakes and explosions. Pr and re denote precision and recall. Eq.F1 and Ep.F1 denote earthquakes F1 and explosions F1. F1 denotes macroF1.

Test Set	Eq. pr	Ep. pr	Eq. re	Ep. re	Eq. F1	Ep. F1	F1
DT	95.7 per cent	90.6 per cent	92.5 per cent	94.6 per cent	0.94	0.93	0.94
DT∗	76.4 per cent	98.5 per cent	92.2 per cent	94.6 per cent	0.84	0.97	0.91
DT ($ML \leq 2$)	100 per cent	38.0 per cent	94.8 per cent	100 per cent	0.97	0.55	0.76
DT ($ML \leq 2$)∗∗	100 per cent	96.0 per cent	95.0 per cent	100 per cent	0.97	0.98	0.98

Table 6. Test results of the transfer learning EQTypeNet on the DT data set: ∗ indicates that this test set involved human intervention and ∗∗ indicates that this test set involved data balancing. Eq. and Ep. denote earthquakes and explosions. Pr and re denote precision and recall. Eq.F1 and Ep.F1 denote earthquakes F1 and explosions F1. F1 denotes macroF1.

Test Set	Eq. pr	Ep. pr	Eq. re	Ep. re	Eq. F1	Ep. F1	F1
DT	96.1 per cent	90.1 per cent	92.0 per cent	95.2 per cent	0.94	0.93	0.94
DT∗	79.2 per cent	98.4 per cent	91.8 per cent	95.4 per cent	0.85	0.97	0.91
DT ($ML \leq 2$)	99.8 per cent	31.1 per cent	93.4 per cent	95.0 per cent	0.97	0.47	0.72
DT ($ML \leq 2$)∗∗	100 per cent	100 per cent	100 per cent	100 per cent	1	1	1

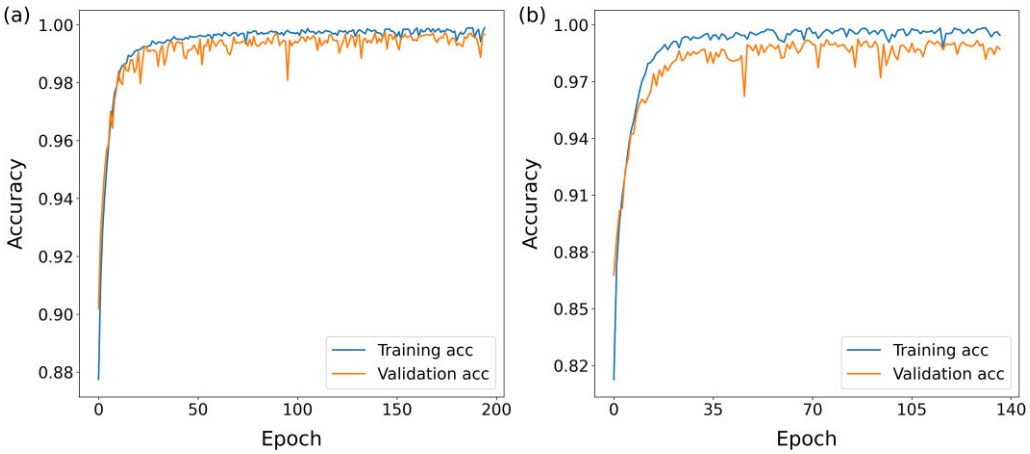

Figure 7. Accuracy curve of model training: (a) Training accuracy curves for the EQTypeNet without transfer learning. (b) Training accuracy curves for the EQTypeNet incorporating transfer learning.

(1) During the transfer process, earthquakes from the CC training set may have caused the model to learn features that are disturbing relative to the DT earthquakes at the frozen layer. In Fig. 7, the initial accuracy of the model incorporating transfer learning is lower than the model without transfer learning. To further test the impact of earthquake events on transfer, we evaluated the performance of the model incorporating transfer learning on the DT∗ test set. When the number of earthquakes in the test set decreases, the transfer learning model identifies earthquakes better compared to the non-transfer learning model.

(2) Transfer learning may be affected by the scope of the data region, where the CC source domain data features differ significantly from the target domain data. Despite the CC data area having strong seismicity and a complex geologic background, the range of this data set (114°–117°W, 41°–51°N) is much smaller than the DT data set (65°–140°E, 15°–50°N), and its data diversity may be weaker than the DT data set.

(3) The size of the source domain data may affect the performance of the transfer learning model. The CC data set is only about twice as large as the DT data set, and compared to the transfer study of Zhu *et al.* (2023), the difference between source and target domain data is less than an order of magnitude. The result of the models with the same framework in the CC test set (Table 2) is significantly higher than in the DT test set. Considering the difference in data volume, increasing data set size might improve the performance of the model combined with transfer learning.

To test the identified ability of the transfer learning model to small magnitude events, we evaluated the model on the DT ($M_L \leq 2$) test set (Table 6, Line 3). The evaluation results exhibit a certain degree of randomness caused by the limited amount of explosion data in the DT ($M_L \leq 2$) test set. The transfer learning model detected fewer explosions than the non-transfer learning model in the small magnitude test set (Table 5, Line 3; Table 6, Line 3). From the confusion matrix of the two models (Fig. 8), all four explosion records were recognized, but 9 or 7 records out of 127 earthquakes were misidentified as explosions. Considering the influence of data imbalance, the transfer learning model significantly improved performance on data balanced DT ($M_L \leq 2$)∗∗ (Table 5, Line 4; Table 6, Line 4)

4.2 Earthquakes, explosions and collapses ternary classification models

The EQTypeNet trained on the DT data set has an F1-score of 0.94 on the DT earthquakes and explosions test set (Table 5, Line 1), and the generalizability of the *P*/*S* ratio branch has also been confirmed.

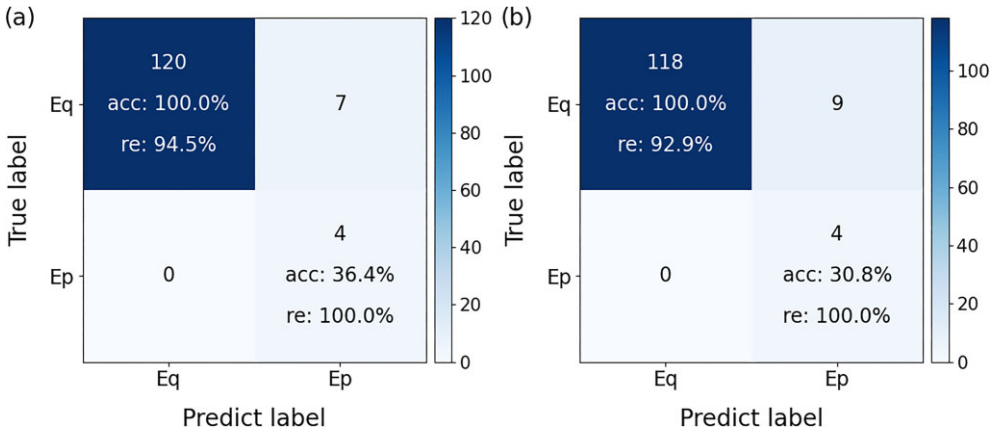

Figure 8. Confusion matrix for model classification test of small-magnitude events: (a) Confusion matrix for the EQTypeNet without transfer learning. (b) Confusion matrix for the EQTypeNet incorporating transfer learning.

Table 7. Classification performance of the EQTypeNet on the DT test set for individual stations and network average: Eq, Ep and Ss denote earthquakes, explosions and collapses. Pr and re denote precision and recall.

Test set	Metric	Eq	Ep	Ss	MacroF1
Station	pr	94.8 per cent	87.8 per cent	87.5 per cent	0.90
	re	92.7 per cent	87.3 per cent	90.6 per cent	
	F1	0.94	0.88	0.89	
Network	pr	97.0 per cent	94.2 per cent	92.3 per cent	0.95
	re	98.0 per cent	90.7 per cent	94.7 per cent	
	F1	0.98	0.93	0.94	

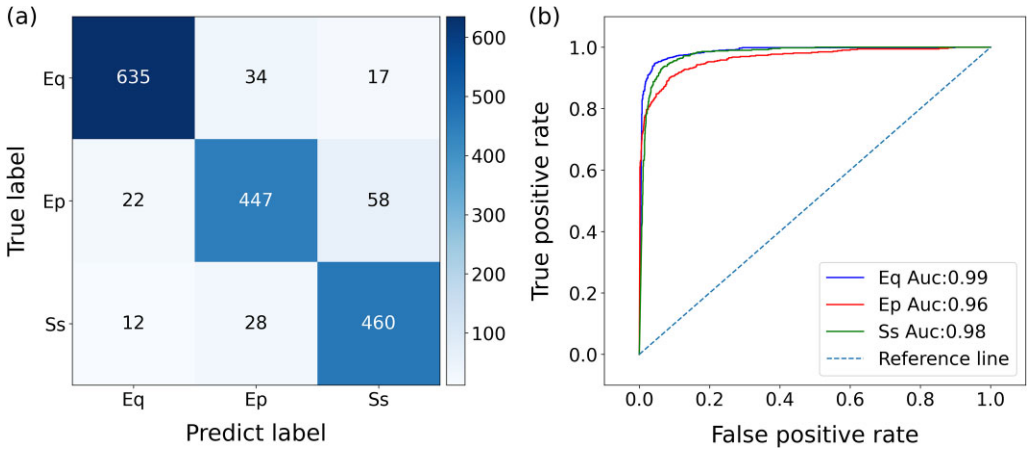

Figure 9. Performance of the ternary classification EQTypeNet based on the DT test set: (a) Confusion Matrix for ternary classification EQTypeNet of earthquakes, explosions and collapses, (b) the ROC curve for the three classification EQTypeNet.

Table 8. The application performance of the EQTypeNet in ShanXi: Eq, Ep and Ss denote earthquakes, explosions and collapses. Pr, re and Acc denote precision, recall and accuracy.

Training set	Metric	Eq	Ep	Ss	Acc	MacroF1
DT	pr	76.6 per cent	24.4 per cent	98.2 per cent	54.3 per cent	0.52
	re	92.7 per cent	93.1 per cent	19.9 per cent		
	F1	0.84	0.39	0.33		
SX	pr	95.0 per cent	69.3 per cent	92.9 per cent	90.5 per cent	0.86
	re	96.6 per cent	75.1 per cent	90.0 per cent		
	F1	0.96	0.72	0.91		
DT and SX	pr	98.9 per cent	96.3 per cent	99.0 per cent	98.6 per cent	0.98
	re	99.6 per cent	96.3 per cent	98.5 per cent		
	F1	0.99	0.96	0.99		

Table 9. The application performance of the EQTypeNet in Northeast China: Eq, Ep and Ss denote earthquakes, explosions and collapses. Pr, re and Acc denote precision, recall and accuracy.

Training set	Metric	Eq	Ep	Ss	Acc	MacroF1
DT	pr	93.6 per cent	27.2 per cent	93.8 per cent	53.9 per cent	0.52
	re	91.3 per cent	96.8 per cent	10.5 per cent		
	F1	0.93	0.43	0.19		
DB	pr	97.6 per cent	92.2 per cent	98.8 per cent	97.2 per cent	0.97
	re	99.1 per cent	98.0 per cent	95.5 per cent		
	F1	0.98	0.95	0.97		
DT and DB	pr	96.2 per cent	92.8 per cent	98.8 per cent	96.8 per cent	0.97
	re	99.2 per cent	98.0 per cent	94.5 per cent		
	F1	0.98	0.95	0.97		

However, our studies performed so far are all binary classification for earthquakes and explosions, which is still far from our goal of ternary classification in China. Although events of a category have some consistent features in waveform and spectrogram, the features of collapses may not be as robust for CNN model classification as earthquakes or explosions, due to the greater regional variability of collapses. The CC data set contains only two categories, earthquakes and explosions, and the transfer learning is only effective in terms of speed. Therefore, we performed ternary classification on the DT data set using the EQTypeNet without transfer learning to evaluate its efficacy in this context.

In the DT test set, the EQTypeNet achieved individual stations precision of 94.8 per cent for earthquakes, 87.8 per cent for explosions and 87.5 per cent for collapses, respectively, and a macroF1 score of 0.90; network precision of 97.0 per cent for earthquakes, 94.2 per cent for explosions, and 92.3 per cent for collapses, respectively, with a macroF1 score of 0.95 (Table 7). Among them, earthquake identification is the best, and explosion identification is slightly lower than collapses (Fig. 9a). Combined with the event distribution of DiTing 2.0 (Fig. 1b), some of the scattered explosions are located in areas where earthquakes are concentrated. Based on an ROC (Receiver Operating Characteristic) curve of the ternary classification, we found that the existence of suitable probability thresholds can successfully distinguish the three classes of events (Fig. 9b).

5 APPLICATION OF THE EQTypeNet TO SHANXI AND NORTHEAST CHINA

To further evaluate the application of the EQTypeNet, we employed the model trained on the DT three-class events in Section 4.2 to the SX and DB test sets. The model was also used as a pre-trained model and fine-tuned using SX and DB training set data to evaluate the performance of transfer learning using Chinese data on small regions in China. To compare the efficacy with and without transfer learning, we re-trained a model, which had the same architecture of the model trained on the DT three-class events, using the SX or DB training sets.

The pre-trained model has a macroF1 of only 0.52 when applied directly on both the SX and DB test sets (Tables 8 and 9, Lines 1–3), and it can only effectively detect earthquake events (Tables 8 and 9, Lines 1–3). Among them, the F1-score of collapses in the DB test set was only 0.19 (Table 9, Line 3), but the F1-score of earthquakes reached 0.93 (Table 9, Line 3). Compared to the CC and the French data set, the relative lack of non-natural events in the DT may have failed to enable the pre-trained model to learn pertinent features of non-natural events throughout China.

The retrained model, based on 1427 waveform records from the SX training set (Table 1), achieves a macroF1 of 0.86 (Table 8, Lines 4–6); based on 18 759 waveform records from the DB training set (Table 1), achieves a macroF1 of 0.97 (Table 9, Lines 4–6). Their classification performance is significantly stronger than the pre-trained model (Tables 8 and 9, Lines 1–3). When we applied transfer learning to the pre-trained model and fine-tuned it using the SX or DB training set, the classification performance of the fine-tuned model significantly improved for SX, with an accuracy of 98.6 per cent and a macroF1 of 0.98 (Table 8, Lines 7–9). However, for DB, the macroF1 of the fine-tuned model remained the same as the re-trained model, with an accuracy of 96.8 per cent (Table 9, Lines 7–9). Combining the size of the SX and DB data sets and the distribution of the data set magnitudes (Table 1) and epicentres (Fig. 3), we found that the EQTypeNet requires relatively small data sizes when it is applied to classify events on a small regional scale with larger magnitudes. The addition of multiple dropout layers to the model framework (Fig. 6) may have improved the model's classification performance on limited-scale data sets with larger magnitudes and a small regional scale (Tables 8 and 9, Lines 4–6). When the SX training set is used for training, the re-trained model achieves a macroF1 of 0.86. When the amount of training data is further increased, such as using the larger DB data set, the re-trained model achieves classification result no less than the fine-tuned model.

6 DISCUSSION

6.1 Affect of *P*/*S* ratio selection on The EQTypeNet's performance

This study shows that the application of the *P*/*S* ratio features significantly improves the generalizability of models. However, in practical applications, previous studies have found that the choice of frequency bands for calculating *P*/*S* ratios can affect the classification effect compared to more primitive data inputs such as waveforms and spectrograms (Pan *et al.* 2007; Wang *et al.* 2020). In addition, the range of the optimum *P*/*S* ratio calculation is also affected by the epicentre distance. Compared to regional distances, when *P*/*S* ratios are applied to local earthquake classification, their effectiveness will experience a significant decrease (O'Rourke *et al.* 2016; Pyle & Walter 2019). Therefore, this paper further discussed the influence of different epicentral distance ranges and *P*/*S* ratio calculation frequency bands on earthquake classification in China.

In this study, all DT waveform data were band-pass filtered at every 2 Hz of 2–24 Hz as the centre frequency (1–3, 3–5, . . . , 23–25 Hz), and *P*/*S* ratios were calculated. Meanwhile, to examine the

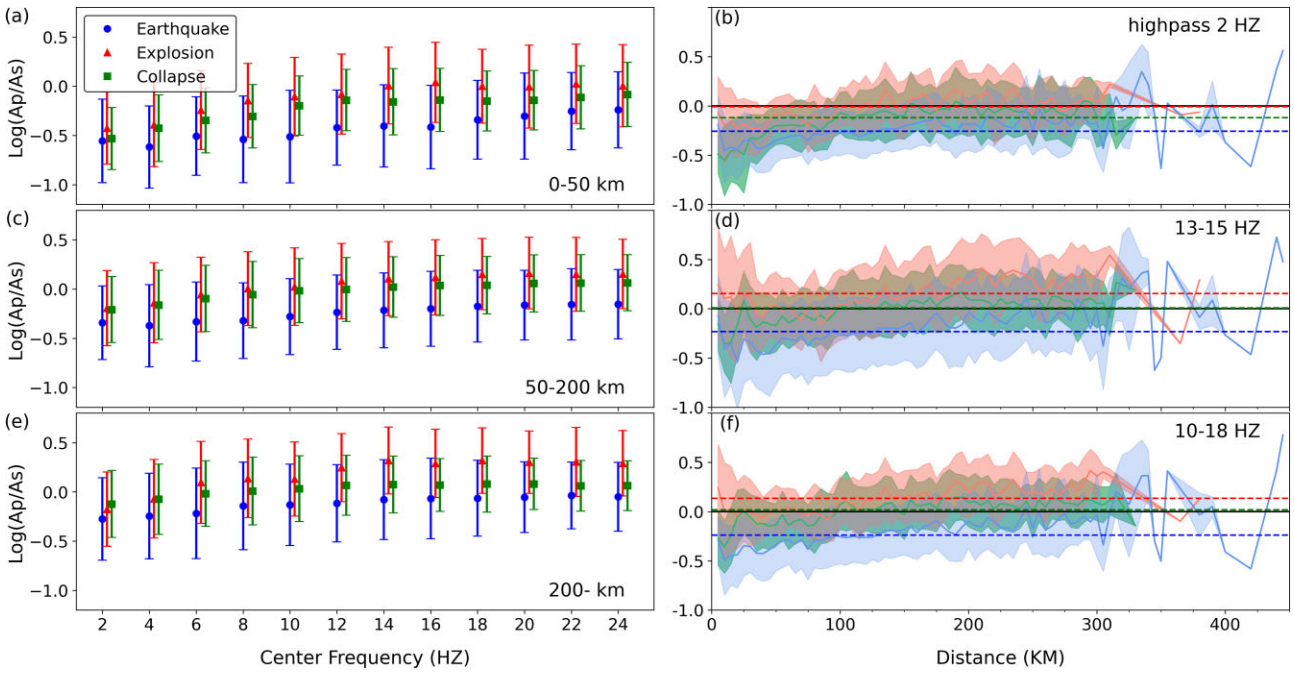

Figure 10. *Z* component *P*/*S* ratios in the DT data set and it as a function of epicentre distance: The mean *P*/*S* ratios are (a) within 50 km of the epicentre, (c) within 50–200 km of the epicentre, (e) outside 200 km of the epicentre. Circles, triangles and squares represent the mean values of earthquakes, explosions and collapses for each frequency band, respectively. Error bars are one standard deviation. The mean *P*/*S* ratios after (b) 2 Hz high-pass filtering, (d) 13–15 Hz band-pass filtering, and (f) 10–18 Hz band-pass filtering. Bold blue, bold red and bold green lines indicate the average *P*/*S* ratios for all earthquakes, explosions and collapses per 5 km range. Dashed blue, dashed red and dashed green lines indicate the average *P*/*S* ratios for all earthquakes, explosions and collapses. The shaded area is a range of standard deviations.

effect of epicenter distance on the *P*/*S* ratio, we counted the mean and standard deviation of the *P*/*S* ratio in different frequency bands according to the range of epicentre distance (Figs 10a, c and e). It can be seen that the mean values of *P*/*S* ratios of all frequency bands with epicentre distances at local (within 50 km, 50–200 km) and regional (200–1500 km) distances are able to distinguish between earthquakes, explosions and collapses. Among them, the *P*/*S* ratio of earthquakes is the lowest, that of explosions is higher, and collapses is slightly lower than explosions, which is consistent with the results of Wang *et al.* (2023). The statistics of the mean *P*/*S* ratio for each frequency band also show that the three classes of events are best categorized in the 13–15 Hz band. However, whether epicentre distances are regional or local, the standard deviations of the *P*/*S* ratios for the three classes of events cross and overlap with each other, and the three classes of events are not separated within a single standard deviation. The large standard deviation of *P*/*S* ratios in the DT data set at local distances (Figs 10a and c) is similar to the study results of Walter *et al.* (2014), who used data from experimental blasting of source physics to classify events by P/S ratio of local distances. At the regional distances, the large standard deviation of *P*/*S* ratio (Fig. 10e), which may be mainly caused by the complex tectonic environment. Since the DiTing 2.0 data set covers the entire China, the strong tectonic activity makes the focal mechanisms in each region relatively complex, which may lead to large differences in the *P*/*S* ratios of earthquakes with different mechanisms. In addition, factors such as the existence of fluids can cause differences in the *P*/*S* ratio characteristics among events within the same category (Kim *et al.* 1994; Pyle & Walter 2019; Zhang *et al.* 2020). In order to verify the influence of event distribution ranges on the *P*/*S* ratio, we reduced the event distribution range. In Shanxi and Northeast China, we found that: at regional distances, the individual deviations of *P*/*S* ratios in the SX data set are similar to those in the DT data set (Fig. 11e), which may be related to the complex tectonic environment of the Shanxi area; but in the DB data set, the *P*/*S* ratios distinguish well between earthquakes and explosions, possibly related to the relatively homogeneous tectonic setting of Northeast China (Fig. 11f). In addition, the relatively large data volume was also an important reason for the large standard deviation of *P*/*S* ratios in the DT data set.

To reduce the impact of artificially selected frequency bands on the calculation of *P*/*S* ratios, *P*/*S* ratios used for model training should be filtered over a broader frequency band (Wang *et al.* 2020). Figs 10a, c and e show that the *P*/*S* ratios at 13–15 Hz provide the best classification for three classes of events. Because this band is narrower, we further examined the distribution of the *P*/*S* ratio at 10–18 Hz as a function of epicentral distance (Fig. 10f). The classification results for the *P*/*S* ratios at 10–18 Hz (Fig. 10f) and 13–15 Hz (Fig. 10d) are similar for three types of events, and both are slightly higher than the *P*/*S* ratios of the DT data set based on the 2 Hz high-pass filtering. Therefore, we calculated the *P*/*S* ratios at 10–18 Hz and used them along with the corresponding waveforms and spectrograms, which were obtained from the pre-processings described in Section 2.4, to form a new DT (10–18 Hz) data set. The new data set was divided into training and test sets in a 9:1 ratio. We retrained the EQTypeNet using the DT (10–18 Hz) training set, and the network performance obtained by the re-trained model on the DT test set is consistent with the original model (Table 10). Inconsistent *P*/*S* ratio calculation bands in the training and testing sets may affect the final results, we evaluated the performance of the re-trained model again on the DT (10–18 Hz) test set. However,

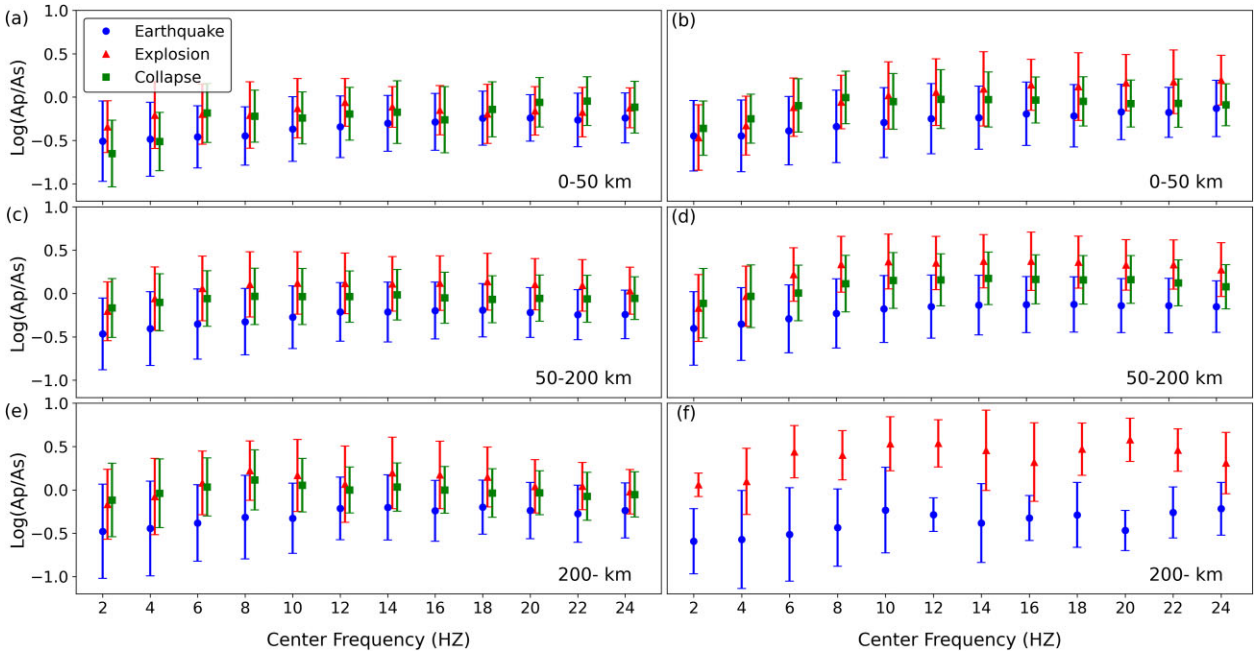

Figure 11. *Z* component *P*/*S* ratios in the SX data set and the DB data set: The mean *P*/*S* ratios at the SX data set are (a) within 50 km of the epicentre, (c) within 50–200 km of the epicentre, (e) outside 200 km of the epicentre. The mean *P*/*S* ratios at the DB data set are (b) within 50 km of the epicentre, (d) within 50–200 km of the epicentre, (f) outside 200 km of the epicentre. Circles, triangles and squares represent the mean values of earthquakes, explosions and collapses for each frequency band, respectively. Error bars are one standard deviation.

Table 10. The network classification performance of EQTypeNet s trained with *P*/*S* ratios from different filter bands and corresponding waveforms and spectrograms on the DT test set: Eq.F1, Ep.F1 and Ss.F1 denote earthquakes F1, explosions F1 and collapses F1. F1 denotes macroF1.

Training set/Test set	Eq. F1	Ep. F1	Ss. F1	F1
DT/DT	0.98	0.93	0.94	0.95
DT(10–18 Hz)/DT	0.98	0.93	0.94	0.95
DT(10–18 Hz)/DT(10–18 Hz)	0.98	0.92	0.92	0.94

the detection of explosions and collapses by the re-trained model decreased compared to the DT test set, and the macroF1 of network classification decreased by 0.01. The average *P*/*S* ratio separation degree between collapses and explosions in the DT (10–18 Hz) data set is similar to that of the DT data set, and the average *P*/*S* ratio separation degree of earthquakes from collapses and explosions is about two to three times higher than that of the DT data set (Figs 10b and f). However, as shown in Table 10, the re-trained model shows no improvement in the detection of seismic events compared to the original EQTypeNet on both the DT (10–18 Hz) and DT test sets. Our tests indicate that differences in filter bands do not significantly affect the classification results of the EQTypeNet.

6.2 Unified features of cross regional events

In Section 4.1.2, we speculated that geological factors have a relatively weak influence on explosions. As increasing the train data volume improves the classification performance of DL models, we expect that a larger data set will enhance our EQTypeNet's application in classifying earthquakes and explosions. To this end, we produced a cross regional explosion event data set incorporating the ComCat and DiTing 2.0 data sets, which includes all explosions from both data sets (Table 1). Based on the magnitude distribution ratio of the explosions, we selected the same number of earthquakes in the DiTing 2.0 data set. We divided the data set into training and test sets in a ratio of 9:1 according to the number of events. The training set was subjected to data balancing and had a total of 157 254 waveform records. In order to simulate the practical application, the test set was not data balanced and consisted of 2225 earthquake waveform records and 1364 explosion waveform records. Since the final target classification area of this paper is China, all explosions for this test set are from the DiTing 2.0 data set. The EQTypeNet trained using this data set achieved the following improvements in individual station metrics compared to the model trained on the DT data set (Table 5): The F1-score of earthquakes increased from 0.94 to 0.99, with 100 per cent precision and 97.9 per cent recall; for explosions, the F1-score improved from 0.93 to 0.98, with 96.8 per cent precision and 100 per cent recall; Additionally, the macroF1 reached as high as 0.99, with a accuracy of 98.7 per cent. The significant improvement in the model's classification performance on this data set confirms that explosions have uniform features across regions.

We further explored whether there are uniform features of collapses across regions. There are sufficient collapses in the DB and DT data sets in Table 1, and we added all the collapses in DT to the DB training set. To ensure that the model can focus on the cross-regional features of collapses, the DB training set also incorporates about 5337 earthquake records from the Northeast region of the DiTing 2.0 data set. The data-enhanced DB earthquake and collapse event training set has a total of 64 403 waveform records, and the test set consists of earthquakes and collapses from the DB test set. The EQTypeNet trained using this data set has the earthquakes F1-score of 0.99, with 98.8 per cent precision and 100 per cent recall; for collapses, the F1-score of 0.99, with 100 per cent precision and 99.1 per cent recall; Additionally, the macroF1 reached as high as 0.99, with a accuracy of 99.5 per cent. Based on the results alone,

Table 11. Classification results using the EQTypeNet, 1D CNN-LSTM, AlexNet architecture and ResNet-18 to classify in three training sets: Acc and F1 denote accuracy and macroF1.

Model	Feature	Training1		Training2		Training3	
		Acc	F1	Acc	F1	Acc	F1
EQTypeNet	WaveformSpectrogram	83.8 per cent	0.83	**86.8 per cent**	0.86	88.0 per cent	0.87
1D CNN-LSTM	Waveform	**84.2 per cent**	0.83	86.7 per cent	0.86	87.7 per cent	0.87
AlexNet	Spectrogram	83.6 per cent	0.83	85.1 per cent	0.85	**88.2 per cent**	0.88
ResNet-18	Waveform	71.6 per cent	0.71	73.5 per cent	0.73	79.7 per cent	0.79
ResNet-18	Spectrogram	81.7 per cent	0.81	86.2 per cent	0.86	87.0 per cent	0.86

the cross-regional characteristics of collapses seem to be stronger than explosions. Data source specific analysis. The two source zones of the explosion data set span very large, while the two source zones of the collapse data set are containment relationships, which may be the reason why the model works better on the collapse data set. On the other hand, since the data set labels were chosen artificially, it cannot be ruled out that the event discrimination criteria may have had an influence in the process of merging the data sets from different countries.

To compare the results of the Northeast China classification application in Section 5, we added explosions to that data set in the same manner as collapses, with explosions from both the DT and DB data sets. The EQTypeNet trained on the three types of data sets has the earthquakes F1-score of 0.96, with 94.9 per cent precision and 97.9 per cent recall; for explosions, the F1-score of 0.95, with 91.4 per cent precision and 98 per cent recall; for collapses, the F1-score of 0.97, with 99.3 per cent precision and 94.3 per cent recall; Additionally, the macroF1 reached 0.96 and the classification accuracy of 96.2 per cent. These results are slightly worse than transfer learning compared to Table 9. The classification effect of the EQTypeNet on the cross-region explosion and collapse data set indicates that the explosions and collapses have uniform characteristics across regions. Considering the time cost and classification performance, the EQTypeNet combined with transfer learning can be better applied to triple classification of earthquakes, explosions and collapses in small regions. And in region-specific binary classification, such as explosion and earthquake, increasing data set diversity and size might be more effective than transfer learning.

6.3 Comparison with previous methods

Liu *et al.* (2020) proposed CNNs to distinguish between tectonic and non-tectonic (Mainly for explosion and collapse) seismicity across most of China. Their data set comprised 3799 tectonic earthquake events and 1043 non-tectonic seismic events, with magnitudes ranging between M_L 2.5–4.7. They reported an individual station macroF1 score of 0.90 and a network macroF1 score of 0.93 for their training set. Comparatively, our EQTypeNet, trained on a more limited data set, achieves individual stations 0.94 macroF1 score (Table 5, Line 1) on a much larger region scale and smaller magnitudes between earthquakes and explosions. And in the triple categorization of explosions, collapses and earthquakes, EQTypeNet achieves individual stations 0.90 macroF1 score and network 0.95 macroF1 score (Table 7).

Wang *et al.* (2023) developed an XGBoost model trained on earthquake, explosion and collapse data from North China. The accuracies of XGBoost with 36-D features data set are 96.41 per cent for earthquakes, 90.38 per cent for explosions, and 94.04 per cent for mining-induced earthquakes. The data in the DB data set is distributed in North China, and EQTypeNet, trained on the DB, achieves a classification accuracy of 97.2 per cent (Table 7, Line 4–6). Since XGBoost uses a one-vs-rest strategy for triple classification, we cannot directly compare the performance of the two models. Numerically, both EQTypeNet and XGBoost obtain good results. Comparatively, EQTypeNet is simpler to train and can be directly triple classified without extracting features artificially.

EQTypeNet is built on the framework of CNN models, and we attempted to compare it with several common DL methods for earthquake classification. Due to the difference in data, we directly chose AlexNet (Krizhevsky *et al.* 2012), ResNet-18 (He *et al.* 2016) and LSTM (Long Short-Term Memory) for comparison. Since AlexNet is not suitable for the waveform input of $6001 \times 1 \times 3$, we only tested the model's effectiveness in classifying spectral data. In the test of waveform data classification, the waveform dimension was changed to 6001×3 to satisfy the LSTM input. Due to the long time-series, we added two 1D convolutional layers (with a kernel size of 3 and 128 and 64 neurons, respectively) before the two LSTM layers (with 128 and 64 neurons, respectively) and one dense layer (with 32 neurons). The same hyperparameters such as 0.001 learning rate etc. were used for all models. We added dropout layers and L2 regularization appropriately to mitigate overfitting. Training sets containing 10 000, 20 000 and 50 000 samples (including earthquakes, explosions and collapses) were created, namely training1, training2 and training3. These training data sets come from a random selection of DT training set, and test set is DT test set.

The results in Table 11 show that the model performance increases with the size of the data set. EQTypeNet, which incorporates both waveform and spectrogram branches, achieved the best results compared to the other models trained with waveform or spectrogram, but only when trained on training2. We believe that the performance of EQTypeNet improves with the addition of amplitude features (Table 7), but the sensitivity of 1-D CNN-LSTM to waveforms deserves attention. The classification results of ResNet-18 are not outstanding, and its convolutional kernel size may not be suitable for our input data. ResNet-18 is better suited for training image data, even though we adjusted the size of the convolutional kernel accordingly, for example, from (7,7) to (7,1), the classification results of ResNet-18 trained with waveforms are still poor. On the other hand, high-complexity models may not be suitable for seismic auto-classification tasks in terms of training time considerations. Simpler models such as EQTypeNet and 1D CNN-LSTM obtained close classification results.

The results of training1 confirm the limitations of EQTypeNet. For classification tasks over a wide range of study areas, a data set of more than 20 000 is desirable for EQTypeNet training. The concatenate layer of EQTypeNet enables it to use a diverse data set, but the classification results of AlexNet and 1-D CNN-LSTM raise a question worth considering: Are waveforms and spectrogram mandatory features for a model in any study area? In contrast,

amplitude ratios may provide a greater boost to the results. In future studies, we suggest that more emphasis can be placed on the physical characteristics of the seismic data than on the model architecture. Additionally, since the data set labels are manually labelled, there is a possibility of mislabeling, for which EQTypeNet has no countermeasure. In the future, we will update the classification categories of EQTypeNet to enable its application in real-time event detection.

7 CONCLUSION

The aim of this study is to develop a deep learning ternary classification model of explosions, collapses and earthquakes that is suitable for China to meet the requirements of automatic seismic classification for large-scale data. Based on previous research, we constructed EQTypeNet, a tri-branch CNN model, using waveform, spectrogram, and event *P*/*S* ratio as features for classification and detection. The training data set for the study was obtained after balancing the data on the DiTing 2.0 data set. On the DiTing 2.0 test set, the individual stations macroF1 achieved 0.90 (Table 7, Lines 1–3), with accuracies of 94.8, 87.8 and 87.5 per cent for earthquakes, explosions and collapses (Table 7, Line 1); the network macroF1 achieved 0.95 (Table 7, Line 4), with accuracies of 97.0, 94.2 and 92.3 per cent for earthquakes, explosions and collapses (Table 7, Lines 4–6). At the same time, the *P*/*S* ratio utilized for the model is not sensitive to the calculation frequency band and epicenter distance. In addition, to alleviate the dependence of deep learning on the amount of training data and to better apply the model to different regions in China with obvious differences in regional characteristics, we pre-trained the model on the training data set. The fine-tuned model, based on the pre-trained model, performs well on small regions. Its macroF1 reached 0.98 (Table 8, Lines 7–9) and 0.97 (Table 9, Lines 7–9) on the Shanxi data set and the DongBei data set, with accuracies of 98.6 per cent (Table 8, Lines 7–9) and 96.8 per cent (Table 9, Lines 7–9). In selecting the source domain data for pre-training, we also tested the possibility of using large-scale earthquake and explosion data sets from the Pacific Northwest of the United States. The transferability analysis showed that the explosions were less affected by the regional geological context. After adding ComCat explosions training, the model has an individual stations macroF1 of up to 0.99 and a accuracy of 98.7 per cent on the explosions and earthquake test sets in China. We also explored whether there are cross-regional features of collapse events that can be learned by EQTypeNet. When trained on the Northeast data set with the addition of the DiTing 2.0 collapses and a corresponding number of earthquakes, EQTypeNet achieveed a classification accuracy of 99.5 per cent with a macroF1 as high as 0.99 on the Northeast test set. The model performed well in binary classification across regions, but when the number of cross-region events was increased to three categories, it did not perform as well as transfer learning. In summary, we reach the following conclusions:

(1)The choice of frequency band for *P*/*S* ratio computation does not affect the classification performance of the EQTypeNet when the median of the *P*/*S* ratios of all waveform recordings from the same event is used as the feature of the event *P*/*S* ratio. At this point, the EQTypeNet can be applied directly without considering the optimal calculation band for the *P*/*S* ratio of the applied region.
(2)We suggest considering the use of explosions or collapses from other regions to alleviate the pressure of data imbalance in the application area, particularly in the classification of earthquakes and explosions or earthquakes and collapses. And in the three-classification task, transfer learning may be more effective.
(3)The EQTypeNet developed in this study is also applicable to the classification of small magnitude earthquakes, collapses and explosions, and has shown better applicability in large areas and regions with data imbalance.
(4)Compared with other deep learning models, we believe that more attention should be paid to the physical features of seismic data than to the model architecture. By increasing suitable physical features in the data set, the EQTypeNet is expected to further improve the accuracy and generalization ability of the classification of the three types of events.

ACKNOWLEDGMENTS

We would like to express our sincere gratitude to Dr Marine Denolle and another anonymous reviewer for their invaluable insights and constructive suggestions. Their meticulous reviews significantly enhanced the quality and clarity of our manuscript. The ComCat data set used in this study was collected and produced by Ni *et al.* (2023) from the Department of Earth and Space Sciences at the University of Washington. The DiTing 2.0 data set was provided by Zhao *et al.* (2023) from the Institute of Geophysics, China Earthquake Administration. The ShanXi data set was provided by Zhang Na from the Earthquake Administration of Shanxi Province, China. The DongBei data set was provided by Sun Li (Liang *et al.* 2023) from the China Earthquake Networks Center. This study was jointly funded by the Basic Scientific Research Business Fee Special Project of the Earthquake Prediction Institute of China Earthquake Administration (CEAIEF20230210, CEAIEF20240405) and the National Natural Science Foundation of China (42174066), and supported by the Shanghai Artificial Intelligence Laboratory. All figures in this article were drawn using Matplotlib (Hunter 2007) and GMT 6 (Wessel *et al.* 2019).

DATA AVAILABILITY

The ComCat data set can be downloaded from https://github.com/niyiyu/PNW-ML, and the DiTing 2.0 data set can be downloaded from https://www.esdc.ac.cn/article/137 by registered users of the Earth Science International Data Center. The EQTypeNets for binary classification of earthquakes and explosions in China and ternary classification of earthquakes, explosions and collapses are stored at https://github.com/mfllwz/China-Earthquake-Classification-Model. The model algorithms are written in Python, with data processing based on the Python libraries ObsPy (Krischer *et al.* 2015) and Numpy (Harris *et al.* 2020), and model parameters are iterated using Keras (Chollet et al. 2015).

REFERENCES

Chen, D., Wang, E. & Li, N., 2021. Study on the rupture properties and automatic identification model of micro-earthquakes and blasting events in a coal mine, *Soil Dyn. Earthquake Eng.,* **146,** 106 759, doi:10.1016/j.soildyn.2021.106759.

Chollet, F. *et al.*, 2015. Keras, *GitHub repository,* GitHub. Retrieved from. https://github.com/fchollet/keras (accessed June 27, 2025).

Cornou, C. *et al.*, 2022. Rapid response to the M w 4.9 earthquake of November 11, 2019 in Le Teil, Lower Rhône Valley, France, *C. R. Géosci.,* **353,** 441–463.

Duan, Y., Shen, Y., Canbulat, I., Luo, X. & Si, G., 2021. Classification of clustered microseismic events in a coal mine using machine learning, *J. Rock Mech. Geotech. Eng.,* **13,** 1256–1273.

Fereidoni, A. & Atkinson, G.M., 2017. Discriminating earthquakes from quarry blasts based on ShakeMap ground-motion parameters, *Bull. seism. Soc. Am.,* **107,** 1931–1939.

Gulia, L. & Gasperini, P., 2021. Contamination of frequency–Magnitude slope (*b*-Value) by quarry blasts: an example for Italy, *Seismol. Res. Lett.,* **92,** 3538–3551.

Gulia, L. & Wiemer, S., 2019. Real-time discrimination of earthquake foreshocks and aftershocks, *Nature,* **574,** 193–199.

Hammer, C., Beyreuther, M. & Ohrnberger, M., 2012. A seismic-event spotting system for volcano fast-response systems, *Bull. seism. Soc. Am.,* **102,** 948–960.

Harris, C.R. *et al.*, 2020. Array programming with NumPy, *Nature,* **585,** 357–362.

He, K., Zhang, X., Ren, S. & Sun, J., 2016. Deep residual learning for image recognition, *2016IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Presented at the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* pp. 770–778, IEEE.

Hinton, G.E., Srivastava, N., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R.R., 2012. Improving neural networks by preventing co-adaptation of feature detectors, Retrieved fromhttp://arxiv.org/abs/1207.0580, (accessed June 27, 2025).

Hoffer, E., Hubara, I. & Soudry, D., 2017. Train longer, generalize better: closing the generalization gap in large batch training of neural networks, *Advances in Neural Information Processing Systems,* **Vol. 30**. Curran Associates Inc., Red Hook, NY, USA.

Hourcade, C., Bonnin, M. & Beucler, É., 2022. New CNN-based tool to discriminate anthropogenic from natural low magnitude seismic events, *Geophys. J. Int.,* **232,** 2119–2132.

Hunter, J.D., 2007. Matplotlib: a 2-D graphics environment, *Comput. Sci. Eng.,* **9,** 90–95.

Jiang, J., Stankovic, V., Stankovic, L., Parastatidis, E. & Pytharouli, S., 2023. Microseismic event classification with time-, frequency-, and wavelet-domain convolutional neural networks, *IEEE Trans. Geosci. Remote Sens.,* **61,** 1–14.

Kim, W.-Y., Simpson, D. & Richards, P.G., 1994. High-frequency spectra of regional phases from earthquakes and chemical explosions, *Bull. seism. Soc. Am.,* **84,** 1365–1386.

Kolaj, M., 2018. Discriminating between low-magnitude shallow earthquakes and road construction blasts near Big Salmon River, New Brunswick, Canada, *Seismol. Res. Lett.,* **89,** 1966–1976.

Kong, Q., Wang, R., Walter, W.R., Pyle, M., Koper, K. & Schmandt, B., 2022. Combining deep learning with physics based features in explosion-earthquake discrimination, *Geophys. Res. Lett.,* **49,** e2022GL098645, doi:10.1029/2022GL098645.

Koper, K.D., Burlacu, R., Armstrong, A.D. & Herrmann, R.B., 2024. Classifying small earthquakes, explosions and collapses in the western United States using physics-based features and machine learning, *Geophys. J. Int.,* **239,** 1257–1270.

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), *Bull. seism. Soc. Am.,* **111,** 558–570.

Kortström, J., Uski, M. & Tiira, T., 2016. Automatic classification of seismic events within a regional seismograph network, *Comput. Geosci.,* **87,** 22–30.

Krischer, L., Megies, T., Barsch, R., Beyreuther, M., Lecocq, T., Caudron, C. & Wassermann, J., 2015. ObsPy: a bridge for seismology into the scientific Python ecosystem, *Comput. Sci. Disc.,* **8,** 014 003, doi:10.1088/1749-4699/8/1/014003.

Krizhevsky, A., Sutskever, I. & Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks, *Adv. Neural Inf. Process. Syst.,* **25,** 1097–1105.

Krizhevsky, A., Sutskever, I. & Hinton, G.E., 2017. ImageNet classification with deep convolutional neural networks, *Commun. ACM,* **60,** 84–90.

Li, J., Stankovic, L., Pytharouli, S. & Stankovic, V., 2020. Automated platform for microseismic signal analysis: denoising, detection, and classification in slope stability studies, *IEEE Trans. Geosci. Remote Sens.,* **59,** 7996–8006.

Liang, H., Sun, L., Chen, S. & Zhi, M., 2023. Research on seismic event classification based on SVM algorithm: an application in Northeast China, *Chin. J. Geophys. (in Chinese),* **66,** 5030–5040.

Liu, X., Ren, T., Chen, H. & Chen, Y., 2020. Classification of tectonic and non-tectonic seismicity based on convolutional neural network, *Geophys. J. Int.,* **224,** 191–198.

Maguire, R., Schmandt, B., Wang, R., Kong, Q. & Sanchez, P., 2024. Generalization of deep-learning models for classification of local distance earthquakes and explosions across various geologic settings, *Seismol. Res. Lett.,* **95,** 2229–2238.

Manley, G.F., Mather, T.A., Pyle, D.M., Clifton, D.A., Rodgers, M., Thompson, G. & Londoño, J.M., 2022. A deep active learning approach to the automatic classification of volcano-seismic events, *Front. Earth Sci.,* **10,** 807 926, doi:10.3389/feart.2022.807926.

Marin, S., 2004. A probabilistic approach to seismic hazard in metropolitan France, *Bull. seism. Soc. Am.,* **94,** 2137–2163.

Mousavi, S.M. & Beroza, G.C., 2023. Machine learning in earthquake seismology, *Annu. Rev. Earth Planet. Sci.,* **51,** 105–129.

Musil, M. & Plešinger, A., 1996. Discrimination between local microearthquakes and quarry blasts by multi-layer perceptrons and Kohonen maps, *Bull. seism. Soc. Am.,* **86,** 1077–1090.

Ni, Y., Hutko, A., Skene, F., Denolle, M., Malone, S., Bodin, P., Hartog, R. & Wright, A., 2023. Curated Pacific Northwest AI-ready seismic dataset, *Seismica,* **2,** doi:10.26443/seismica.v2i1.368.

Oquab, M., Bottou, L., Laptev, I. & Sivic, J., 2014. Learning and transferring mid-level image representations using convolutional neural networks, *2014 IEEE Conference on Computer Vision and Pattern Recognition, Presented at the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* pp. 1717–1724, IEEE.

O'Rourke, C.T. & Baker, G.E., 2017. A spectrogram-based method of *Rg* detection for explosion monitoring, *Bull. seism. Soc. Am.,* **107,** 34–42.

O'Rourke, C.T., Baker, G.E. & Sheehan, A.F., 2016. Using *P* /*S* amplitude ratios for seismic discrimination at local distances, *Bull. seism. Soc. Am.,* **106,** 2320–2331.

Pan, C., Jin, P. & Wang, H., 2007. Applicability of P/S amplitude ratios for the discrimination of low magnitude seismic events, *Acta Seismol. Sin.,* **29,** 521–528.

Pan, S.J. & Yang, Q., 2010. A survey on transfer learning, *IEEE Trans. Knowl. Data Eng.,* **22,** 1345–1359.

Pyle, M.L. & Walter, W.R., 2019. Investigating the effectiveness of P/S amplitude ratios for local distance event discrimination, *Bull. seism. Soc. Am.,* **109,** L14304, doi:10.1785/0120180256.

Rosenstein, M.T., Marx, Z., Kaelbling, L.P. & Dietterich, T.G., 2005. To transfer or not to transfer, *NIPS 2005 Workshop on Transfer Learning,* Vol., **898,** pp. 4.

Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C. & Liu, C., 2018. A survey on deep transfer learning, *Artificial Neural Networks and Machine Learning–ICANN 2018: 27th International Conference on Artificial Neural Networks, Rhodes, Greece, October 4–7, 2018, Proceedings, Part III 27,* pp. 270–279, Springer.

Tang, L., Zhang, M. & Wen, L., 2020. Support vector machine classification of seismic events in the Tianshan orogenic belt, *J. Geophys. Res. Solid Earth,* **125,** e2019JB018132, doi:10.1029/2019JB018132.

Titos, M., Gutiérrez, L., Benítez, C., Rey Devesa, P., Koulakov, I. & Ibáñez, J.M., 2023. Multi-station volcano tectonic earthquake monitoring based on transfer learning, *Front. Earth Sci.,* **11,** 1204 832. doi:10.3389/feart.2023.1204832.

Walter, W., Pyle, M., Ford, S. & Pasyanos, M., 2014. Local distance application of P/S methods of discriminating explosions from earthquakes—How low can we go?, *Presented at Seismol. Soc. Am. Annual Meeting, Anchorage, Alaska, 30 April 2014*.

Walter, W.R., Dodge, D.A., Ichinose, G., Myers, S.C., Pasyanos, M.E. & Ford, S.R., 2018. Body-wave methods of distinguishing between explosions, collapses, and earthquakes: application to recent events in North Korea, *Seismol. Res. Lett.,* **89,** 2131–2138.

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, *Geophys. Res. Lett.,* **48,** e2021GL095721, doi:10.1029/2021GL095721.

Wang, R., Schmandt, B. & Kiser, E., 2020. Seismic discrimination of controlled explosions and earthquakes near Mount St. Helens using *P* /*S* ratios, *JGR Solid Earth,* **125,** e2020JB020338, doi:10.1029/2020JB020338.

Wang, T., Bian, Y., Zhang, Y. & Hou, X., 2023. Classification of earthquakes, explosions and mining-induced earthquakes based on XGBoost algorithm, *Comput. Geosci.,* **170,** 105 242, doi:10.1016/j.cageo.2022.105242.

Wei, Y., Yang, Q., Wang, T., Jiang, C. & Bian, Y., 2019. Earthquake and explosion identification based on deep learning residual network model, *Acta Seismol. Sin.,* **41,** 646–657.

Wenner, M., Hibert, C., Van Herwijnen, A., Meier, L. & Walter, F., 2021. Near-real-time automated classification of seismic signals of slope failures with continuous random forests, *Nat. Hazards Earth Syst. Sci.,* **21,** 339–361.

Wessel, P., Luis, J.F., Uieda, L., Scharroo, R., Wobbe, F., Smith, W.H. & Tian, D., 2019. The generic mapping tools version 6, *Geochem. Geophys. Geosyst.,* **20,** 5556–5564.

Yi, G., Wen, X. & Su, Y., 2008. Study on the potential strong-earthquake risk for the eastern boundary of the Sichuan-Yunnan active faulted-block, China, *Chin. J. Geophys. (in Chinese),* **51,** 1719–1725.

Zhang, Y., Wang, T., Su, L., Yang, Y. & Hui, S., 2020. Study on the regional characteristics of the subsidence seismic records in the northern Shaanxi, *Acta Seismol. Sin.,* **42,** 684–696.

Zhao, M., Chen, S., Zhang, B., Scharroo, R., Wobbe, F., Smith, W.H.F. & Tian, D., 2023. Diting dataset 2.0, China Seismograph Network multifunctional large-scale Artificial intelligence training dataset, doi:10.11998/IESDC.SEIS.D03MOD20230002.

Zhao, M., Xiao, Z., Chen, S. & Fang, L., 2023. DiTing: a large-scale Chinese seismic benchmark dataset for artificial intelligence in seismology, *Earthq. Sci.,* **36,** 84–94.

Zhao, Y., Liu, W. & Gao, Y., 1995. Distinguishing earthquake, explosion and mine earthquake in Beijing area, *Seismological and Geomagnetic Observation and Research,* **16,** 48–54.

Zhu, J., Fang, L., Miao, F., Fan, L., Zhang, J. & Li, Z., 2023. Deep learning and transfer learning of earthquake and quarry-blast discrimination: applications to southern California and eastern Kentucky, *Geophys. J. Int.,* **236,** 979–993.