J Seismol (2022) 26:897–912
https://doi.org/10.1007/s10950-022-10109-5

ORIGINAL ARTICLE

Rapid classification of local seismic events using machine learning

Luozhao Jia · Hongfeng Chen · Kang Xing

Received: 19 January 2022 / Accepted: 24 August 2022 / Published online: 19 September 2022

Abstract Regional seismic networks often observe artificially induced seismic events such as blasting and collapses. Misclassified seismic events in the earthquake catalog can therefore interfere with assessments of natural seismic activity. Traditional methods rely on the period, and phase features of seismic waves to determine the nature of seismic events. We designed three seismic event classifiers with reference to convolutional neural network structures such as VGGnet, ResNet, and Inception. The designed classifiers were tested and compared using three-channel seismic full-waveform time-series data and spectral data. Our classifiers are shown to only require 60 s of full-waveform seismic event data and first-arrival times for alignment; additional phase labeling or numerical filtering is unnecessary. Rapid classification of earthquakes, blasting, and mine collapses can be achieved within approximately 1 min of an event. As a test case, this study uses 6.4 k observations of actual local seismic events with magnitudes ranging from M_L 0.6 to M_L 4.5 obtained from 47 broadband seismic stations in the Henan Regional Network of the China Seismological Network Center; these observations include natural earthquakes, blasting, and collapse events. The results indicate that our classifiers can reach a lower classification magnitude limit of M_L 0.6 and that their recall and accuracy exceed 90%, outperforming manually performed routine classifications and similar approaches. These findings provide an important reference for the rapid classification of small and medium earthquakes.

Keywords Induced earthquake · Blasting · Collapse · Machine learning · Deep learning

Article highlights
- We designed and compared three deep-learning classifiers to distinguish natural earthquakes, blasting, and collapse using convolutional neural network structures.
- The effects of the time-series input and spectrum on the classifiers were compared.
- The designed classifiers do not need to extract waveform features or mark seismic phases in advance and are suitable for moderate and microseismic events.

L. Jia (✉) · K. Xing
Henan Earthquake Agency, Zhengzhou, China
e-mail: Lezhao.jia@gmail.com

H. Chen
China Earthquake Networks Center, Beijing, China

1 Introduction

The rapid classification of seismic events is an important task for seismic networks. However, modern seismic observations of artificially induced events can be misclassified in earthquake catalogs as natural earthquakes, which complicates analyses of seismic activity (Mousavi et al. 2016). In general, larger seismic events have more seismic information recorded in their seismic waveforms and have waveform characteristics that are more obvious and more distinguishable than those of smaller events. For larger seismic events, the type of events can be correctly distinguished manually via the seismic phase characteristics (Blandford 1982; Rodgers et al. 1997). Distinct P and S phases are difficult to detect for small seismic events on regional broadband sensors because of the attenuation of their high frequencies, their low signal-to-noise ratios, and the sensor characteristics. This makes it difficult to quickly and accurately categorize such events, regardless of the use of manual or algorithmic methods. Using limited seismic observation data to quickly and reliably distinguish between natural earthquakes and blasting, collapses, or other seismic events is challenging.

In recent years, the main methods used to automatically classify induced and natural earthquakes using machine learning have been the first motion direction, seismic wave period, P/S-wave amplitude ratio, seismic phase complexity, coda attenuation characteristics, surface wave development, and frequency spectrum characteristics. The identification of induced earthquakes is primarily based on one or more of these characteristics. For example, Yang et al. (2005) used a spectral analysis to distinguish earthquakes and nuclear explosions and Koper et al. (2016) used the coda/duration magnitude difference of local earthquakes to distinguish between artificially created seismic events and naturally occurring tectonic earthquakes in Utah and its surrounding areas. Meanwhile, Tang et al. (2019) used a support vector machine method to distinguish structural earthquakes, quarry blasting, and induced earthquakes that occurred in the Tianshan orogenic belt, using characteristics such as the spectral amplitude and daytime incidence. According to Cho 2014, based on earthquake, explosion, and nuclear test data, there is a clear difference between blasting and seismic frequencies. Scarpetta et al. (2005) proposed a method to distinguish local seismic signals and volcanic tectonic earthquakes by taking the spectral characteristics of signals and the parametric attributes of their waveforms as the input signals of a multilayer perceptron. They designed four types of neural networks and achieved good results. Bregman et al. (2020) proposed the nonlinear dimensionality reduction of P and S waves generated by seismic arrays to identify earthquakes and blasting, the main feature of their method being that the signal amplification of seismic arrays can be used to discriminate smaller seismic events in the far field. Yildinm et al. (2011) used feedforward neural networks, adaptive neural fuzzy inference systems, and probabilistic neural networks to distinguish earthquakes and quarry blasts in Istanbul and nearby areas (Marmara) with a recognition rate of more than 97%. Saad et al. (2019) applied the support vector machine method to distinguish earthquakes and quarry blasts using a wavelet filter bank to extract unique features from data collected 5 s before and after the P wave. They tested 900 events and reached an accuracy rate of 98.5%. Shang et al. (2017) tested 1600 seismic events using principal component analysis and neural network methods and showed that the classification results of artificial neural network classifiers are superior to those of logistic regression and Bayes and Fisher classifiers.

Note that feature extraction from the full waveform is prone to introducing errors and that most feature extraction processes are complicated and difficult to use in real-time systems. With the development of machine-learning technology, the processing of seismic signals using related technologies has shown great potential. Some seismic waveform classification methods developed in recent years do not require the extraction of prior features. For example, Trani et al. (2021) designed two convolutional neural network (CNN) models that use time-series data and the spectrum diagram, respectively, as input to detect seismo-acoustic events and identify their sources in areas with high seismic noise and intense anthropogenic activity. Using their dataset, the application of spectral input was found to result in a better performance than the application of time-series data input. He et al. (2020) used machine learning to detect a slow slip event in seabed pressure data. Their method uses a model combining a CNN and a recurrent neural network to train two types of data and has a high event detection rate. Johnson et al. (2020) used an unsupervised algorithm to cluster the seismic signal and the background noise to accurately identify seismic signals. Seydoux et al. (2020) used a deep scattering network and a Gaussian mixture model to cluster seismic signal segments and detect new structures. This method was used to detect traditionally difficult to identify small seismic events preceding a landslide in Greenland in 2017. The semi-supervised learning method proposed by Linville (2022) explores the classification of earthquakes and blasts in a limited-label seismic dataset and surpasses the performance of supervised classification. Kuyuk et al. (2011) used an unsupervised learning approach, i.e., a self-organizing map, to distinguish between microseisms and quarry blasting near Istanbul, Turkey, using the self-organizing map as a neural classifier and a complementary reliability estimator to distinguish seismic events from the vertical components of seismic waves. This method directly extracts the features of the frequency-domain and time-domain data (e.g., complexity, spectral ratio, S/P-wave amplitude peak ratio) and achieves an accuracy rate of more than 94%. Meier et al. (2018) conducted a detailed comparative analysis with respect to the performances of different types of classifiers in the classification and recognition of seismic signals and noise for early earthquake warning. Their comparison of a fully connected neural network, recurrent neural network, CNN, and a generator plus random forest classifier showed that the CNN-based classifier had the highest precision and recall, outperforming several other neural networks. Dong et al. (2020) used the data from the microseismic monitoring system of a mine to establish a CNN-based microseismic event classifier to distinguish between microseismic events and blasting. Their classifier uses a four-layer convolution structure and a 2×2 convolution kernel and achieved an accuracy of more than 98% on the verification set.

Using the full seismic waveform for classification based on a CNN network essentially involves handing over the process of extracting the waveform features to the CNN network, which may become sensitive to the patterns in the waveforms. This capacity can therefore improve the performance of automatic seismic event classification.

The goal of this paper is to test the performance of different convolutional network structures on the classification of small and medium earthquakes and to provide a reference for similar research. In a seismic network, the volumes of different types of seismic event data are not balanced. In most cases, the volume of the natural earthquake data is far greater than that of other types of data and the volume of available actual non-natural earthquake data is limited. The experiment reported in this paper reveals that, restricted by the characteristics of the seismic waveform itself, the hierarchical structure of a network has a certain relationship with the number of data samples. This study investigates what type of convolution network structure performs best in the classification and recognition of events from a small volume of seismic waveform data.

We use 6.4 k actual local seismic events recorded by the Henan Regional Network of the China Seismic Network Center; this dataset contains 126 k channel data of raw observations. After data augmentation, there were approximately 150 k samples of raw seismic channel data. To reliably distinguish between earthquakes, blasting, and collapse events, we designed and optimized three seismic event classifiers with reference to CNN structures such as VGGnet, ResNet, and Inception. The three designed classifiers were tested and compared using three-channel seismic full-waveform time-series data and spectral data. To ensure a realistic comparison of the classifier performance, we set the classifier parameters and number of training samples to the same order of magnitude and use the same input–output structure and evaluation criteria.

Section 2 describes the datasets used in this study. Section 3 introduces the different classifiers and evaluation methods. Section 4 analyzes and compares the performances of the classifiers. Section 5 presents the potential uses of convolution-based seismic waveform classifiers.

2 Data

The seismic waveform dataset used for the training was taken from actual records of the Henan Regional Seismic Network Center of the China Seismic Networks Center. The dataset includes records of three types of events recorded by the network from June 2007 to March 2020, namely, records of natural earthquakes, artificial blasting events, and collapse events. The events were recorded by a network of 47 broadband seismic observation stations (Fig. 1a). A total of 6.4 k events were used, with magnitudes ranging from M_L 0.6 to M_L 4.5, with the minimum earthquake magnitude recorded by the regional network being M_L 0.6. The epicentral distance has a range of 0–400 km (Fig. 1b). Any peak ground velocity that lies outside six standard deviations of the velocity calculated using the standard ground-motion prediction equation (Bora et al. 2014) was discarded as an outlier. Ultimately, 42 k seismic event waveforms were retained, each event having three channel records for

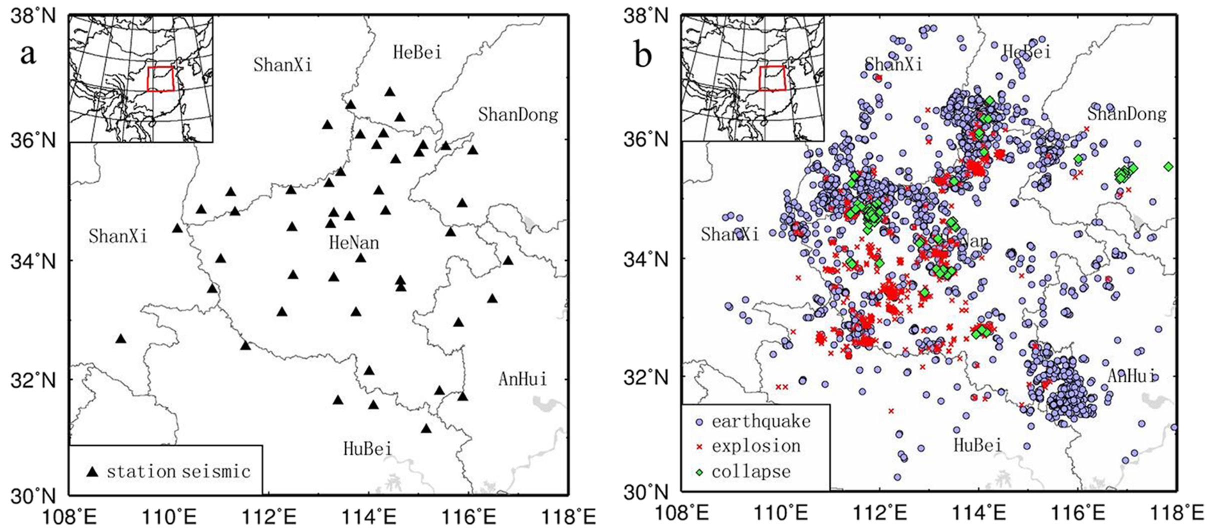


Fig. 1 Distribution of **a** seismic stations and **b** seismic events

a total of 126 k channel samples. All of the seismic waveforms in the dataset were recorded using broadband feedback seismometers. The frequency range of these instruments is 60 s/40 Hz, the sampling rate is 100 Hz, and the noise is lower than the new low-noise model, i.e., 30 s/4 Hz (Peterson, 1993). The seismic events were natural earthquakes, and the ratio of natural earthquakes to artificial blasting to collapses in the dataset is approximately 2:1.5:1.

2.1 Data sample

We scaled the data range to (− 1, 1) by normalizing the samples. The pure seismic waveforms were obtained by processing and removing the instrumental responses of the different stations. We corrected the waveform shift caused by the superposition of long-period seismic waves in the sample data. In this way, all of the seismic waveforms can be compared under a unified offset. Figure 2 shows typical data records for the three categories of events.

To construct the dataset, we set the length of the training sample such that we used the data from 1 s prior to the first P-wave arrival to 59 s after the first arrival for a total waveform fragment of 60 s. An analysis of the nearby seismic events shows that, for local seismic events below a magnitude of 4 within a distance of 400 km, the 60-s length covers the complete waveform of the vast majority of seismic events. A 60-s input signal has 100 sampling points per second; therefore, there are 6000 values per training sample.

Our classifiers use the 60-s full-waveform seismic data and the first wave arrival time and does not need to distinguish other seismic phases. This design makes the seismic event classifiers easy to use.

A broadband seismograph has at least three channels, which correspond to the vertical, east–west, and north–south directions. We formed a single sample for training according to the parallel arrangement of the vertical, east–west, and north–south channels recorded by each seismic station. The study by Kriegerowski et al. (2019) highlighted that high accuracy can be achieved when using this waveform arrangement for the earthquake location; even though this results in fewer training samples, their experiments show that this method increases the training accuracy. We used two types of input data: the time-series seismic waveform and the seismic waveform spectrum calculated by taking the absolute value after a fast Fourier transform.

We did not filter the seismic waveform prior to training and there were no artificial extractions of any prior seismic waveform features provided to the algorithm model.

We evaluated the routine local manual classification of the seismic events; multiple catalogers participated in the manual classification of the seismic

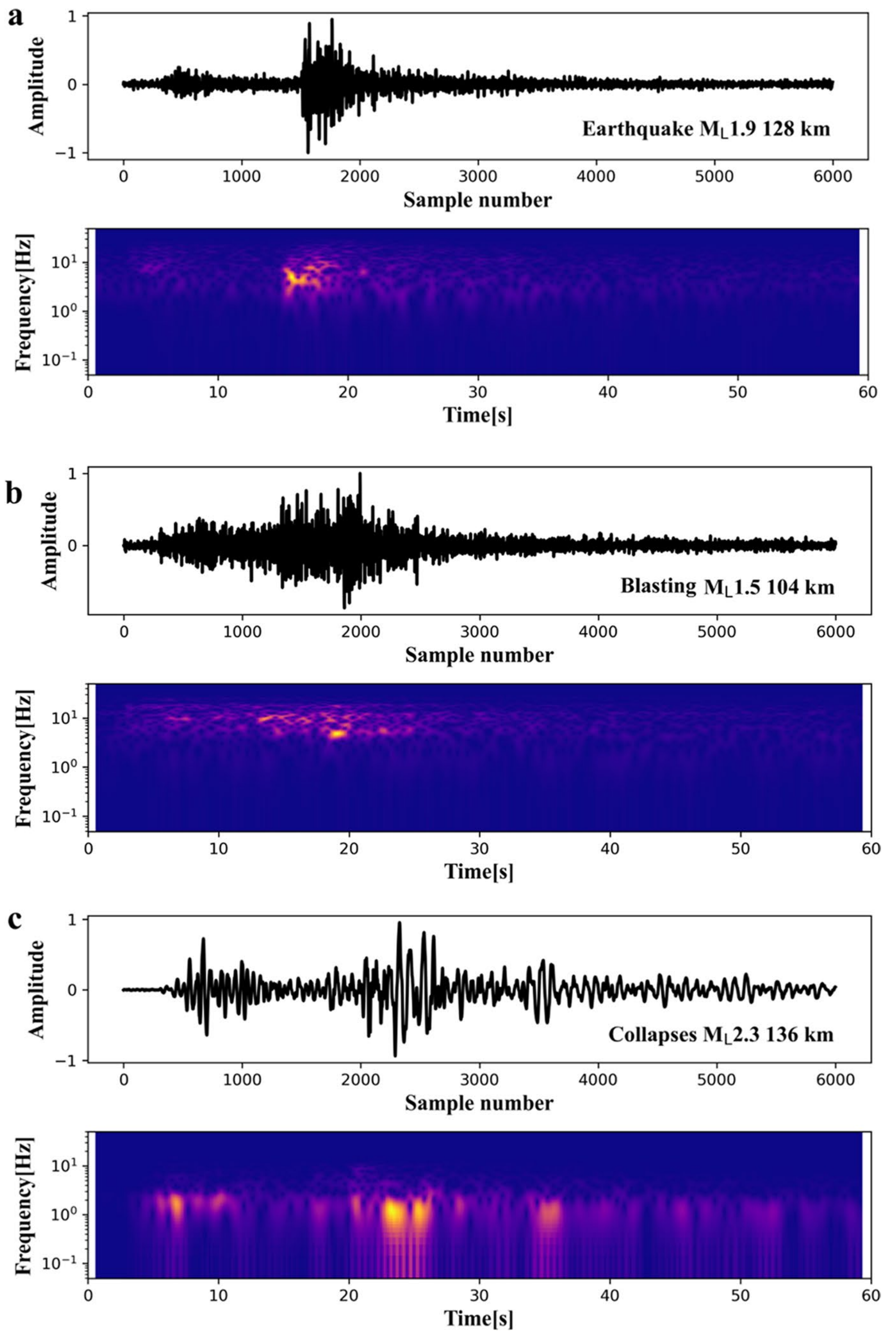


Fig. 2 Time series (upper panels) and spectral series (lower panels) for **a** a typical natural earthquake, **b** a typical blasting event, and **c** a typical collapse event

events in our dataset at different times for the period covered by our dataset. These catalogers manually performed routine classifications of the seismic events according to the seismic wave features (e.g., the amplitude ratio, spectrum, and period). Via consistency analyses, field validations, random testing, and other means, we evaluated the accuracy of the manually performed routine classification in the microseismic dataset ($M_L < 2.5$) as being between 80 and 90%. We manually checked all data samples and discarded misclassified samples.

We used data augmentation methods to effectively increase the training set by 20%. Specifically, we randomly rotated the original wave train by $\pm 5°$ to make the waveform look like a new dataset and we offset the P-wave position of the original wave train to generate new samples. In addition, the Gaussian method was used to randomly add ± 0.3 vertical interference noise to the original wave train, causing the original wave train to move horizontally back and forth to generate new samples. Data augmentation operations on datasets have been shown to improve machine-learning performance (Van Dyk et al. 2001).

2.2 Training set/validation set split

We divided the dataset into independent training (70%) and validation (30%) datasets to evaluate the performances of the classifiers (Fig. 3). In our dataset, there are a large number of natural earthquake data samples and a far fewer number of data records of induced earthquakes (blasting and collapses). Therefore, we randomly discarded part of the natural earthquake dataset.

The three types of data samples maintain the same ratio when split into training and validation sets. Each seismic event in the training and validation sets exists independently, and there is no crossover. The size of the overall sample dataset used in the calculation was 150 k. The sizes of the datasets for the natural earthquakes, blasting events, and collapse events were 57 k, 51 k, and 42 k, respectively. We randomly shuffled the datasets together and then split the overall dataset into a 105 k dataset for training and a 45 k dataset for testing. The seismic stations we used are evenly distributed throughout the training and test datasets. Each dataset entry contains three data channels, and every three channels form a training sample.

3 Method

3.1 Model definition

To define the classifiers, we referred to typical convolutional network structures, such as the LeNet-5 network (Lecun et al. 1998), AlexNet architecture (Krizhevsky et al. 2012), VGGnet architecture (Simonyan et al. 2014), GoogLeNet (Szegedy et al. 2014, 2017), ResNet architecture (He et al. 2016a), and DenseNet (Huang et al. 2016). Three classifiers of CNN structures were then refined and designed. We designed classifier 1 based on the VGGnet model with aserial CNN structure, classifier 2 with reference to the Inception model with a parallel structure, and classifier 3 with reference to the ResNet model with shortcut connections (Fig. 4). The correspondence between the parameter size and the sample size is an important issue in machine-learning research (Kaplan et al. 2020; Halevy, 2009). In our study of the classification of seismic waveforms, the data size of the seismic waveforms is certain; however, an excessive parameter size risks overfitting the model (Goodfellow et al. 2016). Therefore, we optimized the network structure based on our training sample data to highlight the characteristics of the network

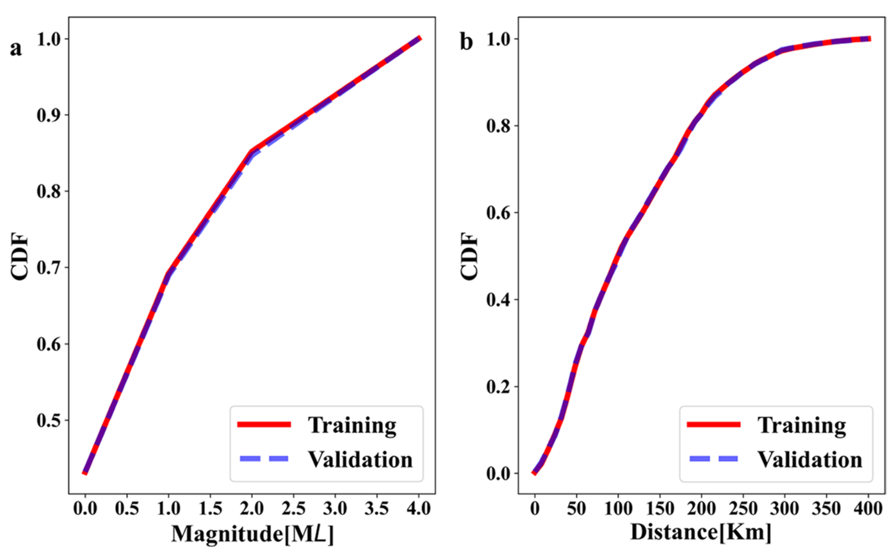
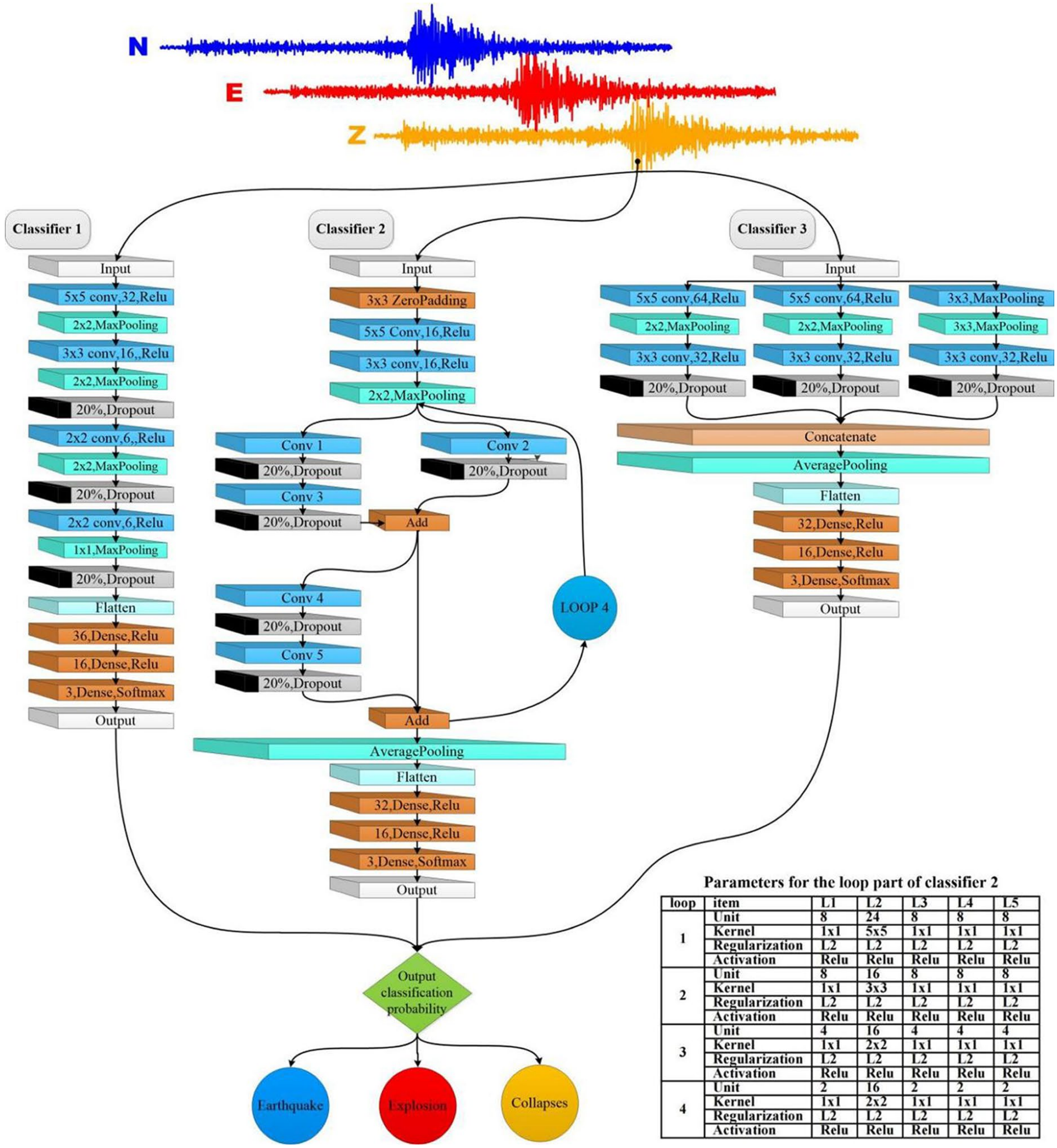


Fig. 3 Comparison of the cumulative distribution functions (CDFs) of the training and validation datasets for **a** the seismic magnitude and **b** the epicentral distance. The distributions of the magnitude and epicentral distance in the training and test sets are basically coincident. Most samples have a magnitude below 2.5 and an epicentral distance within 250 km



Parameters for the loop part of classifier 2

loop	item	L1	L2	L3	L4	L5
1	Unit	8	24	8	8	8
	Kernel	1x1	5x5	1x1	1x1	1x1
	Regularization	L2	L2	L2	L2	L2
	Activation	Relu	Relu	Relu	Relu	Relu
2	Unit	8	16	8	8	8
	Kernel	1x1	3x3	1x1	1x1	1x1
	Regularization	L2	L2	L2	L2	L2
	Activation	Relu	Relu	Relu	Relu	Relu
3	Unit	4	16	4	4	4
	Kernel	1x1	2x2	1x1	1x1	1x1
	Regularization	L2	L2	L2	L2	L2
	Activation	Relu	Relu	Relu	Relu	Relu
4	Unit	2	16	2	2	2
	Kernel	1x1	2x2	1x1	1x1	1x1
	Regularization	L2	L2	L2	L2	L2
	Activation	Relu	Relu	Relu	Relu	Relu

Fig. 4 Three types of classifiers are defined in the figure. Classifier 1 is a serial classifier based on the VGGnet network structure. Classifier 2 is a classifier based on the residual network, in which layers 6–15 are convolutional networks with shortcut connections that are cycled four times. The table in the figure details the parameters of each cycle. Classifier 3 is based on the Inception network structure and uses three groups of convolution modules. All the classifiers output the qualitative probabilities of the three types of seismic events

structure. To evaluate the performances of the different classifiers, we kept the number of training parameters of the different optimized network structures within the same order of magnitude and set the ratio of the network training parameters to the training samples to be less than 1.

Our goal was to test the impact of different network structures on the classifier performance; accordingly, we used commonly employed loss and activation functions. We used the categorical cross entropy function (Zhang and Sabuncu 2018) as the loss function. This function uses the cross entropy of the actual value and the predicted value to evaluate the difference between the current training probability distribution and the actual distribution. A small cross entropy value indicates that two probability distributions are similar.

We used the Adam function (Kingma & Ba, 2014; Keskar & Socher, 2017) as the optimization function for the classifiers. This algorithm is suitable for nonstationary targets and problems with very noisy or sparse gradients. No special optimization is required to use this function for seismic waveform data.

We used a nonlinear modified rectified linear element as the activation function (Glorot et al. 2011), and we specified that all classifiers use the normalized exponential function or softmax function. There are j node output values, S_i is the probability value of the ith element in the sequence, and e^i is the output value of the ith element. Using Eq. (1), the output of the neurons is mapped between (0, 1). The output value represents the qualitative probability of each classification.

$$S_i = \frac{e^i}{\sum_j e^j} \tag{1}$$

When setting the hyperparameters, we adopted the early stopping technique (Raskutti et al. 2013) and set the maximum number of iterations to 800. To achieve optimal accuracy, classifier 3 used 300 iterations, classifier 2 used 150 iterations, and classifier 1 used 180 iterations. In the selection of the batch and learning rate, the alternative batch was set to [30, 50, 100, 200, 500, 800] and the alternative learning rate was set to [0.1, 0.01, 0.001, 0.0001]. Using the grid search method, the batch setting was determined to be 100 and the learning rate was determined to be 0.01.

Classifier 1 was designed based on the VGGnet model. The VGGnet network structure is a deep CNN that was jointly developed by Oxford University and Deep Mind Technologies (Simonyan et al. 2014). This network builds a serial deep network structure by repeatedly stacking a combination of small convolution cores and maximum pooling. To match our data volume, we simplified the VGGnet network structure to 16 layers, including 4 convolution layers, such that there were fewer calculation parameters than training samples. We set the dropout layer probability to 20% to ensure the generalization ability of the model. The fully connected layer has 55 neurons. Finally, a flattening layer transforms the multi-dimensional input of the front node of the neuron into a single dimension and the convolution layer combination is passed to the fully connected layer. A three-layer fully connected layer is used for further feature extraction. Finally, the softmax function is used for the three-class classification output.

Classifier 2 was designed based on the ResNet model. The residual network is a CNN that was proposed by Microsoft Research in 2016 (He et al. 2016b). Its main feature is the use of skipped connection residual blocks to directly map the shallow and deep layers, maximizing the resolution. When the numbers of parameters are the same, the residual network has a greater depth than those of the other classifiers.

In the design of the residual network, we used the convolution layer of the two-layer 2×2 convolution kernel, which is different from traditional convolution kernels that have an odd number of dimensions. He and Jian (2015) conducted experiments using 5×5 and 3×3 convolution kernels and noted that changing a two-layer 3×3 convolution kernel into a four-layer 2×2 convolution kernel did not increase the number of parameters but did improve the accuracy. We adopted this approach in our study. In addition, we used multiple convolution layers of the 1×1 convolution kernel. For one-dimensional seismic waveform data, the use of the 1×1 convolution kernel greatly increases the nonlinear characteristics without decreasing the network resolution, deepens the network, and further reduces the number of parameters.

In the design of a residual network classifier, some scholars believe that using the batch normalization (BN) algorithm (Ioffe & Szegedy, 2015) effectively improves the convergence speed of the network and prevents overfitting. We used the BN algorithm to solve the problem of gradient saturation. In general, a network using the BN algorithm does not require L2 regularization or a dropout operation (Srivastava et al. 2014). Our experiment demonstrates that the BN algorithm is only suitable for preventing gradient saturation after the data samples are overfitted in the network structure. The excessive use of the BN algorithm leads to periodic oscillations of the training loss curve. Therefore, we used L2 norm regularization to suppress overfitting when we designed the classifier based on the residual network.

Classifier 3 was designed based on the Inception model. The Inception network structure, also known as GoogLeNet, is a deep-learning network structure that was proposed by Szegedy et al. (2015). This network structure is characterized by the parallel execution of multiple convolution operations or pooling operations and the splicing of all output results in a deeper feature map. We took the popular Inception V3 model as an example (Szegedy et al. 2016) when considering the effect of the network structure on the model accuracy as a whole.

The basis of Inception is to improve the network performance by increasing the width of the network. In each Inception module, the introduction of large-scale filter convolution has a high calculation cost. In our experiments, we found that it is feasible to use multiple small convolution cores instead of a few large convolution operation modules. In addition, the Inception network uses auxiliary decision branches to remarkably improve the network performance (Szegedy et al. 2016). Accordingly, when we designed the classifier based on the Inception network structure, we introduced a maximum pooling branch as an auxiliary decision branch to accelerate the convergence of the network. In each branch, we use a dropout layer to improve the generalization ability and prevent overfitting. After merging multiple network branches, we add a global maximum pooling layer in front of the network output instead of the BN algorithm, which was used in the original network; this improves not only the convergence speed of the network but also the overall accuracy of the classifier.

3.2 Evaluation method

We evaluated the classifiers using three evaluation methods, namely, numerical evaluation metrics, receiver operating characteristic curves (ROCs), and confusion matrixes. We consider the category with the highest probability given by the classifier as being the classification result predicted by that classifier. We record the classification result as a true positive if the predicted result and the actual result belong to that class. If the predicted result belongs to the class and the actual result does not, we record the classification result as a false positive. If the classifier prediction does not belong to the class and the actual result does, we record a false negative. If the classifier prediction does not belong to the class and actual result does not belong to the class, we record a true negative. Our classifiers are all three-class classification models, and each class has instances of true positives, false positives, false negatives, and true negatives.

We use the accuracy, precision, recall, and F1 values as evaluation indexes. In general, the calculation of these four indicators is only applicable to two categories. However, in our study, we need to evaluate the classifiers for three categories. Here, we introduce the macro-average to obtain multicategory indicators, that is, when calculating the accuracy and recall rate, we first calculate the values for each category, then average the values across all categories, and finally use the average values as the final accuracy and recall of the classifier.

$$\text{Precision} = \frac{TP}{TP+FP} = \frac{\sum_{i=1}^{n} TP_i}{\sum_{i=1}^{n} TP_i + \sum_{i=1}^{n} FP_i} \tag{2}$$

$$\text{Recall} = \frac{TP}{TP+FN} = \frac{\sum_{i=1}^{n} TP_i}{\sum_{i=1}^{n} TP_i + \sum_{i=1}^{n} FN_i} \tag{3}$$

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{4}$$

$$\text{F1} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{5}$$

Here, TP, FP, TN, and FN denote the numbers of true positives, false positives, true negatives, and false negatives, respectively.

The ROC curve is used to evaluate the results (Bachmann et al. 2006; Zhou et al. 2008). The ROC curve combines the sensitivity and specificity in a graphical method, accurately reflects the relationship between the specificity and sensitivity of a classifier, and is a comprehensive index with which to evaluate the classification performance of a classifier. In this specific application, the ROC curve directly provides the recognition ability of the classifier at any boundary value. The best diagnosis boundary value point (i.e., the threshold point with the lowest total number of false positives and false negatives) is directly selected using the Youden index

(Fluss et al. 2005). The ROC curve can be used to compare the performances of two or more classification algorithms. We also calculated the area under the ROC curve for each classifier. The classifier that has the largest area under the ROC curve has the best performance.

The true positive rate (TPR) is the proportion of all positive instances in the positive class predicted by the classifier, whereas the false positive rate (FPR) is the proportion of actual negative instances among all the negative instances in the positive classes predicted by the classifier as follows:

$$\mathrm{TPR} = \frac{TP}{TP + FN} \tag{6}$$

$$\mathrm{FPR} = \frac{FP}{FP + TN} \tag{7}$$

To calculate the ROC curves, we first map the multi-class classification problem into a two-class classification problem and set a probability threshold for each category. If the threshold is exceeded, the sample belongs to the category and the classification is recorded as a positive. If the threshold is not exceeded, the sample does not belong to the category and the classification is recorded as a negative. We then calculate TPR and FPR. On this basis, we transform the probability threshold into an ROC curve. The classification performance of the classifier is better when the ROC curve is closer to the point (0, 1) and deviates from a line along the 45° diagonal.

A confusion matrix is a standard format for accuracy evaluation and is expressed in matrix form with n rows and n columns. The confusion matrix is a visual tool that is suitable for supervising the evaluation of learning results (Sammut et al. 2017). In the application of a confusion matrix, we used rows to represent the truth number of classes and columns to represent the prediction number of the classifiers. Higher values from top left to bottom right are associated with a better classifier performance.

4 Results and discussion

4.1 Results

We evaluated the performances of the different classifiers and analyzed the classification results in detail. We used the same data and output for all classifiers. We tested the classifiers using 15 k of sample data that were not used in the training. Each classifier except the baseline had the same order of magnitude parameters and the same test data. The test data were input into the classifier via the time series and frequency spectrum for evaluation. For each of the sample data, the classifiers give the probability of the event being a natural earthquake, a blasting event, or a collapse. For comparison, we introduced the two-class classifier designed by Dong et al. (2020) as the test baseline. To match our data, we modified the input part of this classifier into a three-channel seismic waveform input and modified the output classification function to have three classes.

The accuracy, precision, recall, and F1 value were used to evaluate the performances of the classifiers. The results are given in Table 1.

The ROC curves of the three classifiers are shown in Fig. 5.

The baseline belongs to the same structure classifier as the VGGnet-based classifier 1 and therefore basically coincides with the ROC curve of the latter.

Table 1 Accuracy, precision, recall, and F1 values of the final classification results are slightly higher for classifier 1 than for the other classifiers. The baseline uses more parameters than the other classifiers

No	Input	Accuracy(%)	Precision (%)	Recall (%)	F1 (%)	Number of parameters
Baseline	Time series	90.802	90.816	91.948	91.297	80 k
	Spectrum	86.840	87.677	88.309	87.961	
1	Time series	92.181	92.102	92.665	92.371	21 k
	Spectrum	88.592	89.096	89.354	89.175	
2	Time series	91.096	90.879	92.176	91.381	17 K
	Spectrum	89.461	90.132	89.740	89.916	
3	Time series	91.632	92.651	91.286	91.822	35 k
	Spectrum	88.043	90.596	86.745	88.227	

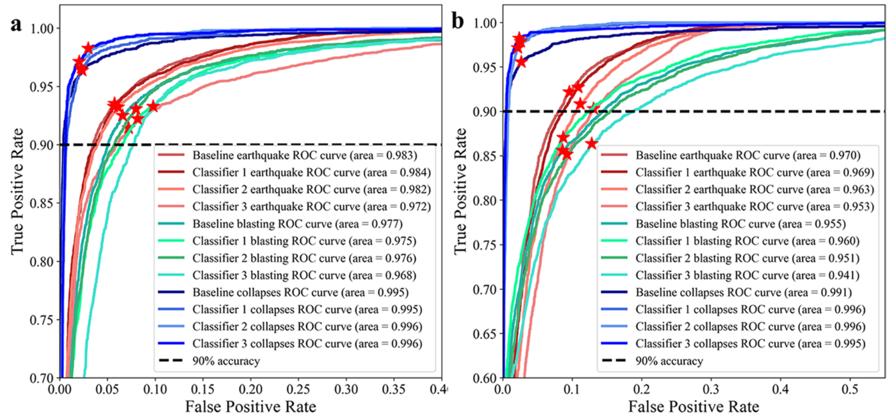


Fig. 5 Receiver operating characteristic (ROC) curves of the three classifiers. The abscissa gives the false positive rate and the ordinate gives the true positive rate. **a** Time-series data input results, **b** spectral input results. Each classifier has three curves representing its classification performance for the three different types of seismic events. The red stars in the figure show the best threshold points

All of the classifiers have the highest classification performance for collapse events (blue line in Fig. 5) and the worst classification performance for blasting events (green line). However, the worst classification performance of the classifiers using the time-series data input was still greater than 0.9 and the performance of the classifier using the spectral input was lower than that of the classifier using the time-series input. In the classification of natural seismic events, classifier 1 had a higher classification performance than the other two classifiers. In the classification of blasting events, classifier 2 had the best classification performance and the ROC curve of classifier 2 was on average between those of the other two classifiers. The baseline had the largest ROC curve area when using time-series data. Classifiers 2 and 3 had the best performance for classifying collapse events. However, the optimal threshold of the CNN automatic classifiers designed in this paper is greater than 90% (black dotted line in Fig. 5).

The results obtained using a confusion matrix to evaluate the classifiers are shown in Fig. 6.

The four classifiers (including the baseline) were found to more frequently misjudge natural earthquake and blasting events than collapse events for both the time-series data input and the spectral input. For incorrect classifications, classifier 1 had nearly identical error rates for natural earthquakes and blasting, classifier 2 was more likely to mistake natural earthquakes for blasting, and classifier 3 was more likely to mistake blasting for natural earthquakes. The four classifiers had a low probability of misidentifying collapse as blasting or blasting as collapse. Overall, the spectral input resulted in higher errors than the time-series data input (Fig. 6).

We analyzed the misclassified samples. There were 231 samples with a common classification error for the four classifiers. The vast majority of these misclassified events are located at the edge of the seismic observation network or in large coal mines in the study area (Fig. 7). The baseline and classifier 1 belong to similar classifier types and are not discussed separately here. There were 166 samples correctly classified by only classifier 1, 184 samples correctly classified by only classifier 2, and 308 samples correctly classified by only classifier 3. The performance of classifier 3 was relatively balanced and showed an advantage in classifying earthquakes of different magnitudes and distances. This may be related to classifier 3 setting up multiple parallel convolutional layers in the initial layer, which may enable the classifier to simultaneously recognize multiple waveform features. Most of the samples correctly classified by classifiers 1 or 2 alone were concentrated below M_L 3.0, with epicenters within 300 km (Fig. 8). From a waveform viewpoint, we believe that classifier 1 has a better classification effect for events with relatively large amplitude of P and S waves; such classifiers may be more sensitive to S/P-wave amplitude ratio features (Bennett et al. 1989; Baumgardt et al. 1990). Classifiers 2 and 3 had better classification effects for small events with less prominent seismic phases.

Some physical insights are as follows.

1. All classifiers have poor performance for identifying earthquakes that deviate from the study area. This is a main reason for the investigated classifiers reaching a maximum accuracy of only 92%. This is because the station records of deviating earthquakes have two characteristics: (1) There are fewer stations recording seismic waveforms and the epicenter is far away, and (2) the

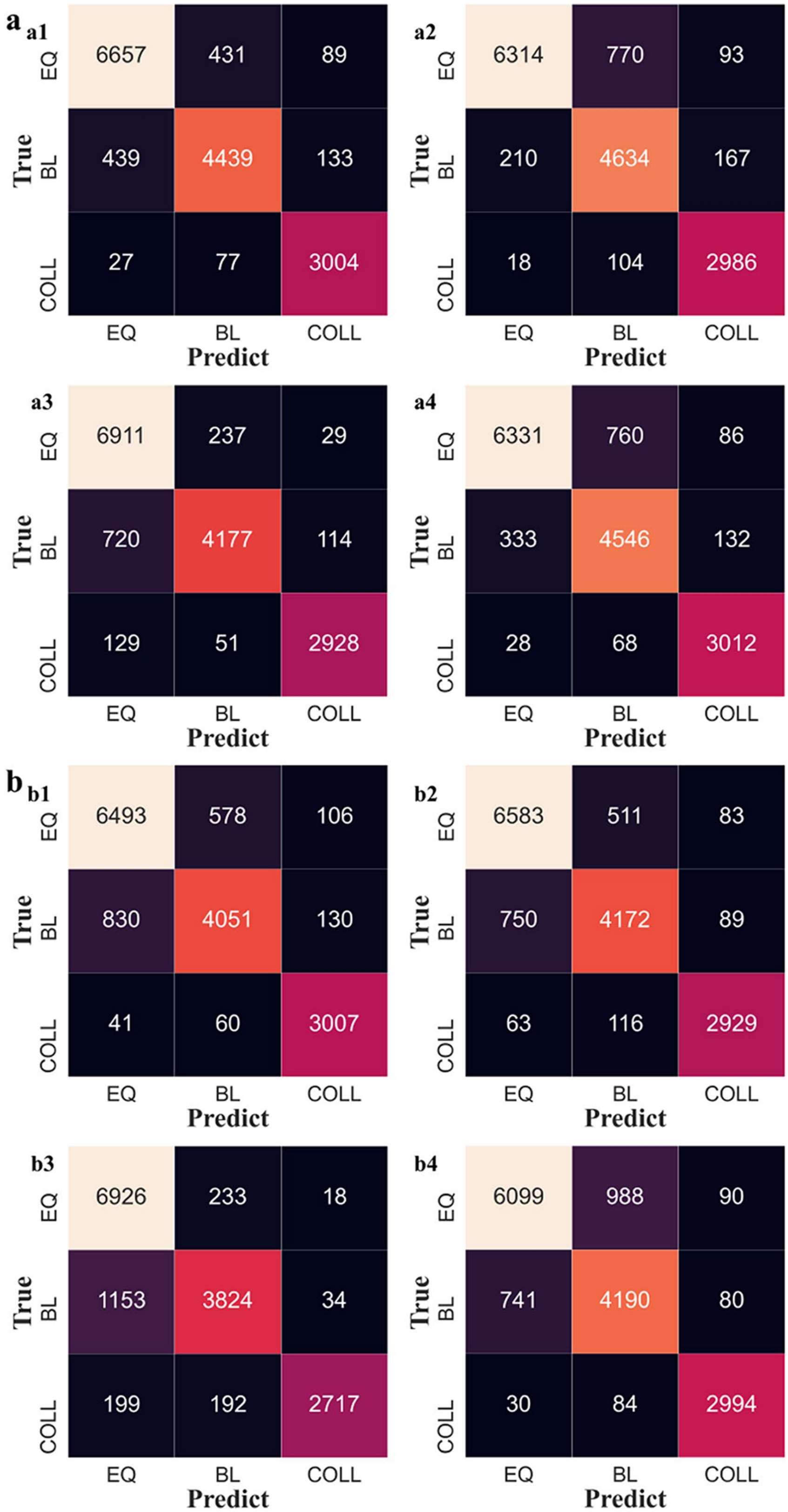


Fig. 6 Confusion matrix evaluation of different classifiers: **a** time-series data input and **b** spectral input, where 1, 2, 3, and 4 of both panels correspond to classifiers 1, 2, 3, and the baseline, respectively. The ordinate corresponds to the real label result, the abscissa corresponds to the predicted value of the classifier, including natural earthquake (EQ), blasting (BL), collapse (COLL), and the grid values from the upper left to lower right reflect the number of samples correctly judged by the classifier

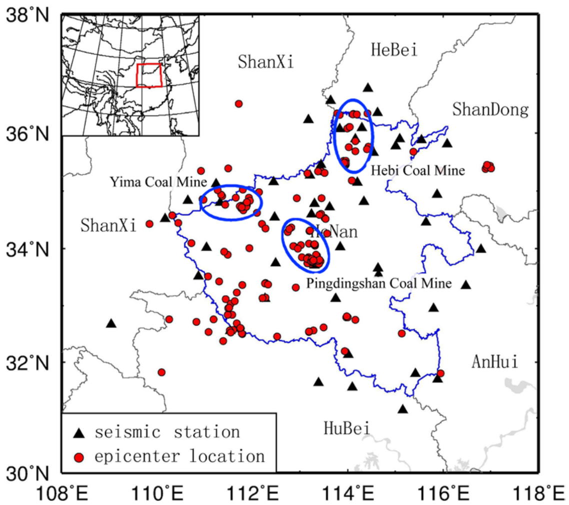


Fig. 7 Spatial distribution of misclassifications near the mining areas of large coal mines (blue ovals). Most misclassified events occur in coal mining areas and areas with weak seismic monitoring capabilities

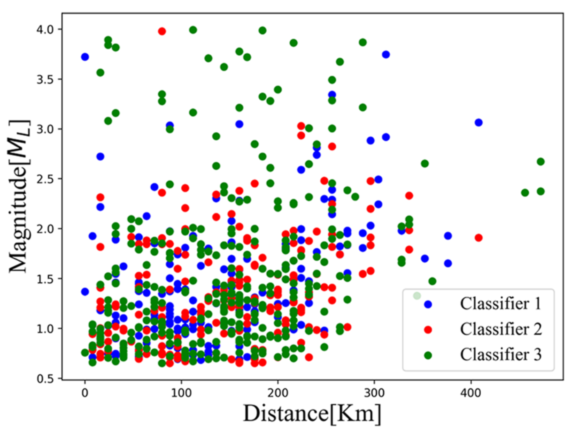


Fig. 8 Classifier advantage classification distribution. The circles represent samples only correctly classified by classifier 1 (blue), classifier 2 (red), or classifier 3 (green)

P and S waves are relatively separated, making all such events similar to natural earthquakes in terms of their waveform characteristics.

2. Seismic waveforms in large coal mining areas are more likely to be miscategorized. This is because a large number of goafs ("holes" created by human excavation or natural geological movement under the surface) form underground after years of mining in coal mining areas. Many of these seismic events occur in goafs and their epicenters are relatively shallow; therefore, seismic waveforms originating from goafs are easily confused with shallow blasting.
3. All of the classifiers classify collapse events well. This is likely because collapse events have larger periods than the other two types of events and visually differ from those of blasting and natural earthquakes. This is consistent with our typical perception of analyzing artificial earthquakes because the collapse period in the Henan region is larger than the periods of the other two event types and is easier to distinguish with the naked eye. Blasting and natural earthquakes are also easily confused in manual analyses of some smaller seismic events in both the frequency domain and the spectral domain.

4.2 Discussion

The baseline and classifier 1 are both serial convolutional network classifiers. Our experiments show that their performances are very similar; the accuracy of the baseline reaches 90.8%, and the accuracy of classifier 1 reaches 92.18%. The higher accuracy of classifier 1 may be because our network is deeper, learns more features during training, and therefore performs better. Note that, compared with the other classifiers, the baseline uses several times the number of parameters. We conjecture that models with fewer parameters may have better generalization ability; however, further research with additional data is required.

The performances of our classifiers are lower using spectral input than using direct input time-series data. This finding is seemingly inconsistent with the study of Trani et al. (2021); we attribute this to two causes. (1) Our classifier structures are different from the classifier structure used in their study, and subtle structural differences may lead to changes in the classification performance. (2) Converting time-series data into a spectrum involves many details; we used a relatively simple fast Fourier transform method. Different conversion methods may lead to differences in the classifier performance; however, this also requires further study.

The method we used only requires 60 s of the seismic waveform and the first-arrival time for alignment; the labeling of additional seismic phases or feature extraction is not required. We therefore believe that a classifier based on CNN may recognize more features, have fewer data processing links, and have stronger adaptability compared with a method that directly uses feature extraction or other machine-learning classification methods, such as a support vector machine or an ordinary multilayer neural network that does not use the convolution operation.

Our experiment shows that the classification ability of classifier 1 based on the VGGnet network structure is slightly better than those of the other two classifiers. Classifier 2 based on ResNet achieves similar accuracy using minimal computational parameters. However, the calculation accuracy does not differ greatly across the different classifiers. This may be related to the training sample size and the group characteristics of the samples. The advantages of the ResNet and Inception networks involve their greater number of convolution operations relating to the characteristics of the network structure and their fewer parameters; these advantages are not obvious when the training sample is not large. The false negative and false positive samples do not completely coincide for the different classifiers, indicating that the waveform features recognized by different classifiers are not completely consistent.

We also randomly selected several seismic events to observe the same seismic event with different epicentral distances. We found that the epicentral distance does not have a significant impact on the classifier performance. For most events, the classification results of the far and near stations were the same; we attribute this to two causes. (1) The seismic waveforms of the remote and near stations of the same event were included during the training. (2) The magnitude of the dataset was relatively small, and most of the seismic stations that recorded waveforms were not far away.

It is undeniable that there are still small errors in the classification of seismic events by the machine-learning-based classifiers examined in this paper. These classification errors generally occur for seismic events with magnitudes below M_L 2.0. The classifiers studied in this paper can be directly applied to data in regions with similar crustal structure; however, for regions with large differences in crustal structure, the introduced classifiers should be used for training with local seismic data to generate a localized model prior to application. Therefore, we suggest that, when using the classifiers designed in this study in different regions, the VGGnet network based on the serial structure be tried first. Then, if the loss curve or another indicator suggests the possibility of overfitting, the other two classifiers can be tried to achieve better results. When a classifier is used in sensitive regions, multiple classifiers should be combined in an integrated approach and a manual audit should be conducted to ensure accuracy.

For online operation, the classifiers only require that a single station receive the seismic event waveform to provide an accurate classification. After a seismic event occurs, generally more than one seismic station receives the seismic waveform. We assign a higher weight to seismic stations within 100 km of the event according to the seismic network layout. This is clearly marked on the output results for events with a consistency of more than 80% of the seismic stations; otherwise, the average probability after weighting by the near stations is output. According to the actual operation, the high-noise conditions of individual stations have little effect on the classifier performance.

Each classifier requires approximately 40 min using a single RTX2080 GPU to train using our method. Seismic waveform detection using the trained model on a common CPU machine requires 1 s of processing time and 60 s of the seismic waveform. According to actual work experience with the Henan seismic network in China, the current seismic waveform length used for accurate seismic positioning plus the positioning time itself takes at minimum approximately 1 min. Adding our classification procedure does not significantly increase the total time and is acceptable for practical application in a regional seismic network.

5 Conclusions

We referred to CNN structures, i.e., VGGnet, ResNet, and Inception, to design and optimize three types of seismic event classifiers. Three-channel seismic full-waveform time-series data and spectral data were used to test and evaluate the three designed classifiers and to describe the advantages, disadvantages, application scope, and suggestions for classifier use. Our classifiers use 60 s of full-waveform seismic data to achieve a recall rate and accuracy rate of more than 90% under the condition that the lower limit of the recognition magnitude reaches M_L 0.6; this method does not require the advance extraction of the waveform features or marking of the seismic phases. The research results surpass those of manually performed routine classifications and similar approaches, and our method can easily be used in actual seismic observation environments, thus providing a valuable reference for similar research.

Acknowledgements The seismic waveform dataset used in this paper was provided by the Henan Regional Center of the China Seismic Network Center. We thank the anonymous reviewers for their useful comments and suggestions. The seismic waveforms used in this paper can be downloaded from https://data.earthquake.cn. The algorithm was implemented using the Python software package Tensorflow (https://tensorflow.org), tflearn (http://tflearn.org), Matplotlib implementation, keras (https://keras.io), and scikit-learn (http://scikit-learn.org).

Author contribution Luozhao Jia: completed the experiments, program design, and writing.

Hongfeng Chen: contributed to the experimental concept and the analysis of the findings.

Kang Xing: participated in the collection of the data and the production of the illustrations.

Funding This research was supported by the Earthquake Science Spark Program of China Earthquake Administration no. XH20036.

Code availability Name of code: Seismic-classifier.

Developer: Luozhao Jia.

Contact address: Henan Earthquake Agency, No. 10 Zhengguang Road, Zhengzhou City, China.

Email: lezhao.jia@gmail.com.

First year available: 2021.

Required hardware: PC or server.

Required hardware for training environment: NVIDIA RTX2080.

The above graphics card, operating system: Windows or Linux.

Programming language: Python 3.7, Tensorflow 2.0.

The source codes are available for download at the link: https://github.com/epnet2018/Seismic-classifier.

Declarations

Competing interests The authors declare no competing interests.

References

Bachmann, LM, Puhan MA, Riet GT, & Bossuyt PM (2006). Sample sizes of studies on diagnostic accuracy: literature survey. BMJ, 332

Baumgardt DR, Young GB (1990) Regional seismic waveform discriminates and case-based event identification using regional arrays. Bull Seismol Soc Am 80:1874–1892

Bennett TJ, Barker BW, McLaughlin KL, Murphy JR, 1989. Regional discrimination of quarry blasts, earthquakes and underground nuclear explosions. Final Report, GL-TR-89–0114, S-Cubed, La Jolla, CA, 146 pp

Blandford R (1982) Seismic event discrimination. Bull Seismol Soc Am 72:569–587

Bora SS, Scherbaum F, Kuehn N, Stafford P (2014) Fourier spectral- and duration models for the generation of response spectra adjustable to different source-, propagation-, and site conditions. Bull Earthq Eng 12(1):467–493

Bregman Y, Lindenbaum O, & Rabin N (2020). Array based earthquakes-explosion discrimination using diffusion maps. Pure and Applied Geophysics(2a)

Cho,& Kwang-Hyun.(2014). Discriminating between explosions and earthquakes. Applied Geophysics(04)

Dong LJ, Tang Z, Li XB, Chen YC, Xue JC (2020) Discrimination of mining microseismic events and blasts using convolutional neural networks and original waveform. J Cent South Univ 27(10):3078–3089

Fluss R, Faraggi D, & Reiser B (2005). Estimation of the Youden index and its associated cutoff point. Biometrical Journal, 47(4)

Glorot X, Bordes A, & Bengio Y (2011). Deep sparse rectifier neural networks. J Mach Learn Res (Vol.15, pp.315–323)

Goodfellow I, Bengio Y, & Courville A (2016). Deep learning. JMLR.org

Halevy A, Norvig P, Pereira F (2009) The unreasonable effectiveness of data. IEEE Intell Syst 24(2):8–12

He K, & Jian S (2015). Convolutional neural networks at constrained time cost. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE

He K, Zhang X, Ren S, et al 2016a Deep residual learning for image recognition [C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 770–778

He K, Zhang X, Ren S, & Jian S (2016b). Identity mappings in deep residual networks. European Conference on Computer Vision

He B, Wei M, Watts DR, & Shen Y (2020). Detecting slow slip events from seafloor pressure data using machine learning. Geophysical Research Letters, 47

Huang G, Liu Z, Laurens V, & Weinberger KQ (2016). Densely connected convolutional networks. IEEE Computer Society

Ioffe S & Szegedy C (2015). Batch normalization: accelerating deep network training by reducing internal covariate shift. Proceedings of the 32nd International Conference on Machine Learning in Proceedings of Machine Learning Research 37:448–456

Johnson CW, Y Ben-Zion, Meng H & Vernon F (2020). Identifying different classes of seismic noise signals using unsupervised learning. Geophysical Research Letters, 47(15)

Kaplan J, Mccandlish S, Henighan T, Brown TB, Chess B, & R Child et al. (2020). Scaling laws for neural language models

Keskar NS, & Socher R (2017). Improving generalization performance by switching from adam to sgd

Kingma D, & Ba J (2014). Adam: a method for stochastic optimization. Computer Science

Koper Keith, D, Hale Mark, J Stein, & Jared et al (2016). Magnitude-based discrimination of man-made seismic events from naturally occurring earthquakes in Utah, USA. Geophysical Research Letters

Kriegerowski M, Petersen GM, Vasyura-Bathke H, Ohrnberger M (2019) A deep convolutional neural network for localization of clustered earthquakes based on multistation full waveforms. Seismol Res Lett 90(2A):510–516
Krizhevsky A, Sutskever I, Hinton GE 2012 ImageNet classification with deep convolutional neural networks [C] Advances in neural information processing systems. 1097–1105
Kuyuk HS, Yildirim E, Dogan E, Horasan G (2011) An unsupervised learning algorithm: application to the discrimination of seismic events and quarry blasts in the vicinity of Istanbul. Nat Hazard 11(1):93–100
Lecun Y, Bottou L (1998) Gradient-based learning applied to document recognition. Proc IEEE 86(11):2278–2324
Linville LM (2022) Event-based training in label-limited regimes. J Geophys Res Solid Earth 127:e2021JB022820
Meier MA, Ross ZE, Ramachandran A, Balakrishna A, Nair S, & Kundzicz P et al (2018). Reliable real-time seismic signal/noise discrimination with machine learning. Journal of Geophysical Research: Solid Earth
Mousavi SM (2016) Comment on "Recent developments of the middle east catalog" by Zare et al. J Seismol 21:1–12
Peterson J 1993,Obervations and modeling of seismic background noise. U.S. Geol. Survey Open-File Report 93–322,95 pp.
Raskutti G,Wainwright MJ, & Yu B (2013). Early stopping and non-parametric regression: an optimal data-dependent stopping rule. J Mach Learn Res, 15
Rodgers AJ, Lay T, Walter WR, Myers S (1997) A comparison of regional-phase amplitude ratio measurement techniques. Bull Seismol Soc Am 87:1613–1621
Saad OM, Shalaby A, & Sayed MS (2019). Automatic discrimination of earthquakes and quarry blasts using wavelet filter bank and support vector machine. J Seismol
Sammut C, & Webb GI (2017). Encyclopedia of machine learning and data mining. https://doi.org/10.1007/978-1-4899-7687-1.
Scarpetta S, Giudicepietro F, Ezin EC, Petrosino S, Del Pezzo E, Martini M, Marinaro M (2005) Automatic classification of seismic signals at Mt. Vesuvius Volcano, Italy, using neural networks. Bull Seismol Soc Am 95(1):185–196
Seydoux L, Balestriero R, Poli P, Hoop MD, Baraniuk R (2020) Clustering earthquake signals and background noises in continuous seismic data with unsupervised deep learning. Nat Commun 11(1):3972
Simonyan K & Zisserman A (2014). Very deep convolutional networks for large-scale image recognition. Computer Science
Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R (2014) Dropout: a simple way to prevent neural networks from over-fitting. J Mach Learn Res 15(1):1929–1958
Szegedy C, Liu W, Jia Y, Sermanet P, & Rabinovich A (2014). Going deeper with convolutions. IEEE Computer Society
Szegedy C,Vanhoucke V, Ioffe S, Shlens J & Wojna Z (2016). Rethinking the inception architecture for computer vision. IEEE, 2818–2826
Szegedy C, Ioffe S, Vanhoucke V & Alemi AA (2017). Inception-v4, inception-ResNet and the impact of residual connections on learning
Tang L, Zhang M, Wen L (2019) Support vector machine classification of seismic events in the Tianshan orogenic belt. J Geophys Res Solid Earth. https://doi.org/10.1029/2019JB018132
Trani L, Pagani GA, Zanetti J, Chapeland C, Evers LG (2021) DeepQuake -- an application of CNN for seismo-acoustic event classification in The Netherlands
Van Dyk DA, Meng XL (2001) The art of data augmentation. J Comput Graph Stat 10(1):1–50
Shang X, Li X, Morales-Esteban A, Chen G (2017) Improving microseismic event and quarry blast classification using artificial neural networks based on principal component analysis. Soil Dynamics & Earthquake Engineering
Yang XH, Shen P, Liu XQ, & Zheng ZZ (2005). Application of method of spectral component ratio of wavelet-packets to discrimination between earthquakes and nuclear explosions. Chinese J Geophys, 48(1)
Yildinm E, Guelbag A, Horasan G, Dogan E (2011) Discrimination of quarry blasts and earthquakes in the vicinity of Istanbul using soft computing techniques. Comput Geosci 37(9):1209–1217
Zhang Z, & Sabuncu M. (2018). Generalized cross entropy loss for training deep neural networks with noisy labels. Advances in neural information processing systems, 31
Zhou XH ,Obuchowski NA, & Mcclish, DK (2008). 2. Measures of diagnostic accuracy. John Wiley & Sons, Inc