

EXPRESS LETTER

Open Access



Discriminating seismic events using 1D and 2D CNNs: applications to volcanic and tectonic datasets

Masaru Nakano^{1*}  and Daisuke Sugiyama²

Abstract

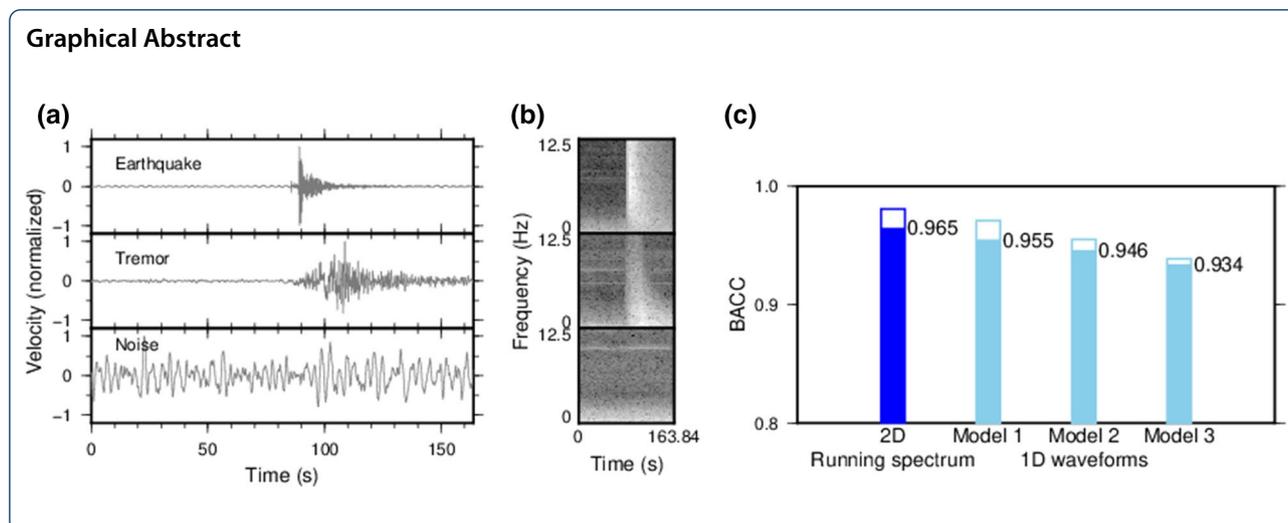
Detecting seismic events, discriminating between different event types, and picking P- and S-wave arrival times are fundamental but laborious tasks in seismology. In response to the ever-increasing volume of seismic observational data, machine learning (ML) methods have been applied to try to resolve these issues. Although it is straightforward to input standard (time-domain) seismic waveforms into ML models, many studies have used time–frequency-domain representations because the frequency components may be effective for discriminating events. However, detailed comparisons of the performances of these two methods are lacking. In this study, we compared the performances of 1D and 2D convolutional neural networks (CNNs) in discriminating events in datasets from two different tectonic settings: tectonic tremor and ordinary earthquakes observed at the Nankai trough, and eruption signals and other volcanic earthquakes at Sakurajima volcano. We found that the 1D and 2D CNNs performed similarly in these applications. Half of the misclassified events were misassigned the same labels in both CNNs, implying that the CNNs learned similar features inherent to the input signals and thus misclassified them similarly. Because the first convolutional layer of a 1D CNN applies a set of finite impulse response (FIR) filters to the input seismograms, these filters are thought to extract signals effective for discriminating events in the first step. Therefore, because our application was the discrimination of signals dominated by low- and high-frequency components, we tested which frequency components were effective for signal discriminations based on the filter responses alone. We found that the FIR filters comprised high-pass and low-pass filters with cut-off frequencies around 7–9 Hz, frequencies at which the magnitude relations of the input signal classes change. This difference in the power of high- and low-frequency components proved essential for correct signal classifications in our dataset.

Keywords: Machine learning, Convolutional neural network, Sakurajima, Explosion earthquake, Nankai trough, Slow earthquake

*Correspondence: mnakano@jamstec.go.jp

¹ Institute for Marine Geodynamics (IMG), Japan Agency for Marine–Earth Science and Technology, 2-15, Natsushima-cho, Yokosuka, Kanagawa 237-0061, Japan

Full list of author information is available at the end of the article



Introduction

Event identification and phase picking are the most fundamental tasks in seismic event monitoring, but are also quite laborious. Because of the ever-increasing volume of seismic observational data from dense arrays covering entire nations and their surrounding oceans, such as USArray (Kerr 2013) and Japan's MOWLAS (Aoi et al. 2020), these tasks now require automated systems. Various machine learning (ML) methods have been applied to resolve this issue (e.g., Dowla et al. 1990; Wang and Teng 1995; Del Pezzo et al. 2003; Scarpetta et al. 2005; Kong et al. 2016), of which convolutional neural networks (CNNs; e.g., LeCun et al. 2015) have frequently been used for seismic signal discriminations and phase picking (e.g., Perol et al. 2018; Ross et al. 2018a, b; Sugiyama et al. 2021). CNNs have been applied to slow earthquakes in subduction zones (Nakano et al. 2019; Takahashi et al. 2021) and volcanic signals (Canário et al. 2020). Recent studies have combined CNNs with other methods to improve the accuracy and efficiency of these tasks (Mousavi et al. 2019, 2020; Soto and Schurr 2021).

Whereas it is straightforward to use waveform traces as the input for ML applications to seismic data (e.g., Perol et al. 2018; Ross et al. 2018a, b), many studies use time–frequency-domain representations (running spectra, wavelet transforms) to classify signals (e.g., Dowla et al. 1990; Wang and Teng 1995; Yoon et al. 2015; Holtzman et al. 2018; Mousavi et al. 2019; Dokht et al. 2019; Nakano et al. 2019; Rouet-Leduc et al. 2020; Takahashi et al. 2021). Because the spectral characteristics of seismic signals depend on their source locations and source mechanisms, the time–frequency-domain representation is expected to improve ML performances. Several studies (Nakano et al. 2019; Rouet-Leduc et al. 2020; Takahashi et al. 2021) have discriminated low-frequency

(2–8 Hz; e.g., Obara 2002) earthquakes in subduction zones from ordinary earthquakes dominated by higher frequency components. Similarly, volcanic earthquakes can also be classified by their dominant frequencies. Iguuchi (1994) classified volcanic earthquakes at Sakurajima volcano into 4 groups: A-type, BH-type, BL-type, and explosion. Of those, A-type earthquakes result from fault ruptures within the edifice and are dominated by energy at 10–20 Hz, whereas the other types are dominated by lower frequency components. Chouet (1996) classified volcanic earthquakes as either volcano-tectonic (VT) associated with shear failures and dominated by higher frequency components, or long-period (LP) events dominated by frequency components lower than several hertz, and Scarpetta et al. (2005) and Canário et al. (2020) developed methods to classify volcanic earthquakes by focusing on differences in their frequency characteristics. Despite the importance of efficiently classifying these signals using automated ML systems for monitoring seismic and volcanic activity, detailed comparisons of the performances of ML methods using time–frequency-domain data to those using time-domain data are lacking.

Canário et al. (2020) made such a comparison, and found that the performances were mostly similar. However, their study was based only on 10 frequency components after performing a wavelet transform of the time–frequency-domain representations. CNN performance might depend on the specific selection of frequency components. In this study, we compared the performances of 1D (time-domain) and 2D (time–frequency-domain) CNNs by using the same time-window width, number of data points, and frequency components in the running spectral images (2D) as those of the 1D waveform traces. We used two datasets from different tectonic settings, one from the Nankai trough and one

from Sakurajima volcano, to gauge whether CNN performance depended on the input dataset. Based on our results, we discuss the possibility of determining the most useful frequency components solely using 1D CNNs.

Methods and data

We compared the performances of 1D and 2D CNNs in discriminating different seismic and volcanic signals; these CNNs use waveform traces and running spectral images, respectively, as input data. For the 2D CNN, we used the model developed in our previous study (Nakano et al. 2019), in which running spectral images of 64×64 pixels are processed using two sets of convolutional and max pooling layers, and then passed into two fully connected layers to classify signals into three classes: earthquake, tectonic tremor, or noise. The dimension of the convolutional kernel (also called convolutional filter) was 5×5 , and the stride of the max pooling layer was 2×2 . We adopted the rectified linear unit (e.g., Fukushima 1980) as an activation function. The first and second convolutional layers had 32 and 16 channels, respectively (Additional file 1: Table S1), and, respectively, used 64 and 28 frequency components, corresponding to the number of pixels on the vertical axis. The lengths of the first and second fully connected layers were 20 and 3, respectively (see details in Nakano et al. 2019).

For comparison, we constructed a 1D CNN by replacing the 2D convolutional and pooling layers with two sets of 1D convolutional and pooling layers each (i.e., four sets of convolutional and max pooling layers), and the output was passed into two fully connected layers (Additional file 1: Table S2). The length of the convolutional kernel was 5, and the stride of the max pooling layer was 2, matching the 2D CNN described above. Again, we adopted the rectified linear unit as an activation function. We tested three models with different channel numbers in each layer (Additional file 1: Table S2). In Model 1, the first and second convolutional layers each had 64 channels, corresponding to the number of frequency components in the first layer of the 2D CNN, and the third and fourth layers each had 28 channels, as in the second layer of the 2D CNN. In Model 2, the first layer had 16 channels, and the number of channels was doubled in each subsequent layer, reaching 128 in the fourth layer, as in the model of Ross et al. (2018b). Model 3 was the same as Model 2, but the first layer had only two channels. The number of trainable parameters for each model is summarized in Additional file 1: Table S3.

To prepare the data for input into the 1D CNN, we decimated the seismic records to 25 Hz and trimmed the data so that the signal to be classified fit within a 163.84-s time window. The waveforms were then normalized to the absolute maximum value in each trace. Each

waveform trace had 4096 data points, and the Nyquist frequency was 12.5 Hz. A running spectral image corresponding to the waveform trace was created by computing the amplitude spectra of the short-term Fourier transforms of 5.12 s (128-point) half-overlapping time windows of the decimated continuous waveform, and then cutting them to the same time window to obtain images of 64×64 pixels. These running spectral images included signals between 0.2 and 12.5 Hz within the 163.84-s time window. Each image was normalized to the maximum value and used as the input to the 2D CNN. The waveform traces and running spectral images prepared in this way have a common number of data points and contain frequency components up to 12.5 Hz, but the running spectral images do not contain phase information or frequency components below 0.2 Hz.

To evaluate whether CNN performance depends on the input dataset, we used two catalogs from different tectonic settings. The first was the catalog of shallow tectonic tremor at the Nankai trough subduction zone used by Nakano et al. (2019) (Additional file 1: Fig. S1), and the other comprised volcanic earthquakes excited by summit explosions at Sakurajima volcano, based on the eruption catalog developed by the Kagoshima Meteorological Office of the Japan Meteorological Agency (JMA) (Additional file 1: Fig. S2).

For the Nankai catalog, we created the input waveforms and running spectral images from DONET three-component broadband seismometer records (Kaneda et al. 2015; Kawaguchi et al. 2015; Aoi et al. 2020) using time windows centered on the origin times of local earthquakes and tectonic tremor events. Noise data were created with the same start time as that of Nakano et al. (2019). From the catalog, we randomly selected events in each signal class and split the dataset into 70% for training, 15% for validation, and 15% for test. In Nakano et al. (2019), input data with low signal-to-noise ratios (S/N) were removed by visual inspection. Here, we evaluated the data quality based on the S/N computed from the peak absolute amplitude and the standard deviation during the first 10 s of the waveform traces, band-passed between 1 and 10 Hz. Signals with S/N in the top 80% were used for local earthquakes and tectonic tremor events, and those in the bottom 80% were used for noise. The criteria for signal selection were the same for the waveform and running spectrum datasets and they included the same data. Figure 1a and b shows representative input waveforms and their corresponding running spectral images, respectively, for ordinary earthquakes, tectonic tremor events, and noise. Compared to local earthquakes, tectonic tremor signals were dominated by lower frequency components and had longer durations. The number of waveforms in each dataset is listed in Table 1.

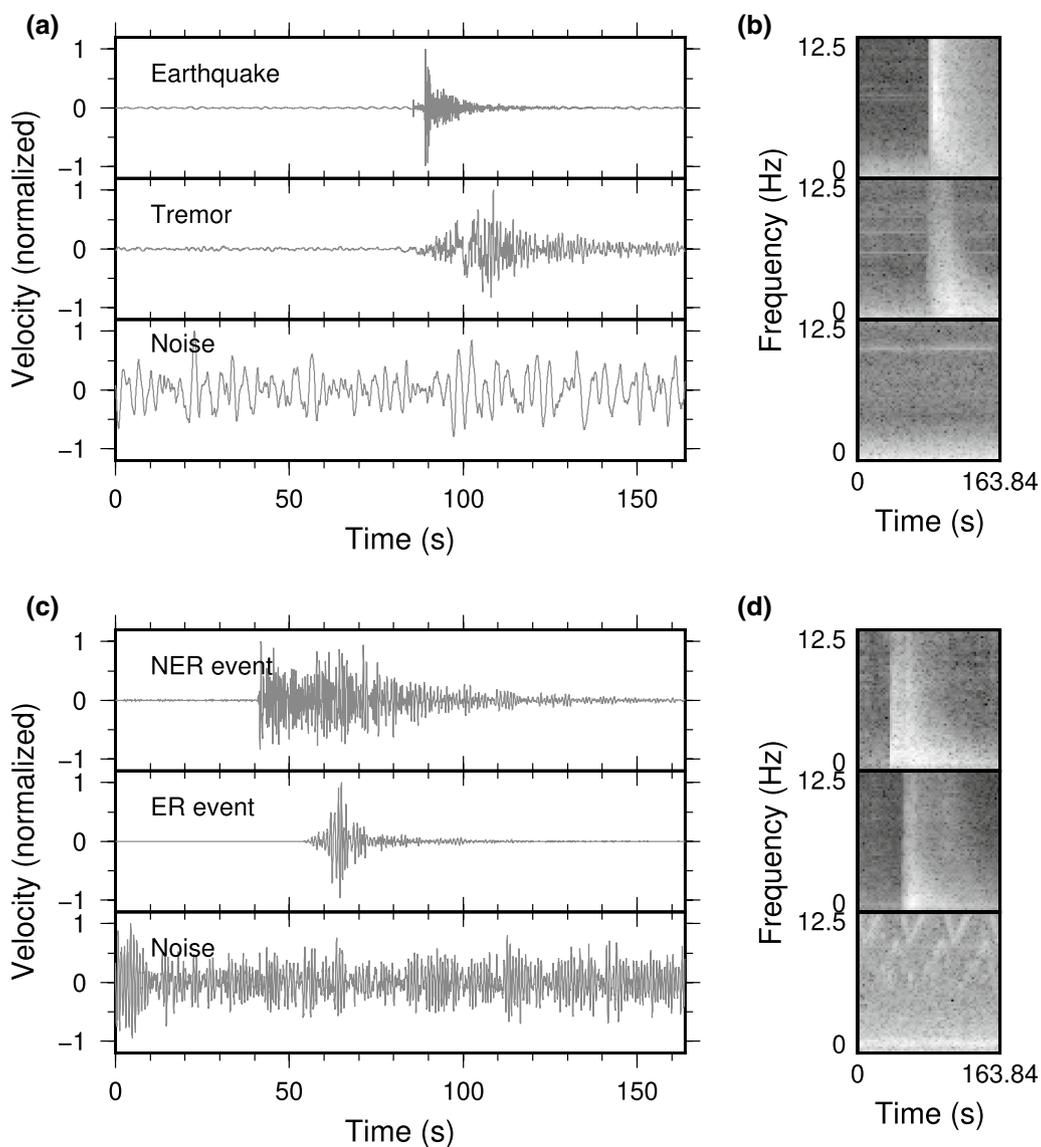


Fig. 1 Representative waveforms and running spectral images used as input data for CNNs. **a** Waveforms and **b** corresponding running spectral images of a local earthquake, a tectonic tremor event, and noise from the Nankai dataset. **c** Waveforms and **d** corresponding running spectral images of a non-eruptive (NER) volcanic event, an eruptive (ER) event, and noise from the Sakurajima dataset. Amplitudes in **a, c** are normalized to the maximum amplitude

Table 1 Numbers of samples used for the training, validation, and test datasets

| Dataset | Nankai trough | | | Sakurajima | | |
|------------|---------------|--------|-------|------------|----------|--------|
| | Earthquake | Tremor | Noise | NER event | ER event | Noise |
| Training | 4068 | 3270 | 4085 | 113,378 | 12,295 | 81,530 |
| Validation | 859 | 603 | 889 | 6,034 | 1,272 | 38,547 |
| Test | 972 | 742 | 876 | 9,104 | 1,367 | 15,718 |

For the Sakurajima catalog, we used volcanic earthquakes that occurred in 2015, when eruptive activity was the most active in recent years; activity ceased by the end of September. At Sakurajima, A-type volcanic earthquakes result from faulting within the edifice, BH-type from the release of volumetric strain due to magmatic intrusions, BL-type are related to ejections of volcanic bombs and ash from the summit, and explosion earthquakes are excited from violent explosive eruptions (Iguchi 1994). Here, we classified seismic signals as eruptive events (ER, including BL-type and explosion earthquakes), non-eruptive events (NER, including BH- and A-type earthquakes), or noise. The aim of this classification was to develop a telemetric seismic monitoring system for volcanic eruptions. Figure 1c and d shows representative waveforms and their corresponding running spectral images, respectively, for our three signal classifications at Sakurajima. ER events are dominated by lower frequency components like tectonic tremor, whereas NER events contain higher frequency components like local earthquakes (Iguchi 1994). Input data for the Sakurajima catalog were created from the vertical components of seismograms of the Sakurajima observation network operated by JMA. Since the eruption catalog only defines event origin times down to the minute, the exact time that each signal appeared was unknown. Therefore, we created three input waveforms and running spectral images for each ER and NER event by randomly shifting the center of time window between ± 65.536 s (i.e., 40% of the time window) from the origin time. Noise data were created from randomly selected time windows, excluding the 327.68 s preceding and 1 h following event origin times. We selected enough time windows of noise to roughly match the number of NER events. The quality of each waveform was evaluated based on the same S/N criterion as in the Nankai dataset. In this way, we obtained 14,934, 128,516, and 125,795 waveforms for ER events, NER events, and noise, respectively. In each class, we used events that occurred during February–August as the training dataset, and events occurring in January and September as the validation and test datasets, respectively. The number of waveforms in each dataset varied according to the event type because, compared to NER events, far fewer ER events occurred during the analysis period and because the monthly number of events varied (Table 1).

We trained the 1D and 2D CNN models using the training dataset for 300 epochs. Because the amount of data in each class was quite variable, especially in the Sakurajima dataset, we allowed the duplication of samples from minority classes to balance the size of the majority class when subdividing the dataset into minibatches. The parameters of the CNN were trained using

a cross-entropy loss function with the Adam optimization algorithm (Kingma and Ba 2015) in minibatches of 64 waveforms. We determine the predicted class as that of the largest output probabilities from the CNN. Model performance was evaluated after each epoch using the validation dataset and the balanced accuracy (BACC) metric, defined for a three-class input as:

$$\text{BACC} = \frac{1}{3} \left(\frac{T_{00}}{N_0} + \frac{T_{11}}{N_1} + \frac{T_{22}}{N_2} \right), \quad (1)$$

where T_{ii} is the number of correctly classified waveforms and N_i the total number of waveforms in the i th class. Based on these results, we selected the model with the highest BACC value and applied it to the test dataset to evaluate model performance upon generalization, again based on BACC.

Results

Figure 2 summarizes the performances of the trained 1D and 2D CNN models when applied to the validation and test datasets (open and closed bars, respectively). Among the 1D models, model 1 achieved the highest BACC value for both the Nankai and Sakurajima validation datasets. However, the 2D CNN achieved the highest

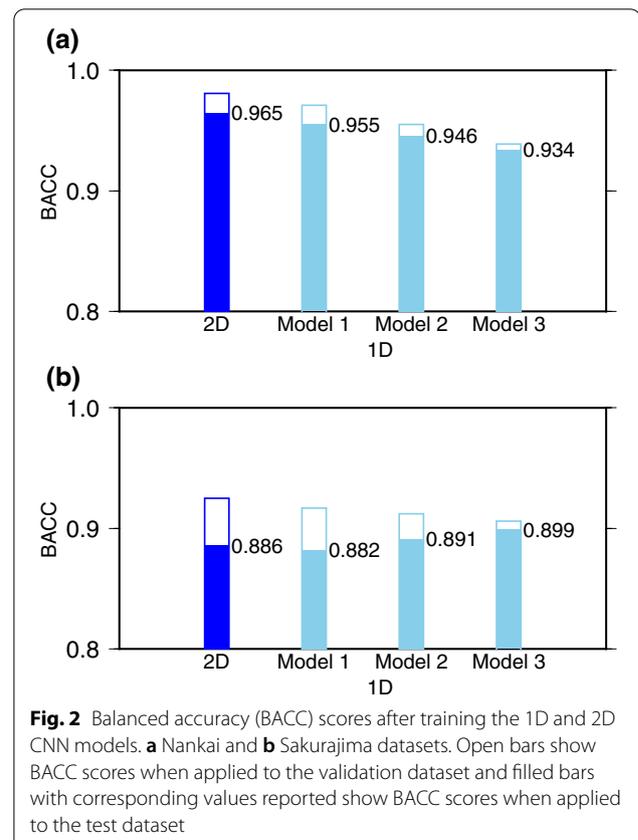


Fig. 2 Balanced accuracy (BACC) scores after training the 1D and 2D CNN models. **a** Nankai and **b** Sakurajima datasets. Open bars show BACC scores when applied to the validation dataset and filled bars with corresponding values reported show BACC scores when applied to the test dataset

overall BACC value for both validation datasets, indicating that the 2D model using running spectral images performed better than the 1D models using waveform traces. However, when generalized to the test dataset, the BACC value of model 1 was very similar to that of the 2D CNN for both datasets. This result implies that, when generalized, the performances of the 1D and 2D CNNs were comparable.

Table 2 shows the confusion matrices of the 2D and 1D CNN model 1 for the Nankai dataset (those for the Sakurajima dataset are reported in Additional file 1: Table S4). In the Nankai dataset, 94 and 121 samples were misclassified by the 2D and 1D CNNs, respectively; of those, 50 samples from 20 earthquakes or tremor events were misclassified to the same class by both CNNs. In the Sakurajima dataset, 1129 and 1257 samples were misclassified by the 2D and 1D CNNs, respectively; of those, 731 samples from 156 ER/NER events were misclassified to the same class by both CNNs. These results imply that the 2D and 1D CNNs learned similar features inherent to the input signals and misclassified these samples similarly.

Computation times taken for training the CNN parameters and classifications of one sample are summarized in Additional file 1: Table S5. The training times for the 1D models were three to four times longer than that for the 2D model. Classification of one sample takes about 1 ms, but the 1D models take two to three times longer than the 2D model.

Discussion

Our results show that the 1D and 2D CNNs performed comparably. Similar results were obtained by Canário et al. (2020), who used only 10 frequency components in their time–frequency-domain data. Although the running spectra (2D) explicitly represent dominant frequency components and durations characteristic to the signals, the 1D CNN seems to have successfully learned the features necessary for effectively discriminating between the signal classes. Therefore, both time-domain and time–frequency-domain data can be used to classify seismic signals and achieve similar predictive performance. However, running spectral images lose some information, especially the long-period components of

seismic data. Very-low-frequency earthquakes occurring at subduction zones (e.g., Ito et al. 2007; Nakano et al. 2018) and very-long-period events at volcanoes (e.g., Chouet 1996) are dominated by signals longer than tens of seconds, which were lost from the running spectra used in this study but retained in the waveform traces. Using 1D CNNs may therefore result in better predictive performance when discriminating these signals. Of course, running spectral images can include long-period components if they are created with a proper time window. Although 1D CNNs take several times longer for signal classifications than 2D CNNs (Additional file 1: Table S5), the computation time of about 1 ms is short enough for real-time applications. Because 1D CNNs do not require the same preprocessing as running spectral computations, they are better suited to real-time seismic monitoring than 2D CNNs.

The convolutional layer of a 1D CNN can be written as:

$$u_{i,k} = f \left(b_k + \sum_{i'=-m}^m \sum_{k'=1}^n a_{i',k',k} z_{i+i',k'} \right), \quad (2)$$

where $z_{i,k}$ and $u_{j,k}$ are the i th data of k th channel in the input and output of the layer, respectively, $a_{i',k',k}$ is a kernel applied to the input, b_k is the bias on the k th channel, m is the kernel size, n is the number of channels in the input layer, and $f()$ is an activation function. By setting $f(x) = x$, $b_k = 0$, and $n = 1$, Eq. (2) simplifies to:

$$u_i = \sum_{i'=-m}^m a_{i',k,k} z_{i+i'}, \quad (3)$$

which constitutes a finite impulse response (FIR) filter of length m . Because the weights and bias are tuned for signal classification during training, 1D CNNs are considered to learn the discriminating frequency components of the input signals.

In our application, the difference in signal power between the low- and high-frequency components may be key to discriminating between the signals because, in both the Nankai and Sakurajima datasets, the two earthquake classes were dominated by high- or low-frequency components. As shown by Eq. (3), the convolutional layer

Table 2 Confusion matrices for 2D CNN and 1D CNN model 1 to the Nankai trough dataset

| Class | 2D model predicted event class | | | 1D model 1 predicted event class | | |
|------------|--------------------------------|--------|-------|----------------------------------|--------|-------|
| | Earthquake | Tremor | Noise | Earthquake | Tremor | Noise |
| Earthquake | 915 | 32 | 25 | 901 | 28 | 43 |
| Tremor | 21 | 717 | 4 | 23 | 712 | 7 |
| Noise | 5 | 7 | 864 | 13 | 7 | 856 |

of a 1D CNN is basically equivalent to a set of FIR filters. In the first convolutional layer, the input seismic waveforms are filtered by the FIR filters, then passed into the following layer after application of the activation function. Therefore, signals in the passband of the FIR filters in the first convolutional layer are used for classifications in the first step. To check this hypothesis, we tested a CNN model with only two channels in the first layer (1D CNN model 3) and computed the FIR filter responses from the channel weights. We note that performance was not significantly degraded for this model (Fig. 2). Figure 3 shows the responses of the FIR filters in the two channels of the first convolutional layer; they were band-rejection filters with a stopband at 4–6 Hz and different amplitude responses for the higher frequency components. This result implies that the differences in the frequency components above and below 4–6 Hz were important for signal discriminations in our dataset.

To visualize the difference in the frequency components of the input signals, we computed average spectra for each signal class in the Nankai and Sakurajima datasets (Fig. 3e, f) by normalizing the waveforms to their absolute maximum amplitude, performing a Fourier transform, and then averaging the signals for each class. Tectonic tremor and ER events were dominated by frequency components lower than 5–7 Hz, whereas local earthquakes and NER events were dominated by frequency components above 7 Hz. Noise dominated frequencies below 1 Hz and above 10 Hz in the Nankai dataset, whereas in the Sakurajima dataset noise showed lower signal levels than NER events below 6 Hz. Therefore, differences in the signal levels at higher and lower frequencies seem key to accurate signal classifications, as in the frequency scanning method to discriminate tectonic tremor from ordinary earthquakes (e.g., Katakami et al. 2017; Sit et al. 2012). We confirmed that this property is retained by CNNs using many channels in the first

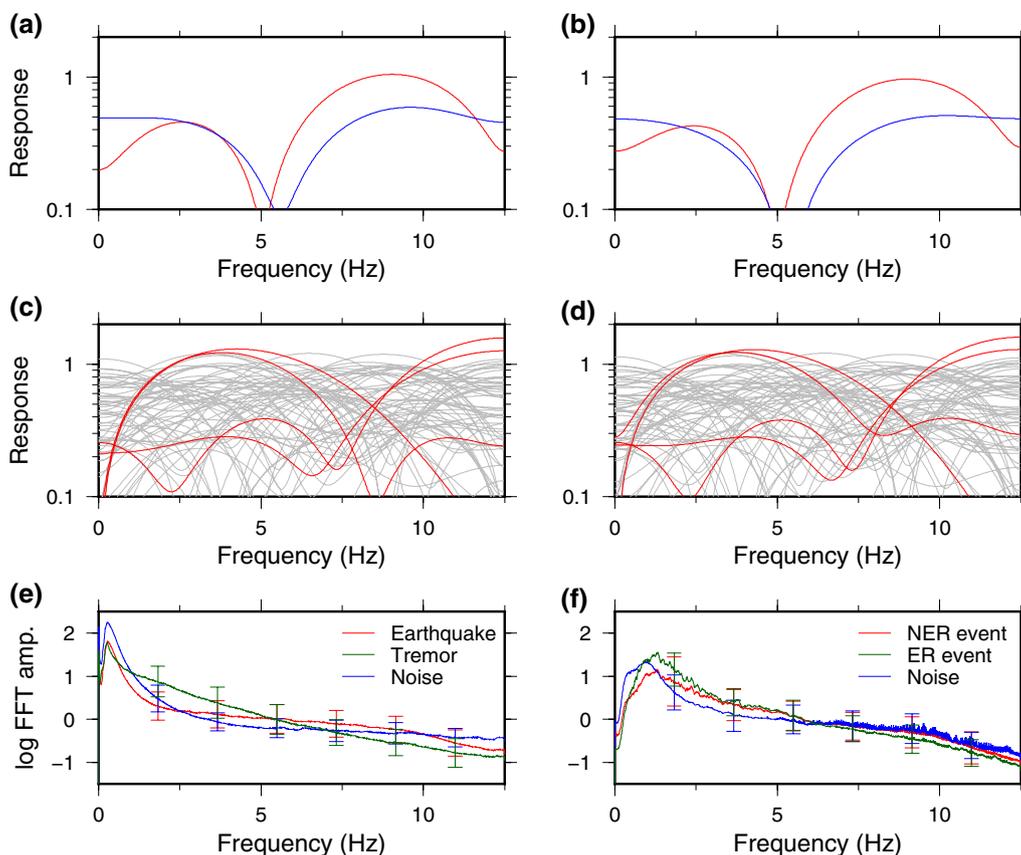


Fig. 3 FIR filter responses in the first convolutional layer and average spectra of each signal class. **a, c, e** Nankai and **b, d, f** Sakurajima datasets. FIR filter response curves are computed from the weights of **a, b** the two channels in the first convolutional layer of 1D CNN model 3 and **c, d** the 64 channels in the first layer of 1D CNN model 1 (see Table S2). In **a, b** red and blue lines indicate response curves for the first and second channels, respectively. In **c, d** response curves for channels with the four largest amplitudes are shown by red lines, and other channels by gray lines. **e, f** Average spectra of each signal class

convolutional layer; Fig. 3 shows the FIR filter responses computed for the first convolutional layer (64 channels) of the 1D CNN model 1. Although the FIR filters showed various responses with different passbands and stopbands that covered the entire frequency range, filters with the largest amplitude responses comprised low- and high-pass filters with cutoffs around 7–9 Hz. Again, the difference between the higher and lower frequency components arises as fundamental information for signal classification.

Although the filter responses of the first convolutional layer correspond to the characteristics of the input signal, interpreting deeper layers is not straightforward. If we similarly compute the filter responses of the deeper convolutional layers, we obtain responses only for low-frequency components because the pooling layer down-samples the data. These filters were applied to the input data after non-linear conversions by the activation function and max pooling, and the outputs from different channels in the previous layer were added. Therefore, certain aspects of neural networks remain a ‘black-box’.

So far, we have focused on the amplitude responses of the filters, but the phase characteristics of the filters and the appearance of signals in different channels should also play important roles for different applications. For example, Ross et al. (2018b) developed a CNN to estimate P- and S-wave arrival times. Although the dominant frequencies of S-waves may be lower than those of P-waves due to attenuation at higher frequencies, these waves generally have different signal durations and appearances in three-component seismograms that may be important for their identification. Therefore, the correspondence between signal characteristics and FIR filter responses may be limited to applications such as in this study.

We used the same Nankai dataset and 2D CNN structure as in our previous study (Nakano et al. 2019), but the performance upon generalization was lower in this study: the earlier study achieved an accuracy of 0.995, whereas we obtained $BACC=0.965$. This difference is because Nakano et al. (2019) fixed the CNN model and used the same dataset for validation and test, resulting in generally better performance. When applied to the validation dataset in this study, the model performance attained $BACC=0.981$ (Fig. 2a). In addition, Nakano et al. (2019) removed low-quality data by visual inspection, whereas we removed low-quality data based on the S/N computed from waveform traces. We note that most of our misclassified data had small S/N . Some local earthquake data that were misclassified as tectonic tremor were dominated by lower frequency components at all stations, and this signal characteristics may be due to source properties or path effects. Such data might have been noticed and removed by visual inspections, but not based on the

S/N as in this study. However, visual inspection is not practical in automated seismic monitoring systems, and the performance attained in this study based by objectively selecting data should be realistic.

The $BACC$ for the Sakurajima dataset was slightly lower than that for the Nankai dataset. When applied to the validation dataset for Sakurajima, the model performance attained $BACC=0.943$. Misclassifications mostly occurred between NER and ER events; noise was rarely misclassified and other events were rarely misclassified as noise (Additional file 1: Table S4). Misclassified NER events had greater lower frequency components than other events in the same class, i.e., they had signal characteristics similar to ER events. It is possible that seismic signals from small, uncatalogued eruptions were included in the NER dataset, or that the NER dataset was more variable because of the variety of possible source processes and path effects at volcanoes. Future work should therefore seek higher performance CNN models to resolve this problem.

Concluding remarks

We compared the performances of 1D and 2D CNNs (using waveform traces and running spectral images as input data, respectively) in classifying seismic signals from two different tectonic settings: tectonic tremor vs. ordinary earthquakes at Nankai trough and eruptive vs. non-eruptive volcanic earthquakes at Sakurajima. In both applications, the 1D and 2D CNNs performed similarly, indicating that the data preprocessing to produce the time–frequency-domain representation is not necessary to achieve high performance signal discriminations. We cannot exclude the possibility that different results will be obtained for different datasets or when other network structures are used because the behaviors of neural networks remain difficult to predict.

Because the signals in our datasets were dominated by high- and low-frequency components, we were able to determine which frequency components were effective for signal classifications using the 1D CNN. The first convolutional layer applies a set of FIR filters to the input seismogram and the filtered seismograms are passed into deeper layers, meaning that the pass-band frequencies of the filters are effective for signal discrimination. Our application to the Nankai and Sakurajima datasets showed that the filters constituted low- and high-pass filters with cut-off frequencies around 7–9 Hz: tectonic tremor events (Nankai) and eruptive earthquakes (Sakurajima) were dominated by lower frequencies, and local earthquakes (Nankai) and non-eruptive earthquakes (Sakurajima) by higher frequencies.

These results may help to reveal useful characteristics of input signal classes and to improve the performance of future signal discrimination models.

Abbreviations

ML: Machine learning; CNN: Convolutional neural network; VT: Volcano-tectonic; LP: Long-period; JMA: Japan Meteorological Agency; S/N: Signal-to-noise ratios; ER: Eruptive; NER: Non-eruptive; BACC: Balanced accuracy; FIR: Finite impulse response.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40623-022-01696-1>.

Additional file 1: Fig. S1 Map showing the distribution of DONET stations (gray triangles), hypocenters of regular earthquakes (circles), and slow earthquakes (inverted triangles) used in this study. **Fig. S2** Map showing the distribution of JMA seismic stations (gray triangles) and location of earthquakes in 2015 determined by JMA (orange dots). Red triangles are the locations of summit craters. **Table S1** Architecture of the 2D CNN model. **Table S2** Architectures of the 1D CNN models. **Table S3** Number of trainable parameters. **Table S4** Confusion matrices for 2D CNN and 1D CNN model 1 to the Sakurajima dataset. **Table S5** Computation times for training and classifications.

Acknowledgements

We used waveform data from DONET and the Japan Meteorological Agency (JMA) observation network at Sakurajima (National Research Institute for Earth Science and Disaster Resilience 2019). We also used the Sakurajima eruption catalog created by the Kagoshima Meteorological Office of JMA. All figures were drawn using Generic Mapping Tools (Wessel and Smith 1998). We thank two anonymous reviewers and the editor N. Uchida for careful review and constructive comments, which have improved the manuscript.

Author contributions

MN designed this paper and performed the analysis. DS developed the computation code for the CNNs. All authors read and approved the final manuscript.

Funding

This study was supported by JSPS KAKENHI Grant Number JP19K04050 and JP21H05205 (to MN).

Availability of data and materials

Seismic waveform data from DONET and the JMA observation network at Sakurajima are available at <http://www.hinet.bosai.go.jp/?LANG=en>. The Sakurajima eruption catalog of the Kagoshima Meteorological Office of JMA is available at https://www.jma-net.go.jp/kagoshima/vol/kazan_top.html.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no conflicts of interest associated with this manuscript.

Author details

¹Institute for Marine Geodynamics (IMG), Japan Agency for Marine-Earth Science and Technology, 2-15, Natsushima-cho, Yokosuka, Kanagawa 237-0061, Japan. ²Research Institute for Value-Added Information Generation (VAiG), Japan Agency for Marine-Earth Science and Technology, 3173-25 Showa-machi, Kanazawa-ku, Yokohama 236-0001, Japan.

Received: 6 July 2022 Accepted: 27 August 2022

Published online: 08 September 2022

References

- Aoi S, Asano Y, Kunugi T, Kimura T, Uehira K, Takahashi N, Ueda H, Shiomi K, Matsumoto T, Fujiwara H (2020) MOWLAS: NIED observation network for earthquake, tsunami and volcano. *Earth Planet Space* 72:126. <https://doi.org/10.1186/s40623-020-01250-x>
- Canário JP, Mello R, Curilemc M, Huenupan F, Rios R (2020) In-depth comparison of deep artificial neural network architectures on seismic events classification. *J Volcanol Geotherm Res* 401:106881. <https://doi.org/10.1016/j.jvolgeores.2020.106881>
- Chouet BA (1996) Long-period volcano seismicity: its source and use in eruption forecasting. *Nature* 380:309–316
- Del Pezzo E, Esposito A, Giudicepietro F, Marinaro M, Martini M, Scarpetta S (2003) Discrimination of earthquakes and underwater explosions using neural networks. *Bull Seismo Soc Am* 93:215–223
- Dokht RMH, Kao H, Visser R, Smith B (2019) Seismic event and phase detection using time–frequency representation and convolutional neural networks. *Seismo Res Lett* 90(2A):481–490. <https://doi.org/10.1785/0220180308>
- Dowla FU, Taylor SR, Anderson RW (1990) Seismic discrimination with artificial neural networks: preliminary results with regional spectral data. *Bull Seismo Soc Am* 80:1346–1373
- Fukushima K (1980) Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol Cybern* 36:193–202. <https://doi.org/10.1007/BF00344251>
- Holtzman BK, Paté A, Paisley J, Waldhauser F, Repetto D (2018) Machine learning reveals cyclic changes in seismic source spectra in Geysers geothermal field. *Sci Adv* 4:eaa02929. <https://doi.org/10.1126/sciadv.aao2929>
- Iguchi M (1994) A vertical expansion source model for the mechanisms of earthquakes originated in the magma conduit of an andesitic volcano: Sakurajima, Japan. *Bull Volcanol Soc Japan* 39:49–67
- Ito Y, Obara K, Shiomi K, Sekine S, Hirose H (2007) Slow earthquakes coincident with episodic tremors and slow slip events. *Science* 315:503–506. <https://doi.org/10.1126/science.1134454>
- Kaneda Y, Kawaguchi K, Araki E, Matsumoto H, Nakamura T, Kamiya S, Ariyoshi K, Hori T, Baba T, Takahashi N (2015) Development and application of an advanced ocean floor network system for megathrust earthquakes and tsunamis. In: Paolo F, Laura B, Angelo DS (eds) *Seafloor observatories*. Springer, Berlin, pp 643–662 (10.1007/978-3-642-11374-1_25)
- Katakami S, Yamashita Y, Yakhara H, Shimizu H, Ito Y, Ohta K (2017) Tidal response in shallow tectonic tremors. *Geophys Res Lett* 44:9699–9706. <https://doi.org/10.1002/2017GL074060>
- Kawaguchi K, Kaneko S, Nishida T, Komine T (2015) Construction of the DONET real-time seafloor observatory for earthquakes and tsunami monitoring. In: Paolo F, Laura B, Angelo DS (eds) *Seafloor observatories*. Springer, Berlin, pp 211–228 (10.1007/978-3-642-11374-1_10)
- Kerr RA (2013) Geophysical exploration linking deep earth and backyard geology. *Science* 340:1283–1285. <https://doi.org/10.1126/science.340.6138.1283>
- Kingma DP, Ba J (2015) Adam: a method for stochastic optimization. In: *Proceedings of the 3rd international conference on learning representations (ICLR2015)*. Hilton San Diego Resort-Spa, San Diego, 7–9 May 2015, Accessed October 2018. Available at <http://arxiv.org/abs/1412.6980>.
- Kong Q, Allen RM, Schreier L, Kwon YW (2016) MyShake: a smartphone seismic network for earthquake early warning and beyond. *Sci Adv* 2:e1501055. <https://doi.org/10.1126/sciadv.1501055>
- LeCun Y, Bengio Y, Hinton G (2015) Deep learning. *Nature* 521:436–444. <https://doi.org/10.1038/nature14539>
- Mousavi SM, Zhu W, Sheng Y, Beroza GC (2019) CRED: a deep residual network of convolutional and recurrent units for earthquake signal detection. *Sci Rep* 9:10267. <https://doi.org/10.1038/s41598-019-45748-1>
- Mousavi SM, Ellsworth WL, Zhu W, Chuang LY, Beroza GC (2020) Earthquake transformer—an attentive deeplearning model for simultaneous earthquake detection and phase picking. *Nat Comm* 11:3952. <https://doi.org/10.1038/s41467-020-17591-w>
- Nakano M, Hori T, Araki E, Kodaira S, Ide S (2018) Shallow very-low-frequency earthquakes accompany slow slip events in the Nankai subduction zone. *Nat Comm* 9:984. <https://doi.org/10.1038/s41467-018-03431-5>

- Nakano M, Sugiyama D, Hori T, Kuwatani T, Tsuboi S (2019) Discrimination of seismic signals from earthquakes and tectonic tremor by applying a convolutional neural network to running spectral images. *Seism Res Lett* 90(2A):530–538. <https://doi.org/10.1785/0220180279>
- National Research Institute for Earth Science and Disaster Resilience (2019) NIED MOWLAS National Research Institute for Earth Science and Disaster Resilience. <https://doi.org/10.17598/NIED.0009>
- Obara K (2002) Nonvolcanic deep tremor associated with subduction in Southwest Japan. *Science* 296:1679–1681. <https://doi.org/10.1126/science.1070378>
- Perol T, Gharbi M, Denolle M (2018) Convolutional neural network for earthquake detection and location. *Sci Adv* 4:e1700578. <https://doi.org/10.1126/sciadv.1700578>
- Ross ZE, Meier MA, Hauksson E (2018a) P wave arrival picking and first-motion polarity determination with deep learning. *J Geophys Res* 123:5120–5129. <https://doi.org/10.1029/2017JB015251>
- Ross ZE, Meier MA, Hauksson E, Heaton TH (2018b) Generalized seismic phase detection with deep learning. *Bull Seism Soc Am* 108:2894–2901. <https://doi.org/10.1785/0120180080>
- Rouet-Leduc B, Hulbert C, McBrearty IW, Johnson PA (2020) Probing slow earthquakes with deep learning. *Geophys Res Lett* 47:e2019GL085870. <https://doi.org/10.1029/2019GL085870>
- 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 Seismo Soc Am* 95:185–196. <https://doi.org/10.1785/0120030075>
- Sit S, Brudzinski M, Kao H (2012) Detecting tectonic tremor through frequency scanning at a single station: application to the Cascadia margin. *Earth Planet Sci Lett* 353–354:134–144. <https://doi.org/10.1016/j.epsl.2012.08.002>
- Soto H, Schurr B (2021) DeepPhasePick: a method for detecting and picking seismic phases from local earthquakes based on highly optimized convolutional and recurrent deep neural networks. *Geophys J Int* 227:1268–1294. <https://doi.org/10.1093/gji/ggab266>
- Sugiyama D, Tsuboi S, Yukutake Y (2021) Application of deep learning-based neural networks using theoretical seismograms as training data for locating earthquakes in the Hakone volcanic region. *Japan Earth Planet Space* 73:135. <https://doi.org/10.1186/s40623-021-01461-w>
- Takahashi H, Tateiwa K, Yano K, Kano M (2021) A convolutional neural network-based classification of local earthquakes and tectonic tremors in Sanriku-oki, Japan, using S-net data. *Earth Planet Space* 73:186. <https://doi.org/10.1186/s40623-021-01524-y>
- Wang J, Teng TL (1995) Artificial neural network-based seismic detector. *Bull Seismo Soc Am* 85:308–319
- Wessel P, Smith WHF (1998) New, improved version of generic mapping tools released. *Eos* 79:579
- Yoon CE, O'Reilly O, Bergen KJ, Beroza CG (2015) Earthquake detection through computationally efficient similarity search. *Sci Adv* 1:e1501057. <https://doi.org/10.1126/sciadv.1501057>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)
