Earth, Planets and Space

FULL PAPER

Open Access

Development of a high-performance seismic phase picker using deep learning in the Hakone volcanic area

Ahyi Kim[1*], Yuji Nakamura[1], Yohei Yukutake[2], Hiroki Uematsu[3] and Yuki Abe[4]

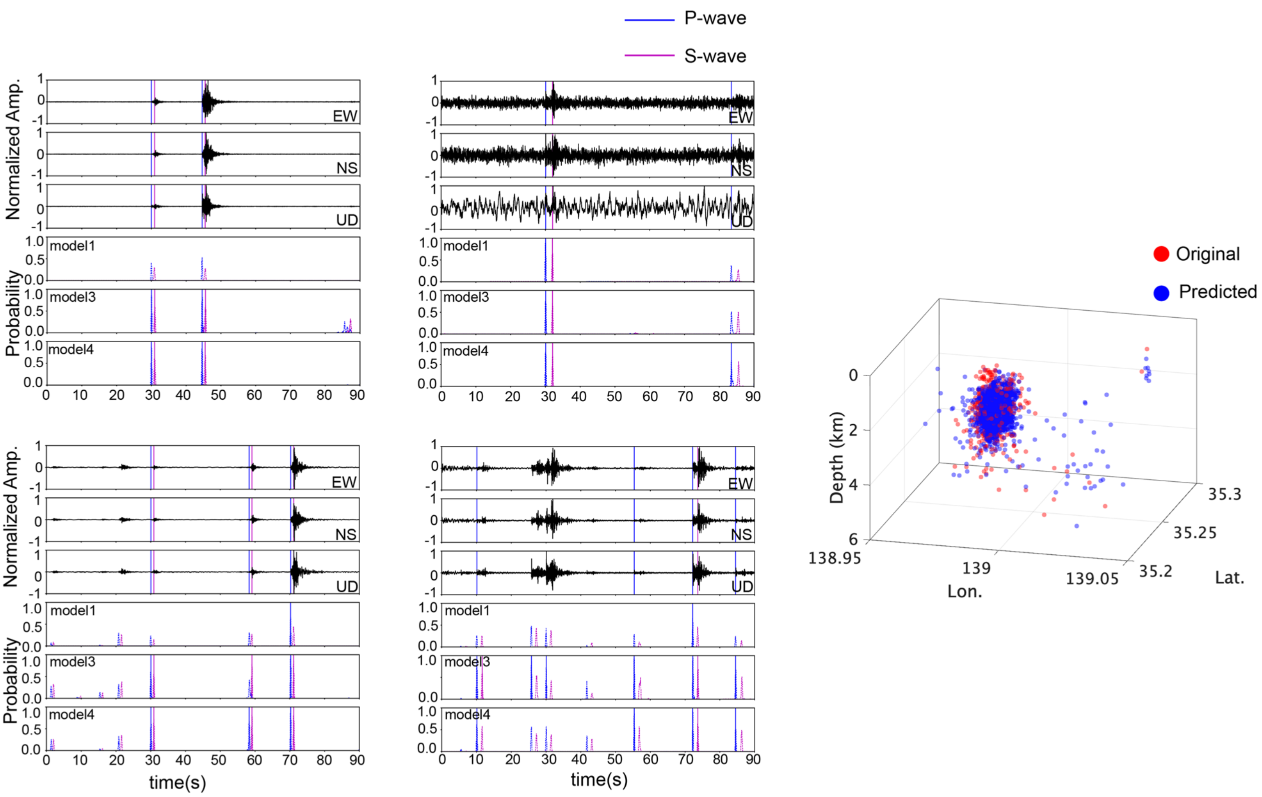
Abstract

In volcanic regions, active earthquake swarms often occur in association with volcanic activity, and their rapid detection and analysis are crucial for volcano disaster prevention. Currently, these processes are ultimately left to human judgment and require significant time and money, making detailed real-time verification impossible. To overcome this issue, we attempted to apply machine learning, which has been successfully applied to various seismological fields to date. For seismic phase pick, several models have already been trained using a large amount of training data (mainly crustal earthquakes). Although there are some cases in which these models can be applied without any problems, regional dependence on pre-trained models has been reported. Since this study targets earthquakes in a volcanic region, applying existing pre-trained models may be difficult. Therefore, in this study, we compared three models; the publicly available trained model (model 0), a model which was trained with approximately 220,000 P- and S-wave onset reading data recorded at the Hakone volcano from 1999 to 2020 with initialized parameters (model 1) using the same architecture, and a model fine-tuned with the aforementioned Hakone data using the parameters of model 0 as initial values (model 2), and evaluated their phase identification performance for the Hakone data. As a result, the seismic phase detection rates of models 1 and 2 were much higher than those of model 0. However, small-amplitude signals are often missed when multiple seismic events occur within a detection time window. Therefore, we created training data with two earthquakes in the same time window, retrained the model using the data, and successfully detected events that previously would have been missed. In addition, it was found that more events were detected by setting the threshold to a low probability value for detection, increasing the number of seismic phase detections, and filtering by phase association and hypocenter location.

Keywords Deep learning, U-Net, Fully convolutional network, Machine learning, Phase pick, Volcanic earthquake

*Correspondence:
Ahyi Kim
ahyik@yokohama-cu.ac.jp
Full list of author information is available at the end of the article

Graphical Abstract



Introduction

In volcanic regions, active earthquake swarms are often associated with volcanic activities, and their rapid detection and analysis are vital for disaster prevention. In addition, understanding seismic activities and their mechanisms is also essential for mitigating future disasters. These tasks begin with detecting and identifying seismic phases but those tasks mainly rely on human labor and are both time-consuming and labor-intensive. The general method for estimating the onset of a seismic phase involves applying the autoregressive Akaike information criterion (AR-AIC) picker (Akazawa 2004) to events detected by the short- (STA) and long-term average (LTA) methods (Allen 1978). However, the final decision is typically left to human labor.

Various machine learning methods—especially deep neural networks—have been developed to detect and pick seismic phase arrivals. Early studies used multilayer perceptron networks to detect earthquakes (e.g., Dysart and Pulli 1990; Musil and Plešinger 1996; Fedorenko et al. 1999; Ursino et al. 2001; Kong et al. 2016), and phase-picking techniques have been developed (e.g., Dai and MacBeth 1995; Tiira 1999; Zhao and Takano 1999; Wiszniowski et al. 2014). Subsequently, seismic event detection and phase-picking techniques were developed using a convolutional neural network (CNN) with convolutional and pooling layers added to deep learning (e.g., Perol et al. 2018; Ross et al. 2018). Zhu and Beroza (2019) used the training data of millions of seismograms picked manually throughout northern California and built a model to detect P- and S-waves using the U-Net architecture (Ronneberger and Fischer 2015), a fully convolutional network (FCN). An EQ transformer was also developed by introducing the attention mechanism used in the transformer (Vaswani et al. 2017) into a conventional FCN to perform the detection of earthquake signals and phase picking simultaneously (Mousavi et al. 2020). Most of the above-described codes and pre-trained models are publicly available on GitHub (https://github.com/) and other websites for use by anyone. However, it has also been known that these models have regional characteristics, where the detection performance varies depending on the region in which they are applied (Münchmeyer et al. 2022).

Furthermore, most of the above pre-trained models use crustal earthquakes as training data, which may make them unsuitable for detecting earthquakes proximate to volcanoes. Earthquakes in volcanic regions are characterized by active earthquake swarms, and it may be necessary to build models suitable for such environments. However, the problem is that few volcanoes have accumulated the vast amount of data required to train using deep learning. For example, Lapins et al. (2021) used transfer learning from the pre-trained model of Ross et al. (2018) to improve the performance of earthquake phase detection in areas with just a small amount of data.

On the other hand, Hakone—the target area of this study—has over 20 years of accumulated observation data to be trained from scratch. Therefore, this study uses the pre-trained model and architecture of Zhu and Beroza (2019), which has shown one of the highest performances in previous studies (Münchmeyer et al. 2022). We created models using the PhaseNet pre-trained model (model 0) and its architecture, and we evaluated their performance under various conditions.

Seismic activities and observations network in Hakone

Hakone volcano is located in central Japan's northern boundary zone of the Izu–Bonin–Mariana volcanic arc. It has a caldera topography surrounded by an outer ring of about 10 km diameter (Fig. 1). The most recent large-scale eruption with lava ejecta was about 3000 years ago. Although no such large-scale eruption has occurred, since some studies indicate that a phreatomagmatic eruption happened between the twelfth and thirteenth centuries (e.g., Kobayashi et al. 2006). In addition, a very small eruption occurred in 2015 at Owakudani for the first time in the Hakone volcano's recorded history (e.g., Mannen et al. 2018). The Hakone volcano is still active today, with earthquakes occurring very superficially in the crust of the caldera and occasional large swarms of earthquakes. In this context, the hot spring research institute of Kanagawa Prefecture (HSRI) installed a short-period seismometer with a 200 Hz sampling rate and has been observing and studying the causes of these seismic events since 1989 (Fig. 1). Typical seismic activity since then includes active earthquake swarms in 2001, 2006, 2008, and 2009 (e.g., Yukutake et al. 2011a), and active seismic swarms associated with phreatic eruptions were observed in 2015 (e.g., Mannen et al. 2018). The mechanism of the seismic swarms at the Hakone volcano has also been extensively studied. The prevailing theory is that they are caused by the migration process of high-temperature, high-pressure hydrothermal water supplied from deep underground through fractures, such as micro-faults (Yukutake et al. 2011a). Seismic activity also increased after the 2011 M9.0 Tohoku–oki earthquake, although with different characteristics from the above earthquake swarm (Yukutake et al. 2011b).

Although seismic observations at Hakone began in 1989, we used data from 1999, because it was from that year that the WIN system (Urabe and Tsukada 1992), the standard used in Japan for continuous waveform recording, began recording and the seismic catalog was properly maintained. In fact, the catalog has been in place since 1995 but the accuracy of the onset readings was not very good initially, and there were fewer earthquakes in Hakone, so we have chosen to use data from 1999 onward. In this study, we used 217,553 seismic waveforms that contain one P- and S-wave onset reading per data set recorded by the above seismic network in April 1999–December 2020, as training data. Figure 1 shows seismic stations, which include the National Research Institute for Earth Science and Disaster Prevention (NIED) Hi-net (Obara et al. 2005) as well as stations operated by HSRI.

Construction of a seismic phase picker

In this study, we used the PhaseNet architecture constructed by Zhu and Beroza (2019). The method is based on the structure of U-Net (Ronneberger and Fischer 2015), a type of encoder–decoder FCN developed for biomedical image processing and modified to handle 1D time-series data. In the decoder part, convolution and upsampling are repeatedly applied to the input data to improve the resolution of the feature map and obtain the desired region extraction results. PhaseNet takes a three-component seismic waveform as input and outputs the probability of P- and S-wave onsets and noise per sample. See Zhu and Beroza (2019) for more details about the model.

While CNN classifies given unknown data, U-Net aims toward segmentation, so it does not matter how many events are in the detection window. In addition, one of the largest differences between CNN and U-Net is that the length of the input data must be the same as the training data for CNN, because it has a fully connected layer at the end but it is variable for U-Net which lacks one.

In this study, we used seismic events recorded from 1999 to 2020 in the Hakone volcanic area mentioned in the previous section. Experts have manually inspected this data for P- and S-waves and we will treat the information as the ground truth. In many cases, the picked onsets are stored with the P-wave at the same position in the time window, and the position could be incorrectly learned as the P-wave onset. To avoid this case, the P- and S-waves were always included in PhaseNet and the waveforms are randomly shifted back and forth

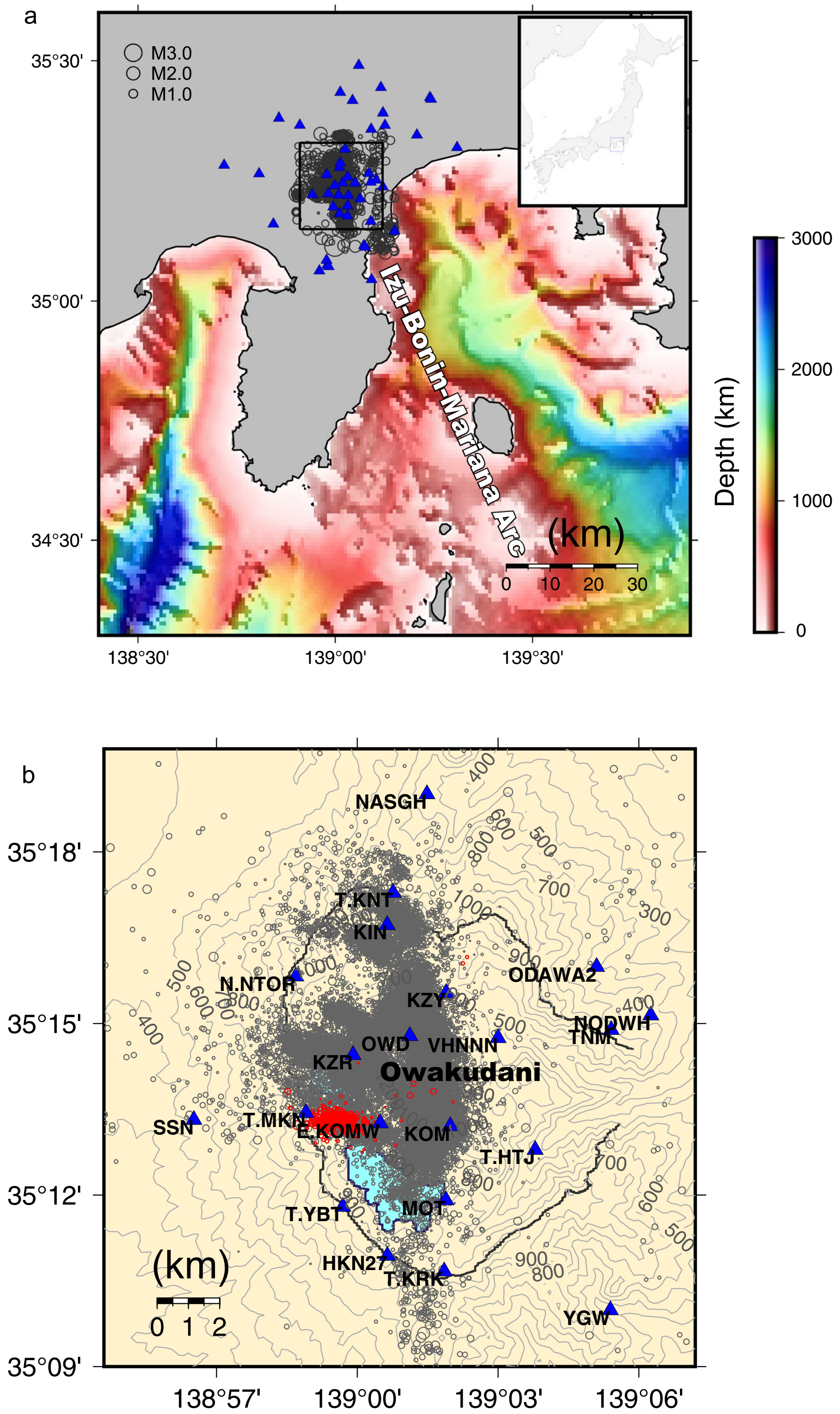


Fig. 1 Earthquakes and the distribution of seismic stations in the Hakone volcanic area. **a** Earthquakes (gray circles) and seismic stations (blue triangles) in May 1999–June 2021. The earthquake symbol size is scaled by the magnitude. **b** Enlarged view of the area enclosed by the square in **a**. The red circles represent the distribution of earthquakes in May 18–20, 2019 used for the test data

in the position of the P-wave onset. In this study, as in the original, each seismic waveform was cut out of a 90-s time window; after shifting them in the above manner, they were cut to 30 s and used as input data. Other than that, the original 200 Hz sampling was downsampled to 100 Hz as preprocessing.

In this study, three models were first created, and their performances were evaluated as follows:

(1) Model 0: A trained model built using training data from millions of crustal earthquakes in Northern California (Zhu and Beroza 2019; publicly available on GitHub).

(2) Model 1: A model trained from scratch using the PhaseNet architecture with the Hakone seismic data.

(3) Model 2: A fine-tuned model with the Hakone volcano training data used in model 1 and the parameters in model 0 as initial values.

In training models 1 and 2, 217,553 data were randomly assigned as follows: 80% as training data and 20% as validation data. To determine the best parameters for each model, their performances were verified by varying the learning rate and batch size and repeating the computation up to epoch 100 (Fig. 2). The results showed that the best parameters for model 1 were those with a batch size of 64 and a learning rate of 0.01, while the best parameters for model 2 were those with a batch size of 16 and a learning rate of 0.1. The best F1 scores were almost the same for two models (Fig. 3). From now on, the models with the best parameters for both will be called models 1 and 2, respectively. Using the same data for validation, the F1 values for model 0 were calculated as 0.832 for the P-wave and 0.707 for the S-wave, mainly showing a significant improvement in the S-wave picking.

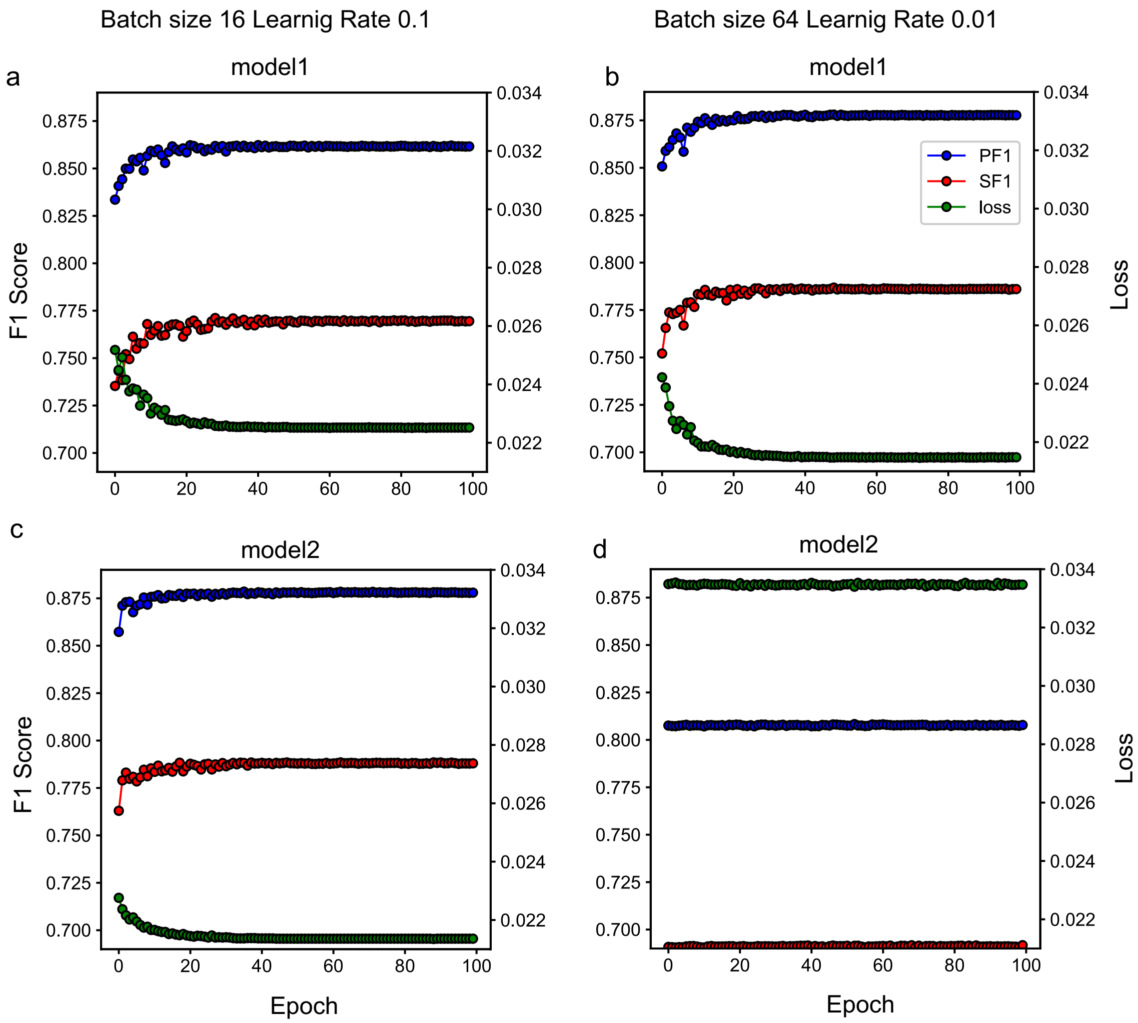


Fig. 2 Examples of F1 score for P-wave (blue) and S-wave (red) and loss (green) for different hyperparameters. **a** Model 1, batch size 16, learning rate 0.1, **b** model 1, batch size 64, learning rate 0.01, **c** model 2, batch size 16, learning rate 0.1, **d** model 2, batch size 64, learning rate 0.01

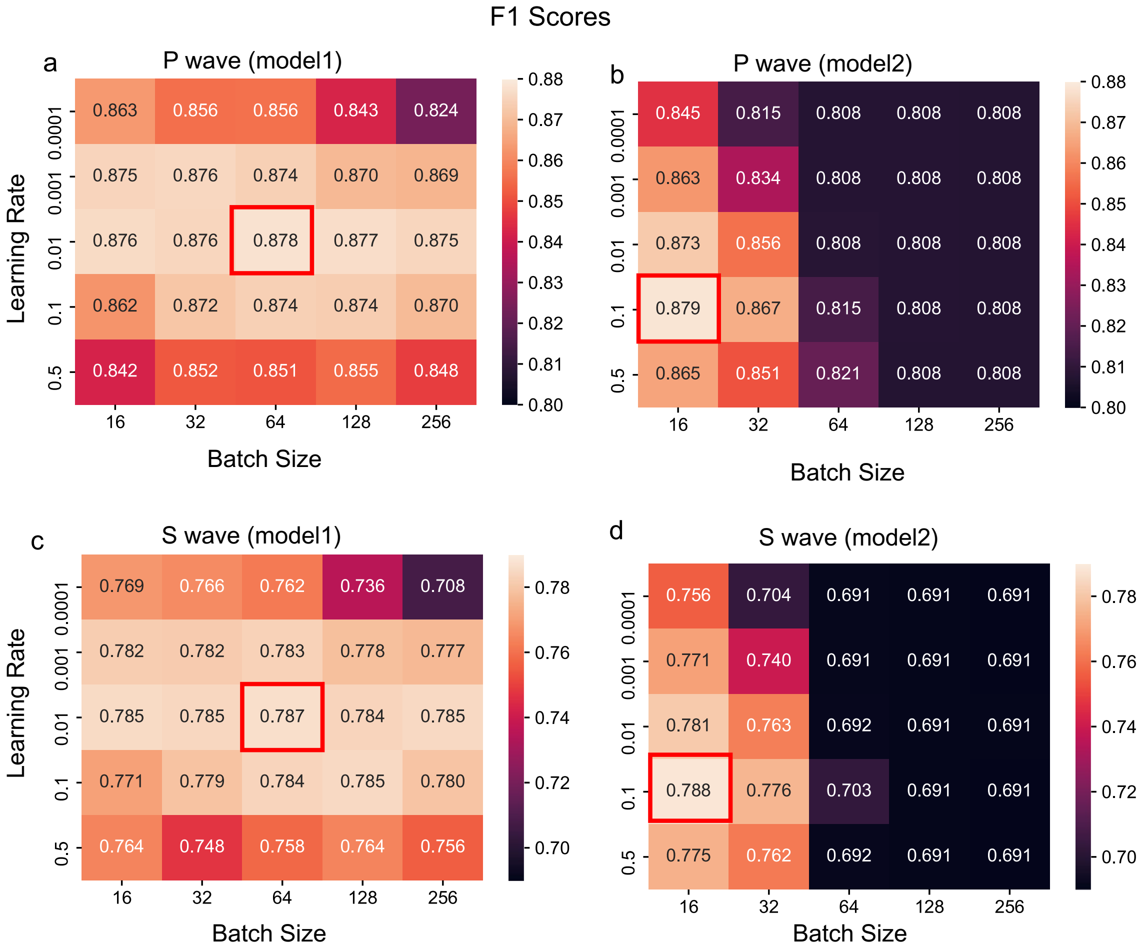


Fig. 3 Heat maps of the highest F1 score. The F1 scores were obtained when the learning rate and batch size were varied and calculated up to epoch 100. **a** model 1 (P-wave), **b** model 2 (P-wave), **c** model 1 (S-wave), **d** model 2 (S-wave)

An examples of the heatmap for F1 scores with respect to various batch size and the learning rate are shown in Fig. 3.

Application to test data

These models were applied to 5559 seismic phase data (441 events) from swarm earthquakes that occurred May 18–20, 2019 but were not used for training or validation. The detection time window to pick the phase onset is 90 s and the detection threshold is a probability of 0.6. The F1 scores of the P-wave for models 0, 1, and 2 are 0.823, 0.860, and 0.857 and those for the S-wave are 0.641, 0.755, and 0.749, respectively. As with the validation data, models 1 and 2 showed a better F1 score, indicating that the two models trained by Hakone data performed better than model 0. There was almost no difference in travel time between the three models compared to human-operated data (Fig. 4). Figure 5 shows an example of phase picking. The results indicate that each model can detect earthquakes with a high probability when there is only one event or when there are multiple events with similar amplitudes in the detection window (Fig. 5a, b). In addition, each model can detect seismic events with a low signal-to-noise ratio (Fig. 5d).

On the other hand, when there were multiple earthquakes with significantly different amplitudes in the detection time window, model 0 detects the event with a larger amplitude and higher probability. Models 1 and 2 still have probability amplification for smaller amplitude events, but the probability of larger amplitude events is often lower than that in model 0 (Fig. 5c, e, and f). The above results suggest that the reason for all models' slightly lower performance in the test data compared to the validation data is that the test data are from one of the most active swarms in Hakone and thus contain many earthquakes in the detection window, missing some earthquakes.

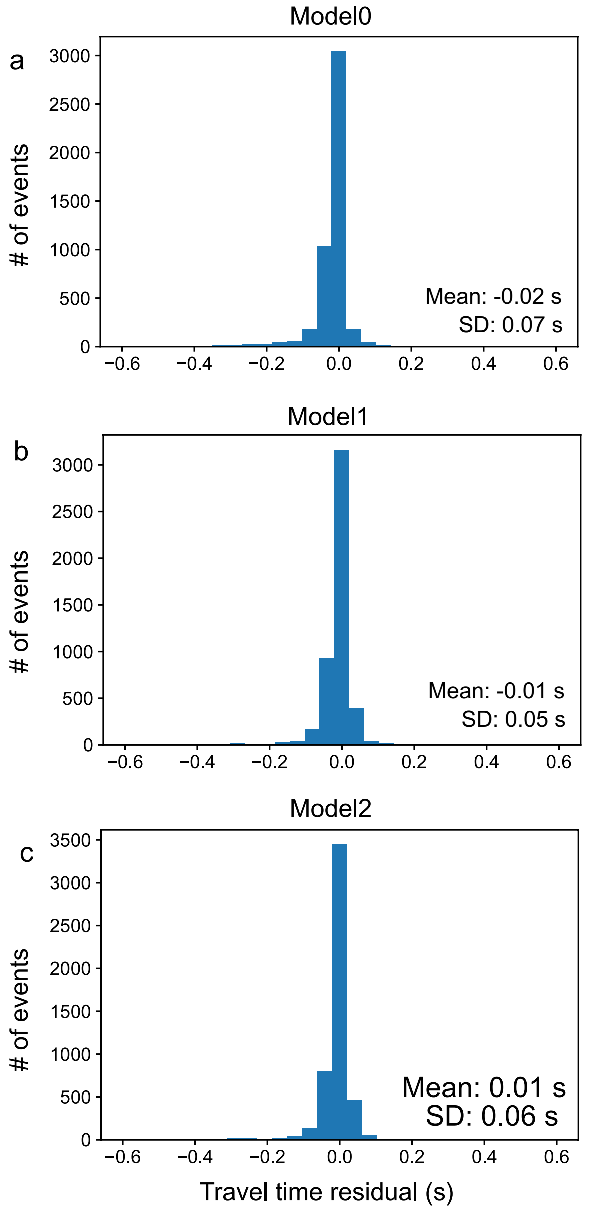


Fig. 4 Travel time residuals of the test data for each model. The width of the bin is 0.02 s and the number of the bin is 30. Values in the lower right corner indicate the mean and standard deviation (SD). **a** model 0, **b** model 1, **c** model 2

Application to continuous data

Then, models 0, 1, and 2 were applied to continuous waveforms for the same period as the test data. The number of manually detected earthquakes by HSRI during this period was 441. After detecting P- and S-waves, we used the rapid earthquake association and location (REAL) (Zhang et al. 2019) for phase association and preliminary location. The parameters used in REAL are 10 km for the search range, horizontally centered at the station recording the initiating phase, and a depth of 20 km. The search grid was set to approximately 2 km horizontally and in-depth. The time windows were tested for 1 h, 30 min, 10 min, 1 min, 30 s, 10 s, and 5 s. The best results were obtained at 30 s, the same data length used in training, which suggests that shortening the time window may not improve the performance (Table 1). In addition, when applying the method to continuous waveform data, as was the case with the above test data, there were many cases in which small-amplitude events were missed when multiple events of largely different amplitudes were included in the detection time window. Zhu et al. (2020) showed that by having multiple seismic events in the training data, the detection performance improved when many earthquakes occurred in a short period. Thus, it may be necessary to insert multiple events into the training data to improve the detection performance of active swarm earthquakes, where many earthquakes occur across a short period near a volcano. Thus, in this study, we verified the performance when the training data contained two events. Here, two events were randomly extracted from the validation data and combined, so that the P-wave of the following event comes 6–25 s after the first arrives, and 100,000 semi-synthetic training data containing two earthquakes in a 90-s time window were created. The model re-trained from scratch with only those data was denoted model 3 and the model trained with the same data as model 3 that uses model 1 as the starting model is denoted model 4 hereafter. As a result, both models 3 and 4 significantly improved the problems observed in models 1 and 2, but model 3 was lower than model 1 in the total number of detections (Table 2; Fig. 6). The number of event detections in model 4 also exceeded that of model 1 (Table 2), suggesting the

(See figure on next page.)

Fig. 5 Examples of raw seismic waveforms recorded under various conditions predicted using models 0, 1, and 2. The upper three figures show the observed waveforms from the top in east–west, north–south, and vertical directions, and the vertical axis shows the normalized amplitude. The lower three figures show the results of model 0, model 1, and model 2 from the top. The vertical axis is probability. The dotted lines represent the probability values of P-wave and S-wave, respectively. When the probability exceeds 0.6, it is indicated by a solid line and is considered to be detected as the respective phase. **a** One earthquake in the detection window, **b** two seismic events of almost the same amplitude in the detection window, **c** two events of very different amplitude in the detection window, **d** poor signal-to-noise ratio, **e**, **f** multiple earthquakes of very different amplitude in the detection window

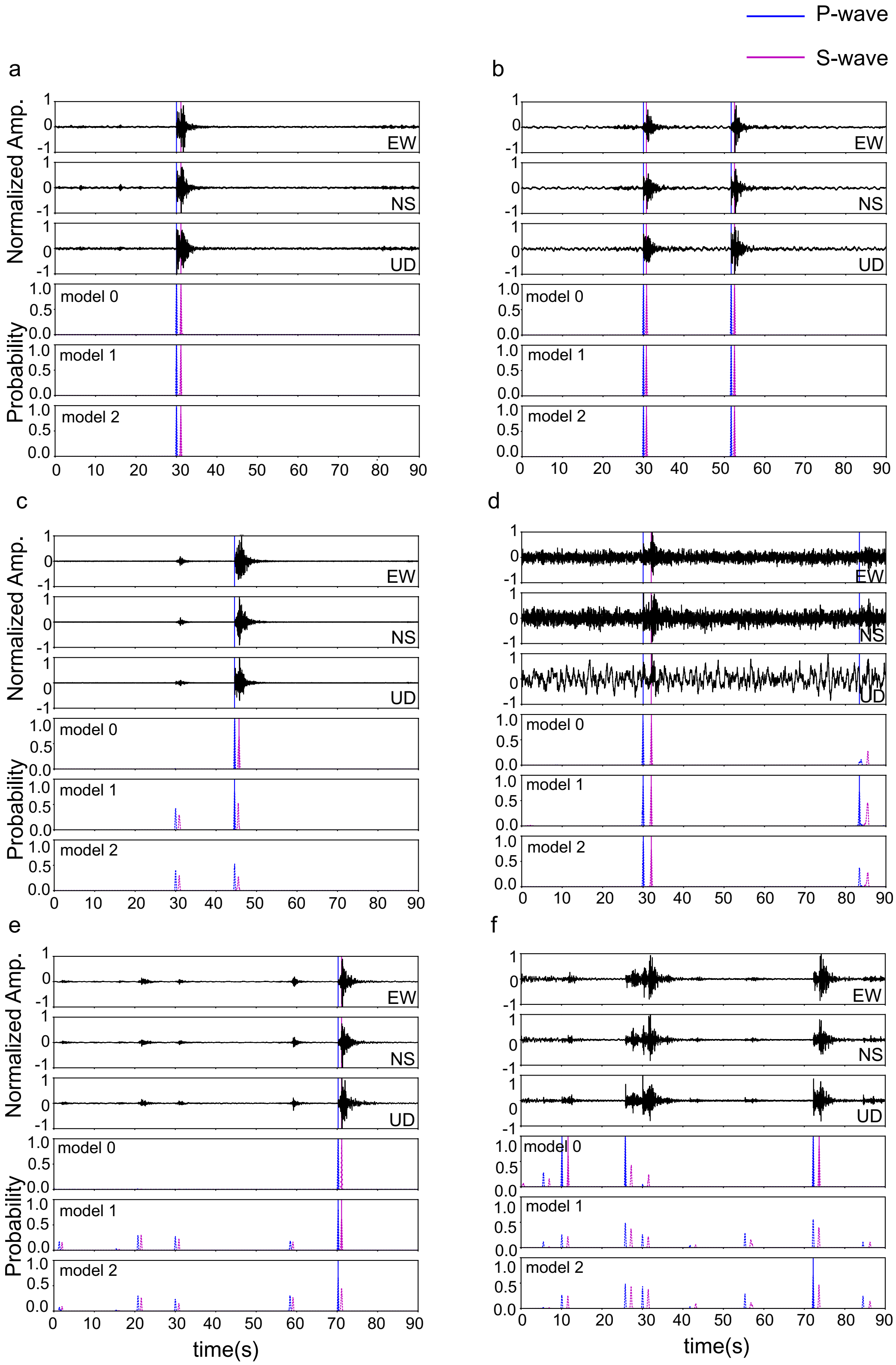


Fig. 5 (See legend on previous page.)

Table 1 Time window and number of earthquakes detected

Time window	# of event detection
10 s	920
20 s	1088
30 s	1244
1 min	1047
10 min	787
30 min	670
1 h	610

Table 2 Numbers of seismic waves were detected for each model with a detection threshold of 0.6

Model	# of event detection
Model0	1028
Model1	1244
Model2	1131
Model3	1120
Model4	1311

importance of selecting training data depending on the seismic activity to be applied (Zhu et al. 2020).

Discussion

In this study, we constructed a model trained from scratch (model 1) using the PhaseNet architecture. We then fine-tuned the model with the Hakone data using model 0 as the initial value to create model 2 and compared the performances of models 1 and 2 with that of model 0.

In model 1, we were able to find the batch size and learning rate that produced the highest F1 score, but we could not find the optimal value for the batch size in model 2. While a smaller batch size could produce a better F1 score, we were unable to conduct further investigation due to the limitations in our computational facilities. Regarding batch size, it is generally believed that as the batch size increases, the features of the input parameters become averaged out and individual features of the data may be lost. Conversely, a smaller batch size can be considered more sensitive to individual data (Keskar et al. 2017). In other words, in the case of model 2, a model that learned more detailed features of the Hakone data from the model 0 with a smaller batch size is considered to have performed better. However, determining the hyperparameters can only be accomplished through detailed benchmark tests in the end. Although detailed hyperparameter testing is computationally demanding, the optimizer used in PhaseNet, called Adam, is known to require only low-level hyperparameter tuning (Kingma and Ba 2014). Since the values chosen for model 1 are not very different from those used in the original model, we believe that the test results are reliable.

Model 1 showed the best performance among the three models. However, when multiple events with different amplitudes were in the same detection time window, the probability for small amplitude events was lower than the threshold for phase pick, and the prediction performance for large amplitude events was often lower. On the other hand, Model 0 did not respond to any small amplitude events in such cases and only tended to detect large amplitude events. The training data contain only one earthquake per data set, the amplitude of which is normalized by the maximum value. The normalization means that small-amplitude events in the same detection time window are considered noise and unlikely to be detected. The probability increased slightly in models 1 and 2, because the seismicity in Hakone has different characteristics from those of the California crustal earthquakes used in model 0 and may have captured those characteristics. To overcome this issue, two earthquakes were randomly taken from the aforementioned validation data and combined to create semi-synthetic training data. Creating such data allows small-amplitude events in the data normalized by the maximum amplitude to be labeled and trained as P- and S-waves. We evaluated the performance of models 3—where all weights are initialized—and 4—taking the parameters of model 1 as initial values. As the result, we found that both models showed an improved ability to detect seismic phases with smaller amplitudes when there were multiple events with different amplitudes in the same detection time window. Model 4 outperformed model 1 in terms of the number of events detected, whereas model 3 was less good than model 1 in terms of the number of events detected. This may be because the number of seismic waves is 100,000, half the size of the data used to train model 1. Deep learning models for seismic phase pick usually use amplitude-normalized data as training data. The amplitude varies with the distance from the epicenter and the magnitude of the earthquake in each seismic data set. Therefore, when using seismic events acquired over a wide area as training data, the amplitude range will cover several orders of magnitude and the model may not converge if trained without normalization. However, in the case of a group of earthquakes that occurred within a narrow magnitude range of magnitude, the inclusion of fluctuations in amplitude values within that range in the training data could improve the model's performance.

Since we used the validation data to create models 3 and 4, it cannot be completely ruled out that the performance improvement is due to the use of validation data.

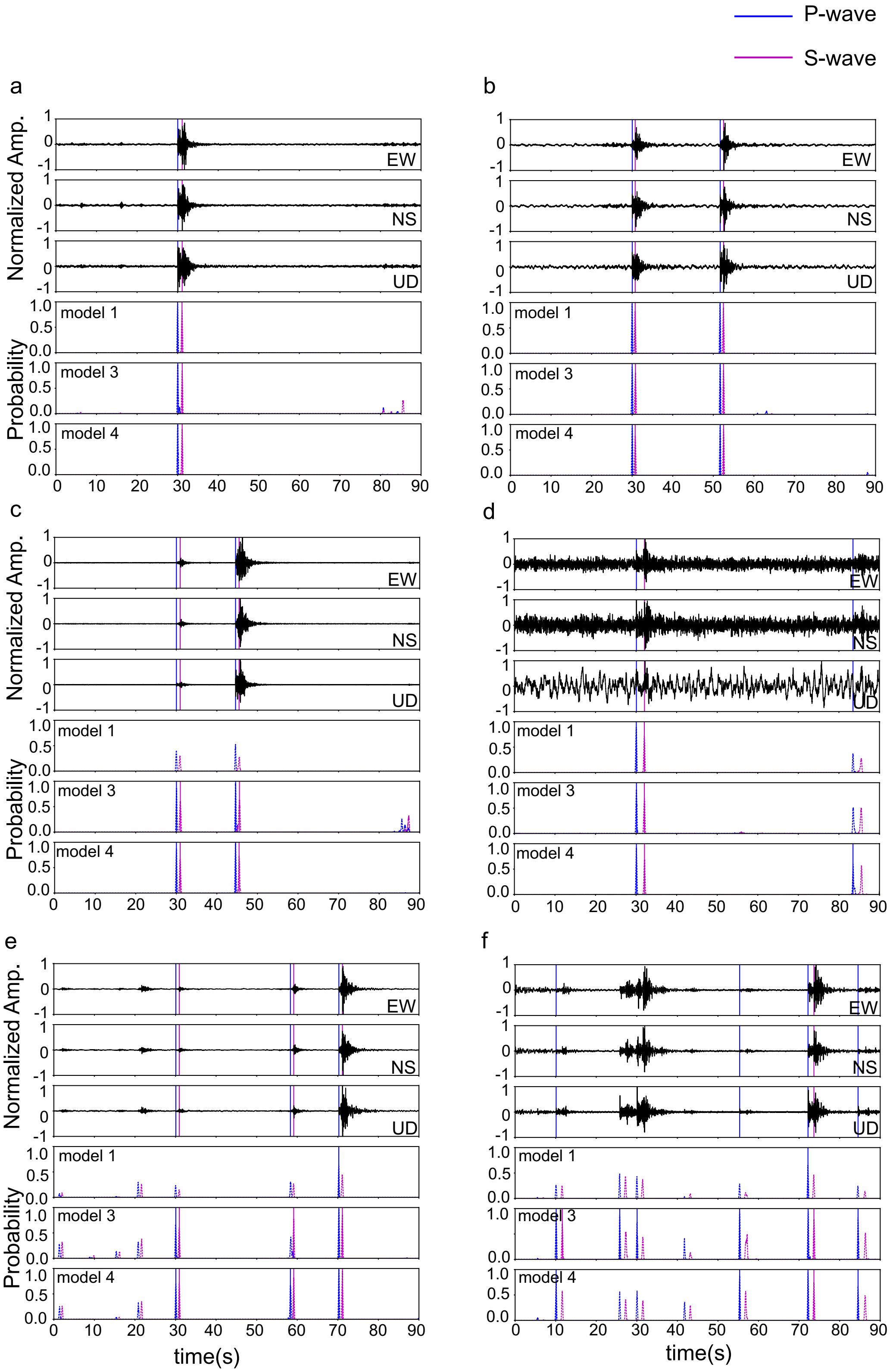


Fig. 6 Same as Fig. 5, but with models 1, 3, and 4

Thus, we performed tests to show that the improved performance of the phase pick is due not to the use of validation data but to the inclusion of two earthquakes in the training data as follows:

1. Create a fine-tuned model of model 1 with validation data consisting of 43,511 seismic data and each data contains a single event (model v).

2. Create ten fine-tuned models (model sN, where $N=1, 2, 3, 4, 5, 6, 7, 8, 9, 10$). To create the models, ten different datasets are prepared using the same method used in model 4. Each dataset consists of 43,511 training data and each training data contain two earthquakes.

Then, we compared the performances of models 1, v, and sN.

In this test, we first compared the number of earthquakes detected by applying the picks obtained from each model to REAL (Table 3). Next, the same earthquakes shown in Fig. 6 were used to compare the detection performance (Fig. 7 and Additional file 1: Fig. S1).

These results show that simply fine-tuning with validation data yields almost the same performance as model 1, whereas model sN clearly improved the phase identification performance as seen in models 3 and 4. Based on the above verification, we believe that the improvement in the performance of model 4 over model 1 is due to the inclusion of two events per unit of training data rather than the use of validation data. However, in some cases, model sn, which uses training data with about half the amount of data than that of model 4, which uses 100,000 training data, improves the detection performance, and there is room for further study on the preparing the training data that includes multiple earthquakes (Figs. 6 and 7).

Table 3 Numbers of seismic waves were detected for each model with a detection threshold of 0.6

Model	# of event detection
Model v	1246
Model s1	1278
Model s2	1280
Model s3	1279
Model s4	1279
Model s5	1281
Model s6	1278
Model s7	1284
Model s8	1278
Model s9	1278
Model s10	1280

Mean and Standard deviation (SD) of model sN are shown at the bottom

Mean: 1279.5

SD: 1.9

Model 4 detected more earthquakes in the continuous seismic waveforms than model 1, but some events were still missed. Although these may be improved by increasing the size of the data, a detailed inspection of the results shows that many events tended to show even a slight increase in probability. Therefore, we predicted the same continuous waveform data by setting the detection threshold of model 4 to a probability of 0.1 and 0.3 and applying REAL to the results. We further applied VELEST (Kissling et al. 1994) for the earthquake location and finally relocated events using HypoDD (Waldhauser and Ellsworth 2000). The events were relocated as 2094, 1296, and 1091 earthquakes, with probability thresholds of 0.1, 0.3, and 0.6, respectively (Fig. 8; Table 4). The total number of earthquakes recorded in the original catalog during this period is 441, so if the detection threshold is set as 0.1, we have detected about 4.7 times as many earthquakes. From the above, it is not necessary to set the threshold high at the time of event detection; instead, it is more effective to set it low and filter out noise in the subsequent phase association, hypocenter location, and relocation.

Regarding the types of earthquakes newly detected by the PhaseNet model, most events are located within the cluster of hypocenters in the original catalog, but some are outliers. Although the magnitudes in the original catalog, which have been edited by HSRI for the test data (441 events), are calculated in different ways and cannot be compared directly, the histogram shown in Fig. 9 indicates that the smaller the probability threshold, the smaller magnitude the picked-up earthquakes. However, as the probability threshold decreases, the number of earthquakes outside the cluster increases. Hence, it is necessary for future work to examine whether these newly detected small earthquakes are real. Yukutake et al. (2022) applied matched filter method (MF) (Gibbons and Ringdal 2006) to this earthquake swarm, and the number of earthquakes detected was 2600 in August 18–20, 2019, which is more than the number detected by model 4 with the probability threshold of 0.1. The hypocenter depth relocated by model 4 is very similar to the depth of the original catalog but generally deeper than that determined by MF (Fig. 10). Possible explanations for this difference are that our hypocenter relocation is not based on waveform correlation and/or that the station correction is performed in MF but not in this study. Regarding the processing time, it is possible to improve the performance of both methods by devising the code. As mentioned in Zhu and Beroza (2019), to improve phase detection in the contentious data, new data sets with more non-seismic signals may be needed for training

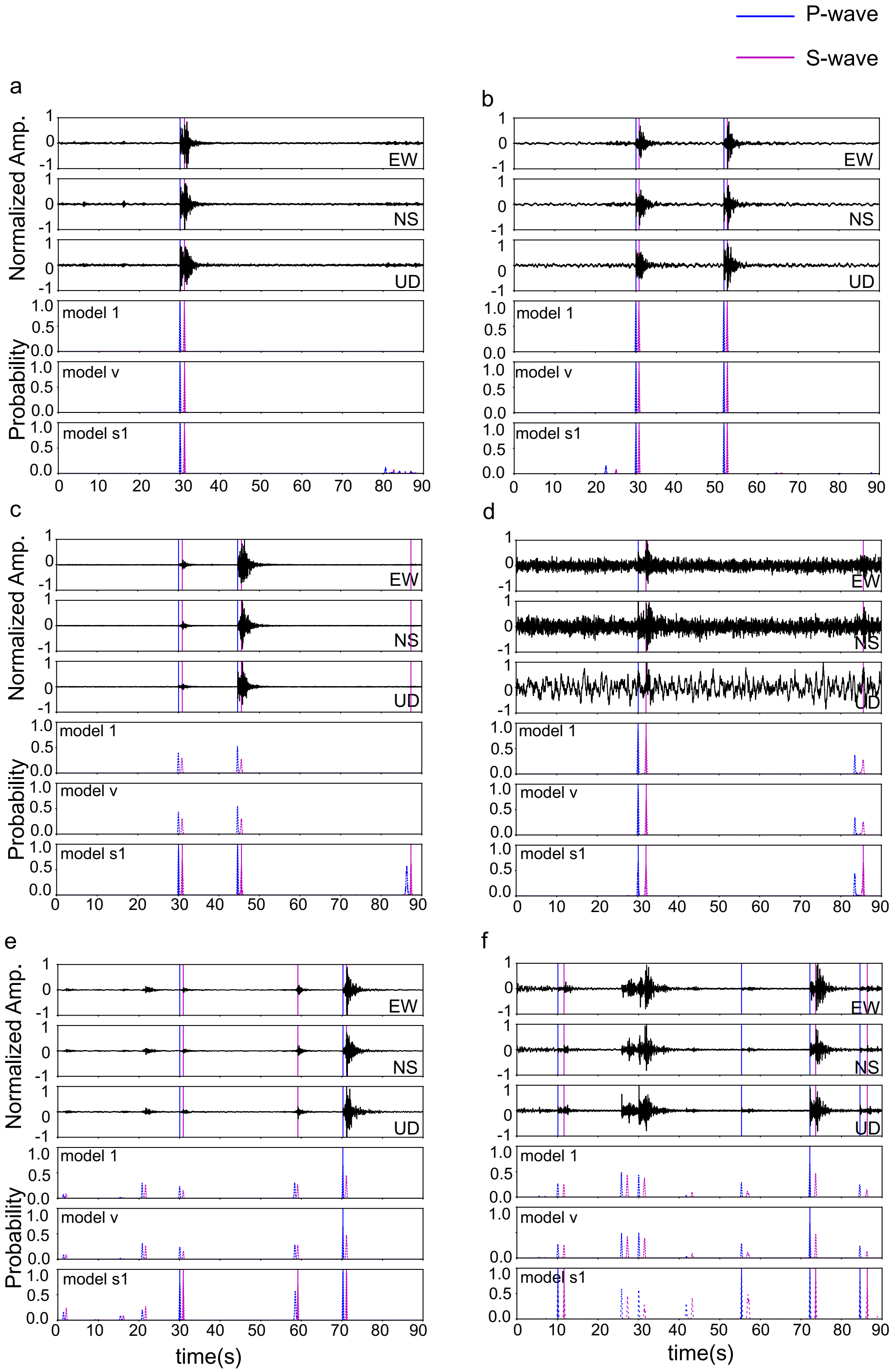


Fig. 7 Same as Fig. 5, but with models 1, v, and s1

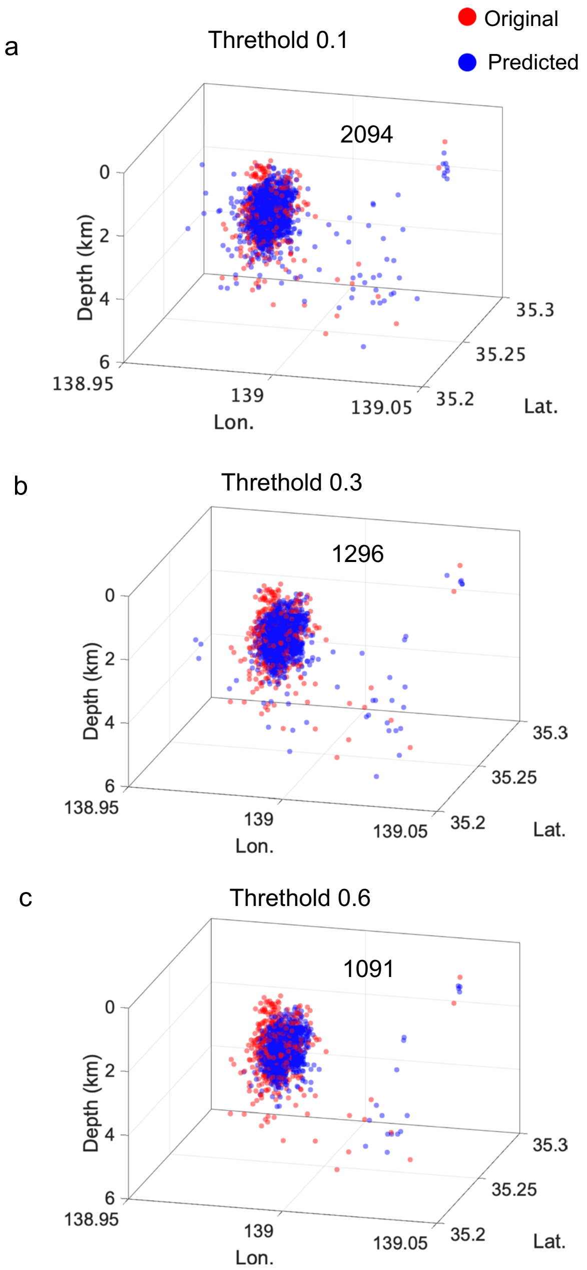


Fig. 8 Hypocenter distribution relocated by hypoDD. PhaseNet detection thresholds are the probability of **a** 0.1, **b** 0.3, and **c** 0.6

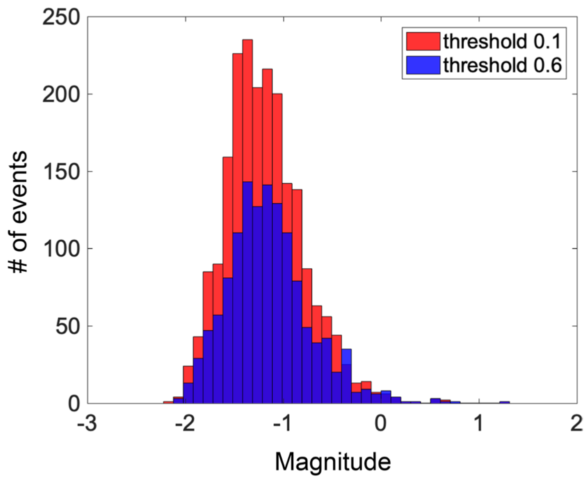


Fig. 9 Histograms of earthquake magnitudes relocated by hypoDD. Red and blue indicate PhaseNet detection thresholds of the probability of 0.1 and 0.6, respectively

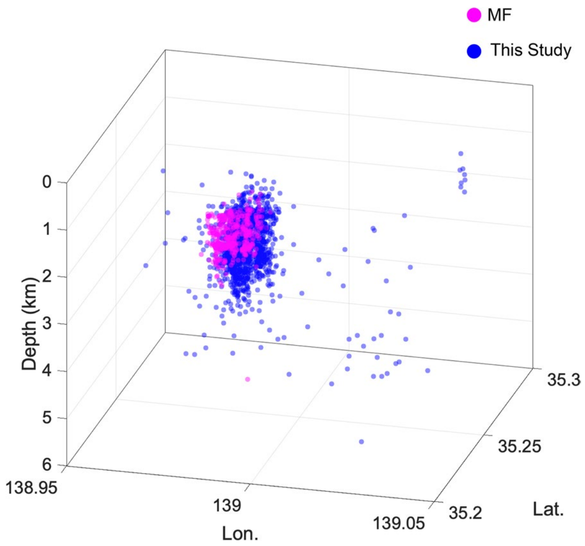


Fig. 10 Hypocenter distribution comparison. The magenta circle indicates the hypocenters determined by matched filter method, and the blue circle indicates those relocated in this study. Here, the PhaseNet detection threshold is the probability of 0.1

Table 4 Number of event detections for each algorithm relative to the detection threshold

Threshold	REAL	VELEST	hypoDD
0.1	7381	6909	2094
0.3	1937	1865	1296
0.6	1311	1302	1091

to learn the features that distinguish noise spikes that resemble seismic phases.

One advantage of PhaseNet over MF is that it does not require template earthquakes. The model we constructed can be generalized to some extent, even if trained on data from other regions with similar geological backgrounds. For example, when model 1 was applied to seismic data from other volcanic areas in Japan, it successfully detected eight times more earthquakes than those in the catalog edited by human labor (Yukutake and Kim 2022). Since not all volcanoes have been monitored with high quality for many years—like Hakone volcano—it is worth aiming to improve the performance of such machine learning models. In the future, it may be possible to create a model specialized for a particular region by transfer learning with a small amount of data based on the model developed at Hakone volcano. In addition, many seismic events may overlap during the very active swarm period. Using the method of training data with multiple earthquakes used in this study, we can artificially create overlapping data and add it to the training data, thereby learning its characteristics and potentially contributing to the event detection capability.

Conclusion

To automate and improve the quality of seismic phase detection at Hakone volcano, a few models were created using the PhaseNet architecture with training data of approximately 220,000 P- and S-wave onset readings in the area. The newly constructed model outperformed model 0 (pretrained open-to-public model) when there was only one earthquake or multiple events with similar amplitudes. When multiple events with different amplitudes existed in the same detection time window, there were many cases, where earthquakes with small amplitudes were missed with models 1 and 2. Fine-tuning with 100,000 semi-synthetic training data—including two events per one data—using the parameters of model 1 initial values significantly improved the above problems and increased the number of seismic detections.

Careful inspection of the detection results showed that even for phases that did not lead to detection, there was often a small amplification of likelihood (Fig. 6). Therefore, we lowered the threshold and passed the data to REAL, VELEST, and HypoDD, which showed that the number of earthquake detections was higher than when the same process was performed with a higher threshold. These results suggest that, although further investigation is needed, it may be possible to detect a large number of earthquakes if we do not set a strict threshold at the time of seismic wave detection and instead use phase association, hypocenter location, and other methods to filter out the false seismic phase detections. In the future, we will apply this model to earthquakes in volcanic regions, where training data are insufficient to verify its generalization. We will also consider creating models specific to individual volcanoes using transfer learning.

Abbreviations

AR	Autoregressive
AIC	Akaike information criterion
STA	Short-term average
LTA	Long-term average
HSRI	Hot spring research institute of Kanagawa Prefecture
NIED	National Research Institute for Earth Science and Disaster Resilience
CNN	Convolutional neural network
FCN	Fully convolutional network
REAL	Rapid earthquake association and location
MF	Matched filter method

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40623-023-01840-5.

Additional file 1: Fig. S1. Comparison of models 1, v, and s2. **Fig. S2.** Comparison of models 1, v, and s3. **Fig. S3.** Comparison of models 1, v, and s4. **Fig. S4.** Comparison of models 1, v, and s5. **Fig. S5.** Comparison of models 1, v, and s6. **Fig. S6.** Comparison of models 1, v, and s7. **Fig. S7.** Comparison of models 1, v, and s8. **Fig. S8.** Comparison of models 1, v, and s9. **Fig. S9.** Comparison of models 1, v, and s10.

Acknowledgements

We are grateful to the two anonymous reviewers for their insightful reviews that improved the quality of the manuscript. We also thank Naoki Uchida for his editorial stewardship of the journal. We thank Chonho Lee for his helpful comments and suggestions on our deep learning analyses. We used waveforms from stations that were part of Hi-net (NIED), the Hot Spring Research Institute of Kanagawa Prefecture. The PhaseNet code and model were downloaded from https://github.com/wayneweiqiang/PhaseNet (last accessed September 15, 2022). The codes for phase association (REAL), hypocenter location (VELEST), and hypocenter relocation (hypoDD) are from the following package: https://github.com/Dal-mzhang/LOC-FLOW/blob/main/README.md (last accessed September 15, 2022). Maps were drawn using Generic Mapping Tools (Wessel and Smith 1998).

Author contributions

AK designed this paper and performed the analysis. YN developed the code for data preparation and analyses. All authors read and approved the final manuscript.

Funding

This study was supported by JSPS KAKENHI Grant Number JP15KK0171 (to AK) and JP22K03752 (to YY).

Availability of data and materials

Seismic waveform data from Hi-net (NIED) is available at https://www.hinet.bosai.go.jp/?LANG=en.

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

[1]Yokohama City University, Yokohama, Japan, 22-2, Seto, Kanazawa-ku, 236-0027. [2]Earthquake Research Institute, The University of Tokyo, Tokyo, Japan, 1-1-1 Yayoi, Bunkyo-ku, 113-0032. [3]The Graduate University for Advanced Studies, National Institute of Informatics, Tokyo, Japan, 2-1-2 Hitotsubashi, Chiyoda-ku, 101-8430. [4]Hot Springs Research Institute of Kanagawa Prefectural Government, Kanagawa, Japan, 586 Iriuda, Odawara, 250-0031.

Received: 10 November 2022 Accepted: 2 May 2023
Published online: 16 May 2023

References

Akazawa T (2004) A technique for automatic detection of onset time of P-and S-Phases in strong motion records. 13th World Conference on Earthquake Engineering

Allen RE (1978) Automatic phase pickers: their present use and future prospects. Bull Seismol Soc Am 72(6B):225–242

Dai H, MacBeth C (1995) Automatic picking of seismic arrivals in local earthquake data using an artificial neural network. Geophys J Int 120:758–774. https://doi.org/10.1111/j.1365-246X.1995.tb01851.x

Dysart PS, Pulli J (1990) An experiment in the use of trained neural networks for regional seismic event classification. Geophys Res Lett 17:977–980. https://doi.org/10.1029/GL017i007p00977

Fedorenko YV, Husebye ES, Ruud BO (1999) Explosion site recognition; neural net discriminator using single three-component stations. Phys Earth Planet in 113(1):131–142

Gibbons SJ, Ringdal F (2006) The detection of low magnitude seismic events using array-based waveform correlation. Geophys J Int 165(1):149–166. https://doi.org/10.1111/j.1365-246X.2006.02865.x

Keskar NS, Mudigere D, Nocedal J, Smelyanskiy M, Tang PTP (2017) On large-batch training for deep learning: generalization gap and sharp minima. arXiv:1609.04836

Kingma DP, Ba J (2014) Adam: a method for stochastic optimization. arXiv:1412.6980

Kissling E, Ellsworth WL, Eberhart-Phillips D, Kradolfer U (1994) Initial reference models in local earthquake tomography. J Geophys Res 99:19635–19646

Kobayashi M, Mannen K, Okuno M, Nakamura T, Hakamata K (2006) The Owakidani Tephra group: a newly discovered post-magmatic eruption product of Hakone volcano, Japan (in Japanese with English abstract). Bull Volcanol Soc Jpn 51:245–256

Kong Q, Richard A, Schreier L, Kwon YW (2016) MyShake: a smartphone seismic network for earthquake early warning and beyond. Sci Adv 2(2):e1501055. https://doi.org/10.1126/sciadv.1501055

Lapins S, Goitom B, Kendall JM, Werner MJ, Cashman KV, Hammond JOS (2021) A little data goes a long way: automating seismic phase arrival picking at Nabro volcano with transfer learning. J Geophys Res 126:e2021JB021910. https://doi.org/10.1029/2021JB021910

Mannen K, Yukutake Y, Kikugawa G, Harada M, Itadera K, Takenaka J (2018) Chronology of the 2015 eruption of Hakone volcano, Japan: geological background, mechanism of volcanic unrest and disaster mitigation measures during the crisis. Earth Planets Space 70:68. https://doi.org/10.1186/s40623-018-0844-2

Mousavi SM, Ellsworth WL, Zhu W, Chuang LY, Beroza GC (2020) Earthquake transformer—an attentive deep-learning model for simultaneous earthquake detection and phase picking. Nat Commun 11:3952. https://doi.org/10.1038/s41467-020-17591-w

Münchmeyer J, Woollam J, Rietbrock A, Tilmann F, Lange D, Bornstein T, Diehl T, Giunchi C, Haslinger F, Jozinovic D, Michelini A, Saul J, Soto H (2022) Which picker fits my data? A quantitative evaluation of deep learning based seismic pickers. J Geophys Res 127:e2021JB023499. https://doi.org/10.1029/2021JB023499

Musil M, Plešinger A (1996) Discrimination between local microearthquakes and quarry blasts by multi-layer perceptrons and Kohonen maps. Bull Seismol Soc Am 86(4):1077–1090. https://doi.org/10.1785/BSSA0860041077

Obara K, Kasahara K, Hori S, Okada T (2005) A densely distributed high-sensitivity seismograph network in Japan: Hi-net by National Research Institute for Earth Science and Disaster Prevention. Rev Sci Instrum 76:021301. https://doi.org/10.1063/1.1854197

Perol T, Gharbi M, Denolle M (2018) Convolutional neural network for earthquake detection and location. Sci Adv 4(2):e1700578. https://doi.org/10.1126/sciadv.1700578

Ronneberger O, Fischer P, Brox T (2015) U-Net: convolutional networks for biomedical image segmentation. Miccai 9351:234–241. https://doi.org/10.1007/978-3-319-24574-4_28

Ross ZE, Meier MA, Hauksson E, Heaton TH (2018) Generalized seismic phase detection with deep learning. Bull Seismol Soc Am 108(5A):2894–2901. https://doi.org/10.1785/0120180080

Tiira T (1999) Detecting teleseismic events using artificial neural networks. Comput Geosci 25(8):929–938. https://doi.org/10.1016/S0098-3004(99)00056-4

Urabe T, Tsukada S (1992) WIN-A workstation program for processing waveform data from microearthquake network (abstract in Japanese), Program and Abstract, Seismo. Soc. Japan, No. 2, 331

Ursino A, Langer H, Scarfi L, Di Grazia G, Gresta S (2001) Discrimination of quarry blasts from tectonic microearthquakes in the Hyblean plateau (southeastern Sicily). Ann Geophys 44:703–722. https://doi.org/10.4401/ag-3569

Vaswani A, Shazeer NM, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I (2017) Attention is All you Need. NIPS'17, Proceedings of the 31st International Conference on Neural Information Processing Systems December 2017 6000–6010.

Waldhauser F, Ellsworth WL (2000) A double-difference earthquake location algorithm: method and application to the northern Hayward fault. Bull Seismol Soc Am 90(6):1353–1368. https://doi.org/10.1785/0120000006

Wessel P, Smith WHF (1998) New, improved version of generic mapping tools released. Eos 79:579

Wiszniowski J, Plesiewicz BM, Trojanowski J (2014) Application of real time recurrent neural network for detection of small natural earthquakes in Poland. Acta Geophys 62(3):469–485. https://doi.org/10.2478/s11600-013-0140-2

Yukutake Y, Ito H, Honda R, Harada M, Tanada T, Yoshida A (2011a) Fluid-induced swarm earthquake sequence revealed by precisely determined hypocenters and focal mechanisms in the 2009 activity at Hakone volcano. Japan J Geophys Res 116:B04308. https://doi.org/10.1029/2010JB008036

Yukutake Y, Honda R, Harada M (2011b) Remotely-triggered seismicity in the Hakone volcano following the 2011 off the Pacific coast of Tohoku Earthquake. Earth Planet Sp 63:41. https://doi.org/10.5047/eps.2011.05.004

Yukutake Y, Kim A (2022) Detection and hypocenter determination of volcanic earthquakes using Machine Learning: Application to Kirishima volcano, Japan Geoscience Union Meeting, 22–27, May 2022

Yukutake Y, Yoshida K, Honda R (2022) Interaction between aseismic slip and fluid invasion in earthquake swarms revealed by dense geodetic and seismic observations. J Geophys Res 127:e2021JB022933. https://doi.org/10.1029/2021JB022933

Zhang M, Ellsworth WL, Beroza GC (2019) Rapid earthquake association and location. Seismol Res Lett 90(6):2276–2284. https://doi.org/10.1785/0220190052

Zhao Y, Takano K (1999) An artificial neural network approach for broadband seismic phase picking. Bull Seismol Soc Am 89(3):670–680. https://doi.org/10.1785/BSSA0890030670

Zhu W, Beroza GC (2019) PhaseNet: a deep-neural-network-based seismic arrival-time picking method. Geophys J Int 216(1):261–273. https://doi.org/10.1093/gji/ggy423

Zhu W, Mousavi SM, Beroza GC (2020) Chapter Four—Seismic signal augmentation to improve generalization of deep neural networks. In: Moseley B, Krischer L (eds) Advances in Geophysics, vol 61. Elsevier, Amsterdam, pp 151–177. https://doi.org/10.1016/bs.agph.2020.07.003

Publisher's Note