

OPEN

Pilot study of eruption forecasting with muography using convolutional neural network

Yukihiro Nomura^{1*}, Mitsutaka Nemoto², Naoto Hayashi¹, Shouhei Hanaoka³, Masaki Murata⁴, Takeharu Yoshikawa¹, Yoshitaka Masutani⁵, Eriko Maeda¹, Osamu Abe³ & Hiroyuki K. M. Tanaka⁶

Muography is a novel method of visualizing the internal structures of active volcanoes by using high-energy near-horizontally arriving cosmic muons. The purpose of this study is to show the feasibility of muography to forecast the eruption event with the aid of the convolutional neural network (CNN). In this study, seven daily consecutive muographic images were fed into the CNN to compute the probability of eruptions on the eighth day, and our CNN model was trained by hyperparameter tuning with the Bayesian optimization algorithm. By using the data acquired in Sakurajima volcano, Japan, as an example, the forecasting performance achieved a value of 0.726 for the area under the receiver operating characteristic curve, showing the reasonable correlation between the muographic images and eruption events. Our result suggests that muography has the potential for eruption forecasting of volcanoes.

Muography is a newly developed imaging technique utilizing high-energy near-horizontally arriving cosmic muons and enables us to visualize the internal structures of large objects. Muography produces a projection image (hereafter, muogram) of a large body by mapping out the number of muons that are transmitted through it. Muograms were first used in 1970 by Alvarez *et al.* to search for hidden chambers in the Chephren's Second Pyramid¹. Almost forty years later, muograms were first used to explore the internal structures of volcanoes². Muograms depicting the internal structures of a volcano are of particular importance because they may be used in the study of eruption dynamics³. The first experimental evidence discovered using muograms was obtained by studying the summit of Mount Asama, Japan². The observation, in conjunction with the observation of a low-density magma pathway imaged underneath a solidified magma deposit, confirmed that muograms could resolve a volcano structure with more precision than the preexisting geophysical techniques. Since this first experimental study on the internal structure of volcanoes in 2007, similar experiments have been carried not only in Japan³⁻⁶ but also in the US⁷ and Europe⁸⁻¹¹.

Eruption forecasting is one of the most critical tasks in modern volcanology^{12,13}. For these tasks, many methods based on statistical algorithms or machine learning have been reported¹⁴⁻¹⁹. These methods use data such as seismic activity, ground deformation, and gas emission. However, to the best of our knowledge, there has been no literature on eruption forecasting using muograms. Muography is conceptually similar to standard X-ray radiography. In the field of medical image analysis, including the analysis of X-ray radiographs, deep learning has been shown to achieve remarkable results^{20,21}. We expect that volcanic eruptions can also be forecasted by applying deep learning to muograms.

The purpose of this study is to show the feasibility of eruption forecasting using muograms with deep learning. We focused on muographic data acquired at Sakurajima volcano, Japan, between 2014 and 2016 when it was most activated in the last eruption episode (2009–2017).

¹Department of Computational Diagnostic Radiology and Preventive Medicine, The University of Tokyo Hospital, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8655, Japan. ²Faculty of Biology-Oriented Science and Technology, Kindai University, Nishimitani 930, Kinokawa, Wakayama, 649-6493, Japan. ³Department of Radiology, The University of Tokyo Hospital, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8655, Japan. ⁴Department of Management, Japan University of Economics, 3-11-25 Gojo, Dazaifu-shi, Fukuoka, 818-0197, Japan. ⁵Graduate School of Information Sciences, Hiroshima City University, 3-4-1 Ozuka-Higashi, Asaminami-ku, Hiroshima, 731-3194, Japan. ⁶Earthquake Research Institute, The University of Tokyo, 1-1-1 Yayoi, Bunkyo, Tokyo, 113-0032, Japan. *email: nomuray-tyk@umin.ac.jp

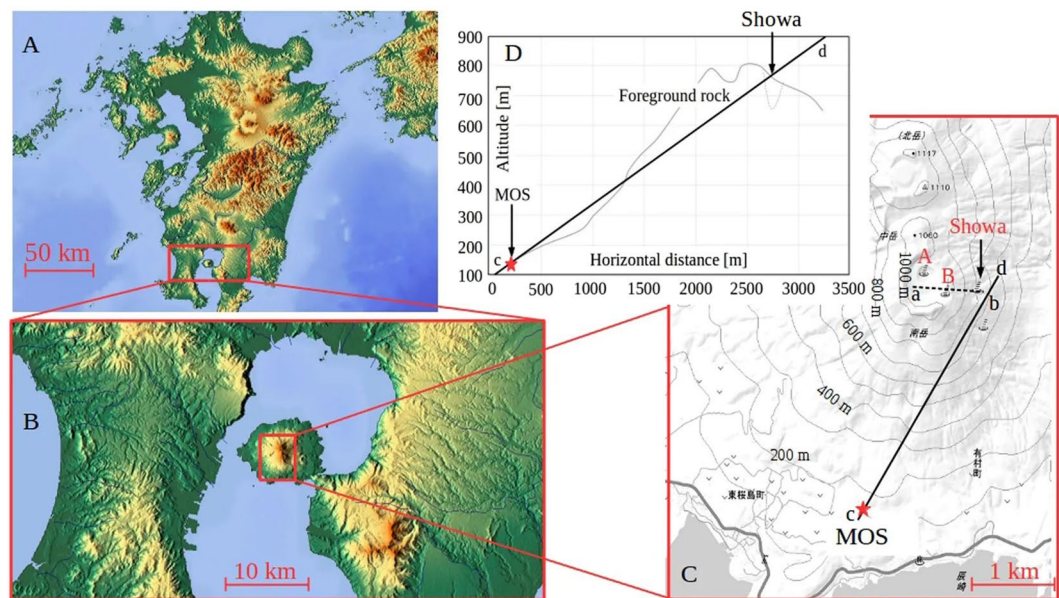


Figure 1. Schematic drawing of our muography experiment. (A,B) Location of Sakurajima volcano in Kyushu, Japan (<https://maps-for-free.com/>). (C) Elevation map and the schematic of measurement site. The muography observation system (MOS, red star) was installed in the Southwest direction at distances of 2.8 km, 2.7 km, and 2.6 km from the Showa crater, the Minamidake A crater, and the B crater, respectively. The elevation map was created from data of Geospatial Information Authority of Japan (<http://www.gsi.go.jp/>) and edited by the authors. (D) Cross-sectional view of Sakurajima volcano along line c-d in (C). (Source: Oláh, L. *et al.*, *Sci. Rep.* 8, 3207, pp. 26, cc BY 4.0. We have changed “mMOS” to “MOS” and “Crater Showa” to “Showa”).

Results

Muography observation system at Sakurajima volcano. The muography observation system (MOS)⁴ was installed at Sakurajima Muography Observatory (SMO)²². Figure 1(A–D) respectively show the location of the measurement site, a topographic map of the measurement site, and a cross-sectional view of Sakurajima volcano. A more detailed description of this MOS can be found elsewhere⁴; thus, the MOS will be briefly introduced here. The system consists of five 10-cm-thick lead plates and six layers of scintillation position-sensitive planes. Each position-sensitive plane consists of $N_x = 15$ and $N_y = 15$ adjacent scintillator strips, which together form a segmented plane with 15×15 segments; thus, the total active area for collecting muons is 2.25 m². The observation system (red star in Fig. 1(C,D)) was installed in the southwest direction at distances of 2.8 km, 2.7 km, and 2.6 km from the Showa crater, the Minamidake A crater, and the B crater, respectively⁶. The three craters were located within the field of view of the MOS. The elevation of the measurement site was ~150 m above sea level (ASL). The angular resolution of the muography observation system was 33 milliradians (mrad).

Eruption forecasting using convolutional neural network. We investigated the effectiveness of a convolutional neural network (CNN) for eruption forecasting at the Showa crater of Sakurajima volcano based on muograms. The CNN, which is one of the prominent deep learning models, has shown success in image classification and object detection²³.

Figure 2(A) shows the relationship between the input data for the CNN model and the prediction term for an eruption. The muograms used in this study were plots of the daily muon count (observation period: 00:00:00 to 23:59:59). We inputted muograms obtained for seven consecutive days into the CNN to compute the probability of eruptions for the eighth day, called the “prediction day.” The number of muograms was determined experimentally. If at least one eruption occurred at Showa crater during the prediction day, the day was labeled as an eruption. The eruption times were based on the data from the website of Kagoshima Meteorological Office, Japan (https://www.jma-net.go.jp/kagoshima/vol/kazan_top.html, in Japanese). The total number of eruptions during the prediction day was not referred to. From the above, the CNN model predicts whether an eruption will occur at Showa crater on the eighth day using muograms obtained for seven consecutive days.

We used 464 sets of seven consecutive daily muograms and a label of the prediction day, which were obtained between 7 November 2014 and 12 May 2016. We excluded seven consecutive daily muograms that include unobserved periods due to maintenance of the MOS. We split the dataset into three subsets: a training set, a validation set, and a test set. The training set was used to train the model and included 382 sets of muographic data (prediction days, 20 November 2014–28 January 2016; number of eruption days, 191). The validation set was used to calculate the evaluation criterion of the hyperparameter tuning and includes 40 sets of muographic data (prediction days, 8 February 2016–18 March 2016; number of eruption days, 20). The test set was used to evaluate the best model throughout the hyperparameter tuning, which includes 42 sets of muographic data (prediction days,

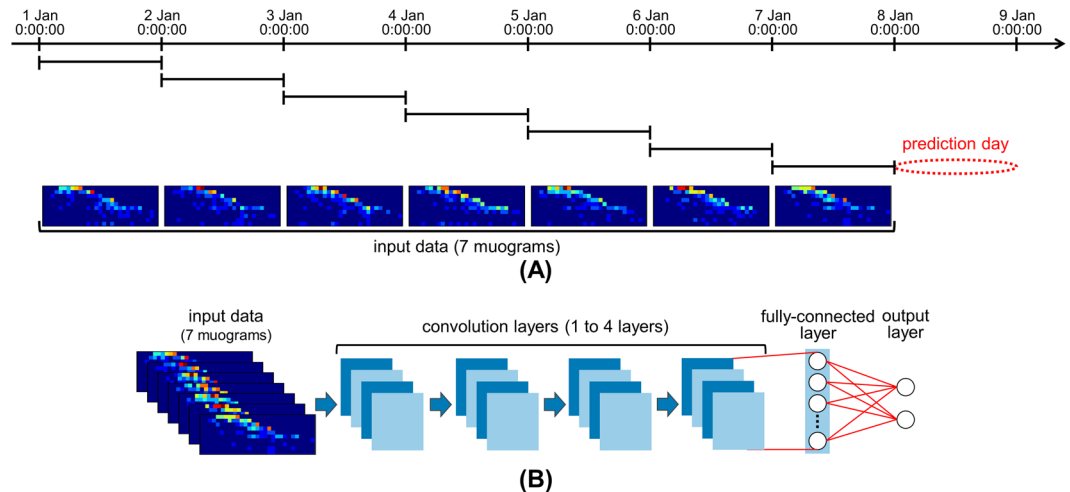


Figure 2. (A) Relationship between input data for CNN model and prediction term for eruption, (B) configuration of CNN model. Muograms are plotted in a color scale.

25 March 2016–12 May 2016; number of eruption days, 21). During the period of the three datasets, 1,439 and 10 eruptions respectively occurred at the Showa and Minamidake craters.

Figure 2(B) shows the configuration of our network. Our network consisted of one to four convolutional layers, one fully connected layer, and one output layer with two units with softmax activation. We employed a rectified linear unit (ReLU) function²⁴ as the activation function for all layers except the output layer. Batch normalization²⁵ was performed before each ReLU function. The dropout strategy²⁶ was adopted for all layers except the output layer to avoid overfitting. We utilized the Adam method²⁷ to optimize the network weights. The training procedure of the CNN model is described in Methods.

We compared the following four input regions:

- Showa crater region: $165 \text{ mrad} < \theta < 297 \text{ mrad}$, $-66 \text{ mrad} < \phi < 66 \text{ mrad}$ (5×5 segments)
- Minamidake crater region: $198 \text{ mrad} < \theta < 330 \text{ mrad}$, $-330 \text{ mrad} < \phi < 99 \text{ mrad}$ (5×8 segments)
- surface region: $33 \text{ mrad} < \theta < 198 \text{ mrad}$, $132 \text{ mrad} < \phi < 264 \text{ mrad}$ (5×5 segments)
- all segments: $33 \text{ mrad} < \theta < 330 \text{ mrad}$, $-462 \text{ mrad} < \phi < 462 \text{ mrad}$ (10×29 segments)

where θ and ϕ indicate the elevation angle and azimuth angle, respectively. The Minamidake crater region included both the Minamidake craters A and B. The surface region included 5×5 segments of the mountain surface excluding the three craters as a baseline. The values of the segments outside the volcano were set to 0. For standardization, each segment value was multiplied by 0.001. By assuming the average rock thickness of 1 km in these regions, ~ 50 and ~ 90 muon events/day were respectively expected in the Showa and Minamidake crater regions.

Figure 3 shows an example of three consecutive daily muogram images used for the training set. Figure 4 shows the relative muon counts averaged over 15 events for the training set that were selected to satisfy our condition: there is no eruption at least two days before the eruption, where the “relative muon count” is obtained by dividing by the average of daily muon count acquired during the period of no eruption (from 1 November 2015 to 30 November 2015). The four plots in this figure correspond to the data acquired from the aforementioned four regions ((A) Showa crater region, (B) Minamidake crater region, (C) surface region, and (D) all segments). From Figs. 3,4(A), the muon count of the Showa crater region tended to decrease on the day before the eruption.

For comparison, we also evaluated three types of classifiers: a simple rule-based method, a support vector machine (SVM) model using the radial basis function (RBF) kernel²⁸, and a neural network (NN) model. The simple rule-based method classifies an eruption occurring on at least four out of seven consecutive days as an eruption. This threshold was selected to maximize the accuracy of the training data. The SVM model and the NN model were applied only to the Showa crater region. For the feature values of the two models, we compared 175 segment values (25 segment values \times 7 days) and seven summed values (the summation of 25 segment values for each day). We selected the seven summed values as the feature values with the best performance. The training procedures of the SVM model and the NN model are described in Methods.

We evaluated our method using receiver operating characteristic (ROC) analysis^{29,30}, and the area under the curve (AUC) was calculated, which ranges between 0.0 and 1.0, with values of 0.5 for random classification and 1.0 for perfect classification. If the AUC is less than 0.5, the classification results are meaningless. Figure 5 shows the ROC curves for the test set for each input data. As shown in Fig. 5, the AUC values were 0.726 for the Showa crater region, 0.678 for the Minamidake crater region, 0.444 for the surface region, and 0.544 for all segments. Figure 6 shows the ROC curves for the test set for the SVM and NN models of the Showa crater. As shown in Fig. 6, the AUC values were 0.569 for the SVM model and 0.499 for the NN model.

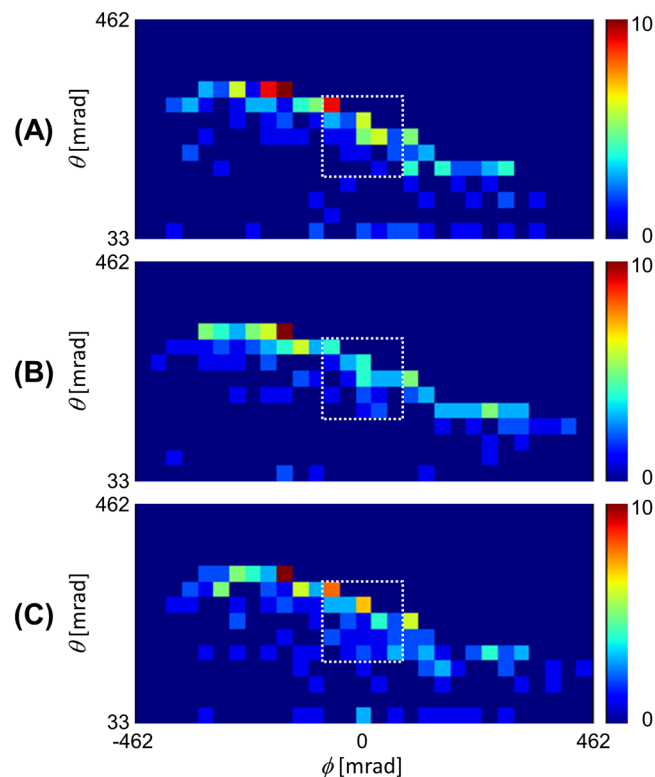


Figure 3. Example of three consecutive daily muograms (observed from (A) 25 July 2015 00:00:00 to 23:59:59, (B) 26 July 2015 00:00:00 to 23:59:59, (C) 27 July 2015 00:00:00 to 23:59:59). The angular resolution was 33 mrad per segment and the data shows 14×29 segments. Muograms are plotted in a color scale (range: 0–10). The segment outside of the volcano was set to 0. White dotted squares indicate the Showa crater region. During the observation period, an eruption occurred at 27 June 2015 01:53:00.

Table 1 shows prediction performance at the optimal cutoff point for each input data. The optimal cutoff point of the ROC curve was chosen using Youden's index³¹. The CNN model of the Showa crater showed the highest accuracy and specificity. In contrast, the SVM model of the Showa crater showed the highest sensitivity.

Discussion

We have shown that our method may achieve moderate performance for day-level eruption forecasting at the Showa crater of Sakurajima volcano. The AUC, accuracy, and specificity were highest when the input to the CNN model was limited to the segments of the Showa crater region. The sensitivity was highest when the input to the CNN model was the all segments. In addition, the muon count of the Showa crater region tends to decrease before the eruption, which was considered to be due to the plugging of a magma pathway by magma deposits on the crater floor. These results suggest that the muographic data of the segments around Showa crater contributed to the forecasting performance. By contrast, the AUC was less than 0.5 when the input to the CNN model was from the segments of the surface region. Consequently, muograms have potential use for eruption forecasting of a volcano by analyzing temporal changes in its internal structures.

The CNN model of the Showa crater region was superior to the SVM model and the NN model in the AUC, accuracy, and specificity. In the CNN model, the convolutional layer works as the feature extractor, extracting local features of each layer. In addition, a deeper convolutional layer can detect more complex features. These roles of the CNN model make it more promising than other machine learning models.

The Showa crater region was limited to 5×5 segments since the angular resolution of the muography observation system was 33 mrad per segment. To overcome this problem of low resolution, a high-definition muography observation system was developed, with which muograms have acquired since January 2017⁶. The angular resolution of the high-definition muography observation system is 2.7 mrad per segment, and it can be expected to image more detailed internal structures of the volcano. We plan to investigate the use of high-resolution muograms when a sufficient number of muograms are collected. We also plan to investigate the eruption forecasting of Minamidake craters, because most of the eruptions have occurred from these craters since November 2017.

The consecutive muographic images are time series data. Recurrent neural networks (RNNs), especially long short-term memory (LSTM)³², are an effective model for processing time series data³³. A combination of CNNs and RNNs is expected to improve the forecasting performance. We plan to investigate their combination in our future work.

At Sakurajima volcano, other observation data, such as volcanic earthquake, volcanic tremor, tilt, and GPS data, are reported on the website of Japan Meteorological Agency (<http://www.data.jma.go.jp/svd/vois/data/>)

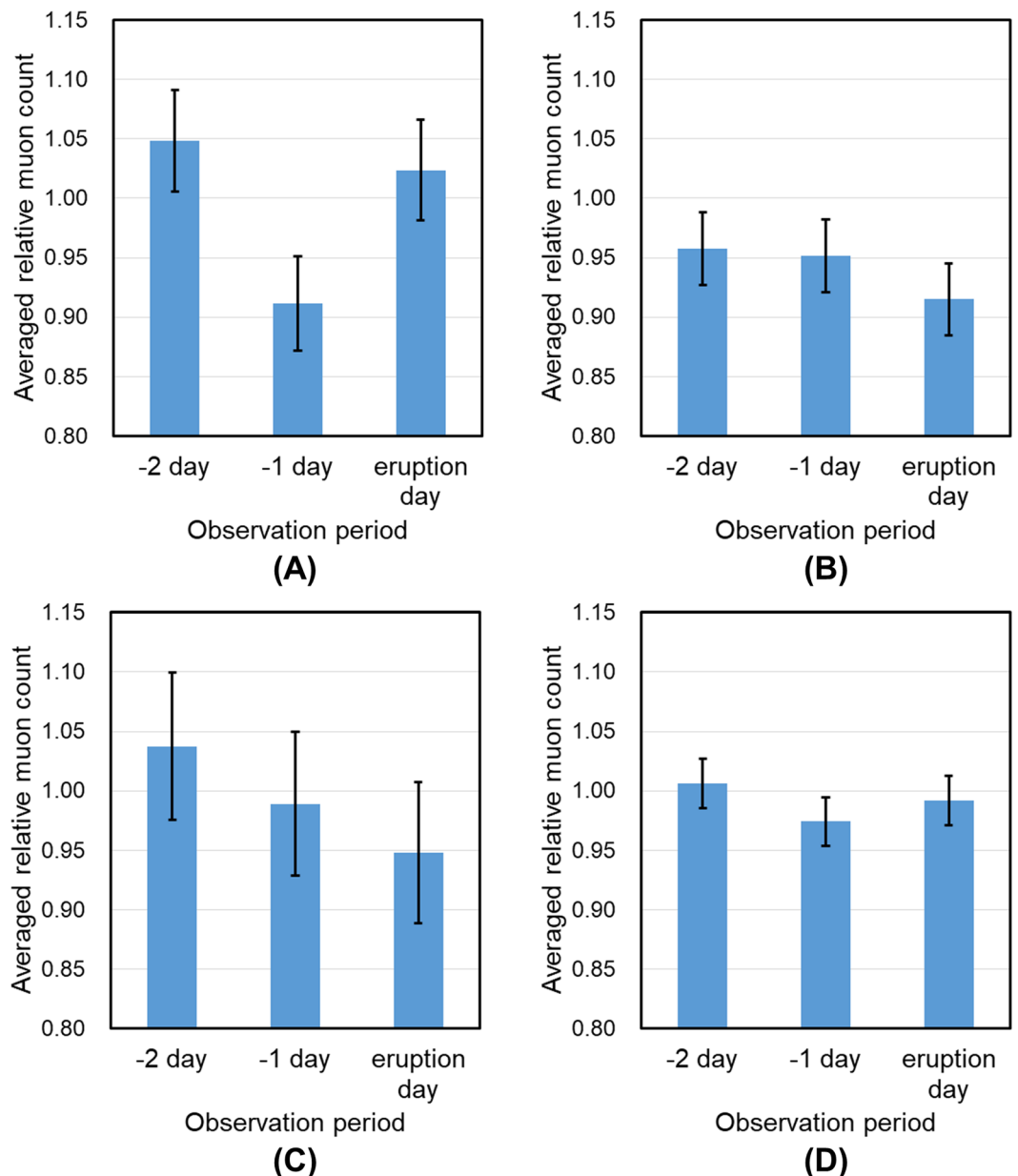


Figure 4. Averaged relative muon counts for each input region ((A) Showa crater region, (B) Minamidake crater region, (C) surface region, (D) all segments). Error bars represent standard deviations.

[tokyo/STOCK/bulletin/index_vcatalog.html](https://www.tokyostock.com/bulletin/index_vcatalog.html), in Japanese). However, these data are only available until April 2016 at the time of writing. The forecasting performance may be further improved by adding these observations to the input of the CNN.

There are two limitations to this study. First, in the labeling of the eruption day, we have excluded ten eruptions that occurred at the Minamidake craters during the period of analysis. This is because during the training data period, only two eruptions occurred at the Minamidake crater, both of which were erupted at the Showa crater on the same day. Second, an estimation of the uncertainty or confidence in the output decisions of our model was not carried out. For practical eruption forecasting using our method, it is crucial to estimate the uncertainty or confidence of the model³⁴.

Methods

Training of CNN. In the training of the CNN, the optimization of numerous hyperparameters has a strong effect on the performance of the CNN model. Strategies for hyperparameter optimization include a grid search, random search³⁵, and the Bayesian optimization (BO) algorithm³⁶. BO is a framework for the optimization of black-box functions whose derivatives and convexity properties are unknown. BO is expected to optimize hyperparameters more efficiently than a random search.

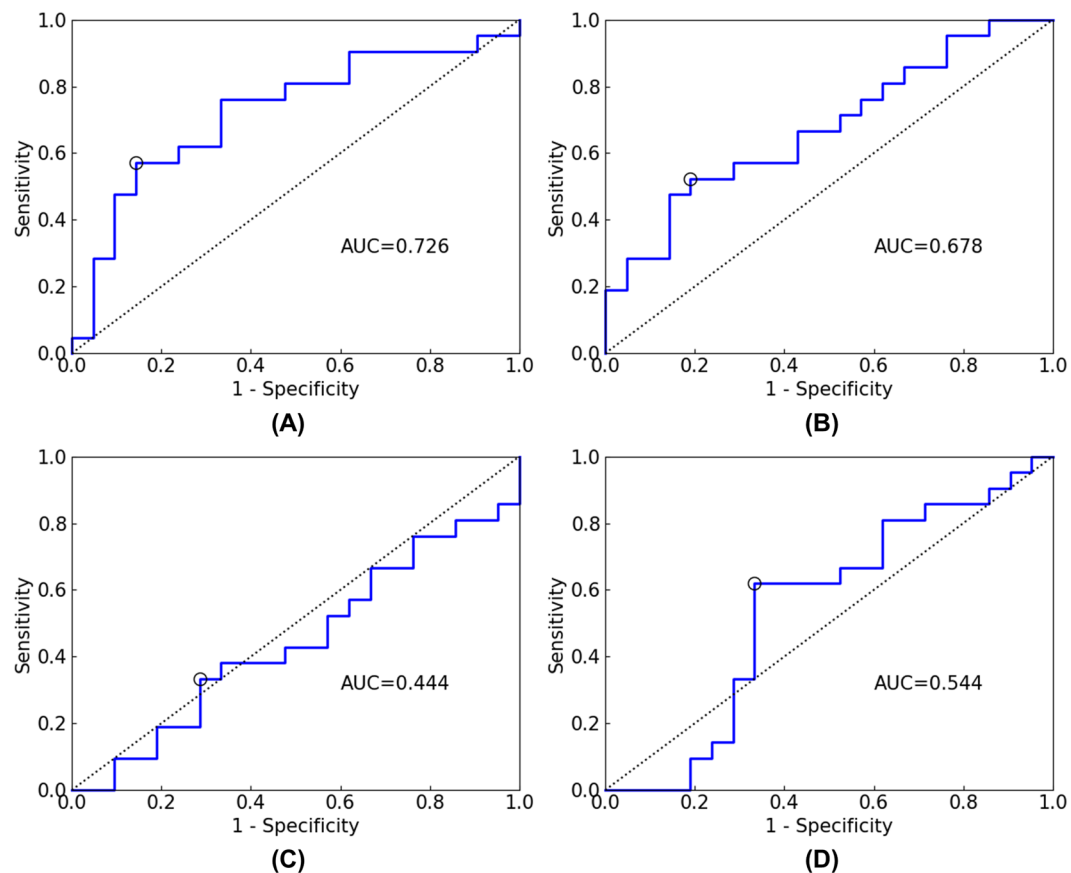


Figure 5. ROC curves for test set for each input data ((A) Showa crater region, (B) Minamidake crater region, (C) surface region, (D) all segments). The circle indicates the cutoff point chosen using Youden's index. AUC is the area under the curve.

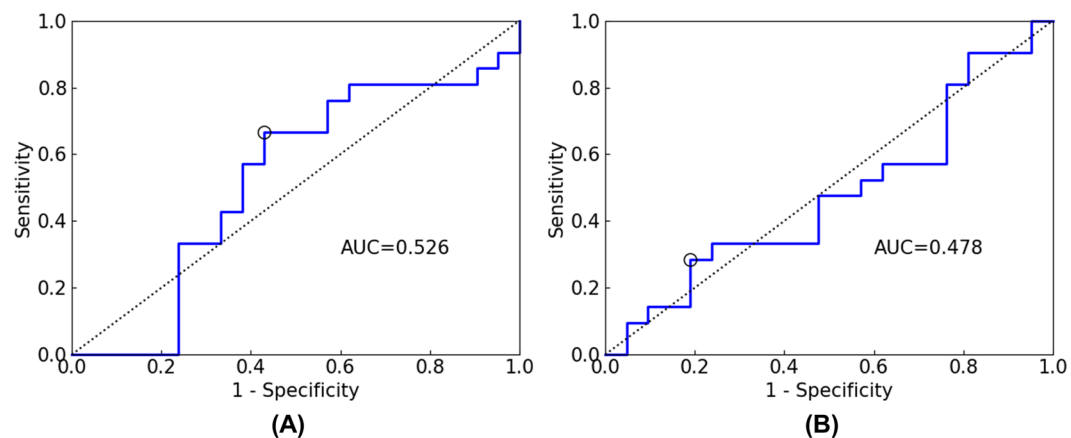


Figure 6. ROC curves for test set for each machine learning model of the Showa crater region ((A) SVM, (B) NN). The circle indicates the cutoff point chosen using Youden's index. AUC is the area under the curve.

In this study, we carried out 200 trials of hyperparameter tuning with the BO algorithm. The tuned hyperparameters of the CNN were the number of filters of each convolution layer (2^c , $c = 2-6$), the number of units of the fully connected layer (2^f , $f = 1-7$), the batch size, two parameters of the Adam method ($\alpha = 10^{-3}-10^{-6}$, $\beta_1 = 0.9-0.99$), and the ratio of dropout (0–0.5). We utilized the AUC of the ROC curve as an evaluation criterion for hyperparameter tuning. The numbers of maximum epochs and the patience of early stopping³⁷ were set to 200 and 20, respectively.

The CNN model was implemented using Keras³⁸ version 2.2.4 with TensorFlow³⁹ version 1.10.0 backend. We trained the network on a GeForce GTX TITAN X (NVIDIA Corporation, Santa Clara, CA) graphics processing

Method	Input region	Accuracy	Sensitivity	Specificity
CNN model	Showa crater	0.714	0.571	0.857
	Minamidake craters	0.667	0.523	0.810
	Surface	0.524	0.333	0.714
	All segments	0.643	0.619	0.667
Rule-based method		0.571	0.523	0.619
SVM model (Showa crater region)		0.619	0.667	0.571
NN model (Showa crater region)		0.548	0.286	0.810

Table 1. Prediction performance at the optimal cutoff point for each input data.

Parameter		Showa crater	Minamidake craters	Surface	All segments
Number of convolution layers		3	2	4	4
Number of filters	first layer	64	8	8	32
	second layer	64	16	8	8
	third layer	4	0	16	64
	fourth layer	0	0	32	8
Number of units of fully connected layer		32	16	64	8
Batch size		8	8	128	16
Parameter of Adam	α	0.002455	0.008680	0.002707	0.003291
	β_1	0.991646	0.986270	0.981427	0.998521
Ratio of dropout		0.486933	0.487076	0.325312	0.106837

Table 2. Selected sets of hyperparameters of the CNN model for each input data.

unit (GPU) with 12 GB memory. Table 2 shows the selected sets of hyperparameters of the CNN model for each input data.

Training of SVM. In the training of the SVM model using the RBF kernel, we also carried out 200 trials of hyperparameter tuning with the BO algorithm. The tuned hyperparameters of the SVM model using the RBF kernel were the regularization parameter ($C = 2^{-5} - 2^{15}$) and the kernel parameter ($\gamma = 2^{-15} - 2^3$)²⁸. We utilized the AUC of the ROC curve as an evaluation criterion for hyperparameter tuning.

The SVM model was implemented using scikit-learn version 0.21.3. The selected hyperparameters of the SVM model were $C = 0.0659$ and $\gamma = 7.237$.

Training of NN. In the training of the NN model, we also carried out 200 trials of hyperparameter tuning with the BO algorithm. The tuned hyperparameters of the NN model were the number of hidden layers (1–3), the number of units of each hidden layer (2^h , $h = 2-8$), the batch size, two parameters of the Adam method ($\alpha = 10^{-3} - 10^{-6}$, $\beta_1 = 0.9 - 0.99$), and the ratio of dropout (0–0.5). We utilized the AUC of the ROC curve as an evaluation criterion for hyperparameter tuning. The maximum number of epochs and the patience of early stopping were set to 100 and 10, respectively.

The NN model was implemented using Keras version 2.2.4 with TensorFlow version 1.10.0 backend. We trained the network on a GeForce GTX TITAN X GPU with 12 GB memory. The selected hyperparameters of the NN model were as follows: number of hidden layers, 3; number of units of first hidden layer, 32; number of units of second hidden layer, 128; number of units of third hidden layer, 16; batch size, 32; α of Adam method, 0.000287; β_1 of Adam method, 0.983763; ratio of dropout, 0.442019.

Received: 27 November 2019; Accepted: 6 March 2020;

Published online: 24 March 2020

References

- Alvarez, L. W. *et al.* Search for hidden chambers in the pyramids. *Science* **167**, 832–839, <https://doi.org/10.1126/science.167.3919.832> (1970).
- Tanaka, H. K. M. *et al.* High resolution imaging in the inhomogeneous crust with cosmic-ray muon radiography: The density structure below the volcanic crater floor of Mt. Asama, Japan. *Earth Planet. Sci. Lett.* **263**, 104–113, <https://doi.org/10.1016/j.epsl.2007.09.001> (2007).
- Tanaka, H. K. M., Kusagaya, T. & Shinohara, H. Radiographic visualization of magma dynamics in an erupting volcano. *Nat. Commun.* **5**, 3381, <https://doi.org/10.1038/ncomms4381> (2014).
- Tanaka, H. K. M., Uchida, T., Tanaka, M., Shinohara, H. & Taira, H. Cosmic-ray muon imaging of magma in a conduit: degassing process of Satsuma-Iwojima volcano, Japan. *Geophys. Res. Lett.* **36**, L01304, <https://doi.org/10.1029/2008gl036451> (2009).
- Tanaka, H. K. M. Instant snapshot of the internal structure of Unzen lava dome, Japan with airborne muography. *Sci. Rep.* **6**, 39741, <https://doi.org/10.1038/srep39741> (2016).
- Oláh, L., Tanaka, H. K. M., Ohminato, T. & Varga, D. High-definition and low-noise muography of the Sakurajima volcano with gaseous tracking detectors. *Sci. Rep.* **8**, 3207, <https://doi.org/10.1038/s41598-018-21423-9> (2018).

7. Lesparre, N. *et al.* Density muon radiography of La Soufrière de Guadeloupe volcano: comparison with geological, electrical resistivity and gravity data. *Geophys. J. Int.* **190**, 1008–1019, <https://doi.org/10.1111/j.1365-246X.2012.05546.x> (2012).
8. Cârloganu, C. *et al.* Towards a muon radiography of the Puy de Dôme. *Geosci. Instrum. Methods Data Syst.* **2**, 55–60, <https://doi.org/10.5194/gi-2-55-2013> (2013).
9. Carbone, D. *et al.* An experiment of muon radiography at Mt Etna (Italy). *Geophys. J. Int.* **196**, 633–643, <https://doi.org/10.1093/gji/ggt403> (2013).
10. Noli, P. *et al.* Muography of the Puy de Dôme. *Ann. Geophys.* **60**, S0105, <https://doi.org/10.4401/ag-7380> (2017).
11. Tioukov, V. *et al.* First muography of Stromboli volcano. *Sci. Rep.* **9**, 6695, <https://doi.org/10.1038/s41598-019-43131-8> (2019).
12. Amezquita-Sanchez, J. P., Valtierra-Rodriguez, M. & Adeli, H. Current efforts for prediction and assessment of natural disasters: Earthquakes, tsunamis, volcanic eruptions, hurricanes, tornados, and floods. *Sci. Iranica* **24**, 2645–2664, <https://doi.org/10.24200/sci.2017.4589> (2017).
13. Ramis, R. O. *et al.* Volcanic and volcano-tectonic activity forecasting: a review on seismic approaches. *Ann. Geophys.* **62**, VO06, <https://doi.org/10.4401/ag-7655> (2019).
14. Newhall, C. & Hoblitt, R. Constructing event trees for volcanic crises. *Bull. Volcanol.* **64**, 3–20, <https://doi.org/10.1007/s004450100173> (2002).
15. Marzocchi, W., Sandri, L. & Selva, J. BET_EF: a probabilistic tool for long- and short-term eruption forecasting. *Bull. Volcanol.* **70**, 623–632, <https://doi.org/10.1007/s00445-007-0157-y> (2008).
16. Lindsay, J. *et al.* Towards real-time eruption forecasting in the Auckland Volcanic Field: application of BET_EF during the New Zealand National Disaster Exercise 'Ruauumoko'. *Bull. Volcanol.* **72**, 185–204, <https://doi.org/10.1007/s00445-009-0311-9> (2010).
17. Segall, P. Volcano deformation and eruption forecasting. *Geol. Soc., London, Spec. Publ.* **380**, 85–106, <https://doi.org/10.1144/sp380.4> (2013).
18. Brancato, A., Buscema, P., Massini, G. & Gresta, S. Pattern recognition for flank eruption forecasting: an application at Mount Etna volcano (Sicily, Italy). *Open J. of Geol.* (2016).
19. Brancato, A. *et al.* K-CM application for supervised pattern recognition at Mt. Etna: an innovative tool to forecast flank eruptive activity. *Bull. Volcanol.* **81**, 40, <https://doi.org/10.1007/s00445-019-1299-4> (2019).
20. Litjens, G. *et al.* A survey on deep learning in medical image analysis. *Med. Image Anal.* **42**, 60–88, <https://doi.org/10.1016/j.media.2017.07.005> (2017).
21. Sahiner, B. *et al.* Deep learning in medical imaging and radiation therapy. *Med. Phys.* **46**, e1–e36, <https://doi.org/10.1002/mp.13264> (2019).
22. Oláh, L., Tanaka, H. K. M., Ohminato, T., Hamar, G. & Varga, D. Plug formation imaged beneath the active craters of Sakurajima volcano with muography. *Geophys. Res. Lett.* **46**, 10417–10424, <https://doi.org/10.1029/2019gl084784> (2019).
23. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444, <https://doi.org/10.1038/nature14539> (2015).
24. Nair, V. & Hinton, G. E. Rectified linear units improve restricted Boltzmann machines. *Proceedings of the 27th International Conference on Machine Learning*. 807–814 (2010).
25. Ioffe, S. & Szegedy, C. Batch normalization: accelerating deep network training by reducing internal covariate shift. *arXiv preprint*. arXiv:1502.03167 (2015).
26. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **15**, 1929–1958 (2014).
27. Kingma, D. P. & Ba, J. Adam: a method for stochastic optimization. *arXiv preprint*. arXiv:1412.6980 (2014).
28. Hsu, C. W., Chang, C. C. & Lin, C. J. A practical guide to support vector classification, <http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf> (2016).
29. Obuchowski, N. A. ROC analysis. *AJR Am. J. Roentgenol.* **184**, 364–372, <https://doi.org/10.2214/ajr.184.2.01840364> (2005).
30. Fawcett, T. An introduction to ROC. *analysis. Pattern Recognit. Lett.* **27**, 861–874, <https://doi.org/10.1016/j.patrec.2005.10.010> (2006).
31. Youden, W. J. Index for rating diagnostic tests. *Cancer* **3**, 32–35, <https://doi.org/10.1002/1097-0142> (1950).
32. Hochreiter, S. & Schmidhuber, J. Long short-term memory. *Neural Comput* **9**, 1735–1780, <https://doi.org/10.1162/neco.1997.9.8.1735> (1997).
33. Chen, J., Yang, L., Zhang, Y., Alber, M. & Chen, D. Z. Combining fully convolutional and recurrent neural networks for 3D biomedical image segmentation. *In Advances in Neural Information Processing Systems* **29**, 3036–3044 (2016).
34. Kendall, A. & Gal, Y. What uncertainties do we need in bayesian deep learning for computer vision? *In Adv. Neural Inf. Process. Syst.* **30**, 5574–5584 (2017).
35. Bergstra, J. & Bengio, Y. Random search for hyper-parameter optimization. *J. Mach. Learn. Res.* **13**, 281–305 (2012).
36. Snoek, J., Larochelle, H. & Adams, R. P. Practical Bayesian optimization of machine learning algorithms. *In Adv. Neural Inf. Process. Syst.* **25**, 2951–2959 (2012).
37. Prechelt, L. Early stopping – but when? *In Neural Networks: Tricks of the Trade: Second Edition* (eds. Montavon, G., Orr, G. B. & Müller, K.-R.) 53–67 (Springer, Berlin, Heidelberg, 2012).
38. Chollet, F. *et al.* Keras, <https://www.keras.io> (2015).
39. Abadi, M. *et al.* TensorFlow: Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint*. arXiv:1603.04467 (2016).

Acknowledgements

The Department of Computational Radiology and Preventive Medicine, The University of Tokyo Hospital, is sponsored by HIMEDIC Inc. and Siemens Healthcare K.K. This study was supported by the Earthquake Research Institute Joint Usage/Research Program in The University of Tokyo (project number: 2017-B-02). We would like to thank two the anonymous reviewers for their helpful comments.

Author contributions

N.H., T.Y., Y.M., E.M., O.A. and H.K.M.T. conceived and designed the study. Y.N., M.N. and H.K.M.T. prepared the dataset. Y.N., S.H. and M.M. performed data analysis. Y.N., M.N., N.H., E.M., and H.K.M.T. wrote the main manuscript. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Y.N.

Reprints and permissions information is available at www.nature.com/reprints.

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020