scientific reports

Check for updates

OPEN

Automatic classification between COVID-19 pneumonia, non-COVID-19 pneumonia, and the healthy on chest X-ray image: combination of data augmentation methods

Mizuho Nishio[1]✉, Shunjiro Noguchi[2], Hidetoshi Matsuo[1] & Takamichi Murakami[1]

This study aimed to develop and validate computer-aided diagnosis (CXDx) system for classification between COVID-19 pneumonia, non-COVID-19 pneumonia, and the healthy on chest X-ray (CXR) images. From two public datasets, 1248 CXR images were obtained, which included 215, 533, and 500 CXR images of COVID-19 pneumonia patients, non-COVID-19 pneumonia patients, and the healthy samples, respectively. The proposed CADx system utilized VGG16 as a pre-trained model and combination of conventional method and mixup as data augmentation methods. Other types of pre-trained models were compared with the VGG16-based model. Single type or no data augmentation methods were also evaluated. Splitting of training/validation/test sets was used when building and evaluating the CADx system. Three-category accuracy was evaluated for test set with 125 CXR images. The three-category accuracy of the CAD system was 83.6% between COVID-19 pneumonia, non-COVID-19 pneumonia, and the healthy. Sensitivity for COVID-19 pneumonia was more than 90%. The combination of conventional method and mixup was more useful than single type or no data augmentation method. In conclusion, this study was able to create an accurate CADx system for the 3-category classification. Source code of our CADx system is available as open source for COVID-19 research.

The outbreak of the novel coronavirus disease (COVID-19) started in Wuhan, Hubei province, China at the end of 2019[1], and COVID-19 spread across the world in 2020. COVID-19 is caused by a strain of coronavirus called Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2)[2]. The World Health Organization declared COVID-19 as a pandemic on March 11, 2020[3]. COVID-19 can be detected with the use of real-time polymerase chain reaction (RT-PCR) test of SARS-CoV-2. Although the specificity of RT-PCR was sufficiently high for COVID-19, its sensitivity was relatively low in detecting COVID-19[4]. Chest computed tomography (CT) was useful for detecting abnormal findings of COVID-19 pneumonia. It was characterized by ground-glass opacity distributed predominantly on lung peripherals[4,5]. The CT findings of COVID-19 could be regarded as distinct from viral and bacterial pneumonia.

Although the usefulness of CT for detecting COVID-19 pneumonia was shown in several studies, CT is not suitable for COVID-19 screening due to its cost and radiation exposure[6]. On the other hand, chest X-Ray imaging (CXR) is cost-effective and commonly used for screening purposes. CXR findings of COVID-19 pneumonia is characterized by the following; consolidation was the most common finding, followed by ground glass opacity; distribution of CXR abnormalities could be peripheral and lower zone distribution with bilateral involvement; pleural effusion was uncommon[7]. Compared with chest CT, the sensitivity of CXR is generally low for

[1]Department of Radiology, Kobe University Graduate School of Medicine, 7-5-2 Kusunoki-cho, Chuo-ku, Kobe 650-0017, Japan. [2]Department of Diagnostic Imaging and Nuclear Medicine, Kyoto University Graduate School of Medicine, 54 Shogoin Kawaharacho, Sakyo-ku, Kyoto 606-8507, Japan. ✉email: nishiomizuho@gmail.com

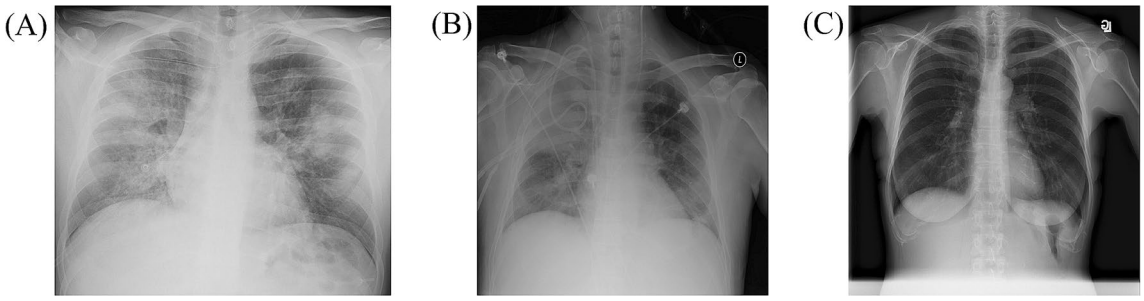

Figure 1. Representative CXR images of COVID-19 pneumonia, non-COVID-19 pneumonia, and the healthy. *CXR* chest X-ray imaging, *COVID-19* novel coronavirus disease. (**A**) COVID-19 pneumonia of 30-year-old male. (**B**) Non-COVID-19 pneumonia of 56-year-old male. (**C**) No pneumonia of 60-year-old female.

pulmonary diseases. Therefore, accurate diagnosis of COVID-19 pneumonia can be more challenging on CXR than on chest CT.

Computer-aided diagnosis (CADx) is being used for detection and diagnosis in several medical fields. CADx utilizes artificial intelligence methods for improving its diagnostic accuracy and robustness. Recent advances in machine learning, particularly deep learning with convolutional neural network (CNN), have shown promising performance of CADx in classifying disease patterns on medical images, such as CXR and chest CT[8–11].

The purpose of this study was to develop CADx system for classification between COVID-19 pneumonia, non-COVID-19 pneumonia, and the healthy using CXR images and CNN. Since the number of publicly available CXR images of COVID-19 pneumonia was limited, we developed the CNN model which could be accurate and robust even if the training data of CNN was small. The proposed method included the transfer learning, in which CNN models pre-trained on a large dataset is used for the improvement of accuracy and robustness[9,12]. Although this study mainly utilized a commonly used pre-trained model (VGG16), the latest CNN model (EfficientNet) was also used for transfer learning. Next, the combination of data augmentation methods was used for improving model's robustness. In addition to conventional data augmentation method (such as flipping, shifting, rotating, and etc.), mixup, and Random Image Cropping and Patching (RICAP) were used in this study[13–15]. Finally, the model was examined to evaluate whether it distinguishes COVID-19 pneumonia from both non-COVID-19 pneumonia and the healthy on CXR images.

Material and methods

Our study used anonymized data collected from public datasets. Therefore, institutional review board approval was waived according to the regulations of our country. No informed consent was required.

Dataset. Two datasets were used: (I) one dataset for CXR images of COVID-19 and non-COVID-19 pneumonia and (II) the other for CXR images of the healthy and non-COVID-19 pneumonia. (I) The COVID-19 image data collection repository on GitHub is a growing collection of CXR and CT images of COVID-19 pneumonia[16]. In addition to COVID-19 pneumonia, this repository contains a small number of CXR and CT images of non-COVID-19 pneumonia. (II) The RSNA Pneumonia Detection Challenge dataset available on Kaggle contains CXR images of non-COVID-19 pneumonia and the healthy[17]. Figure 1 shows representative CXR images of COVID-19, non-COVID-19 pneumonia, and the healthy.

From the dataset (I), CXR images of lateral view and CT images were excluded, and CXR images of both posterior-anterior and anterior–posterior views were included. Based on these criteria, 215 and 33 CXR images were included from the dataset (I) for COVID-19 and non-COVID-19 pneumonia, respectively. In addition, 500 and 500 CXR images were randomly selected from the dataset (II) for the healthy and non-COVID-19 pneumonia, respectively. In order to avoid strong class imbalance, the 1000 CXR images were selected from the dataset (II). In total, 215, 533, and 500 CXR images of COVID-19 pneumonia, non-COVID-19 pneumonia, and the healthy were used for development and validation of the proposed method.

From the two datasets, patient's age and sex were collected. Table 1 summarizes the patients' characteristics and CXR attributes. The 1248 CXR images were divided into 998, 125, and 125 images for training, validation, test sets, respectively. For image normalization, CXR images were divided by 255, and pixel values of them ranged from 0 to 1.

Deep learning model and data augmentation. VGG16[18] was mainly used as deep learning model for the proposed method, and transfer learning was performed for the classification of CXR images of COVID-19, non-COVID-19 pneumonia, and the healthy. Based on our preliminary experiments, VGG16 without transfer learning easily led to overfitting and performance degradation. To search for optimal hyperparameters of the VGG16-based model and combination of data augmentation methods, random search was performed[19]. The outline of deep learning model is shown in Fig. 2.

Publicly available weights of VGG16 obtained by pre-training on ImageNet dataset were used for transfer learning. The layers of VGG16 were sorted in the order of image processing, and all trainable parameters of the 1st–10th layers in VGG16 was frozen for transfer learning.

Category	Value
Number of images	1248
Sex	
Male	512
Female	702
Not available	34
Age	
Available	1205
Not available	43
Mean ± SD of age (years)	48.1 ± 17.5
Diagnosis	
COVID-19	215
Non-COVID-19 pneumonia	533
The healthy	500
CXR view	
PA	666
AP	582

Table 1. Patients' characteristics and CXR attributes. *CXR* chest X-ray imaging, *COVID-19* novel coronavirus disease, *SD* standard deviation, *PA* posterior–anterior view, *AP* anterior–posterior view.

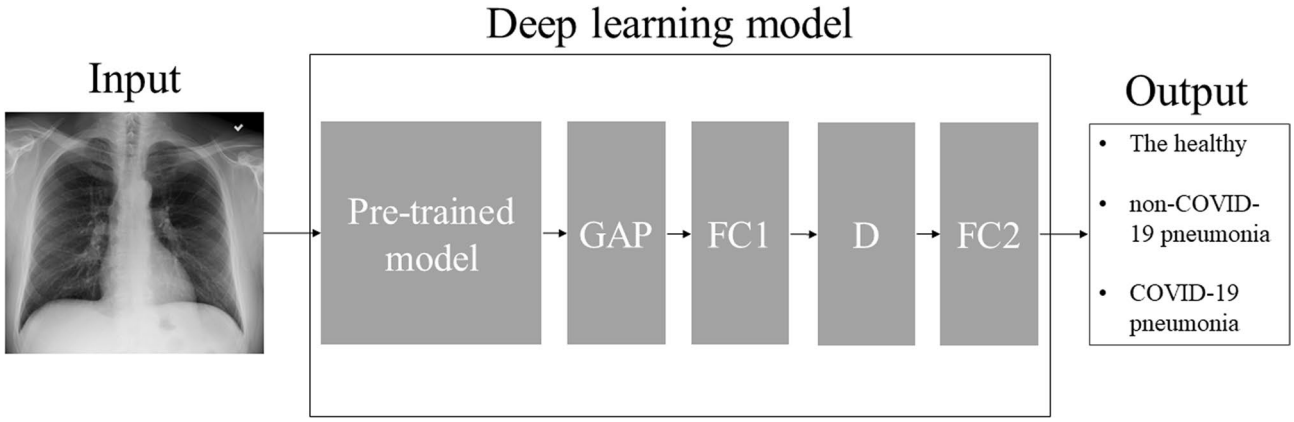

Figure 2. Outline of deep learning model of the proposed method. Note: For pre-trained models, VGG16, Resnet-50, MobileNet, DenseNet-121, and EfficientNet were used in the current study. Activation function is omitted for brevity. *GAP* global averaging pooling layer, *FC* fully-connected layer, *D* dropout layer.

After the convolution layers of VGG16, the global averaging pooling layer, fully-connected layer, and dropout layer were added to VGG16. For the 3-category classification, the final 3-unit fully-connected layer was added after the dropout layer. Activation functions of the first and last fully-connected layer were rectified linear unit and softmax, respectively. Hyperparameters obtained by the random search of the VGG16-based model were as follows. The probability of the dropout layer was 0.1, and the number of units in the first fully-connected layer was 416. RMSprop with learning rate of 1.0×10^{-4} was used as the optimizer, and cross entropy loss between class labels and outputs of the model was reduced by optimizing trainable parameters of the non-frozen layers in VGG16 and the fully-connected layers. The input image size of VGG16 was changed to 220×220 pixels. The network was trained using a batch size of 8, and the number of training epochs was set to 100. Early stopping was enabled using validation loss, and the patience of early stopping was set to 7. Summary of the optimal VGG16-based model is shown in Doc S1 of Supplementary information.

To prevent overfitting in the model training, optimal combination of the three types of data augmentation methods (conventional method, mixup, and RICAP) was also examined by the random search, and combination of conventional method and mixup were used in the proposed method of the VGG16-based model. The conventional data augmentation method included ± 15° rotation, ± 15% x-axis shift, ± 15% y-axis shift, horizontal flipping, and 85–115% scaling and shear transformation. The parameters of mixup was set to 0.1[13].

The training of the model was performed using a PC with a discrete GPU (Nvidia RTX 2080 Ti, RAM 11 GB). Python (version 3.7, https://www.python.org/) was used as the programing language, and Keras (version 2.2.4, https://keras.io/) and TensorFlow (version 1.13.1, https://tensorflow.org/) were used as deep learning frameworks.

Models	Loss of test set	3-category accuracy of test set (%)
VGG16 (proposed method)	0.4682 ± 0.0289	83.68 ± 2.00
Resnet-50	0.5237 ± 0.0161	77.76 ± 1.18
MobileNet	0.4919 ± 0.0300	78.72 ± 3.22
DenseNet-121	0.5276 ± 0.0082	78.24 ± 2.23
EfficientNet	0.5206 ± 0.0177	78.40 ± 1.82

Table 2. Results of five pre-trained models. Value of each cell was mean ± standard deviation of 5 trials.

		Prediction by the proposed model		
		The healthy	Non-COVID-19 pneumonia	COVID-19 pneumonia
Ground truth	The healthy	43	7	0
	Non-COVID-19 pneumonia	9	41	3
	COVID-19 pneumonia	2	0	20

Table 3. Representative confusion matrix of 3-category classification in test set. Accuracy was 83.2% (104/125).

Models	Loss of test set	3-category accuracy of test set (%)
Proposed method	0.4682 ± 0.0289	83.68 ± 2.00
No data augmentation with layer freezing	0.9009 ± 0.1967	78.72 ± 1.65
Conventional data augmentation method only with layer freezing	0.4863 ± 0.0274	82.56 ± 2.45
Mix-up only with layer freezing	0.6407 ± 0.0674	79.20 ± 1.75
Conventional data augmentation method and mixup without layer freezing	0.5143 ± 0.0179	79.04 ± 2.60

Table 4. Results of ablation study of the proposed method for data augmentation methods and layer freezing. Value of each cell was mean ± standard deviation of 5 trials.

Comparison with other pre-trained models and ablation study. To compare with the VGG16-based model, the following four pre-trained models were used for the transfer learning: Resnet-50[20], MobileNet[21], DenseNet-121[22], and EfficientNet[23]. In the transfer learning using these four pre-trained models, the trainable parameters were not frozen; it was found that freezing trainable parameters in these models degraded the model performance. For the four pre-trained models, random search was also done for optimal hyperparameters and combination of data augmentation methods. For EfficientNet, the best model was selected from B0–B7 by the random search.

To evaluate the effectiveness of data augmentation methods and the freezing of trainable parameters in the VGG16-based model, the following modified models were also evaluated for VGG16-based model: (i) no data augmentation with the freezing, (ii) the conventional method only with the freezing, (iii) mixup only with the freezing, and (iv) conventional method and mixup without the freezing.

Performance evaluation. For each model, performance evaluation was done using the 3-category classification (ternary classification) accuracy of the test set with 125 CXR images. To assure robustness of the models, the 3-category accuracy was calculated 5 times by changing random seed, training the models, and evaluating the test set. In addition, sensitivity of COVID-19 pneumonia was also calculated using the VGG16-based model.

Results

Table 2 shows the results of 3-category classification between COVID-19 pneumonia, non-COVID-19 pneumonia, and the healthy for the five pre-trained models including the proposed method. The results were obtained by the random search to find the optimal hyperparameters and combination of data augmentation methods. Training time per one epoch was less than 20 s in the optimal VGG16-based model. As shown in Table 2, the mean accuracy of the VGG16-based model of proposed method was 83.7%. The mean accuracies of Resnet-50, MobileNet, DenseNet-121, and EfficientNet were lower than the VGG16-based model of propose method; the mean accuracies of these four models were less than 80%. The mean sensitivity of COVID-19 pneumonia was 90.9% for the VGG16-based model. Table 3 shows representative confusion matrix of 3-category classification in the test set.

In Table 4, the effectiveness of the data augmentation methods and the freezing of trainable parameters were evaluated in the VGG16-based model. Table 4 shows that the layer freezing was effective. The combination of two types of data augmentation methods in the proposed method was more effective than single type or no data augmentation methods.

Table S1 of Supplementary information shows the effect of RICAP obtained by the random search. Although the combination of conventional method, mixup, and RICAP was also evaluated, the combination of three methods was not as good as that of proposed method based on the results of random search. The combination of conventional method and RICAP was slightly inferior to the combination of conventional method and mixup. Therefore, the combination of conventional method and RICAP was not examined intensively in the current study.

Discussion

The results of this study indicate that it was possible to construct an accurate CNN model by using both the transfer learning with VGG16 and the combination of data augmentation methods. Our results show that diagnostic accuracy of the 3-category classification between COVID-19 pneumonia, non-COVID-19 pneumonia, and the healthy was more than 80% in the proposed method. In addition, the sensitivity of COVID-19 was more than 90%.

Table 4 shows that the combination of two types of data augmentation methods was more effective than single type or no data augmentation methods. Our results were consistent with the results of previous study done for bone segmentation with CNN[15]. Because the dataset of the current study was relatively small-sized (number of CXR images was 1248), it was necessary to improve the robustness of CNN models. For this purpose, the current study used the combination of data augmentation methods. The combination of conventional method and mixup was most effective.

Among the several types of pre-trained models, VGG16 was the most accurate for the 3-category classification. Although the classification accuracy of ImageNet dataset was higher in other models than that in VGG16, our results were not compatible with the results of ImageNet dataset. Since the other models are more complicated (e.g., residual learning) and/or have a large number of trainable parameters, overfitting may have occurred in the current study with the small-sized dataset. In addition, because of number of trainable parameters and/or complex structure of networks, hyperparameter tuning was more difficult in Resnet-50, MobileNet, DenseNet-121, and EfficientNet than VGG16. The effectiveness of pre-trained models in a small-sized dataset should be further investigated.

The layer freezing of the trainable parameters was effective only in VGG16. Network architecture of VGG16 was simpler than the other models. For example, skip connection for residual learning in the other networks may hinder the layer freezing. This may affect the usefulness of layer freezing in CNN models.

According to an article of towardsdatascience.com[24], several previous studies constructed datasets by adding pediatric CXR images of non-COVID-19 pneumonia to adult CXR images of COVID-19 pneumonia. However, when a CNN model is trained by these datasets, the model may try to distinguish between non-COVID-19 pneumonia and COVID-19 pneumonia by checking age differences between children and adults rather than disease differentiation. Therefore, this study only included adult CXR images of the healthy and non-COVID-19 pneumonia from the RSNA dataset.

There are some limitations in our study. First, we developed and validated the proposed method using the public datasets. The results of the current study only show that our CADx system could achieve high accuracy in the public datasets. Characteristics of the public datasets may be different from that of clinical data. In this case, overfitting may have occurred in external validation. It is necessary to investigate the usefulness of our CADx system using clinical data. Second, our CADx system was not used by clinicians. Clinical usefulness of our CADx system was not validated.

Conclusion

In conclusion, it is possible to build an accurate CADx system for 3-category classification of COVID-19 pneumonia, non-COVID-19 pneumonia, and the healthy using the proposed method. The combination of two types of data augmentation methods was more useful than single type or no data augmentation methods. We will investigate performance of our CADx system when clinical CXR images with COVID-19 pneumonia, non-COVID-19 pneumonia, and the healthy were fed to the system.

Code availability

Source code and dataset of the current study are available at https://github.com/jurader/covid19_xp.

Received: 16 June 2020; Accepted: 1 October 2020
Published online: 16 October 2020

References

1. WHO | Novel Coronavirus—China. *WHO* (2020). https://www.who.int/csr/don/12-january-2020-novel-coronavirus-china/en/.
2. Stoecklin, S. B. *et al.* First cases of coronavirus disease 2019 (COVID-19) in France: surveillance, investigations and control measures, January 2020. *Eurosurveillance* Vol. 25 (2020).
3. COVID-19 situation reports. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports.
4. Fang, Y. *et al.* Sensitivity of chest CT for COVID-19: comparison to RT-PCR. *Radiology* https://doi.org/10.1148/radiol.2020200432 (2020).
5. Bai, H. X. *et al.* Performance of radiologists in differentiating COVID-19 from viral pneumonia on chest CT. *Radiology* https://doi.org/10.1148/radiol.2020200823 (2020).

6. Sodickson, A. *et al.* Recurrent CT, cumulative radiation exposure, and associated radiation-induced cancer risks from CT of adults. *Radiology* **251**, 175–184 (2009).
7. Wong, H. Y. F. *et al.* Frequency and distribution of chest radiographic findings in COVID-19 positive patients. *Radiology* https://doi.org/10.1148/radiol.2020201160 (2019).
8. Yamashita, R., Nishio, M., Do, R. K. G. & Togashi, K. Convolutional neural networks: an overview and application in radiology. *Insights Imaging* **9**, 611–629 (2018).
9. Nishio, M. *et al.* Computer-aided diagnosis of lung nodule classification between benign nodule, primary lung cancer, and metastatic lung cancer at different image size using deep convolutional neural network with transfer learning. *PLoS ONE* **13**, 1–12 (2018).
10. Chilamkurthy, S. *et al.* Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study. *Lancet* **392**, 2388–2396 (2018).
11. Rajpurkar, P. *et al.* CheXNet: radiologist-level pneumonia detection on chest X-rays with deep learning. (2017).
12. Tan, C. *et al.* A survey on deep transfer learning. In *Lecture Notes in Computer Science* (*including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics*) Vol. 11141. LNCS, 270–279 (Springer, 2018).
13. Zhang, H., Cisse, M., Dauphin, Y. N. & Lopez-Paz, D. Mixup: beyond empirical risk minimization. *Polit. Afr.* 1–13 (2017).
14. Takahashi, R., Matsubara, T. & Uehara, K. Data augmentation using random image cropping and patching for deep CNNs. *IEEE Trans. Circuits Syst. Video Technol.* https://doi.org/10.1109/tcsvt.2019.2935128 (2019).
15. Noguchi, S., Nishio, M., Yakami, M., Nakagomi, K. & Togashi, K. Bone segmentation on whole-body CT using convolutional neural network with novel data augmentation techniques. *Comput. Biol. Med.***121** (2020).
16. Cohen, J. P., Morrison, P. & Dao, L. COVID-19 image data collection (2020).
17. RSNA Pneumonia Detection Challenge | Kaggle. https://www.kaggle.com/c/rsna-pneumonia-detection-challenge.
18. Simonyan, K. & Zisserman, A. Very deep convolutional networks for large-scale image recognition. In *3rd International Conference on Learning Representations, ICLR 2015—Conference Track Proceedings* (International Conference on Learning Representations, ICLR, 2015).
19. Bergstra, J., Ca, J. B. & Ca, Y. B. *Random Search for Hyper-Parameter Optimization Yoshua Bengio*. *Journal of Machine Learning Research*, Vol. 13. https://scikit-learn.sourceforge.net (2012).
20. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 2016-December, 770–778 (IEEE Computer Society, 2016).
21. Howard, A. G. *et al.* MobileNets: efficient convolutional neural networks for mobile vision applications (2017).
22. Huang, G., Liu, Z., Van Der Maaten, L. & Weinberger, K. Q. Densely connected convolutional networks. In *Proceedings—30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, Vol. 2017-January 2261–2269 (Institute of Electrical and Electronics Engineers Inc., 2017).
23. Tan, M. & Le, Q. V. EfficientNet: rethinking model scaling for convolutional neural networks. *36th Int. Conf. Mach. Learn. ICML 2019* 2019-June, 10691–10700 (2019).
24. Explainable AI and COVID-19 Chest X-rays | Towards Data Science. https://towardsdatascience.com/investigation-of-explainable-predictions-of-covid-19-infection-from-chest-x-rays-with-machine-cb370f46af1d.

Author contributions

Conceptualization: M.N. Data curation: M.N. Formal analysis: M.N. Funding acquisition: M.N. Investigation: M.N. Methodology: M.N. Project administration: M.N. Resources: M.N. Software: M.N., S.N. Supervision: T.M. Validation: M.N., H.M. Visualization: M.N. Writing—original draft: M.N. Writing—review and editing: M.N., S.N., H.M., T.M.

Funding

The present study was supported by JSPS KAKENHI (Grant Number 19H03599 and JP19K17232).

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41598-020-74539-2.

Correspondence and requests for materials should be addressed to M.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.