A Deep Learning System to Diagnose COVID-19 Pneumonia Using Masked Lung CT Images to Avoid AI-generated COVID-19 Diagnoses that Include Data outside the Lungs

Takashi NAGAOKA,*, # Takenori KOZUKA,** Takahiro YAMADA,*** Hitoshi HABE,† †† Mitsutaka NEMOTO,††† Masahiro TADA,‡ ‡‡ Koji ABE,‡ ‡‡ Hisashi HANDA,‡ ‡‡ Hisashi YOSHIDA,† Kazunari ISHII,** ‡‡‡ Yuichi KIMURA*, ‡‡‡

Abstract

Objective: The objective of the current study was to develop a novel, artificial intelligence (AI)-based system to diagnose coronavirus disease (COVID-19) using computed tomography (CT) slice images. Prior research has demonstrated that, if not focused on the lungs, AI diagnoses COVID-19 using information outside the lungs. The inclusion of CT training data from multiple facilities and CT models may also cause AI to diagnose COVID-19 with features that are irrelevant to COVID-19. Thus, the objective of the current study was to evaluate a combination of lung mask images and CT slice images from a single facility, using a single CT model, and use AI to differentiate COVID-19 from other types of pneumonia based solely on information related to the lungs.

Method: By superimposing lung mask images on image feature output using an existing AI structure, it was possible to exclude image features other than those around the lungs. The results of this model were also compared with the slice image findings from which only the lung region was extracted. The system adopted an ensemble approach. The outputs of multiple AIs were averaged to differentiate COVID-19 cases from other types of pneumonia, based on CT slice images.

Results: The system evaluated 132 scans of COVID-19 cases and 62 scans of non-COVID-19 cases taken at the single facility using a single CT model. The initial sensitivity, specificity, and accuracy of our system, using a threshold value of 0.50, was shown to be 95%, 53%, and 81%, respectively. Setting the threshold value to 0.84 adjusted the sensitivity and specificity to clinically usable values of 76% and 84%, respectively.

Conclusion: The system developed in the current study was able to differentiate between pneumonia due to COVID-19 and other types of pneumonia with sufficient accuracy for use in clinical practice. This was accomplished without the inclusion of images of clinically meaningless regions and despite the application of more stringent conditions, compared to prior studies.

Keywords: COVID-19, deep learning, artificial intelligence, lung, attention, xAI.

Adv Biomed Eng. 11: pp. 76–86, 2022.
puted tomography (CT) data obtained from patients with pneumonia due to COVID-19 were evaluated. CT is an invaluable tool for identifying the presence of and measuring the progress of pneumonia symptoms, and if CT data were able to differentiate pneumonia due to COVID-19 from other types of pneumonia, the clinical impact would be significant. In this regard, a diagnostic algorithm is needed since a diagnosis based on CT is subjective and is usually performed by radiologists. The effectiveness of using deep learning to read CT scans is receiving attention [1], and deep learning has already been used to interpret other types of medical images [2].

Previous studies have attempted to differentiate pneumonia due to COVID-19 from other types of pneumonia using CT images and deep learning. A full bibliography of related studies is provided in a review [3]. A key implication of these findings is that it is relatively easy for deep learning systems to differentiate between pneumonia due to COVID-19 and other types of pneumonia based on the CT data, and this has been demonstrated by the development of multiple systems that have done so with over 90% accuracy [4–11]. However, prior research has treated deep learning as complex, with inner workings that are perceived as mysterious or unknown, and it is unclear what evidence deep learning systems use to differentiate pneumonia due to COVID-19 from other types of pneumonia, and this is required if this application is to be expanded to a wide range of CT machines. Focusing AI on a specific area is called attention. Therefore, the present study sought to determine the evidence applied by deep learning to identify COVID-19 cases by controlling the focus of AI attention and evaluating changes in the performance as attention shifts.

Several explainable AI (XAI) technologies has been developed in recent years. Of these, gradient-weighted Class Activation Mapping (Grad-CAM) is frequently used to visualize deep learning decision criteria [12]. Grad-CAM measures the influence of a pixel by calculating each pixel’s gradient in each AI output. If a pixel has a large gradient in a given output, this means that the AI is paying attention to that pixel. Previous studies have used Grad-CAM to visualize the evidence behind AI decisions [7–11], but none of them have quantitatively evaluated Grad-CAM measurements. They simply present the Grad-CAM results and report the broad trends only. Previous reports of COVID-19 diagnostic AI-based methods revealed high Grad-CAM values outside the lungs [7, 10, 11]. Therefore, even if a deep learning system could differentiate between pneumonia due to COVID-19 and other types of pneumonia with high accuracy, a risk exists that the system may use information outside the lungs. To the best of our knowledge, previous studies have not focused on the use of AI in relation to the lungs to make a clinically valid diagnosis; neither have they analyzed the Grad-CAM outputs quantitatively.

Methods such as attention branch network [13] and squeeze-and-excitation networks [14], for which image identification and attention are trained simultaneously, have been reported to be effective in this regard. Since it has been established that the lungs can be easily separated from the CT data, in the current study, attention was not trained at the same time as image identification; instead, a separately prepared lung mask image was used.

Studies have developed COVID-19 diagnostic systems using CT slice images, but few have used well-conditioned CT data or extracted lung images to train the system. If possible, the conditions of the two groups (cases of COVID-19 due to pneumonia and cases of other types of pneumonia) should be the same; for example, with reference to table shape and machine-specific noise, to ensure that the AI does not use features other than the symptoms of COVID-19 pneumonia to reach a diagnosis. Since deep learning greedily searches for differences in the input data, obviously if multiple facilities or multiple CT models are used, the deep learning model has the potential to learn the differences between the facilities or CT models without learning the features related to COVID-19. Therefore, in the present study, the data were collected from patients with pneumonia due to COVID-19 and from patients with other types of pneumonia using a single CT machine. This meant that the conditions for all patients were almost identical, except for the cause of pneumonia.

The system had three features. First, it differentiated between pneumonia due to COVID-19 and other types of pneumonia using CT data and a widely used two-dimensional deep learning model. The deep learning model was trained using each CT slice image to ensure that a sufficient number of images were used to train the system. Therefore, fine-tuning was employed instead of transfer learning or a support vector machine which is often used when the number of training images is small. Second, since there was no need to read healthy CT slice images in a clinical setting, the CT slice images of healthy lungs were excluded from the training. Therefore, the output of our system is whether or not the input CT slice image is COVID-19. Third, the image features extracted by deep learning were multiplied by the lung mask to exclude image features other than those related to the lungs. This meant that no data other than those pertaining to the lungs were used to differentiate COVID-19 from types of pneumonia. The system achieved over 80% accuracy, which is considered sufficient for clinical use.
2. Methodology

This study was conducted retrospectively using data obtained for clinical purposes. The research was conducted according to a protocol approved by the institutional review board of the Faculty of Biology-Oriented Science and Technology at Kindai University, Kinokawa, Japan, and by the research ethics committee of the Faculty of Medicine at that same university, Osaka-Sayama, Japan. To provide the patients with an opportunity to withdraw their data from the study, information on the withdrawal procedure was posted on our website, and the data of subjects who refused to participate in the study were removed from the analysis and destroyed.

2.1 COVID-19 classification grades

From the collected CT data, the slices to be read were selected by a radiologist with sufficient experience in reading chest CTs, including those of COVID-19 patients. The radiologist subjectively classified the COVID-19 status of each scan according to seven grades of illness (G0–G6). The classification criteria are detailed in Table 1.

2.2 Subjects

All the CT slice images used in the study were 512 pixels wide and 512 pixels high. The voxel intensity of the CT data was measured using Hounsfield units (HUs); all images had a value of 16 bits. The CT data were exported in the digital imaging and communications in medicine (DICOM) format. In this study, the CT data were measured using a single CT machine installed at the university (Optima® CT660, GE Healthcare, USA) to prevent the deep learning system from being trained to recognize differences in images taken by different CT machines. Ultimately, 132 scans and 4,665 slices of CT data from 74 COVID-19 patients and 62 scans and 1,834 slices from 58 non-COVID-19 patients were analyzed. The numbers of patients, scans, and slices included in the study and the radiologist’s categorization of the images are shown in Table 2. All the COVID-19 patients were diagnosed as positive using a RT-PCR test. Non-COVID patients included 51 patients with pneumonia attributable to causes other than COVID-19, and seven patients

| Table 1 | COVID-19 classification grades and criteria. |
|---------|---------------------------------------------|
| Grade   | Criteria                                    |
| G0      | Symptoms of pneumonia were not observed.    |
| G1      | Symptoms of pneumonia, other than pneumonia due to COVID-19, were observed. |
| G2      | Symptoms of pneumonia that could not be ruled out as due to COVID-19 were observed. |
| G3      | A case of COVID-19 considered to be unsuitable for analysis or did not require a diagnosis (i.e., evidence of a faint shadow due to treatment or other reasons). |
| G4      | A case of COVID-19 considered to be unsuitable because the image quality was inadequate and could not be examined correctly. |
| G5      | A case of mild to moderate COVID-19 considered by the radiologist to be difficult to diagnose. |
| G6      | A typical case of COVID-19 considered by the radiologist to be easy to diagnose. |

| Table 2 | Subjects analyzed in the study. |
|---------|----------------------------------|
| Patients Scans CT slices Analyzed |
| Pneumonia due to COVID-19 74 132 8,833 4,665 |
| G0 0 0 0 |
| G1 4 269 100 |
| G2 0 0 0 |
| G3 43 2,824 1,417 |
| G4 2 152 71 |
| G5 43 2,881 1,534 |
| G6 40 2,707 1,543 |
| Grades |
| Other types of pneumonia 58 62 4,346 1,834 |

COVID-19: coronavirus disease 2019, CT: computed tomography.
with extensive shadowing suggestive of pulmonary edema. As stated previously, the system was developed to support radiologists in their task of diagnosing whether or not the pneumonia was due to COVID-19. Therefore, healthy subjects without pneumonia were excluded from the study.

2.3 Image preprocessing
Before inputting into the deep learning model, the CT slice images were preprocessed. Initially, the lung area was extracted from the DICOM volume data. An image, hereinafter referred to as a mask image, was generated in a binary format. Lungmask version 0.2.4 [15] was used to extract the lung images. This software is a lung extractor that uses a deep learning model called U-net [16]. It was trained using 231 chest CTs. Next, the DICOM image and the mask image were combined to generate an image of only the lung. This image was referred to as the lung image, and the image without lung extraction was referred to as the whole image. All the images were saved as 8-bit portable network graphics (PNG) images. When the DICOM image was converted to a PNG image, the window level and width were set to values close to those in common lung conditions (−700 and 1,400), and the image was normalized so that the minimum value was 0 and the maximum value was 255. In the lung images, all areas outside the lungs were filled with zeroes. Finally, the images were resized to 224 pixels wide and 224 pixels high to prepare them for input into the deep learning model.

2.4 Identifying COVID-19 using a deep learning model
In this study, two types of deep learning models and two types of input images were used. These two models differed in the use of mask images. The effectiveness of attention was verified by these models. The two inputted images differed in their use of lung extraction. The difference between attention and lung extraction in the mask images was validated by these images. The first model is depicted in Fig. 1.

From the models trained on ImageNet [17] and provided by TensorFlow, v. 1.15.2 [18], InceptionV3 [19] was selected for use in this study. InceptionV3 outputted 2,048 sets of image features with 5 × 5 pixels. The image features were then passed through a global average pooling (GAP) layer and a 256-node fully concatenated (FC) layer. They were then outputted from the final two-node FC layer as a two-class classification: pneumonia due to COVID-19 and other types of pneumonia. Hereinafter in this paper, this model is referred to as InceptionV3 without the mask. Next, the CT slice and mask images were inputted in a deep learning model. The second model is depicted in Fig. 2.

A mask image, resized to 5 × 5 pixels, with binary information in one channel (the same size as the image feature output by InceptionV3), was inputted into the network derived from the CT slice images using Lungmask. The mask image was then multiplied by the image features of the CT slice image. The multiplied image features were processed in the same network structure as InceptionV3 without the mask. Hereinafter in this paper, this model is referred to as InceptionV3 with the mask. The original mask image was 224 pixels wide, 224 pixels high, and has one channel of binary information. When the mask image was resized to 5 × 5 pixels, the value of each pixel indicated the probability of the presence of a lung within the area measuring 44.8 (224/5) pixels wide and 44.8 pixels high in the original mask image. When the image features were multiplied by the resized mask image, any image features other than those around the lungs were excluded. The expectation was that this process would enable the deep learning system to identify images of COVID-19 cases based only on the image features pertaining to the lungs. In the final layer of the deep learning model, the features of the image that were as independent of positional information as possible were extracted. It was possible to multiply the mask in shallower (higher resolution) layers, but the possibility that the subsequent layers would again direct attention to something other than the lungs could not be ruled out. Therefore, the mask in the final layer was multiplied.

Fig. 1  Deep learning model of InceptionV3 without the mask.
The CT slice images were normalized and then inputted into the deep learning model. The hyperparameters for the deep learning model were 10 epochs, batch size of 64, and Adam optimization algorithm [20]. The initial learning rate was set at 0.001; this was multiplied by $\frac{1}{10}$ for the progress of one third of the epochs.

An AI system designed to differentiate between pneumonia due to COVID-19 and other types of pneumonia is naturally expected to focus only on the lungs. In this study, the Grad-CAM xAI technique was used to assess the attention of the AI system. The layer just before the GAP layer was set as the feature map layer for Grad-CAM. In other words, two layers were selected as the feature map layer: the layer that outputted the image features in the InceptionV3 without the mask and the layer that outputted the weighted image features multiplied by the mask image in the InceptionV3 with the mask. Grad-CAM generated the gradient information for each inputted CT slice image; the inputted images were the same size as the image features ($5 \times 5$). The location of the pixels with the highest gradient value was then identified from the gradient information calculated by Grad-CAM. An image of the average of all the CT slice images resized to $5 \times 5$ pixels was generated and compared with the Grad-CAM output.

2.5 Analysis

Repeated crossover validation was performed to ensure validity. The CT data from patients with pneumonia due to COVID-19 and patients with other types of pneumonia were randomly assigned to training or evaluation dataset in ratio of 2:1 or 4:1 according to an ensemble approach [21]. Since the number of scans and slices varied from patient to patient, the deep learning loss was scaled by the number of slices in each class. Data augmentation methods were not used. The model that was trained to the final epoch was then used to analyze the evaluation data using an ensemble approach.

The assignment of patients to either the training or evaluation dataset was changed randomly, and three- and five-way crossover validations were conducted thrice and once, respectively. This meant that every CT slice image was evaluated by a total of four deep learning models. The average values of these four evaluations were then used to evaluate the system developed. The model used in this study produced a final output of two values: the probability that the inputted CT slice image was from a patient with pneumonia due to COVID-19 (hereinafter referred to in this paper as COVID-19 probability) and the probability that it was from a patient with other types of pneumonia (hereinafter referred to in this paper as non-COVID-19 probability). Note that the sum of the two probabilities is 1. Finally, each scan was identified as pneumonia due to COVID-19 or other types of pneumonia based on the average COVID-19 probability of all the CT slices analyzed from that scan. Hereinafter, in this paper, the analysis results for each CT slice image and each scan are referred to as “slice by slice” and “scan by scan,” respectively. When COVID-19 probability exceeded a certain threshold value, the CT slice image or scan was identified as diagnostic of pneumonia due to COVID-19.

To compare the effectiveness of our model with those applied in previous studies and to confirm the use-
fulness of the masked images, four types of input images and models were assessed: unmasked whole image, masked whole image, unmasked lung image, and masked lung image. For the unmasked whole image, the deep learning model was trained and evaluated by inputting the whole image into InceptionV3 without the mask. Similarly, for the masked lung image, the deep learning model was trained and evaluated by inputting the lung image into InceptionV3 with the mask. In terms of the unmasked images, the deep learning model was trained and evaluated according to the methods described in previous studies [4–11]. The use of a lung mask to limit the input to lung images was a novel approach in the present study. Whole and lung images were also inputted into the two models to measure the effectiveness of the model and evaluate the impact of lung extraction on COVID-19 differentiation using deep learning.

In the current study, sensitivity, specificity, accuracy, and the area under the receiver operating characteristic curve were used to assess the effectiveness of the system. Student’s t-test was utilized to evaluate significant differences, with statistical significance set at 5%. The statistical analysis was performed using GraphPad Prism 6 (GraphPad Software, California, USA).

3. Results

Examples of the CT slice images (whole images), lung

images, and masked images inputted into the deep learning model, including both a case of pneumonia due to COVID-19 and a case of other type of pneumonia, are shown in Fig. 3. The masked images were enlarged to the same size as the other images. In certain cases, Lung-mask was unable to extract nodules around the chest wall. Since the lungs were not manually extracted, it was difficult to quantitatively evaluate the extraction via Lungmask. However, the radiologist confirmed that the extracted lung images were suitable.

Tables 3 and 4 depict the slice-by-slice and scan-by-scan COVID-19 differentiation performance of the proposed deep learning model. Differentiation performance did not differ significantly between with masks and without masks.

Figures 4 and 5 show the receiver operating characteristic curve and how false positive rate (specificity of −1) and false negative rate (sensitivity of −1) changed with different threshold values for InceptionV3 with masks (whole image input). Of the four types of images used, InceptionV3 masked whole images gave the best performance. As shown in Figs. 4 and 5, the balance between specificity and sensitivity depended on the threshold value. When the threshold value increased to 0.84, the balanced error ratio or the average value of the false positive and false negative rates was minimized, and sensitivity and specificity were 76% and 84%, respectively.

| Input          | Network        | Sensitivity (%) | Specificity (%) | Accuracy (%) | AUC   |
|----------------|----------------|-----------------|-----------------|--------------|-------|
| Whole images   | Without masks  | 86.3            | 57.0            | 75.4         | 0.825 |
| Whole images   | With masks     | 88.9            | 55.1            | 76.3         | 0.823 |
| Lung images    | Without masks  | 87.8            | 41.3            | 70.5         | 0.767 |
| Lung images    | With masks     | 88.3            | 42.6            | 71.2         | 0.764 |

AUC: area under the receiver operating characteristic curve.
Table 5 shows the sensitivity for each grade using InceptionV3 with masks (whole image input) in cases where the threshold values were set at 0.50 and 0.75. For most trials, sensitivity remained above 70%; the exceptions were two grades of illness for which there were markedly fewer scans (G1 and G4).

The impact of the use of masks on AI attention is shown in Fig. 6, which reflects a typical Grad-CAM output. The pixels with the highest gradient values; that is, the locations where the attention of the AI was focused, are shown in red.

![Image](image_url)

**Fig. 4** ROC curve for InceptionV3 with masks (whole image input, scan by scan).

**Fig. 5** Changes in the false positive and false negative rates by threshold value (whole image input, scan by scan).

The spatial distribution of the pixels with the highest gradient values (red shaded pixels) for all CT slice images in all four models is illustrated in Fig. 7. With the use of masks, the pixels with the highest gradient values were located around the lungs, because all the image features other than those around the lungs were set to zero. However, without the use of masks, the pixels with the highest gradient values were mostly found outside the lungs. The majority were located in the fifth row corresponding to the image of the table on which the patient lay during the CT scan. In the unmasked lung image, most pixels with the highest gradient values were located around the lungs; however, in several images, they occurred outside the lungs, where the pixel value was zero.

![Image](image_url)

**Table 4** Scan-by-scan performance in COVID-19 differentiation.

| Input       | Network    | Sensitivity (%) | Specificity (%) | Accuracy (%) | AUC  |
|-------------|------------|-----------------|-----------------|--------------|------|
| Whole images| Without masks | 91.7            | 53.2            | 79.4         | 0.872|
| Whole images| With masks     | 94.7            | 53.2            | 81.4         | 0.859|
| Lung images | Without masks | 95.5            | 41.9            | 78.4         | 0.848|
| Lung images | With masks     | 94.7            | 48.4            | 79.9         | 0.845|

AUC: area under the receiver operating characteristic curve.

**Table 5** Sensitivity for each grade using InceptionV3 with masks (whole image input).

| Grades | Scans | Threshold value of 0.50 | Threshold value of 0.75 |
|--------|-------|-------------------------|-------------------------|
|        | n (%) | n (%)                  | n (%)                  |
| G1     | 4     | 3 (75)                  | 2 (50)                  |
| G3     | 43    | 40 (93)                 | 30 (70)                 |
| G4     | 2     | 1 (50)                  | 1 (50)                  |
| G5     | 43    | 42 (98)                 | 37 (86)                 |
| G6     | 40    | 40 (100)                | 33 (83)                 |
4. Discussion

In clinical practice, the use of CT slice images is extremely useful when differentiation between pneumonia due to COVID-19 and other types of pneumonia is needed in the initial stages of diagnosis. In the current study, the proposed diagnostic system achieved accuracy of 80% despite the application of more stringent study conditions (the inclusion of a single facility and use of a single CT model) compared to the conditions used in previous studies [4–11]. In the current research, images taken by a single CT machine were used to differentiate cases of pneumonia due to COVID-19 and other types of pneumonia, and accuracy in this regard was slightly lower or nearly the same as that obtained in previous studies. Even for G5 images, which radiologists typically find challenging when making a diagnosis, our system obtained sensitivity of 86% (with a threshold value of 0.75). Since the system yielded high sensitivity, even for grade G5 images, it would effectively assist radiologists engaged in daily readings owing to the spread of COVID-19 to make accurate COVID-19 diagnosis. A prototype of our system has already been introduced in clinical practice. We will report on its clinical effectiveness in a future paper. Similar to systems developed in previous studies, the specificity of the proposed system was not very high, and this can be explained by the fact that the CT images of COVID-19 cases are similar to those of cases with other infections.

The specificity of the proposed system is extremely low at the threshold value of 0.50, but specificity can be increased by increasing the threshold value provided that it is not raised too much since this negatively impacts sensitivity. Our understanding is that appropriate threshold values should be chosen when developing a clinically useful AI system, and that a standard threshold value of 0.50 may not be appropriate for all systems. In addition, the low specificity of the proposed system could be improved by including more training data and incorporating various data augmentation methods in the training.

Fig. 6  Typical Grad-CAM outputs for a case of pneumonia due to COVID-19 and a case of other type of pneumonia.

Fig. 7  Distribution of pixel locations with the highest gradient values.
A distinctive and significant finding of the current study was the effect of lung extraction and the use of masks on AI-based diagnosis. A significant difference in diagnostic accuracy was not observed by implementing lung extraction or use of masks. However, the Grad-CAM results strongly suggested that without the use of masks, the AI tended to use information from outside the lungs to differentiate pneumonia due to COVID-19 from other types of pneumonia. A similar trend has been reported in previous studies [7, 10, 11]. Making a diagnosis of pneumonia based on images of regions other than the lungs contradicts clinical practice standards. Although all the scans used in the current study were taken with a single CT device, there was a slight possibility that the AI might have diagnosed COVID-19 cases based on differences in imaging parameters or slight changes in patient position. Regarding the Grad-CAM output for images without masks, the highest gradient values were often associated with the lower part of the CT slice image. It is unlikely that anything in this region was attributable to COVID-19, as this region corresponded to the table, not the patient, in most cases. Thus, the AI may have recognized differences in the patient position, body shape, and table shape, rather than focusing attention on the medical conditions caused by COVID-19 when making the diagnosis. Regarding the Grad-CAM output for images with masks, pixels with the highest gradient values were in the lungs, which indicated that the AI might have identified COVID-19 cases based on lung-related data. A similar trend was seen with the lung images. In the unmasked lung image model, considerable attention was focused on areas outside the lungs without luminance values. This meant that the AI tried to diagnose COVID-19 using the shape of the region outside the lungs where the luminance value was zero. In deep learning models, zero is treated like any other value. In addition, even if a pixel has a value of zero, the feature value of that pixel output by the deep learning model could be non-zero. Since Grad-CAM was created from the feature map, it was quite possible that increased attention was associated with the area with a luminance value of zero.

Diagnostic performance using lung images was also slightly lower than that using whole images. The fact that the lungs were not extracted accurately may have affected the results. In certain cases, the location of increased attention differed between masked whole images and masked lung images. These and other aspects of the relationship between AI output and clinical findings will be reported in future papers.

In the current study, there remains a slight possibility that the AI differentiates the COVID-19 images from the other images using other information unrelated to COVID-19. Therefore, further research into and the development of XAI technology is warranted. Another limitation of this study was that the number of cases evaluated was limited owing to the inclusion of a single facility and the use of a single CT model. The authors believe that AI should be trained for each CT model given the differences in CT models and facilities and the noise inherent in CT models. A more detailed study involving more cases will be performed in the future.

5. Conclusion

A COVID-19 diagnostic support system that utilized CT slice images was developed in the current study. The proposed system, which employed deep learning for lung extraction and differentiation, achieved 81% accuracy. A unique feature of this system was that the AI combined deep learning with lung extraction to differentiate between pneumonia due to COVID-19 and other types of pneumonia using information around the lungs. The methods used in previous studies that involved training the AI using whole CT slice images or whole lung images may have trained the AI to use unrelated parts of the image, specifically those outside the lung, to differentiate pneumonia due to COVID-19 from other types of pneumonia. Although the method used in the current study did not improve the diagnostic performance significantly, nonetheless, it demonstrated that the system was able to successfully differentiate between pneumonia due to COVID-19 and other types of pneumonia without utilizing data pertaining to clinically irrelevant regions.

Conflicts of interest

The authors have no conflicts of interest to declare that are relevant to the content of this article.

Acknowledgement

This study was financially supported by JSPS KAKENHI (JP20K12698) and All-Kindai University support project against COVID-19 provided by the Kindai University.

References

1. Kozuka T, Matsukubo Y, Kadoba T, Oda T, Suzuki A, Hyodo T, Im S, Kaida H, Yagyu Y, Tsurusaki M, Matsuki M, Ishii K: Efficiency of a computer-aided diagnosis (CAD) system with deep learning in detection of pulmonary nodules on 1-mm-thick images of computed tomography. Jpn J Radiol. 38(11), 1052–1061, 2020.
2. Nagaoka T: Recent advances in diagnostic technologies for melanoma. Adv Biomed Eng. 5, 111–117, 2016.
3. Shi F, Wang J, Shi J, Wu Z, Wang Q, Tang Z, He K, Shi Y, Shen D: Review of artificial intelligence techniques in imaging data acquisition, segmentation, and diagnosis for COVID-19. IEEE Rev Biomed Eng. 14, 4–15, 2020.
4. Song Y, Zheng S, Li L, Zhang X, Zhang X, Huang Z, Chen J, Wang R, Zhao H, Zha Y, Shen J, Chong Y, Yang Y: Deep learning enables accurate diagnosis of novel coronavirus (COVID-19) with CT images. IEEE/ACM Trans Comput Biol Bioinform. 18(6), 2775–2780, 2021.

5. Soares E, Angelov P, Biaso S, Froes MH, Abe DK: SARS-CoV-2 CT-scan dataset: A large dataset of real patients CT scans for SARS-CoV-2 identification. MedRxiv. 2020.04.24.20078584, 2020.

6. Jaiswal A, Gianchandani N, Singh D, Kumar V, Kaur M: Classification of the COVID-19 infected patients using DenseNet201 based deep transfer learning. J Biomol Struct Dyn. 39(15), 5682–5689, 2021.

7. Yu Z, Li X, Sun H, Wang J, Zhao T, Chen H, Ma Y, Zhu S, Xie Z: Rapid identification of COVID-19 severity in CT scans through classification of deep features. BioMed Eng OnLine. 19(1), 1–13, 2020.

8. Ko H, Chung H, Kang WS, Kim KW, Shin Y, Kang SJ, Lee JH, Kim YJ, Kim NY, Jung H, Lee, J: COVID-19 pneumonia diagnosis using a simple 2D deep learning framework with a single chest CT image: model development and validation. J Med Internet Res. 22(6), e19569, 2020.

9. Wang S, Zha Y, Li W, Wu Q, Li X, Niu M, Wang M, Qiu X, Li H, Yu H, Gong W, Bai Y, Li L, Zhu Y, Wang L, Tian J: A fully automatic deep learning system for COVID-19 diagnostic and prognostic analysis. Eur Respir J. 56, 2000775, 2020.

10. Alom MZ, Rahman MM, Nasrin MS, Taha TM, Asari VK: COVID_MTNet: COVID-19 detection with multi-task deep learning approach. arXiv preprint arXiv:2004.03747, 2020.

11. He X, Yang X, Zhang S, Zhao J, Zhang Y, Xing E, Xie P: Sample efficient deep learning for COVID 19 diagnosis based on CT scans. medRxiv 2020.04.13.20063941, 2020.

12. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D: Grad-CAM: Visual explanations from deep networks via gradient-based localization. 2017 IEEE International Conference on Computer Vision (ICCV). 618–626, 2017.

13. Fukui H, Hirakawa T, Yamashita T, Fujiyoshi H: Attention branch network: learning of attention mechanism for visual explanation. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 10705–10714, 2019.

14. Hu J, Shen L, Sun G: Squeeze-and-excitation networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 7132–7141, 2018.

15. Hofmanninger J, Prayer F, Pan J, Röhrich S, Prosch H, Langs G: Automatic lung segmentation in routine imaging is primarily a data diversity problem, not a methodology problem. Eur Radiol Exp. 4, 50, 2020.

16. Ronneberger O, Fischer P, Brox T: U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab N, Hornegger J, Wells W, Frangi A (eds) Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science. vol 9351, Springer, Cham, 2015.

17. Deng J, Dong W, Socher R, Li L, Li K, Li F: ImageNet: A large-scale hierarchical image database. 2009 IEEE Conference on Computer Vision and Pattern Recognition. 248–255, 2009.

18. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, Corrado GS, Davis A, Dean J, Devin M, Ghemawat S, Goodfellow I, Harp A, Irving G, Isard M, Jozefowicz R, Jia Y, Kaiser L, Komlo M, Levenberg J, Mané D, Schuster M, Monga R, Moore S, Murray D, Olah C, Shlens J, Steiner B, Sutskever I, Talwar K, Tucker P, Vanhoucke V, Vasudevan V, Viégas F, Vinyals O, Warden P, Wattenberg M, Wicke M, Yu Y, Zheng X: TensorFlow: Large-scale machine learning on heterogeneous systems. Software is available from tensorflow.org, 2015.

19. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z: Rethinking the Inception architecture for computer vision. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2818–2826, 2016.

20. Kingma DP, Ba J: Adam: A method for stochastic optimization. arXiv, 1412.6980, 2014.

21. Qummar S, Khan FG, Shah S, Khan A, Shamshirband S, Rehman ZU, Ahmed Khan IA, Jadoon W: A deep learning ensemble approach for diabetic retinopathy detection. IEEE Access. 7, 150530–150539, 2019.

Takashi NAGAOKA

Takashi NAGAOKA received the Ph.D. degree in engineering from Waseda University, Japan, in 2005. He is lecturer at Faculty of Biology-Oriented Science and Technology, Kindai University, Japan, from 2015. His research fields are dermatology, hyperspectral imaging, and medical systems. He is a member of the Japanese Society for Medical and Biological Engineering, the Society of Instrument and Control Engineers, and the Japanese Skin Cancer Society.

Takahiro YAMADA

Takahiro YAMADA received the Ph.D. degree in medicine from Kindai University, Japan, in 2021. He is a lecturer at the Department of Radiology, Kindai University Faculty of Medicine, Japan, from 2010. His research field is diagnostic radiology. He is a member of the Japan Radiological Society, the Japanese Respiratory Society, and the Japan Lung Cancer Society.
Hitoshi Habe

Hitoshi Habe received his Ph.D. degree in intelligence science and technology from Kyoto University, Japan, in 2006. After working for Mitsubishi Electric Corporation from 1999 to 2002, Kyoto University from 2002 to 2006, NAIST from 2006 to 2011 and Osaka University from 2011 to 2012, he is now an associate professor at Department of Informatics, Faculty of Science and Technology, Kindai University, Japan. From 2010 to 2011, he was a visiting researcher at Department of Engineering, University of Cambridge, UK. His research interests include computer vision, pattern recognition, and image processing. He is a member of IEEE, ACM, IEICE, IPSJ, JSFS and JSAI.

Mitsutaka Nemoto

Mitsutaka Nemoto received the Ph.D. degree in engineering from Tokyo University of Agriculture and Technology, Japan, in 2007. He is a lecturer at Faculty of Biology-Oriented Science and Technology, Kindai University, Japan, from 2017. His research fields are medical image analysis, computer-aided diagnosis / detection, and clinical application of artificial intelligence. He is a member of the Japan Radiological Society, the Japanese Society of Medical Imaging Technology, the Institute of Electronics, Information and Communication Engineers, Japanese Society for Medical and Biological Engineering, and the Japanese Society of Nuclear Medicine.

Masahiro Tada

Masahiro Tada received the Ph.D. degree in engineering from Chuo University, Japan, in 2001. He is associate professor at Department of Informatics, Kindai University, Osaka, Japan, from 2018. His research fields are human interfaces and interactions, social infrastructure, civil engineering, and intelligent informatics. He is a member of the Architectural Institute of Japan, the Japanese Association of Traffic Psychology, the Japan Association of Applied Psychology, the Institute of Image Information and Television Engineers, the Japan Society of Traffic Engineers, the Institute of Electronics, Information and Communication Engineers, and the Information Processing Society of Japan.

Koji Abe

Koji Abe received the Ph.D. degree in engineering from Kanazawa University, Japan, in 2001. He is associate professor at Department of Informatics, Kindai University, Osaka, Japan, from 2010. His research fields are database science, agricultural environmental and information engineering, medical systems, and perceptual information processing. He is a member of the Institute of Electrical Engineers of Japan, the Institute of Image Electronics Engineers of Japan, the Institute of Electronics, Information and Communication Engineers, and IEEE.

Hitoshi Habe

Hitoshi Habe received the Ph.D. degree in informatics from Kyoto University, Japan, in 2001. He is professor at Department of Informatics, Kindai University, Osaka, Japan, from 2019. His research fields are Computational Intelligence and Medical AI. He is a member of the ACM, IEEE, the Japanese Society for Evolutionary Computation, and so on.

Hisashi Handa

Hisashi Handa received the Ph.D. degree in engineering from Hosei University, Japan, in 1995. He has been a professor at Faculty of Biology-Oriented Science and Technology, Kindai University, Japan, since 2012. His research interests include statistical biosignal processing, intelligent informatics, measurement engineering, biomedical engineering, and biological, health, and medical informatics. He is a member of the Institute of Image Information and Television Engineers, the Society of Instrument and Control Engineers, the Japanese Society for Medical and Biological Engineering, IEEE, the Institute of Electronics, Information and Communication Engineers.

Kazunari Ishii

Kazunari Ishii received the Ph.D. degree in medicine from Kobe University, Japan, in 1995. He is professor at Department of Radiology, Kindai University Faculty of Medicine, Japan, from 2012. His research field is radiology and nuclear medicine. He is a member of the Japan Radiological Society, the Japanese Society of Nuclear Medicine, the Japanese Society of Interventional Radiology, the Japanese Society of Neuroradiology, the Japanese Society for Magnetic Resonance in Medicine, the Japan Radiisotope Association, the Radiological Society of North America, the Society of Nuclear Medicine, the European Society of Radiology and the European Association of Nuclear Medicine.

Yuichi Kimura

Yuichi Kimura was born in Nishinomiya, Japan, in 1961. He received the Ph.D. degree from Waseda University, Tokyo, Japan, in 1991, and Doctor of Medical Sciences from Tokyo Medical and Dental University in 2001. He is Professor of Department of Computational Systems Biology, Faculty of Biology-Oriented Science and Technology, Kindai University, Wakayama, Japan, and is concurrently appointed to Cyber Informatics Research Institute. His major field is data and image analysis of PET imaging, and he is interested in AI approaches to clinical images in nuclear medicine and medical imaging. He was President of the Japanese Society for Medical and Biological Engineering from 2018 to 2020. And he is a member of the Japanese Society of Nuclear Medicine, a senior member of IEEE, and other related societies. He received the awards from JSNM and JAMIT in 2000 and 1996, and he was Chair of IEEE EMBS Japan Chapter from 2002 to 2004.