



OPEN

Segmentation of lung lobes and lesions in chest CT for the classification of COVID-19 severity

Prachaya Khomduean^{1,3,4}, Pongpat Phuadomcharoen^{2,3,4}, Totsaporn Boonchu^{2,3,4}, Unchalisa Taetragool³, Kamonwan Chamchoy⁴, Nat Wimolsiri², Tanadol Jarrusrojwuttikul^{5,6}, Ammarut Chuajak^{5,6}, Udomchai Techavipoo⁶ & Numfon Tweeatsani⁶✉

To precisely determine the severity of COVID-19-related pneumonia, computed tomography (CT) is an imaging modality beneficial for patient monitoring and therapy planning. Thus, we aimed to develop a deep learning-based image segmentation model to automatically assess lung lesions related to COVID-19 infection and calculate the total severity score (TSS). The entire dataset consisted of 124 COVID-19 patients acquired from Chulabhorn Hospital, divided into 28 cases without lung lesions and 96 cases with lung lesions categorized severity by radiologists regarding TSS. The model used a 3D-UNet along with DenseNet and ResNet models that had already been trained to separate the lobes of the lungs and figure out the percentage of lung involvement due to COVID-19 infection. It also used the Dice similarity coefficient (DSC) to measure TSS. Our final model, consisting of 3D-UNet integrated with DenseNet169, achieved segmentation of lung lobes and lesions with the Dice similarity coefficients of 91.52% and 76.89%, respectively. The calculated TSS values were similar to those evaluated by radiologists, with an R2 of 0.842. The correlation between the ground-truth TSS and model prediction was greater than that of the radiologist, which was 0.890 and 0.709, respectively.

The rapid pandemic-level outbreak of coronavirus disease 2019 (COVID-19) has caused a wide range and degree of illnesses, predominated by respiratory tract infection^{1–4}. Although most infected patients show asymptomatic or mild clinical manifestations, further investigation beyond real-time reverse transcriptase polymerase chain reaction (RT-PCR) or rapid COVID-19 tests such as chest radiographs is routinely indicated in worsening cases that require hospitalization^{5,6}. Characteristic findings in chest radiographs of COVID-19 related pneumonia are bilateral patchy and/or confluent and bandlike ground-glass opacity or consolidation in a peripheral and mid-to-lower lung zone distribution. By contrast, several studies have found almost one-half of normal chest radiographs at initial presentation disagree with clinical symptoms^{7–10}.

Because of its higher sensitivity, specificity, and speed, chest computed tomography (CT) has become more useful than RT-PCR in early detection, to obtain more information about chest pathology, and to evaluate the severity of lung involvement. Moreover, it can assist triage, especially when hospitalization is required but there is a shortage of healthcare personnel, inpatient beds, and medical equipment, and it may be useful as a standard modality for the rapid diagnosis of COVID-19-related pneumonia^{11–15}. The chest CT findings are peripheral, bilateral, ground-glass opacity (GGO) with some round shapes with or without consolidation or intralobular lines, a reverse halo sign, or other findings of organizing pneumonia^{16–19}.

The total severity score (TSS) has been proposed by Chung et al.²⁰. It is calculated from the summation of lesion scores in five lung lobes and is used to categorize the severity of lung involvement and help determine the

¹Centre of Learning and Research in Celebration of HRH Princess Chulabhorn's 60th Birthday Anniversary, Chulabhorn Royal Academy, Bangkok, Thailand. ²Chulabhorn Hospital, Chulabhorn Royal Academy, Bangkok, Thailand. ³Department of Computer Engineering, Faculty of Engineering, King Mongkut's University of Technology Thonburi, Bangkok, Thailand. ⁴Princess Srisavangavadhana College of Medicine, Chulabhorn Royal Academy, Bangkok, Thailand. ⁵Queen Savang Vadhana Memorial Hospital, Chonburi, Thailand. ⁶Faculty of Health Science Technology, HRH Princess Chulabhorn College of Medical Science, Chulabhorn Royal Academy, Bangkok, Thailand. ✉email: numfon.twe@kra.ac.th

proper therapeutic management and prognosis^{21,22}. TSS reflects the clinical classification of COVID-19²². It has also been shown to provide high specificity in the detection of severe cases and high inter-observer reliability with a short interpretation time compared to other severity scoring system²³. It has been used in many studies, such as the comparison of patients with and without vaccination²⁴, and the viral load factor for hospitalization and mortality of patients²⁵.

To reduce the amount of time required for interpretation and increase the accuracy of lesion detection, deep learning has been used to efficiently analyze medical images by performing tasks such as semantic segmentation. Deep learning was also used in the automated assessment of CT severity scores in COVID-19 patients. Lessmann et al.²⁶ applied deep-learning algorithms that automatically segment the five pulmonary lobes and abnormalities and then predict the severity scores for patients suspected of having COVID-19. The results showed good agreement with the results from independent observers. Chaganti et al.²⁷ automatically computed the percentage of opacity and lung severity score by applying deep reinforcement learning for lung lobe segmentations and using the U-Net model for a semantic segmentation of GGO and consolidations. The results correlated well with the ground truth.

The U-Net model is a convolutional neural network-based model that was originally used for the semantic segmentation of biomedical images and is now one of the most utilized image segmentation techniques. The model structure is U-shaped and consists of two parts: a contracting path (encoder) and an expanding path (decoder)²⁸. Subsequently, a U-Net model was created to support three-dimensional (3D) matrices and is called 3D-Unet²⁹. The 3D-Unet model was used to develop a more efficient 3D imaging model for the segmentation of lesions and lung tissue^{30,31}. Cropping the lung area before lesion segmentation can improve accuracy³². Enshaei et al.³³ developed a model for predicting the lesion area of COVID-19 patients from CT-scan images, using a model to predict the lung area before the lesion regions were considered. This method enables the lesion model to predict lesions more accurately. In another study, a deep learning model was applied to lung lobe segmentation. The model is capable of accurately segmenting each lung lobe from lung CT scans³⁴. It is also utilized in lung lobe segmentation analysis for lung segmentation research to improve segmentation accuracy in multiple diseases such as chronic obstructive pulmonary disease (COPD), lung cancer, and COVID-19-related pneumonia³⁵.

Many studies have used deep learning models for computer-aided diagnostics to determine the intensity of infections. For instance, Aswathy A. L. and Vinod Chandra S. S.³⁶ employed 3D-UNet models to effectively segment the lung parenchyma and infected regions in lung CT scans. Additionally, a previous study demonstrated that the effectiveness of these models for medical image segmentation can improve sensitivity performance³⁷. In another study, the U-Net model combined with the dense convolutional network (DenseNet) was effectively employed to develop a program for classifying the severity of lung CT in COVID-19 by analyzing the lesion area and comparing it with the lung area in lung CT scan images³⁸. They calculated the percentage of infection (PI) using a U-Net model combined with pre-trained models such as residual neural networks (ResNet) and DenseNet. ResNet was first presented by He et al.³⁹ to solve the vanishing gradient problem of deeper networks by adding feedforward links across some layers, resulting in residual optimization of those layers. DenseNet was first presented by Huang et al.⁴⁰ to learn more features by using deeper convolutional layers with many feedforwards linking across layers. For this reason, this knowledge can be applied to lung lobe segmentation and lesion segmentation in CT scan images.

In this study, deep learning semantic segmentation was used for the lung severity scoring of the COVID-19 infection. The proposed method utilized a combination of 3D-UNet models integrated with pre-trained models, DenseNet and ResNet, to compute the PI from the lung lobe and lesion segmentation results and estimate the TSS automatically. The aim was to alleviate the radiologist's workload and time spent on imaging diagnostics, as well as improve reporting accuracy.

Materials and methods

Datasets

Due to its retrospective nature, informed consent was waived, and all data were anonymized. This project was approved by the human research ethics committee of the Chulabhorn Research Institute (research project code 167/2564) and complied with the Declaration of Helsinki. These COVID-19 patients were confirmed by RT-PCR acquired from Chulabhorn Hospital who underwent non-contrast enhanced axial chest CT as a part of routine clinical care throughout the pandemic.

In this study, we randomly selected 124 cases from the database. The selection contained 28 cases without lung lesions and 96 cases with lung lesions. According to TSS, experienced radiologists classified the cases with lung lesions as mild, moderate, and severe. We divided the selection into 3 groups, i.e., training set, test set 1, and test set 2. The training set was used in model training and validation for lung segmentation and lesion segmentation; test set 1 was for segmentation performance evaluation; and test set 2 was for TSS prediction evaluation. We also randomly selected these cases for each group. In addition, for the training set and test set 1, the numbers of cases across different severity types were set to be equal to prevent class imbalance in the training set (the class imbalance causing a potential bias in the trained model) and for a fair comparison in test set 1. The number of CT slices in these cases ranged from 92 to 208. This information was described in Table 1.

Data preprocessing

The lung CT data were saved in JPEG format with a resolution of 512 × 512 pixels and labeled by a program called LabelME⁴¹ (version 4.5.12). The resulting labels were in JavaScript Object Notation (JSON) format. All labeled data were validated by four radiologists and then converted into matrices for model training and evaluation.

In the data preprocessing phase, CT scan images (JPEG format) and labeled data (JSON format) were resized using the “cv2.resize” function from 512 × 512 pixels to 256 × 256 pixels to minimize the required memory

Case type	Training set			Test set 1			Test set 2		
	No. of cases	Total slices	Avg ± std per case	No. of cases	Total slices	Avg ± std per case	No. of cases	Total slices	Avg ± std per case
No lesions	8	985	123 ± 22	5	705	141 ± 39	15	1953	130 ± 27
Mild	8	907	113 ± 12	5	707	141 ± 48	16	2119	132 ± 27
Moderate	8	941	118 ± 13	5	575	115 ± 20	27	3179	118 ± 27
Severe	8	919	115 ± 23	5	597	119 ± 36	14	1469	105 ± 12
Total	32	3752	117 ± 18	20	2584	129 ± 36	72	8720	121 ± 26

Table 1. Summary of axial lung CT scan datasets.

resources (RAM). The interpolation parameter was set to “INTER_AREA” for the CT scan images and “INTER_NEAREST” for the label data because this solution prevented any alteration of the values specified in each pixel. In addition, our model input shape was fixed at a size of $128 \times 256 \times 256$. The CT volumes were adjusted to 128 images per patient according to the following three conditions (Fig. 1).

The first condition, if the CT volume comprised 128 or fewer images, a 256×256 zero-padding matrix was added to increase the volume to 128 images.

The second condition, if the CT volume had between 129 and 175 CT images, 128 images from the CT volume’s middle range were selected to train the model because both the lung parenchyma and lesions appear in this range.

The third condition, if the CT volume contained more than 175 CT images, we skipped the CT slice by selecting only odd-numbered images and adding a 256×256 zero-padding matrix to reach a total of 128 images.

A color adjustment method was applied to improve image contrast by using the contrast-limited adaptive histogram equalization (CLAHE) technique⁴², which is available in the OpenCV library⁴³. The CLAHE parameters were set to a clipLimit of 3 and a tileGridSize of (8, 8). Models were trained/tested in two experiments: the first with original images (no color adjustment) and the second with CLAHE-adjusted images.

Imaging protocol

A 256-slice dual-energy CT scanner (Revolution CT with Gemstone Spectral Imaging (GSI) Xtream, GE Healthcare) at Chulabhorn Hospital was used in this study. An axial chest CT scan without contrast agent was applied. The protocol started with a scout view from lung apices to lung bases in anterior–posterior (AP) and lateral views, and followed by an axial chest scan covering lung apices through bases from inferior to superior. The parameters were quiet breath inspiration, 1.25 mm thickness, 0.28 s/rotation, 0.992 pitch, GSI calculated kVp, 190 mA, lung window of (1550, – 700), soft tissue window of (400, 40), and postprocessing multiplanar reconstruction. The scan time was less than 1.6 s.

Model training

Two models were used in this study: (1) a lung lobe segmentation model and (2) a lesion segmentation model. Training set: 32 cases were split into 24 cases (75%) for model training and 8 cases (25%) for validation, where the dataset was divided equally at each severity type to prevent overfitting. According to related studies, a model that

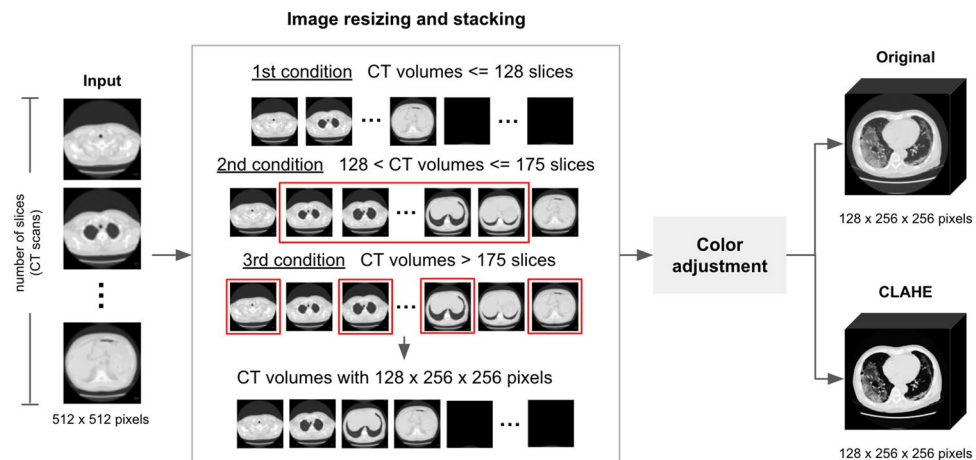


Figure 1. Overall pipeline of the data preprocessing.

combines a 3D-UNet structure with a DenseNet or ResNet is effective in segmenting parts of the image precisely. Therefore, pre-trained DenseNet121, DenseNet169, DenseNet201, ResNet18, ResNet34, ResNet50, ResNet101, and ResNet152 models were obtained through a segmentation-models-3D package from Solovyev et al.⁴⁴.

Lung lobe segmentation. A multiclass semantic segmentation model was used to segment the five lung lobes. Annotated labels consisted of six categories: 0, 1, 2, 3, 4, and 5, which indicate the background, right upper lobe (RUL), right lower lobe (RLL), right middle lobe (RML), left upper lobe (LUL), and left lower lobe (LLL), respectively.

Lesion segmentation. The lesion model was developed from a binary semantic segmentation model that outputs the value 1 for lesion areas and 0 for background areas. Images without extrapulmonary regions are preferred for lesion model training. The dataset used for model training was preprocessed as described in the data preparation section.

In the model training process, lung lobe and lesions segmentation models were trained on servers equipped with an Intel(R) Xeon(R) Gold 6126 CPU at 2.60 GHz, 40 GB of RAM, and an NVIDIA Tesla V100 SXM2 GPU. Figure 2 shows the overall workflow. The model's output is the predicted class for each pixel, which is then used to compute the percentage area of lesions in each lung lobe for the CT score. This score is then used to calculate the TSS value for diagnosing the severity of the current pathology. For both models, Adam optimization was used, the loss function was a hybrid loss function (focal loss + Dice loss), the learning rate was set to 0.0001, a regularizer that applies L2 regularization was used with a value of 0.01, the batch size was set to 1, and the maximum number of epochs was 200. The lesion model activation function was set to sigmoid with a dropout rate of 0.4, whereas the pulmonary lobe model activation function was set to SoftMax with a dropout rate of 0.2. The hybrid loss technique⁴⁵, which combined focal loss and Dice loss, was used to improve model performance.

PI

The PI in each lung lobe was calculated by dividing the number of predicted lesion pixels by the total number of lung lobe pixels in the CT volume. The predicted lesion pixels were obtained from the output of the lesion segmentation model, whereas the predicted lung lobe pixels were derived from the output of the lung lobe segmentation model, in which the value of each pixel identifies the lobar type in the lung CT image. Therefore, the PI was calculated by performing the following equation.

$$\text{Percentage of Infection}_{\text{lobe}} = \frac{\text{Lesion Area (pixels)}}{\text{Lung Lobe Area (pixels)}} \times 100$$

TSS

The TSS proposed by Chung et al.²⁰ was calculated from the sum of the five-lobe CT score, which was calculated from PI based on the criteria listed in Table 2. The severity of COVID-19 patients can be classified from the TSS value based on the severity criteria in Table 3 and the following equation:

$$\text{TSS} = \text{CT Score}_{\text{RUL}} + \text{CT Score}_{\text{RML}} + \text{CT Score}_{\text{RLL}} + \text{CT Score}_{\text{LUL}} + \text{CT Score}_{\text{LLL}}$$

Dice similarity coefficient

The most commonly utilized measurement to evaluate image segmentation is the Dice similarity coefficient (DSC). The DSC calculated the relative overlap between the predicted area and ground truth, and it was used to choose the most appropriate model. The DSC was defined as follows:

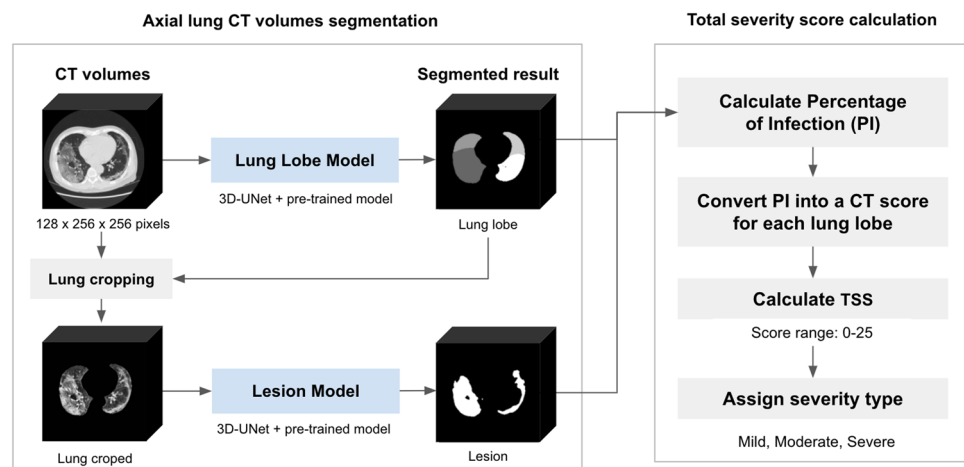


Figure 2. Segmentation Model Workflow and Total Severity Score Calculation Protocol for Lung CT Scans.

PI	Score
No lesion	0
< 5%	1
6–25%	2
26–50%	3
51–75%	4
> 75%	5

Table 2. CT-score criteria based on the percentage of infection (PI) for the diagnosis of pneumonia in COVID-19 patients proposed by Chung et al.²⁰.

TSS	Severity type
≤ 7	Mild
8–17	Moderate
≥ 18	Severe

Table 3. Severity type classification based on the total severity score (TSS).

$$DSC = \frac{2TP}{2TP + FP + FN}$$

where the term “true positive” (TP) refers to an outcome such that the model correctly predicts the positive class, “false positive” (FP) is an outcome such that the model incorrectly predicts the positive class, and “false negative” (FN) is an outcome such that the model incorrectly predicts the negative class.

Hausdorff distance

Hausdorff distance was proposed by Felix Hausdorff in 1914^{46,47}. The measure was applied to evaluate the model’s performance by measuring the distance between two images in pixels. The distance is defined as

$$H(A, B) = \max(h(A, B), h(B, A))$$

$$h(A, B) = \max_{a \in A} \min_{b \in B} |a - b|,$$

where A is a set containing p points (pixels on image A): $\{a_1, a_2, \dots, a_p\}$ and B is a set containing q points (pixels on image B): $\{b_1, b_2, \dots, b_q\}$. For implementation, we applied the function implemented in SciPy package⁴⁸. Since the images used in this research were 256×256 pixels, the Hausdorff distance range was $[0, 256\sqrt{2}]$.

Results

Lung lobe segmentation

The testing of lung lobe segmentation models was divided into two sections, one for the models trained on the original images and one for the models trained on the lung CT images processed by the CLAHE technique. The model’s test results for test set 1 were given in Supplementary Tables S1 and S2. The 3D-UNet + DenseNet169 model trained with the original CT data was found to obtain the best lung lobe segmentation with a DSC of 91.52% and an average Hausdorff distance of 12.9 pixels.

For lung lobe segmentation, the results indicated that the model trained on the original images outperformed the model trained on the CLAHE-adjusted images, as indicated in Table 4. Figure 3 presented the image segmentation results for both the original and CLAHE-adjusted images. In the case of the middle lobe (blue) of the no-lesion group, the model trained on the original images performed better than the model trained on CLAHE-adjusted images, which erroneously labeled as the right upper lung (red) and the background area.

Segmentation results for each severity level indicated that the mild group underperformed the other groups for the model trained on the original images. The DSC values for each lobe segment were presented in Table 4 and Supplementary Tables S3, revealing that the left lobe (upper and lower) segmentation was more accurate than the right lobe segmentation. However, the right and left lower lobes (RLL and LLL) exhibited lower boundary distances, with HD values of 9.41 and 9.36 pixels, respectively.

Lesion segmentation

The results of the lesion model on test set 1 were divided into two sections, one for the original images and one for the images processed using the CLAHE technique (Supplementary Tables S4 and S5). The models trained with the CT images processed by CLAHE were found to be the most effective for lesion segmentation, with the 3D-UNet + DenseNet169 model obtaining a DSC of 76.89% and an average Hausdorff distance of 37.82 pixels.

	Non-CLAHE		CLAHE	
	DSC (%)	HD (pixels)	DSC (%)	HD (pixels)
Lung lobe segmentation				
(Divided into lobes)				
RUL	89.52 ± 4.58	13.38 ± 7.23	89.95 ± 4.74	13.55 ± 7.40
RML	90.56 ± 4.22	16.46 ± 8.90	88.91 ± 5.14	17.45 ± 9.97
RLL	91.29 ± 4.20	9.41 ± 5.37	91.21 ± 4.64	10.91 ± 5.71
LUL	92.72 ± 3.16	15.91 ± 5.29	92.26 ± 3.40	17.51 ± 7.29
LLL	93.53 ± 2.33	9.36 ± 6.80	92.32 ± 4.67	9.28 ± 4.66
(Divided into severity types)				
No lesion	91.86 ± 0.52	12.94 ± 2.35	91.42 ± 2.47	14.78 ± 6.54
Mild	90.36 ± 1.35	15.11 ± 3.45	90.03 ± 2.92	15.31 ± 5.10
Moderate	92.70 ± 2.39	12.09 ± 5.23	92.47 ± 3.14	11.91 ± 4.98
Severe	91.16 ± 4.09	11.48 ± 5.96	89.80 ± 5.60	12.95 ± 6.33
Overall	91.52 ± 2.44	12.90 ± 4.35	90.93 ± 3.61	13.74 ± 5.49
Lesion segmentation				
No lesion*	93.02 ± 8.24	6.53 ± 14.59	90.3 ± 14.49	11.75 ± 26.28
Mild	75.92 ± 13.17	47.80 ± 34.19	77.78 ± 6.92	40.19 ± 29.33
Moderate	65.18 ± 10.32	57.88 ± 11.66	67.85 ± 8.00	55.08 ± 14.10
Severe	67.77 ± 12.58	42.98 ± 5.99	71.62 ± 16.87	44.25 ± 24.69
Overall	75.47 ± 15.2	38.80 ± 26.89	76.89 ± 14.28	37.82 ± 27.68

Table 4. Lung lobe and Lesion segmentation results of the 3D-UNet + DenseNet169 model with test set 1 (n = 20) showing Dice similarity coefficient (DSC), Hausdorff distance (HD), and their ± standard deviations. The overall performance values are in bold. *Predicting lung area without lesions instead of predicting lesion area.

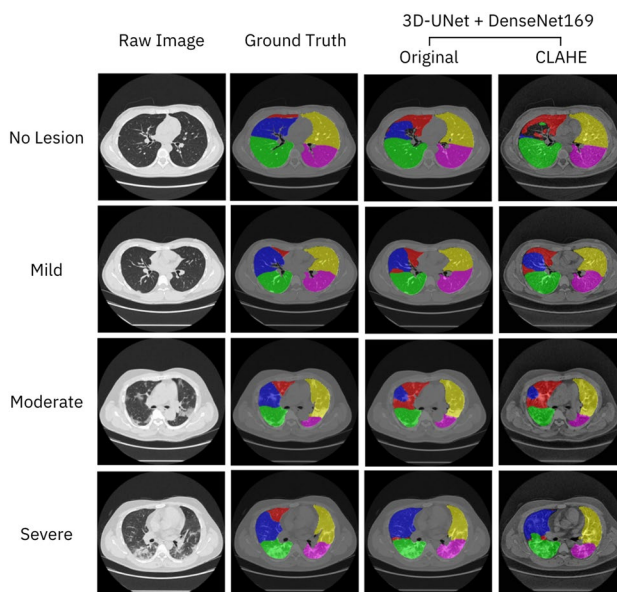


Figure 3. Examples lung lobe segmentation results for each severity level obtained by the 3D-UNet + the DenseNet169 model trained on the original images and images processed by CLAHE. Each lobe of the lung, right upper lobe (RUL), right lower lobe (RLL), right middle lobe (RML), left upper lobe (LUL), and left lower lobe (LLL), is indicated by the colors red, blue, green, yellow, and pink, respectively.

According to the test results, the model trained on the CLAHE-adjusted images performed better than the model trained on the original images. The best model (3D-UNet + DenseNet169) was evaluated on test set 1, and the results were shown in Table 4 and Fig. 4. These results implied that lesion models often accurately predicted lesions when images were completely free of lesions (no-lesion images) and obvious lesions like consolidation, while faint lesions such as GGO were often less accurate in prediction.

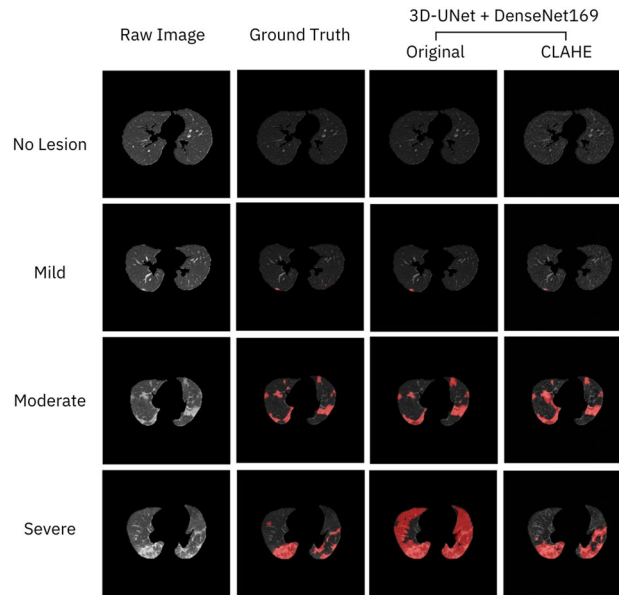


Figure 4. Example of segmentation results obtained by models trained on the original images and images processed by the contrast-limited adaptive histogram equalization (CLAHE) at each level of severity. The red pixels indicate the lesion areas.

TSS regression analysis

The best-performing lung and lesion models with the highest DSC were applied to perform image segmentation for 72 cases (test set 2). The segmentation results were used to calculate the PI in each lung lobe, which was then used to compute the TSS. To analyze the trends and correlation of the TSS results, we compared the radiologist measurements with TSS values calculated based on our approach using Pearson correlation coefficients (r) and R^2 values. Figure 5 presented the results of the analysis of the TSS values obtained by models and those obtained by radiologists. According to the appropriate statistical analysis, the Pearson correlation coefficient of 0.918 indicated that the model-based predictions and the radiologist's diagnosis were positively correlated with an R^2 of 0.842.

To compare the outcomes of TSS that our approach predicted and those that radiologists diagnosed, we used the Bland–Altman plot. The x-axis represented the range of the TSS score and the y-axis represents the difference in the score of TSS between the two methods. It was found that the 95% limits of agreement, or the range of values within -3.318 to 7.652. The mean difference between the two methods was 2.167, which means that the radiologists diagnosed more than TSS predicted on average by 2.167, as shown in Fig. 5A.

Additionally, three techniques for measuring the TSS value were compared: using the label mask of test set 1 as the ground truth, the radiologist's diagnostic TSS (Radiologist), and the model-based TSS (Prediction). TSS (Radiologist) indicated the radiologist determined the TSS value. TSS (GT) indicated the TSS value obtained by using the ground truth for the calculation. TSS (prediction) was the TSS result obtained using the model predictions for the computation. The outcomes of the investigation were presented in Fig. 6. The TSS (Prediction) values were correlated with TSS (GT) values, resulting in an R^2 of 0.890. Furthermore, the R^2 of 0.709 from the

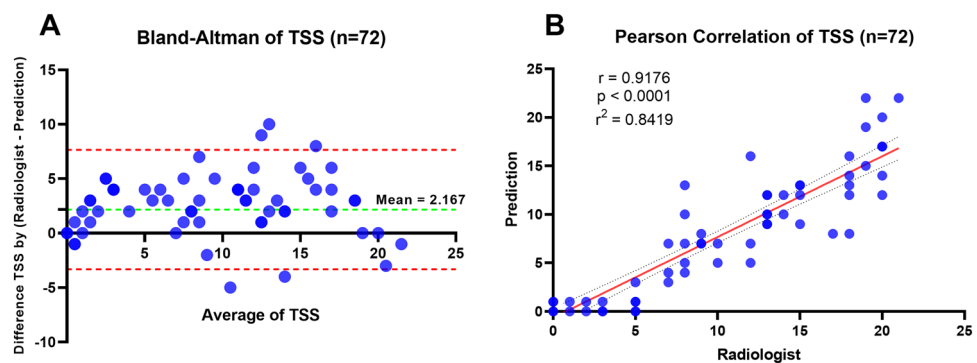


Figure 5. Distribution of TSS values obtained by our model prediction and a radiologist for test set 2 ($n = 72$). (A) Bland–Altman plot illustrating the comparison of TSS values between the radiologist and prediction. (B) Regression plot depicting the correlation between TSS calculated from the radiologist and prediction.

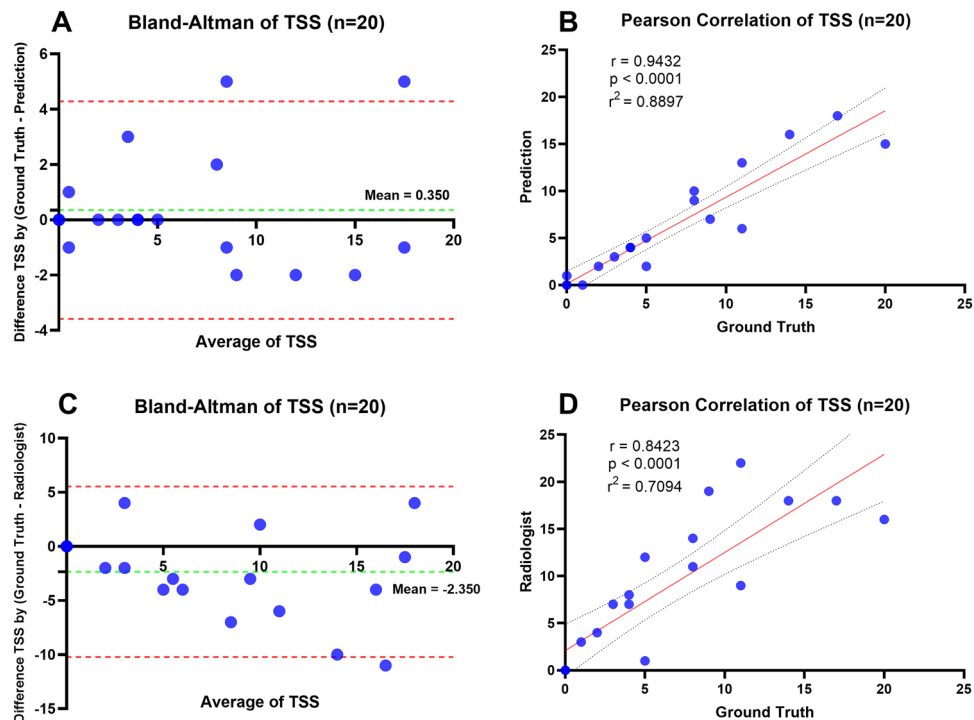


Figure 6. Distribution of TSS values for method comparison with ground truth using Test Set 1. **(A)** Bland–Altman plot illustrating the comparison of TSS values between the ground truth and prediction. **(B)** Regression plot depicting the correlation between TSS calculated from the ground truth and TSS predicted. **(C)** Bland–Altman plot demonstrating the comparison of TSS values between the ground truth and radiologist. **(D)** Regression plot analysis showcasing the correlation between TSS values calculated from the ground truth and those provided by the radiologist.

comparison between TSS (Ground Truth) and TSS (Radiologist) indicated a higher correlation for the model compared to a human.

Ethical approval

Ethical approval was obtained from the human research ethics committee of the Chulabhorn Research Institute (research project code 167/2564).

Informed consent

Informed consent was waived by the human research ethics committee of the Chulabhorn Research Institute due to retrospective study and all data were anonymized.

Discussion

We compared the proposed lung lobe segmentation model's performance with the performance of models in other related studies. Tang et al.⁴⁵ reported a DSC of 91.48% on the LUNA16 dataset and 94.17% on the Tianchi dataset, indicating that our model underperformed theirs. However, Tang et al. employed CT datasets of lungs without lesions for their test datasets, and the models were trained on 40 cases, whereas our model utilized the no-lesion training dataset, which consisted of eight cases. The performance of our model (3D-UNet + DenseNet169) was equivalent to that of Tang et al., with a DSC of 91.86% in the no-lesions group (Table 4). This demonstrates that our approach can be trained using a small dataset. Furthermore, a comparison with a lung segmentation algorithm based on the lung fissure surface yielded a DSC of 84.00%⁴⁹. This indicated that a deep learning approach could efficiently enhance the performance of lung lobe segmentation, particularly due to the continuous training of 3D-UNet on a complete set of data by inputting entire CT slices per case. As a result, the 3D-UNet model was suitable for precise lung lobe segmentation in continuous CT images.

For the lesion segmentation studies, Xiao et al.³⁰ reported a 3D-UNet model with a DSC of 89.12%, and Qiblawey et al.³⁸ reported a DSC of 94.13% for the Feature pyramid network (FPN) combined with the DenseNet201 model. Both models outperformed ours because training a model required a considerable variety of data. For example, Qiblawey et al.³⁸ used 15,698 images for training, whereas our model training used a limited number of 3752 images. Despite the smaller dataset, our lesion model's results demonstrated that it was capable of accurately predicting lesions in the no-lesion and obvious lesions, despite the inaccuracies in lesion segmentation for faint lesions. Many approaches that should improve the performance of lesion segmentation include increasing the amount of training data, applying data augmentation to increase the variety of lesions, enhancing the contrast rate, and utilizing the recent developments in state-of-the-art segmentation models,

e.g. PaddleSeg⁵⁰, to achieve improvements in accurately segmenting lesions from medical imaging data. These approaches would be potential solutions included in our future work.

Image enhancement using CLAHE could not be suitable for the lung lobe segmentation model. The results showed that the lung model trained on the original image accurately segmented better than the model trained on CLAHE images. It might be possible that CLAHE locally emphasizes many details on the image, and thus making the lung fissures more difficult to detect. In contrast, for the lesion segmentation model, using CLAHE provided a small improvement. Other enhancement methods such as the Balance Contrast Enhancement Technique (BCET)⁵¹ could be a potential solution for enhancing contrast quality and preserving the histogram pattern of the image. As shown in a previous study focusing on COVID-19 classification using lung X-ray images, BCET outperformed CLAHE⁵².

Additionally, the TSS regression analysis showed that our model can be utilized to segment lung CT scans effectively. An analysis of the radiologist's diagnostic TSS and model-based TSS results revealed that our method produced observation results that were more accurate with respect to the ground truth than human observation. This demonstrates that the measurement software could be used to eliminate human error in the estimation of infection rates in the lung area. Furthermore, our approach could perform lesion area segmentation and calculate the infection rate automatically. This could dramatically reduce the radiologist's workload and enhance the efficiency of diagnosing COVID-19 severity levels.

TSS is a rapid and objective assessment method for radiologists that provides information in each lobe and is more feasible for manual evaluation. However, it grades each lung lobe equally significantly, without concerning about their volume differences. In contrast, PI calculates the entire affected volume, resulting in a more accurate evaluation of infected lung volume. Nevertheless, some remote hospitals lack an automated program, and TSS appears to be an effective evaluation instrument.

Conclusion

Constructing a model for the automatic segmentation and scoring of COVID-19 infection in chest CT was accomplished through the application of deep learning techniques. According to the findings, the combination of 3D-UNet and DenseNet169 achieved the highest level of performance when it came to the segmentation of lung lobes and lesions. The projected severity score had a strong correlation with the visual assessments made by radiologists. This accurate model provided a dependable method for quantifying the extent of lung involvement. The proposed model was helpful in determining the extent of the lower respiratory tract infection and monitoring the disease in COVID-19 patients.

Data availability

The datasets used and/or analyzed during the current study available from the corresponding author on reasonable request.

Code availability

The source code used for this research work is made publicly available in the GitHub repository <https://github.com/hds-69/csc-app>.

Received: 11 January 2023; Accepted: 17 November 2023

Published online: 28 November 2023

References

- Guan, W.-J. *et al.* Clinical characteristics of coronavirus disease 2019 in China. *N. Engl. J. Med.* **382**, 1708–1720 (2020).
- Zhu, N. *et al.* A novel coronavirus from patients with pneumonia in China, 2019. *N. Engl. J. Med.* **382**, 727–733. <https://doi.org/10.1056/NEJMoa2001017> (2020).
- Huang, C. *et al.* Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* **395**, 497–506. [https://doi.org/10.1016/S0140-6736\(20\)30183-5](https://doi.org/10.1016/S0140-6736(20)30183-5) (2020).
- Berlin, D. A., Gulick, R. M. & Martinez, F. J. Severe Covid-19. *N. Engl. J. Med.* **383**, 2451–2460. <https://doi.org/10.1056/NEJMp2009575> (2020).
- ACR recommendations for the use of chest radiography and computed tomography (CT) for suspected COVID-19 infection. *Acr.org* <https://www.acr.org/Advocacy-and-Economics/ACR-Position-Statements/Recommendations-for-Chest-Radiography-and-CT-for-Suspected-COVID19-Infection> (2020).
- Rubin, G. D. *et al.* The role of chest imaging in patient management during the COVID-19 pandemic: A multinational consensus statement from the Fleischner society. *Chest* **158**, 106–116. <https://doi.org/10.1016/j.chest.2020.04.003> (2020).
- Wong, H. Y. F. *et al.* Frequency and distribution of chest radiographic findings in patients positive for COVID-19. *Radiology* **296**, E72–E78. <https://doi.org/10.1148/radiol.2020201160> (2020).
- Smith, D. L., Grenier, J.-P., Batte, C. & Spieler, B. A characteristic chest radiographic pattern in the setting of the COVID-19 pandemic. *Radiol. Cardiothorac. Imaging* **2**, e200280. <https://doi.org/10.1148/ryct.2020200280> (2020).
- Litmanovich, D. E., Chung, M., Kirkbride, R. R., Kicska, G. & Kanne, J. P. Review of chest radiograph findings of COVID-19 pneumonia and suggested reporting language. *J. Thorac. Imaging* **35**, 354–360. <https://doi.org/10.1097/rti.0000000000000541> (2020).
- Weinstock, M. B. *et al.* Chest x-ray findings in 636 ambulatory patients with COVID-19 presenting to an urgent care center: A normal chest x-ray is no guarantee. *J. Urgent Care Med.* **14**, 13–18 (2020).
- Li, Y. & Xia, L. Coronavirus disease 2019 (COVID-19): Role of chest CT in diagnosis and management. *Am. J. Roentgenol.* **214**, 1280–1286. <https://doi.org/10.2214/AJR.20.22954> (2020).
- Zhou, H. *et al.* Coronavirus disease 2019 (COVID-19): Chest CT characteristics benefit to early disease recognition and patient classification—A single center experience. *Ann. Transl. Med.* **8**, 679 (2020).
- Borakati, A., Perera, A., Johnson, J. & Sood, T. Diagnostic accuracy of X-ray versus CT in COVID-19: A propensity-matched database study. *BMJ Open* **10**, e042946. <https://doi.org/10.1136/bmjopen-2020-042946> (2020).

14. Sverzellati, N. *et al.* Chest x-ray or CT for COVID-19 pneumonia? Comparative study in a simulated triage setting. *Eur. Respir. J.* **58**, 2004188 (2021).
15. Kim, H., Hong, H. & Yoon, S. H. Diagnostic performance of CT and reverse transcriptase-polymerase chain reaction for coronavirus disease 2019: A meta-analysis. *Radiology* <https://doi.org/10.1148/radiol.2020201343> (2020).
16. Zhou, S., Wang, Y., Zhu, T. & Xia, L. CT features of coronavirus disease 2019 (COVID-19) pneumonia in 62 patients in Wuhan, China. *Ajr Am. J. Roentgenol.* **214**, 1287–1294 (2020).
17. Bao, C., Liu, X., Zhang, H., Li, Y. & Liu, J. Coronavirus disease 2019 (COVID-19) CT findings: A systematic review and meta-analysis. *J. Am. Coll. Radiol.* **17**, 701–709 (2020).
18. Niu, R. *et al.* Chest CT features associated with the clinical characteristics of patients with COVID-19 pneumonia. *Ann. Med.* **53**, 169–180. <https://doi.org/10.1080/07853890.2020.1851044> (2021).
19. Simpson, S. *et al.* Radiological society of North America Expert Consensus statement on reporting chest CT findings related to COVID-19. Endorsed by the society of thoracic radiology, the American College of Radiology, and RSNA—secondary publication. *J. Thorac. Imaging* **35**, 219–227. <https://doi.org/10.1097/rti.0000000000000524> (2020).
20. Chung, M. *et al.* CT imaging features of 2019 novel coronavirus (2019-nCoV). *Radiology* **295**, 202–207 (2020).
21. Yang, R. *et al.* Chest CT severity score: An imaging tool for assessing severe COVID-19. *Radiol. Cardiothorac. Imaging* **2**, e200047. <https://doi.org/10.1148/ryct.2020200047> (2020).
22. Li, K. *et al.* CT image visual quantitative evaluation and clinical classification of coronavirus disease (COVID-19). *Eur. Radiol.* **30**, 4407–4416. <https://doi.org/10.1007/s00330-020-06817-6> (2020).
23. Elmokadem, A. H., Mounir, A. M., Ramadan, Z. A., Elsedeiq, M. & Saleh, G. A. Comparison of chest CT severity scoring systems for COVID-19. *Eur. Radiol.* <https://doi.org/10.1007/s00330-021-08432-5> (2022).
24. Russo, G. *et al.* Total Severity Score (TSS) comparison in vaccinated and unvaccinated patients during the fourth wave (December 2021–January 2022) of COVID-19 in Italy. *Eur. Rev. Med. Pharmacol. Sci.* **26** (2022).
25. Yagci, A. K. *et al.* Relationship of the cycle threshold values of SARS-CoV-2 polymerase chain reaction and total severity score of computerized tomography in patients with COVID 19. *Int. J. Infect. Dis.* **101**, 160–166 (2020).
26. Lessmann, N. *et al.* Automated assessment of COVID-19 reporting and data system and chest CT severity scores in patients suspected of having COVID-19 using artificial intelligence. *Radiology* **298**, E18–E28 (2021).
27. Chaganti, S. *et al.* Automated quantification of CT patterns associated with COVID-19 from chest CT. *Radiol. Artif. Intell.* **2**, e200048 (2020).
28. Ronneberger, O., Fischer, P. & Brox, T. 234–241 (Springer International Publishing).
29. Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T. & Ronneberger, O. 424–432 (Springer International Publishing).
30. Xiao, Z., Liu, B., Geng, L., Zhang, F. & Liu, Y. Segmentation of lung nodules using improved 3D-UNet neural network. *Symmetry* **12**, 1787 (2020).
31. Zhou, L. *et al.* An interpretable deep learning workflow for discovering subvisual abnormalities in CT scans of COVID-19 inpatients and survivors. *Nat. Mach. Intell.* **4**, 494–503 (2022).
32. Mansoor, A. *et al.* A generic approach to pathological lung segmentation. *IEEE Trans. Med. Imaging* **33**, 2293–2310 (2014).
33. Enshaei, N. *et al.* COVID-rate: An automated framework for segmentation of COVID-19 lesions from chest CT images. *Scientific Reports* **12**, 1–18 (2022).
34. Gerard, S. E. & Reinhardt, J. M. Pulmonary lobe segmentation using a sequence of convolutional neural networks for marginal learning. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)* (eds Gerard, S. E. & Reinhardt, J. M.) 1207–1211 (IEEE, 2019).
35. Gerard, S. E. *et al.* CT image segmentation for inflamed and fibrotic lungs using a multi-resolution convolutional neural network. *Sci. Rep.* **11**, 1–12 (2021).
36. Aswathy, A. & SS, V. C. Cascaded 3D UNet architecture for segmenting the COVID-19 infection from lung CT volume. *Sci. Rep.* **12**, 3090 (2022).
37. Noguchi, S. *et al.* Deep learning-based algorithm improved radiologists' performance in bone metastases detection on CT. *Eur. Radiol.* **32**, 7976–7987. <https://doi.org/10.1007/s00330-022-08741-3> (2022).
38. Qiblawey, Y. *et al.* Detection and severity classification of COVID-19 in CT images using deep learning. *Diagnostics* **11**, 893 (2021).
39. He, K., Zhang, X., Ren, S. & Sun, J. in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 770–778.
40. Huang, G., Liu, Z. & Weinberger, K. Q. Densely Connected Convolutional Networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2261–2269 (2017).
41. Russell, B. C., Torralba, A., Murphy, K. P. & Freeman, W. T. LabelMe: A database and web-based tool for image annotation. *Int. J. Comput. Vis.* **77**, 157–173 (2008).
42. Zuiderveld, K. Contrast limited adaptive histogram equalization. In *Graphics gems* (ed. Zuiderveld, K.) 474–485 (Elsevier, 1994).
43. Histograms—2: Histogram Equalization. *Opencv.org* https://docs.opencv.org/4.x/d5/daf/tutorial_py_histogram_equalization.html (2023).
44. Solovyev, R., Kalinin, A. A. & Gabruseva, T. 3D convolutional neural networks for stalled brain capillary detection. *Comput. Biol. Med.* **141**, 105089 (2022).
45. Tang, H., Zhang, C. & Xie, X. Automatic pulmonary lobe segmentation using deep learning. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. 1225–1228.
46. Jungeblut, P., Kleist, L. & Miltzow, T. The complexity of the Hausdorff distance. Preprint at <https://arXiv.org/quant-ph/arXiv:2112.04343> (2021).
47. Hausdorff, F. Grundzüge der Mengenlehre. Veit and Company, Leipzig, 1914. <https://archive.org/details/grundzgedermen00hausoft> (1949).
48. Taha, A. A. & Hanbury, A. An efficient algorithm for calculating the exact Hausdorff distance. *IEEE Trans. Pattern Anal. Mach. Intell.* **37**, 2153–2163 (2015).
49. Chen, X., Zhao, H. & Zhou, P. Lung lobe segmentation based on lung fissure surface classification using a point cloud region growing approach. *Algorithms* **13**, 263 (2020).
50. Liu, Y. *et al.* Paddleseg: A high-efficient development toolkit for image segmentation. Preprint at <https://arXiv.org/quant-ph/2101.06175> (2021).
51. Guo, L. J. Balance contrast enhancement technique and its application in image colour composition. *Int. J. Remote Sensing* **12**, 2133–2151. <https://doi.org/10.1080/01431169108955241> (1991).
52. Rahman, T. *et al.* Exploring the effect of image enhancement techniques on COVID-19 detection using chest X-ray images. *Comput. Biol. Med.* **132**, 104319. <https://doi.org/10.1016/j.combiomed.2021.104319> (2021).

Acknowledgements

The authors would like to express our deepest appreciation to everyone who offered advice on this project. We are thankful to the Faculty of Health Science Technology, HRH Princess Chulabhorn College of Medical Science, Chulabhorn Royal Academy, for providing the dataset. We are grateful to the Princess Srisavangavadhana College of Medicine, Chulabhorn Royal Academy for providing knowledge and equipment for this project. A

special thanks to the Department of Computer Engineering, Faculty of Engineering, King Mongkut's University of Technology Thonburi for providing the server used to train the model. Lastly, we thank Kimberly Moravec, PhD, from Edanz (<https://www.edanz.com/ac>) for editing a draft of this manuscript.

Author contributions

P.K., P.P., T.B., U.C.T., K.C., and N.T. designed the study. T.B. mainly performed data annotation. A.C., N.W., T.J., and N.T. checked and validated the annotation results. P.K. and P.P. developed and analyzed the lung lobe and lesion model. P.K. and P.P. wrote the main manuscript text. U.C.T.; K.C., U.D.T., and N.T. revised the manuscript. All authors reviewed and approved the submitted manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-47743-z>.

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

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 licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023