Online COVID-19 diagnosis with chest CT images: Lesion-attention deep neural networks

Bin Liu[1], Xiaoxue Gao[1], Mengshuang He[1], Fengmao Lv[1], and Guosheng Yin[2*]

[1]Center of Statistical Research, School of Statistics, Southwestern University of Finance and Economics, Chengdu, China.
[2]Department of Statistics and Actuarial Science, The University of Hong Kong, Hong Kong, China
[*]Corresponding author: gyin@hku.hk

May 11, 2020

Abstract

Chest (computed tomography) CT scanning is one of the most important technologies for COVID-19 diagnosis in the current clinical practice, which motivates more concerted efforts in developing AI-based diagnostic tools to alleviate the enormous burden on the medical system. We develop a lesion-attention deep neural network (LA-DNN) to predict COVID-19 positive or negative with a richly annotated chest CT image dataset. The CT image dataset contains 746 public chest CT images of COVID-19 patients collected from over 760 preprints, and the data annotations are accompanied with the textual radiology reports. We extract two types of important information from these annotations: One is the flag of whether an image indicates a positive or negative case of COVID-19, and the other is the description of five lesions on the CT images associated with the positive cases. The proposed data-driven LA-DNN model focuses on the primary task of binary classification for COVID-19 diagnosis, while an auxiliary multi-label learning task is implemented simultaneously to draw the model's attention to the five lesions of COVID-19 during the training. The joint task learning process makes it a highly sample-efficient deep model that can learn COVID-19 radiology features effectively with very limited samples. The experimental results show that the area under the curve (AUC) and sensitivity (recall) for the diagnosis of COVID-19 patients are 91.2% and 85.7% respectively, which reach the clinical standards for practical use. An online system has been developed for fast online diagnoses using CT images at the web address https://www.covidct.cn/.

1 Background

The novel coronavirus disease 2019 (COVID-19) is undergoing an unprecedented global outbreak. By May 11, 2020, more than 200 countries or territories have been affected by COVID-19 with a total of more than four million confirmed cases and over 280,000 deaths. Both the numbers of confirmed cases and deaths continue climbing up quickly. The COVID-19 has been declared as an international public health emergency by the World Health Organization (WHO).

The fast increasing numbers of COVID-19 cases and deaths have made the local medical systems in many countries overburdened. The reverse-transcription polymerase chain reaction (RT-PCR) test is the current standard method for detecting the coronavirus for COVID-19 patients. However, the laboratory RT-PCR test usually takes days to deliver the final result, although more efficient tests have been developed to shorten the time of diagnosis. Many countries are experiencing a backlog of test results due to a lack of diagnostic kits at their facilities, and the test results may even take longer than anticipated because of increased demands for testing globally. Not only are these tests insufficient to meet the urgent and vast demands in many countries, but they are also inefficient as the time lag of test reporting may cause treatment delay, especially for patients with critical conditions. Moreover, the sensitivity of the current RT-PCR testing kits is low; that is, a large number of COVID-19 patients cannot be identified accurately after their first tests. As a result, it usually requires several tests to make a final confirmation. Hence, patients may

not receive appropriate treatment and necessary quarantine during the RT-PCR testing period. On the other hand, chest CT scan is another critical tool for COVID-19 diagnosis. According to many existing studies [1, 3], CT scanning serves as a important and necessary supplement for the RT-PCR test and sometimes can even outperform the laboratory test in COVID-19 diagnosis. In contrast to the RT-PCR test, the chest CT scans and the corresponding diagnostic results can be obtained in a much faster way.

To improve the efficiency of the CT-based diagnosis, automatic diagnostic systems have been developed with AI technologies by reading patients' chest CT images as inputs and then output the diagnosis reports [6, 8, 10, 12]. These AI-based methods have demonstrated very promising results on COVID-19 prediction. However, most of the existing work do not share the training data publicly except for the work of He et al. [6], who constructed the first publicly available COVID-19 chest CT dataset by extracting the diagrams from over 760 preprints in *medRxiv* and *bioRxiv*. This public dataset contains 746 samples, among which 349 of them are COVID-19 positive and the rest 397 are negative.

As shown in Figure 1, the CT images in the dataset are annotated with professional textual analysis. The text reports usually narrate the results on whether the patients are COVID-19 positive or not. In addition to the flag of COVID-19, the text also contains information on lesion descriptions of COVID-19 patients. Based on our comprehensive statistical analysis over the entire text annotations, there are five different lesions associated with COVID-19, including the Ground-glass opacities (GGO), Consolidation (Csld), Crazy paving appearance (CrPa), Air bronchograms (AirBr), and Interlobular septal thickening (InSepThi). Figure 1 shows that each of the confirmed COVID-19 cases is attached with one to five lesion labels. Our experiments corroborate that the auxiliary information on the CT image lesions is extremely valuable for COVID-19 diagnosis and can greatly improve the diagnostic accuracy. However, the pioneering work [6] only focused on the COVID-19 diagnosis by conducting a binary classification task on predicting the flag of COVID-19, but ignored the significant amount of information on lesion details.

We develop a highly accurate COVID-19 diagnosis system based on both the chest CT images and the corresponding rich annotations on the five lesions. Our model is a double-task learning process which contains a primary binary classification task on the flag of COVID-19 and an auxiliary multi-label attention learning task on the five lesions. Both tasks are trained synchronously, while it shows that the auxiliary task promotes the primary task to focus its attention on the lesion areas and, as a result, the diagnostic accuracy of COVID-19 is improved drastically. Due to the incorporation of the attention mechanism through five lesions, we refer to our new model as the lesion-attention deep neural network (LA-DNN).

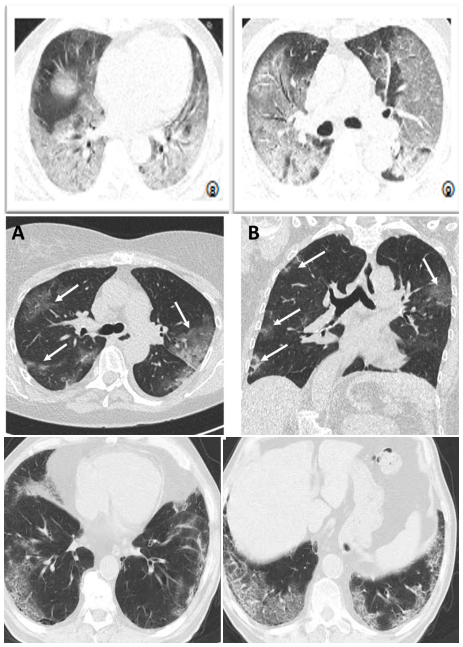
Experimental results demonstrate that our LA-DNN model can achieve great improvements by using the textual information. The area under the curve (AUC), recall (sensitivity), F1, and accuracy for predicting the diagnosis of COVID-19 are 91%, 86%, 85%, and 85%, respectively. These results reach the clinical standards for COVID-19 diagnosis and thus our system can be deployed for practical use to alleviate the enormous burdens of COVID-19 diagnosis [3]. The annotated lesion label file and the implementation codes can be found at `https://github.com/xiaoxuegao499/LA-DNN-for-COVID-19-diagnosis`. An online system has been developed for fast online COVID-19 diagnosis using chest CT images at the web address `https://www.covidct.cn/`. Our online system also welcomes medical staffs to upload their new local patients' image data to validate the diagnostic result, as well as keeping the system updated and the data shared publicly.

2 Methodology

2.1 Motivations and model

There have been an increasing amount of work on developing an AI-based COVID-19 diagnostic system using patients' chest CT scans [10]. Unfortunately, most of the data used in the deep learning models are not publicly available, which makes the existing models and results difficult to verify and reproduce. This situation is improved by He et al. [6], who published the first open-access COVID-19 chest CT scans dataset.

In a supervised learning fashion, classification based on deep learning typically requires a relatively large number of annotated samples to train the model for accurate prediction. However, the current publicly available dataset [6] only contains no more than 800 samples. The shortage of labeled samples and the urgency for the development of automated COVID-19 diagnostic tools



(**Top**) Representative chest CT images with various radiographic abnormalities: bilateral diffuse consolidation with air bronchograms.

(**Middle**) Example images from a 60 years old female patient with clinical and CT findings suggestive of COVID-19 infection, but with repeated negative RT-PCR analysis. Axial (A) and coronal (B) CT images show typical bilateral subpleural areas of ground-glass opacities. The patient was considered to be probably COVID-19 positive and quarantined. Note incidental finding of moderate thoracic dextroscoliosis.

(**Bottom**) The progress of CT findings in a patient with COVID-19. On the tenth day of admission, a follow-up chest CT scan showed an increase of extent of GGO with crazy-paving appearance.

Figure 1: Illustration of the multi-label chest CT images collected from different online papers. All of them have been confirmed as COVID-19 positive. We highlight the six labels with different color: COVID-19 in red, Ground-glass opacitites (GGO) in blue, Consolidation (Csld) in brown, Crazy paving appearance (CrPa) in violet, Air bronchograms (AirBr) in orange, and Interlobular septal thickening (InSepThi) in magenta. Taking the CT scans of the second patient (Middle) as an example, these chest CT images have been annotated with three labels: COVID-19, GGO, and InSepThi.

motivate us to derive a sample-efficient model that can integrate all sources of information for better decision making.

The textual annotations of the patients' chest CT scans contain two valuable tags: One is the flag of COVID-19, and the other is the reports of five lesions. In the pioneering work by He et al. [6], they trained a binary classification model only based on the COVID-19 flag while ignoring all the lesion information. To improve the performance of diagnosis, we propose to integrate the information on the lesion descriptions into the classification of the flag of COVID-19.

Similar to the work of He et al. [6], our goal is to make accurate classification of the COVID-19 positive or negative cases. However, different from their work which focused on making a complex knowledge transfer, we aim to fully exploiting the richly annotated textual information in the data. After annotating the five-category lesions on the COVID-19 positive images, we propose an auxiliary multi-label learning model based on the summarized five different lesion labels, in addition to the primary objective of the binary classification. The auxiliary task applies multi-label learning over the five lesions annotated based on the radiology reports: Ground-glass opacities (GGO), Consolidation (Csld), Crazy paving appearance (CrPa), Air bronchograms (AirBr), and Interlobular septal thickening (InSepThi), As shown in Figure 2, the primary and auxiliary tasks are trained synchronously. The auxiliary multi-label learning task promotes the fine-grained information on the radiology lesions to be integrated into the primary task, which makes the primary task pay more attention to the lesion areas rather than other uninteresting areas when making a final decision. This lesion-attention mechanism drastically improves the diagnostic accuracy to the level of clinical standards by human experts.

2.2 Implementation of LA-DNN

Figure 2 shows the architecture of the proposed LA-DNN model. A pre-trained backbone network takes the patient's chest CT images as inputs, for which seven popular backbone networks are used in the experiments, including VGG16 [9], ResNet18 [5], ResNet50 [5], DenseNet-121 [7], DenseNet-169 [7], EfficientNet-b0 [11], and EfficientNet-b1 [11]. The output of the last layer of the backbone

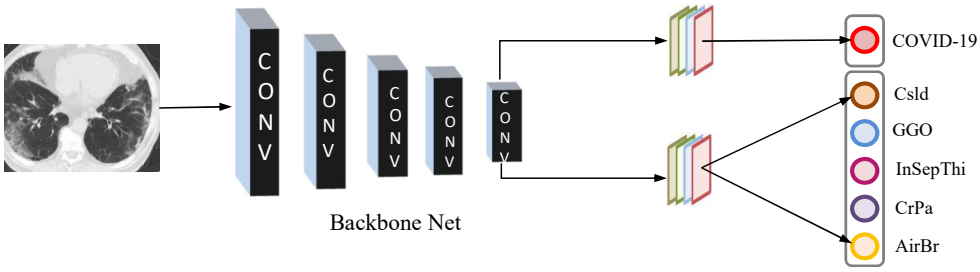


Figure 2: The architecture of the proposed lesion-attention deep neural networks.

networks is carried forward to two branches. One is used to predict whether a patient is COVID-19 positive or not, which is the primary task. Simultaneously, the other branch aims to make a multi-label prediction on the five lesion labels. The prediction errors from both the primary task and the auxiliary multi-label task are fed back to fine-tune the backbone network.

2.3 COVID-19 chest CT scans

We train the proposed LA-DNN model on the public dataset collected by He et al. [6]. This dataset contains 349 samples of COVID-19 positive and 397 COVID-19 negative CT scans, which are collected from 760 preprints about COVID-19 from *medRxiv* and *bioRxiv*, posted from January 19th to March 25th 2020. The negative samples include CT scans of normal individuals or contain other types of diseases. The COVID-19 positive CT images are further annotated with the corresponding radiology reports, which are textual clinical conclusions for the patients. This information is shown to be extremely valuable for the COVID-19 diagnosis. From the textual annotations, we identify and extract two types of important information:

- The first type of information is whether the patients are diagnosed positive or negative for COVID-19, which corresponds to the binary classification labels;
- The second significant but often ignored information is a common description of the five lesions associated with COVID-19. Each CT image of a confirmed patient has been identified at one to five lesions as shown in Figure 1.

3 Results

We split the dataset into a training set, a validation set, and a testing set by patients' IDs with the same ratios of 60%, 15%, 25% respectively as those in [6]. The only difference is that we need to ensure the COVID-19 patients in the training, validation, and testing sets cover all the five lesion labels, the new annotation based on radiology reports.

Given the limited training samples, we first take a classical neural network that has been well pre-trained on a large dataset ImageNet [4] as a feature extraction function, and then fine-tune the weights with the COVID-19 chest CT dataset. We select seven popular architectures as the backbone networks, including VGG16 [9], ResNet18 [5], ResNet50 [5], DenseNet-121 [7], DenseNet-169 [7], EfficientNet-b0 [11], and EfficientNet-b1 [11].

3.1 Overall performance

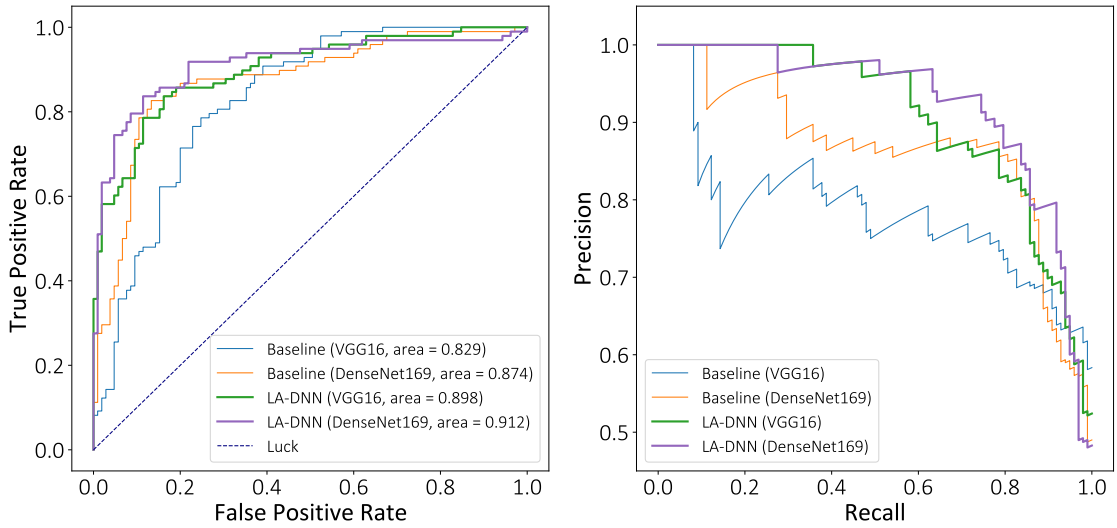
We take the pioneering work of He et al. [6] as the baseline for comparison. We train the baseline and the proposed LA-DNN model on the aforementioned dataset with the same data splitting strategies. The backbone networks for both the baseline and our LA-DNN are the seven classical architectures. Table 1 summarizes the overall performances of the baseline and our LA-DNN model with different settings of the backbone networks. Table 1 shows that the proposed LA-DNN model significantly improves the prediction of COVID-19 on all of the metrics, including accuracy, F1, AUC, and sensitivity.

Table 1: Performance comparisons between the baseline (Base) model and our lesion-attention deep neural network (LA-DNN) model on accuracy, F1, AUC, and sensitivity.

	Accuracy		**F1**		**AUC**		**Sensitivity**	
	Base	LA-DNN	Base	LA-DNN	Base	LA-DNN	Base	LA-DNN
VGG-16	76	83	76	83	82	90	75	85
ResNet-18	74	80	73	80	82	88	71	81
ResNet-50	80	85	81	83	88	91	67	79
DenseNet-121	79	82	79	81	88	89	77	80
DenseNet-169	83	85	81	85	87	91	77	86
EfficientNet-b0	77	84	78	84	89	90	68	88
EfficientNet-b1	79	81	79	80	84	88	70	78

3.2 Curves for assessment

To further assess the proposed LA-DNN model, a receiver operating characteristic (ROC) curve and a precision-recall curve are exploited to evaluate performances under different threshold values when interpreting probabilistic predictions. Figure 3 (a) exhibits the ROC curves of the baseline and our LA-DNN model by selecting the backbone network as VGG16 [9] and DenseNet169 [7] respectively. The results show that our model can predict the COVID-19 based on patients' chest CT scans with an area under the ROC curve (AUC) of 0.912 (when choosing DenseNet169 as the backbone net) and 0.898 (when choosing VGG16 as the backbone net), which demonstrate significant improvements from the corresponding AUC values of 0.874 and 0.829 from the baseline. Figure 3 (b) shows the precision-recall curves of the baseline and our LA-DNN model when choosing the VGG16 [9] and DenseNet169 [7] as the backbone networks respectively. The precision-recall curve plots the precision and the recall under different settings of the thresholds. The ideal model with a perfect prediction corresponds to the point with coordinates of $(1, 1)$. As shown in Figure 3 (b), the curves of our LA-DNN models using the backbone nets VGG16 and DenseNet169 bow towards the point $(1, 1)$ much closer than those of the baseline.



(a) ROC curves of the baseline and LA-DNN

(b) Precision-recall curves of the baseline and LA-DNN

Figure 3: Performance of our proposed LA-DNN model for COVID-19 diagnosis in comparison with the baseline.

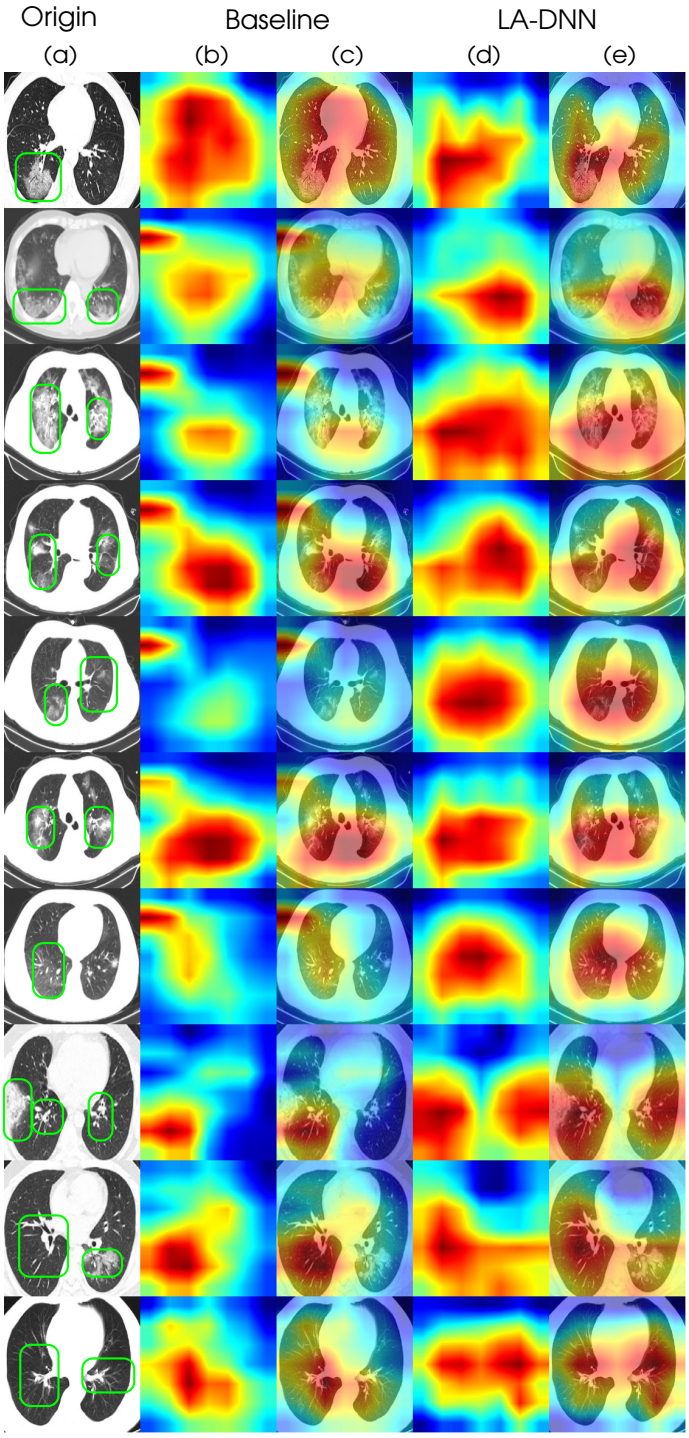


Figure 4: Grad-CAM++ visualization for the baseline and our LA-DNN model with the backbone net of DenseNet-169; Column (a) represents the original CT scans; Columns (b) and (c) are the class activation maps of the baseline [6]; Columns (d) and (e) are the class activation maps of our LA-DNN model. The color from deep red to dark blue corresponds to the activation values from large to small.

4 Analysis

4.1 Lesion attention map

We visualize the lesion attention map (i.e., the class activation map) concerning the five lesion labels with a convolutional neural network (CNN) visualization tool, Grad-CAM++ [2]. Grad-CAM++ can localize lesions in a CT image even if there are multiple occurrences of one lesion. Subsequently, Grad-CAM++ renders a class activation map, which illustrates the importance of each pixel in a feature map towards the final classification result. The attention heat-map exhibits the pixel-wise weighting of the gradients back-propagated from the output with respect to a particular spatial position in the final convolutional feature map of the CNN. In other words, the class activation map is a saliency map indicating which areas the model has paid attention to.

Figure 4 shows the class-specific attention maps for the baseline and our LA-DNN model with both choosing DenseNet169 as the backbone network. The first column represents the original COVID-19 CT images, and the lesion areas of these images are bounded with green boxes. Columns (b) and (c) in Figure 4 show the results of the baseline. In particular, column (b) is the class-specific attention map learned by the baseline. In column (c), the class attention map of the baseline is superimposed on the original images to show the activated areas. The colour of the maps from deep red to dark blue corresponds to the values of pixels' class-specific saliency from large to small. Columns (d) and (e) exhibit the results of the proposed LA-DNN model. The class-specific saliency map is a visual explanation of the lesions of COVID-19 CT scans that are predicted by the network. By comparing the results of the baseline with our model, we observe that the proposed LA-DNN model can capture almost all the salient areas for the COVID-19 prediction.

4.2 Visualization of the primary vs. auxiliary tasks

During the testing, the primary task of our LA-DNN model outputs a binary label on COVID-19 diagnosis, and meanwhile the auxiliary task outputs a five-dimensional vector to predict the five lesions. As shown in Figure 5, we plot the numeric components of these five-dimensional vectors paired with the COVID-19 classfication labels. The paired plot creates a grid of axes for the five lesions including GGO, Csld, CrPa, AirBr and InSepThi. The diagonal figures are treated differently to show the distributions of values of each lesion from the auxiliary task. The off-diagonal axes display the distribution of each lesion over the two categories: COVID-19 or NonCOVID-19. The paired plots in Figure 5 corroborate that the three lesions of CrPa, AirBr, and InSepThi are more important than GGO and Csld in distinguishing COVID-19 from NonCOVID-19.

5 Conclusion

To accommodate the urgent demands for COVID-19 testing, we develop a multi-lesion attention deep neural network for automating the COVID-19 diagnosis using a richly annotated chest CT image dataset. The samples of this dataset are collected from over 760 preprinted papers. To overcome the limitation of the samples, we extract two types of supervised information from these preprinted papers: One is the flag of COVID-19, and the other is the multiple labels for the five lesions of COVID-19. The rich annotations allow us to propose a sample-efficient deep neural network to learn valuable features with a limited number of samples. The proposed highly data-driven deep model contains a primary task on the binary classification for COVID-19 and an auxiliary multi-label attention task that forces the primary task to pay attention to the five lesions of COVID-19 during the training process. The experimental results demonstrate that the proposed LA-DNN model is capable of achieving the current clinical standards for diagnostic testing and thus our system should be broadly deployed for practical use. Currently, an online version of the COVID-19 AI-diagnostic system is set up for validation and continual collection of the data. All our codes and annotated data are publicly available to help other researchers further develop better systems to conquer the COVID-19.

References

[1] T. Ai, Z. Yang, H. Hou, C. Zhan, C. Chen, W. Lv, Q. Tao, Z. Sun, and L. Xia. Correlation of chest ct and rt-pcr testing in coronavirus disease 2019 (covid-19) in china: a report of 1014

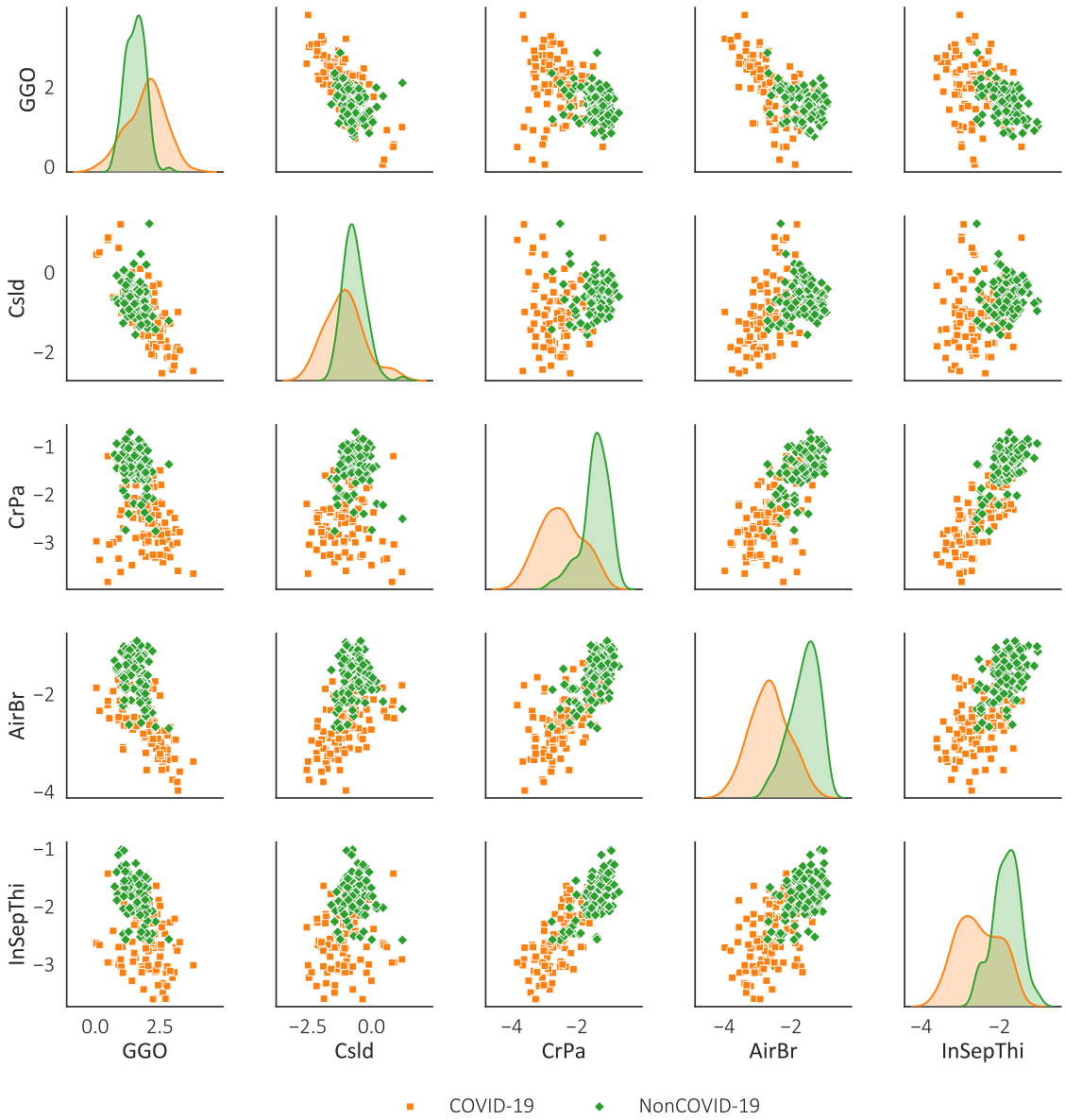


Figure 5: Plots of the pairwise relationships among the five lesions on making the final binary classification of COVID-19.

cases. *Radiology*, page 200642, 2020.

[2] A. Chattopadhay, A. Sarkar, P. Howlader, and V. N. Balasubramanian. Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 839–847, 2018.

[3] A. Dangis, C. Gieraerts, Y. D. Bruecker, L. Janssen, H. Valgaeren, D. Obbels, M. Gillis, M. V. Ranst, J. Frans, A. Demeyere, and R. Symons. Accuracy and reproducibility of low-dose submillisievert chest ct for the diagnosis of covid-19. *Radiology: Cardiothoracic Imaging*, 2(2):e200196, 2020.

[4] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[5] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[6] X. He, X. Yang, S. Zhang, J. Zhao, Y. Zhang, E. Xing, and P. Xie. Sample-efficient deep learning for covid-19 diagnosis based on ct scans. *medRxiv*, 2020.

[7] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017.

[8] F. Shi, J. Wang, J. Shi, Z. Wu, Q. Wang, Z. Tang, K. He, Y. Shi, and D. Shen. Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for covid-19. *IEEE Reviews in Biomedical Engineering*, 2020.

[9] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[10] Y. Song, S. Zheng, L. Li, X. Zhang, X. Zhang, Z. Huang, J. Chen, H. Zhao, Y. Jie, R. Wang, et al. Deep learning enables accurate diagnosis of novel coronavirus (covid-19) with ct images. *medRxiv*, 2020.

[11] M. Tan and Q. V. Le. Efficientnet: Rethinking model scaling for convolutional neural networks. *arXiv preprint arXiv:1905.11946*, 2019.

[12] S. Wang, B. Kang, J. Ma, X. Zeng, M. Xiao, J. Guo, M. Cai, J. Yang, Y. Li, X. Meng, et al. A deep learning algorithm using ct images to screen for corona virus disease (covid-19). *MedRxiv*, 2020.