

# A Fast Online COVID-19 Diagnostic System with Chest CT Scans

Bin Liu, Xiaoxue Gao, Mengshuang He, Lin Liu  
Center of Statistical Research, School of Statistics,  
Southwestern University of Finance and Economics  
Chengdu, China

Guosheng Yin  
Department of Statistics and Actuarial Science,  
The University of Hong Kong  
Hong Kong, China

## ABSTRACT

Chest computed tomography (CT) scanning is one of the most important technologies for COVID-19 diagnosis and disease monitoring, particularly for early detection of coronavirus. Recent advancements in computer vision motivate more concerted efforts in developing AI-driven diagnostic tools to accommodate the enormous demands for the COVID-19 diagnostic tests globally. To help alleviate burdens on medical systems, 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. Based on the textual radiological report accompanied with each CT image, we extract two types of important information for the annotations: One is the indicator of 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-efficient 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 attention of DNN to the five lesions associated with COVID-19. The joint task learning process makes it a highly sample-efficient DNN that can learn COVID-19 radiology features more effectively with limited but high-quality, rich-information samples. The experimental results show that the sensitivity (recall), area under the curve (AUC), F1 score, and accuracy for COVID-19 diagnosis are 88.8%, 94.7%, 87.9%, and 89.0%, respectively. A free online system is currently alive for fast diagnosis using CT images at the website <https://www.covidct.cn/>, and all codes and datasets are freely accessible at our github address.

## KEYWORDS

Attention, COVID-19, CT image, diagnosis, deep neural network

### ACM Reference Format:

Bin Liu, Xiaoxue Gao, Mengshuang He, Lin Liu and Guosheng Yin. 2020. A Fast Online COVID-19 Diagnostic System with Chest CT Scans. In *Proceedings of KDD 2020*. ACM, New York, NY, USA, 8 pages. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>

## 1 BACKGROUND

The novel coronavirus disease 2019 (COVID-19) is undergoing an unprecedented global outbreak. On March 11, 2020, COVID-19 was declared as an international public health emergency by the

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

KDD 2020, August 23–27, 2020, San Diego, CA

© 2020 Association for Computing Machinery.

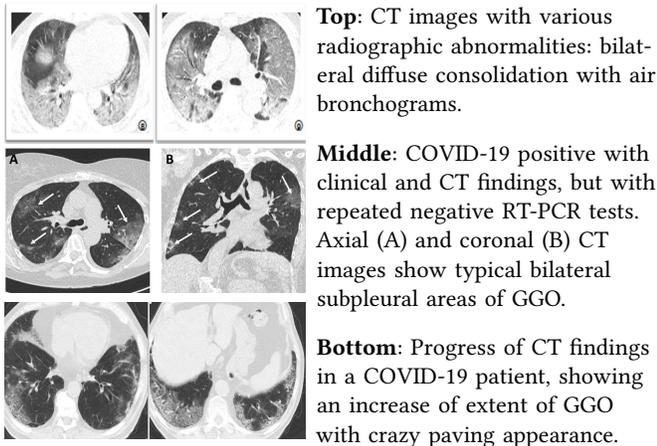
ACM ISBN 978-x-xxxx-xxxx-x/YY/MM... \$15.00

<https://doi.org/10.1145/nnnnnnn.nnnnnnn>

World Health Organization (WHO). By July 24, 2020, more than 200 countries or territories have been affected by COVID-19 with a total of more than 15 million confirmed cases and over 630,000 deaths. Both the numbers of confirmed cases and deaths continue climbing up quickly worldwide.

The fast increasing numbers of COVID-19 cases and deaths have caused overburdens on many local medical systems across the world. Currently, the reverse-transcription polymerase chain reaction (RT-PCR) test is the standard method for detecting the coronavirus in COVID-19 patients. However, the laboratory RT-PCR test usually takes a rather long time (e.g., in days) to deliver the final result. To shorten the time of diagnosis, the real time RT-PCR test is recommended which can deliver a reliable diagnosis result much faster (e.g., in 6–8 hours). Such tests need a collection of a sample with a swab that goes deep in a person’s nose or throat, i.e., parts of the body where the coronavirus gathers. However, if the swabbed areas do not have coronavirus accumulated, the tests may fail to identify a COVID-19 patient correctly, yielding a false negative result. Many countries are experiencing a backlog of test results due to a lack of diagnostic kits at their medical facilities, and the test results may even take longer time than anticipated due to the increasing demands for testing globally. Not only are these tests insufficient to meet the urgent and vast demands in many countries (particularly those with poor medical infrastructures), 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 not high; that is, a large number of COVID-19 patients cannot be identified accurately after their first tests due to false negatives. As a result, it usually requires several tests to make a final confirmation and it is not uncommon to have several negative test results followed by a positive one. Hence, patients may not receive appropriate treatment and necessary quarantine during the RT-PCR testing period. This is particularly true for patients with no or minimal symptoms.

On the other hand, chest CT scans provide another critical tool for COVID-19 diagnosis and disease monitoring, especially for early detection when the symptoms are yet onset. After entering the body, coronavirus often attacks the lung first and thus certain lesions would manifest in the lung’s CT image, before a swab can collect an adequate amount of virus in the nose or throat for testing. According to many existing studies [1, 3], CT scanning serves as an important and necessary supplement for the RT-PCR test and sometimes can even outperform the laboratory test for COVID-19 diagnosis. In contrast to the RT-PCR test, the chest CT scans have been commonly used in many disease areas (e.g., cancer) for diagnosis and monitoring, and the corresponding diagnostic results can be obtained in a much faster way.



**Top:** CT images with various radiographic abnormalities: bilateral diffuse consolidation with air bronchograms.

**Middle:** COVID-19 positive with clinical and CT findings, but with repeated negative RT-PCR tests. Axial (A) and coronal (B) CT images show typical bilateral subpleural areas of GGO.

**Bottom:** Progress of CT findings in a COVID-19 patient, showing an increase of extent of GGO with crazy paving appearance.

**Figure 1: Illustration of the multi-label chest CT images of COVID-19 positive patients collected from online papers. Besides the flag of COVID-19, five lesions labels are annotated: ground glass opacities (GGO), consolidation (Csld), crazy paving appearance (CrPa), air bronchograms (AirBr), and interlobular septal thickening (InSepThi).**

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 results [6, 9, 11, 13]. These AI-driven methods have demonstrated very promising performances on COVID-19 prediction. However, most of the existing work do not share the training data publicly, while He et al. [6] constructed the first openly accessible COVID-19 chest CT dataset by extracting the CT images from over 760 preprints in *medRxiv* and *bioRxiv*. This publicly available dataset contains 746 samples, among which 345 of them are COVID-19 positive and the rest 401 are negative<sup>1</sup>. We continue expanding the dataset by collecting new samples appeared in the latest online publications on COVID-19.

Figure 1 shows several examples of the CT images in the dataset annotated with professional textual analysis, accompanied with the radiological reports on the right side. The text reports typically narrate the results on whether the patients are COVID-19 positive or not, as well as descriptions of lesions 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 opacity (GGO), consolidation (Csld), crazy paving appearance (CrPa), air bronchograms (AirBr), and interlobular septal thickening (InSepThi). Figure 1 shows that most of the confirmed COVID-19 cases is attached with one to five lesion labels<sup>2</sup>. Our experiments corroborate that the auxiliary information on these 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 the common lesions which are distinctive from other types of pneumonia.

We develop a highly accurate COVID-19 diagnosis system based on the chest CT images as well as the corresponding rich annotations on the five lesions. Our model adopts 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 above the level of the state-of-the-art method. Due to the incorporation of the attention mechanism on the 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 sensitivity (recall), area under the receiver operating characteristic (ROC) curve (AUC), F1 score, and accuracy for COVID-19 diagnosis are 88.8%, 94.7%, 87.9%, and 89.0%, respectively. These results improve drastically over the existing work and reach the clinical standards for COVID-19 diagnosis [3, 8]. Therefore, our system can be deployed for practical use to alleviate the enormous burdens of COVID-19 diagnostic tests [3]. The annotated lesion label files and the implementation codes in Python (PyTorch 1.4) can all be freely accessed at <https://github.com/xiaoxuegao499/LA-DNN-for-COVID-19-diagnosis>. An online system has been developed and is openly available for fast COVID-19 diagnosis using chest CT images at the website <https://www.covidct.cn/>.

## 2 METHODOLOGY

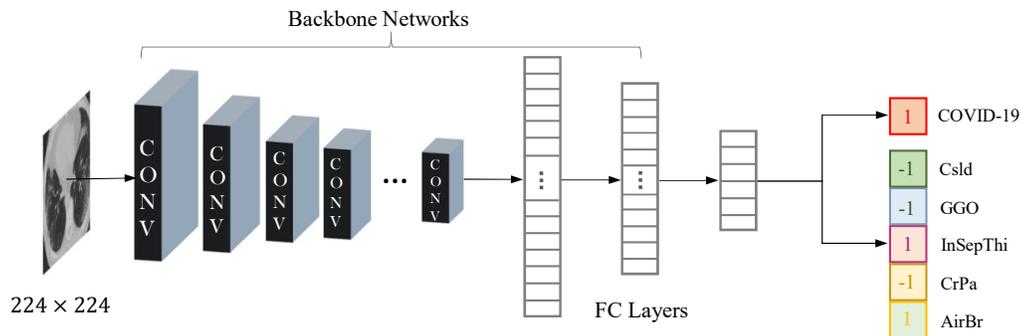
### 2.1 Motivation and model

There has been an increasing amount of work on developing an AI-based COVID-19 diagnostic system using patients’ chest CT scans [11]. 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. He et al. [6] published the first open-access COVID-19 chest CT image dataset by collecting the lung CT images appeared in the online preprints of research work on COVID-19. In a supervised learning process, classification based on deep learning models 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 746 published chest CT images. The shortage of labeled samples and the urgency for the development of automated COVID-19 diagnostic tools motivate us to derive a sample-efficient deep neural network that can integrate all sources of information for optimal decision making.

Through careful studies on the preprints associated with the patients’ chest CT scans, we can extract valuable textual annotations of these CT images. One is the flag of COVID-19, positive or negative, and the other is the radiological reports on five potential lesions in the lungs for COVID-19 positive patients (no information on the lesions for COVID-19 negative patients). In the pioneering work by He et al. [6], they trained a binary classification model based on the COVID-19 flag only, while ignoring all the lesion information which requires further annotations. To improve the

<sup>1</sup>The original dataset [6] has 349/397 positive/negative samples, while the authors assigned 4 positive labels to negative samples by mistake.

<sup>2</sup>We could not find any lesion information for 18 COVID-19 positive cases from the radiological reports.



**Figure 2: The architecture of the proposed lesion-attention deep neural networks with a primary task of binary classification and an auxiliary task of multi-label learning of five lesions; the last three layers are fully connected (FC) layers.**

performance of COVID-19 diagnosis, we propose to integrate the information on the lesion descriptions into the classification of the flag of COVID-19.

Our goal is to make accurate classification of the COVID-19 positive or negative directly using the chest CT images. However, in contrast to the work of He et al. [6] which focused on making a complex knowledge transfer, we aim to fully exploit 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 for COVID-19. The auxiliary task applies multi-label learning over the five lesions annotated using the corresponding radiological reports, which include ground glass opacities (GGO), consolidation (Cslid), 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 in our LA-DNN model, so that the unknown parameters can be learned more effectively than those by only training the primary task for binary classification. The auxiliary multi-label learning task promotes the fine-grained information on the radiology-revealed 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 classification decision. This lesion-attention mechanism drastically improves the diagnostic accuracy up to the level of clinical standards by medical experts. Although our LA-DNN is trained jointly by binary classification and five-label learning with lesion annotations, for practical use it does not require any annotations on the lesions in the lung image. Our system simply takes a plain lung CT image as an input and directly outputs the binary classification as COVID-19 positive or negative.

## 2.2 Implementation of LA-DNN

Figure 2 shows the architecture of the proposed LA-DNN model. Using the ImageNet, we first pre-train the backbone networks, and then via the idea of transfer learning, a pre-trained backbone

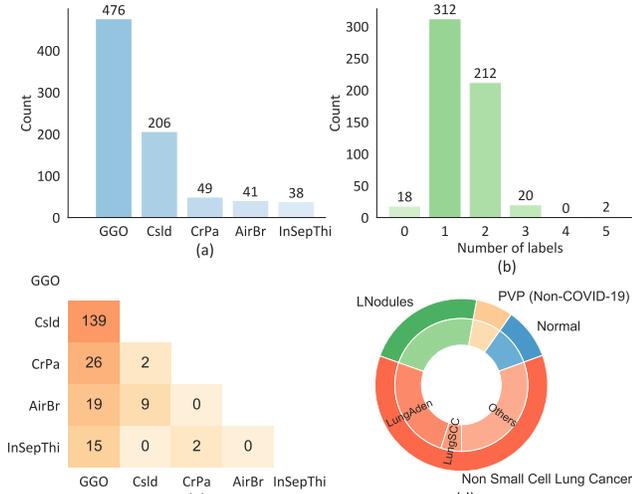
network takes the patient’s chest CT images as inputs. Seven well-known deep neural networks are explored one at a time to be used as the backbone network in the experiment, including VGG-16 [10], ResNet-18 [5], ResNet-50 [5], DenseNet-121 [7], DenseNet-169 [7], EfficientNet-b0 [12], and EfficientNet-b1 [12]. These backbone networks typically involve millions of parameters which are pre-trained using the ImageNet. The output of the last layer of dimension 6 is carried forward to two branches. One branch is used to predict whether a patient is COVID-19 positive or not, for which a primary loss function  $L_P$  is calculated based on the cross entropy of the binary classification task. Simultaneously, the other branch aims to make a multi-label prediction on the five lesions, which brings in an auxiliary loss function  $L_A$ . The multi-label learning of the five lesions are predicted using the *sigmoid* function  $\sigma(x) = 1/(1 + e^{-x})$ . The final prediction error is the summation of  $L_P$  from the primary binary task and  $L_A$  from the auxiliary multi-label task. In other words, we have two different sources of feedback to fine-tune the backbone networks.

## 2.3 COVID-19 chest CT scans

We train the proposed LA-DNN model on the dataset published by He et al. [6] in conjunction with our newly collected dataset. The original dataset contains 345 samples of COVID-19 positive and 401 COVID-19 negative CT scans, which are collected from 760 research preprints related to COVID-19 from *medRxiv* and *bioRxiv*, posted online from January 19th to March 25th 2020. We continue expanding the dataset of chest CT images by further collecting 219 new samples from 57 COVID-19 positive patients and 259 new negative samples from 164 non-COVID-19 patients from the newly appeared publications up to May 21th 2020. We denote the original dataset with 746 samples as  $\mathcal{D}_0$ , and denote the dataset with newly collected 478 samples as  $\mathcal{D}_1$ . The combined dataset  $\mathcal{D}_0 + \mathcal{D}_1$  contains a total of 564 positive samples from 269 COVID-19 patients and 660 negative samples from 339 non-COVID-19 individuals. Table 1 shows the details of the two datasets and the split of the data into training, validation and testing sets. All results reported below are based on the testing sets.

**Table 1: Descriptions of the datasets ( $\mathcal{D}_0$  by He et al. [6] and our newly collected samples in  $\mathcal{D}_1$ ) stratified by COVID-19 positive or negative, with the number of samples (or CT images) and the number of patients given in parentheses in each dataset, and the split of the data based on the patient ratios of 60%, 15%, 25% for training, validation, and testing sets.**

Datasets	COVID-19		COVID-19 Positive	COVID-19 Negative
	Positive	Negative	Training/Validation/Testing	Training/Validation/Testing
$\mathcal{D}_0$	345(212)	401(175)	187(126)/60(32)/98(54)	238(109)/58(24)/105(42)
$\mathcal{D}_1$	219(57)	259(164)	–	–
$\mathcal{D}_0+\mathcal{D}_1$	564(269)	660(339)	340(169)/81(34)/143(66)	398(215)/97(44)/165(80)



**Figure 3: Descriptions of COVID-19 positive and negative samples: (a) Histogram of the numbers of each lesion; (b) Histogram of the numbers of labels of samples in the COVID-19 positive set; (c) Lesion concordance matrix; and (d) Composition of the COVID-19 negative samples.**

The COVID-19 negative samples in the combined dataset include the CT scans of normal individuals as well as patients with other types of diseases, e.g., lung cancer or other types of pneumonia. The COVID-19 positive CT images are further annotated with the corresponding radiological reports, which are the textual clinical information for the patients. The radiological information is proven to be extremely valuable for the COVID-19 diagnosis. From the textual annotations, we extract two types of important information:

- The first type of information is whether patients are diagnosed positive or negative for COVID-19, which corresponds to the binary classification labels.
- The second significant but often ignored information is descriptions of the five common lesions associated with COVID-19. As shown in Figure 3 (b), most of the lung CT images of COVID-19 positive patients are identified with one or two lesions in the lung, and two samples even have five lesions, while only 18 samples have no lesions found in the CT images.

Figure 3 shows the visualization of the COVID-19 positive and negative samples in our combined dataset. Specifically, the statistics of frequency of the five lesions, GGO, Cslid, CrPa, AirBr, and InSepThi, in our dataset are as shown in Figure 3 (a). Clearly, GGO is the most common lesion and consolidation ranks the second. Figure 3 (b) and (c) exhibit the distribution of numbers of lesion

labels for COVID-19 positive samples and the lesion concordance matrix (i.e., the number of times a pair of lesions appeared in the same sample), respectively. From the histogram of the numbers of labels for all COVID-19 positive samples, we find that most of them have either one or two lesion labels. The paired-label concordance matrix of the 212 COVID-19 positive samples, each of which have two lesion labels, demonstrates that GGO and consolidation often appear together in the CT images, as shown in Figure 3 (c). Moreover, GGO is the lesion that has the most frequent interactions with all the other lesions. Figure 3 (d) illustrates the composition of the COVID-19 negative samples. The negative sample set contains lung cancer (Non Small Cell Lung Cancer), Lung Nodules (LNodules), Pulmonary Viral Pneumonia (PVP) which is non-COVID-19 pneumonia, and normal samples from healthy individuals. The lung cancer covers many types of common cancers, including lung adenocarcinomas (LungAden), lung squamous cell carcinoma (LungSCC), and others. These non-COVID-19 samples contain a variety of lung diseases, which can help to train our LA-DNN to distinguish COVID-19 from other lung diseases including other types of pneumonia.

## 3 RESULTS

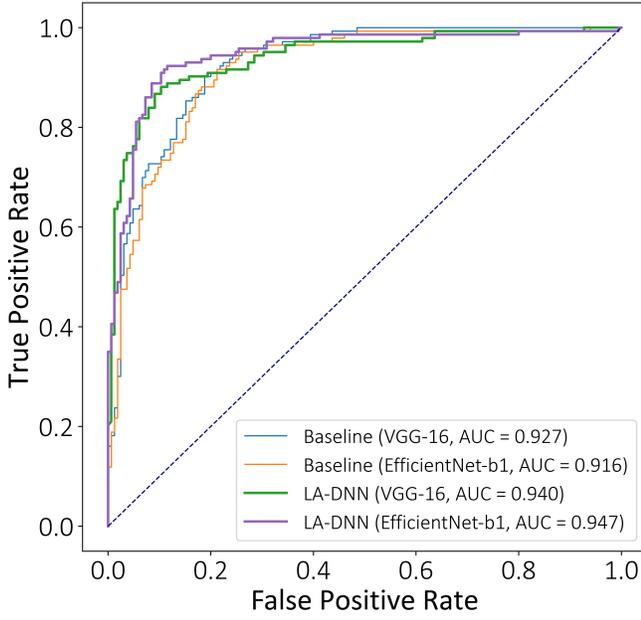
### 3.1 Evaluation metrics

To compare the LA-DNNs under different backbone networks with the corresponding baselines without multi-label lesion learning, we select four performance metrics: sensitivity, the area under the ROC curve (AUC), the F1 score, and accuracy for model evaluation.

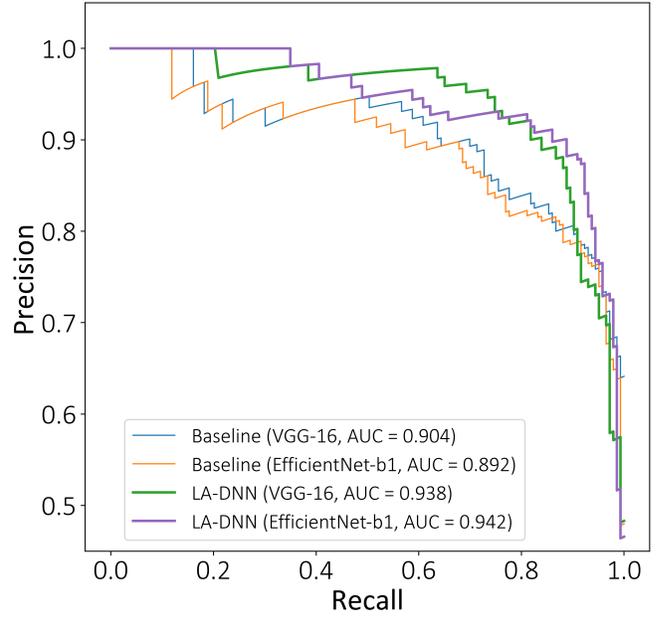
**Sensitivity**, also known as recall or the true positive rate (TPR), is a statistical measure of the performance of a binary classification test that is widely used in medical diagnostic tests. Sensitivity is the probability of diagnostic test/classification positive using LA-DNN for a truly COVID-19 positive patient, which can be calculated as the fraction of the number of true positives (retrieved by our LA-DNN model) over the number of all the positive cases in ground-truth,

$$\text{Sensitivity} = \frac{\#\text{True Positive}}{\#\text{True Positive} + \#\text{False Negative}}$$

The AUC, the area under an ROC curve, is a performance measurement for classification problems, where the ROC curve is created by plotting the TPR against the false positive rate (FPR) at various threshold settings. It delivers the information on how much a classification model is capable of distinguishing different classes. The larger the AUC, the better is the model in predicting positive samples as 1’s (positive labels) and negative samples as -1’s (negative labels). In our case, the higher the AUC, the better is the LA-DNN in distinguishing patients with COVID-19 and those of non-COVID-19.



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



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

**Figure 4: Performances of our LA-DNN model for COVID-19 diagnosis under the ROC and precision-recall curves in comparison with the baseline on the combined dataset  $\mathcal{D}_0 + \mathcal{D}_1$ .**

**Table 2: Comparisons of the sensitivity, AUC, F1 score, and accuracy (%) between the baseline model and our LA-DNN model on the independent testing sets using the original dataset  $\mathcal{D}_0$  and the combined dataset  $\mathcal{D}_0 + \mathcal{D}_1$ , respectively.**

Metrics	Backbone Net	$\mathcal{D}_0$		$\mathcal{D}_0 + \mathcal{D}_1$	
		Baseline	LA-DNN	Baseline	LA-DNN
Sensitivity	VGG-16	75.5	84.7	85.3	<b>88.8</b>
	ResNet-18	71.4	80.6	75.5	82.5
	ResNet-50	67.3	78.6	82.5	87.4
	DenseNet-121	77.6	79.6	86.0	86.0
	DenseNet-169	77.6	85.7	85.3	87.4
	EfficientNet-b0	68.4	<b>87.8</b>	86.7	88.1
	EfficientNet-b1	70.4	77.6	76.2	86.0
	AUC	VGG-16	81.3	89.8	92.7
ResNet-18		83.0	88.2	93.1	93.1
ResNet-50		87.6	90.5	91.6	93.8
DenseNet-121		86.0	88.7	92.8	93.2
DenseNet-169		88.2	<b>91.2</b>	93.1	93.3
EfficientNet-b0		87.7	90.0	92.0	92.4
EfficientNet-b1		84.1	87.6	91.6	<b>94.7</b>
F1 score		VGG-16	74.8	83.0	84.1
	ResNet-18	74.0	79.8	80.6	84.6
	ResNet-50	75.9	83.2	82.8	86.8
	DenseNet-121	80.0	81.2	84.2	85.4
	DenseNet-169	81.3	<b>84.8</b>	86.2	86.8
	EfficientNet-b0	74.9	83.9	85.2	86.6
	EfficientNet-b1	73.4	80.0	79.8	<b>87.9</b>
	Accuracy	VGG-16	75.4	83.2	85.1
ResNet-18		75.9	80.2	83.1	86.0
ResNet-50		79.3	84.7	84.1	87.7
DenseNet-121		81.3	82.3	85.1	86.4
DenseNet-169		82.8	<b>85.2</b>	87.3	87.7
EfficientNet-b0		77.8	83.7	86.0	87.3
EfficientNet-b1		75.4	81.2	82.1	<b>89.0</b>

The **F1 score** is defined as an average of the precision and recall,

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}},$$

where

$$\text{Precision} = \frac{\#\text{True Positive}}{\#\text{True Positive} + \#\text{False Positive}}.$$

The relative contributions of precision and recall to the F1 score are equal, and an F1 score reaches its best value at 1 and the worst score at 0.

**Accuracy** is the rate of the correct predictions over the number of both positive and negative samples in the test set,

$$\text{Accuracy} = \frac{\#\text{True Positive} + \#\text{True Negative}}{\#\text{Positive} + \#\text{Negative}}.$$

Besides the aforementioned four metrics, we use **Hamming loss** to evaluate the performance of the multi-label prediction on the five lesions. The Hamming loss measures the fraction of the incorrect labels with respect to the total number of labels, which is defined as

$$\text{Hamming loss} = \frac{1}{p} \sum_{i=1}^p I(y_i \neq \hat{y}_i)$$

where  $y_i$  and  $\hat{y}_i$  are respectively the  $i$ th component of  $\mathbf{y}$  and  $\hat{\mathbf{y}}$ , the  $p$ -dimensional binary vectors of observed and predicted values, and  $I(\cdot)$  is the indicator function. Hence, the optimal value of the Hamming loss is zero.

### 3.2 Overall performance

In our experiments, we split both the original dataset  $\mathcal{D}_0$  and the combined dataset  $\mathcal{D}_0 + \mathcal{D}_1$  respectively into a training set, a validation set, and a testing set by patients' IDs with the same ratios of

60%, 15%, 25% as those in [6]. One patient may possibly has multiple CT images in the dataset, while all the CT images belonging to the same patient would be allocated together to the same set. Through splitting the dataset by patients, we can avoid information leaking from the training to the testing set, as CT images from the same patient are highly correlated and contain largely overlapping information. We need to ensure that the COVID-19 positive patients in the training, validation, and testing sets cover the lesion labels based on the annotations using the radiological reports.

Due to the limited training samples, we adopt a classical deep neural network that has been well pre-trained on a large dataset ImageNet [4] as a feature extraction function. Through transfer learning, we further fine-tune the weights in the last several fully connected layers with the COVID-19 chest CT image dataset, while keeping the majority part of the pre-trained model intact. We select seven popular deep architectures as the backbone networks, including VGG-16 [10], ResNet-18 [5], ResNet-50 [5], DenseNet-121 [7], DenseNet-169 [7], EfficientNet-b0 [12], and EfficientNet-b1 [12].

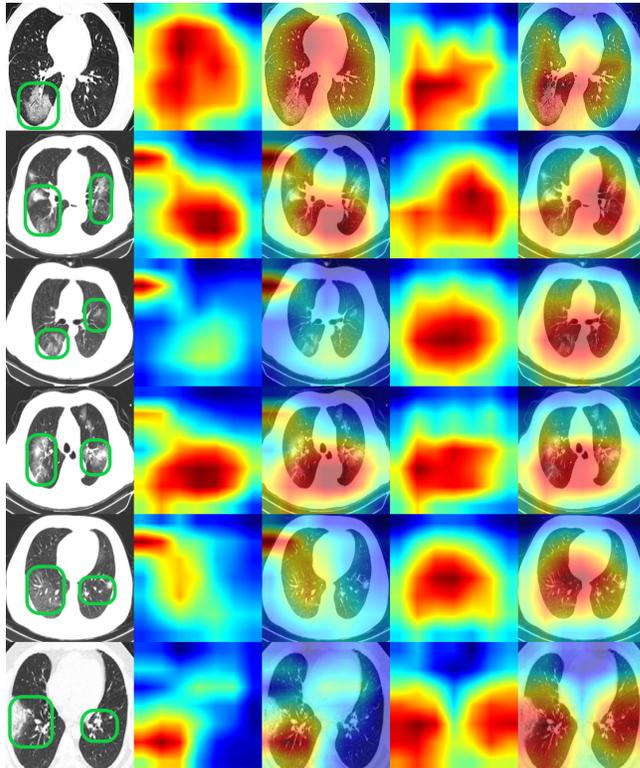
We take the pioneering work of He et al. [6] as the baseline for comparison. We train the baseline and the proposed LA-DNN model on the original dataset  $\mathcal{D}_0$  as well as the combined dataset  $\mathcal{D}_0 + \mathcal{D}_1$  with the same data splitting strategy. Table 2 summarizes the overall performances of the baseline and our LA-DNN model under different backbone networks on the testing sets. Clearly, the proposed LA-DNN model significantly improves the prediction accuracy of COVID-19 positive or negative on all of the four metrics with both the original and combined datasets. Experimental results demonstrate that the auxiliary task learning process by using the five lesions from the textual information can greatly improve the primary task for binary classification of COVID-19. Moreover, the additional dataset  $\mathcal{D}_1$  can further improve the model’s prediction accuracy. Among all the methods considered, the best performance are delivered by the LA-DNN using the VGG-16 and EfficientNet-b1 as the backbone nets with the combined dataset.

### 3.3 Curves for assessment

To further assess the proposed LA-DNN model, an ROC curve and a precision-recall curve are exploited to evaluate the performances under different threshold values when interpreting probabilistic predictions. Figure 4 shows the plots and AUC values on the dataset  $\mathcal{D}_0 + \mathcal{D}_1$ . Figure 4 (a) exhibits the ROC curves of the baseline and our LA-DNN model by selecting VGG-16 [10] and EfficientNet-b1 [12] as the backbone networks, respectively. The results show that our model can predict the COVID-19 based on patients’ chest CT scans with an AUC of 0.947 when choosing EfficientNet-b1 as the backbone net and 0.940 when choosing VGG-16 as the backbone net, which improve substantially over the corresponding AUC values of 0.916 and 0.927 from the baseline. Figure 4 (b) shows the precision-recall curves of the baseline and our LA-DNN model when choosing the VGG-16 [10] and EfficientNet-b1 [12] as the backbone networks, respectively. The precision-recall curve plots the precision against recall under different threshold values. The ideal model with a perfect prediction corresponds to the point with the coordinates of (1, 1). As shown in Figure 4 (b), the curves of our LA-DNN models using the backbone nets VGG-16 and EfficientNet-b1 bend towards the point (1, 1) much closer than those of the baseline.

**Table 3: Performance of multi-label classification on five lesions under the proposed LA-DNN with the dataset  $\mathcal{D}_0 + \mathcal{D}_1$ .**

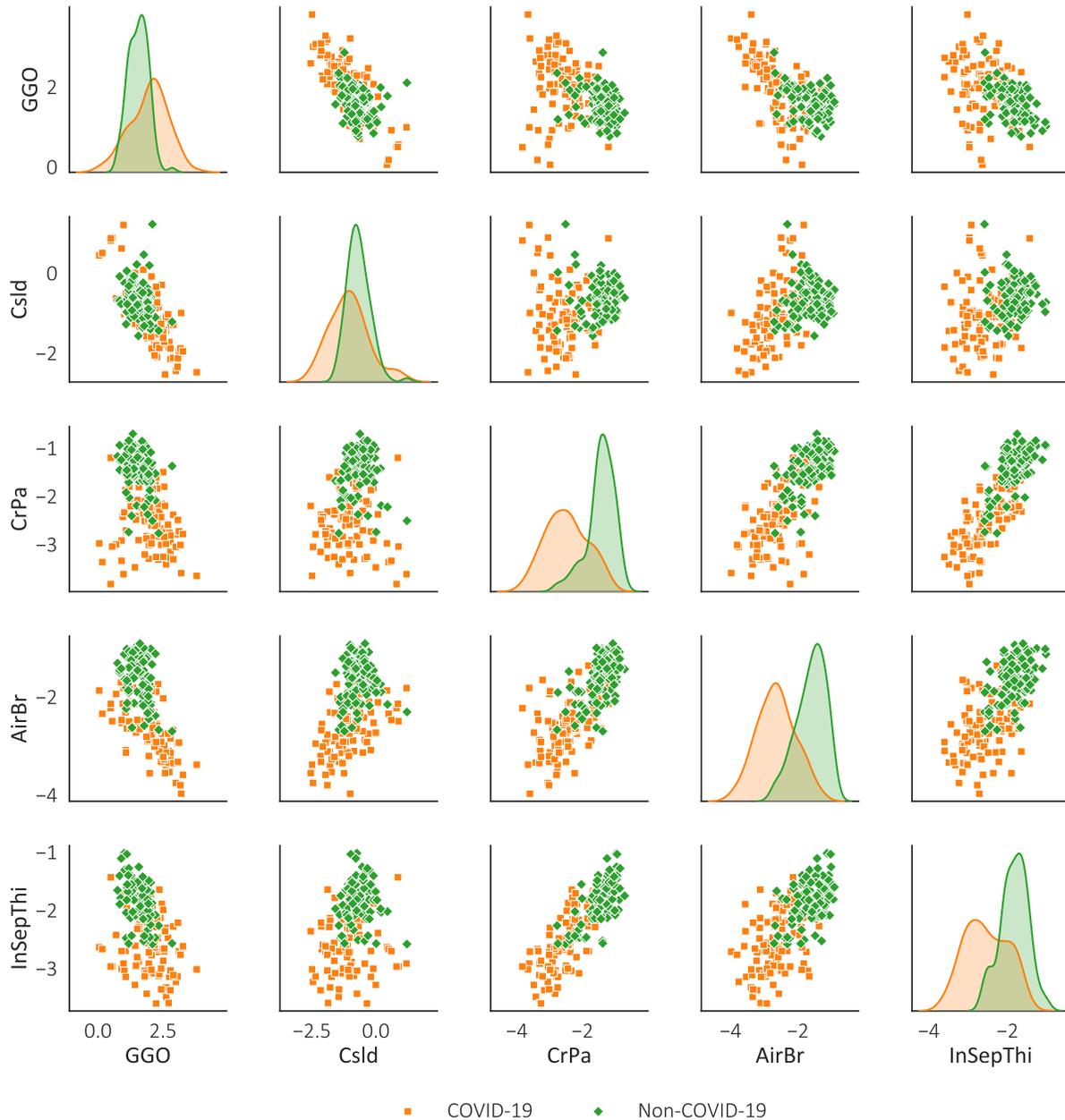
Backbone Net	Accuracy (%)	Hamming loss
VGG-16	87.3	0.127
ResNet-18	85.9	0.141
ResNet-50	88.8	0.112
DenseNet-121	88.5	0.115
DenseNet-169	86.6	0.134
EfficientNet-b0	87.6	0.124
EfficientNet-b1	85.6	0.144



**Figure 5: Grad-CAM++ visualization for the baseline and our LA-DNN model with the backbone net of DenseNet-169. The first column represents the original CT scans; Columns 2 and 3 are the class activation maps of the baseline [6]; Columns 4 and 5 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.**

### 3.4 Multi-label classification on five lesions

The proposed method is a double-task learning approach, which contains the primary task of classifying the flag of COVID-19 and an auxiliary multi-label classification task on predicting the five lesions. The accuracy and hamming loss of the multi-label classification are as shown in Table 3. We observe that the performances of the LA-DNN under the seven different backbone networks are comparable with each other. Among all seven backbone networks, ResNet-18 and DenNet-50 achieve the better results than other networks.



**Figure 6: Plots of the pairwise relationships among the five lesions based on the numerical values of the last fully connected layer of LA-DNN on making the binary classification of COVID-19 or non-COVID-19.**

## 4 ANALYSIS

### 4.1 Lesion attention map

To gain more insights into our LA-DNN model, we can 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 the lesions in a CT image even if there are multiple occurrences of the same 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.

Based on the selected six chest CT image samples, Figure 5 shows the class-specific attention maps for the baseline and our LA-DNN model when both choose DenseNet-169 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 2 and 3 in Figure 5 show the results of the baseline. In particular, column 2 is the class-specific attention map learned by the baseline. In column 3, the class attention map of the baseline is superimposed on the original images to show the activated areas. The color of the maps from deep red to dark blue corresponds to the values of pixels' class-specific saliency from large to small. Columns 4 and 5 in Figure 5 exhibit the corresponding 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 between the baseline and 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 and 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 labels of five lesions. Figure 6 shows the numeric values of these five-dimensional vectors in the last fully connected layer of LA-DNN paired with the COVID-19 classification labels. The paired plot creates a matrix for the five lesions of GGO, CslD, CrPa, AirBr and InSepThi. The diagonal plots exhibit the distributions of values of each lesion from the auxiliary task in distinguishing COVID-19 positive or negative. The off-diagonal axes display the distribution of each paired lesions over the two categories: COVID-19 or non-COVID-19. Not only are the GGO and consolidation lesions common in COVID-19 but they also frequently appear in non-COVID-19 cases (e.g., other types of pneumonia or lung cancer). As a result, both lesions are less powerful in helping to triage COVID-19 or non-COVID-19. In summary, the paired plots in Figure 6 corroborate that the three lesions of CrPa, AirBr, and InSepThi are more important factors than GGO and CslD in distinguishing COVID-19 from non-COVID-19.

## 5 CONCLUSION

To accommodate the urgent and enormous demands for COVID-19 testing, we develop a multi-lesion attention deep neural network for automating the COVID-19 diagnosis using richly annotated chest CT image data. The samples of our dataset are a combination of those collected from over 760 online preprints [6] and the newly added images from online publications. To overcome the limitation of the sample size, we extract two types of supervised information from the radiological text: 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 which forces the model to pay close attention to the five lesions of COVID-19 during the training process. It is worth emphasizing that the labels for five lesions are only required for training our LA-DNN model, which are not required for COVID-19 diagnosis of new testing samples. In other words,

our diagnostic system does not need annotation or segmentation of the lesions in the CT images by clinical experts.

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 AI-driven COVID-19 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 to further develop more accurate systems to defeat the COVID-19.

## ACKNOWLEDGEMENT

We would like to thank two referees for their helpful comments which greatly improved the paper. The research was supported by the Fundamental Research Funds for the Central Universities, NO. JBK1806002, and the Research Grants Council of Hong Kong, NO. 17307218.

## REFERENCES

- [1] Tao Ai, Zhenlu Yang, Hongyan Hou, Chenao Zhan, Chong Chen, Wenzhi Lv, Qian Tao, Ziyong Sun, and Liming Xia. 2020. Correlation of chest CT and RT-PCR testing in coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases. *Radiology* (2020), 200642.
- [2] A. Chattopadhyay, A. Sarkar, P. Howlader, and V. N. Balasubramanian. 2018. Grad-CAM++: Generalized Gradient-Based Visual Explanations for Deep Convolutional Networks. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*. 839–847.
- [3] Anthony Dangis, Christopher Gieraerts, Yves De Bruecker, Lode Janssen, Hanne Valgaeren, Dagmar Obbels, Marc Gillis, Marc Van Ranst, Johan Frans, Annick Demeyere, and Rolf Symons. 2020. Accuracy and reproducibility of low-dose submillisievert chest CT for the diagnosis of COVID-19. *Radiology: Cardiothoracic Imaging* 2, 2 (2020), e200196. <https://doi.org/10.1148/ryct.2020200196>
- [4] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li-Fei-Fei. 2009. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*. 248–255.
- [5] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.
- [6] Xuehai He, Xingyi Yang, Shanghang Zhang, Jinyu Zhao, Yichen Zhang, Eric Xing, and Pengtao Xie. 2020. Sample-Efficient Deep Learning for COVID-19 Diagnosis Based on CT Scans. *medRxiv* (2020). <https://doi.org/10.1101/2020.04.13.20063941>
- [7] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. 2017. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 4700–4708.
- [8] Xueyan Mei, Hao-Chih Lee, Kai-yue Diao, Mingqian Huang, Bin Lin, Chenyu Liu, Zongyu Xie, Yixuan Ma, Philip M Robson, Michael Chung, et al. 2020. Artificial intelligence-enabled rapid diagnosis of patients with COVID-19. *Nature Medicine* (2020), 1–5.
- [9] Feng Shi, Jun Wang, Jun Shi, Ziyang Wu, Qian Wang, Zhenyu Tang, Kelei He, Yinghuan Shi, and Dinggang Shen. 2020. Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for COVID-19. *IEEE Reviews in Biomedical Engineering* (2020).
- [10] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).
- [11] Ying Song, Shuangjia Zheng, Liang Li, Xiang Zhang, Xiaodong Zhang, Ziwang Huang, Jianwen Chen, Huiying Zhao, Yusheng Jie, Ruixuan Wang, et al. 2020. Deep learning Enables Accurate Diagnosis of Novel Coronavirus (COVID-19) with CT images. *medRxiv* (2020). <https://doi.org/10.1101/2020.02.23.20026930>
- [12] Mingxing Tan and Quoc V Le. 2019. Efficientnet: Rethinking model scaling for convolutional neural networks. *arXiv preprint arXiv:1905.11946* (2019).
- [13] Shuai Wang, Bo Kang, Jinlu Ma, Xianjun Zeng, Mingming Xiao, Jia Guo, Mengjiao Cai, Jingyi Yang, Yaodong Li, Xiangfei Meng, et al. 2020. A deep learning algorithm using CT images to screen for Corona Virus Disease (COVID-19). *MedRxiv* (2020). <https://doi.org/10.1101/2020.02.14.20023028>