Dataset

We used public CXR datasets, whose characteristics are summarized in Table I and Table II. In particular, the data in Table I are used for training and validation of the segmentation networks, since the ground-truth segmentation masks are available. The curated data in Table II are from some of the data in Table I as well as other COVID-19 resources, which were used for training, validation, and test for the classification network. More detailed descriptions of the dataset are follows.

TABLE I Segmentation Dataset Resources

Dataset	Class	#	Bits	Mask	
				Lung	Heart
Training					
JSRT/SCR [19] [20]	Normal/Nodule	197	12	O	O
Validation					
JSRT/SCR	Normal/Nodule	50	12	O	O
NLM(MC) [21]	Normal	73	8	O	-

TABLE II Classification Data Set Resources

Dataset	Class	#	Bits	Mask	
				Lung	Heart
JSRT/SCR	Normal	20	12	O	O
NLM(MC)	Normal	73	8	O	-
CoronaHack [22]	Normal	98	24	-	-
NLM(MC)	Tuberculosis	57	8	O	-
CoronaHack	Pneumonia (Bacteria)	21	24	-	-
	Pneumonia (Bacteria)	33	24	-	-
Cohen et al [23]	Pneumonia (Virus)	20	24	-	-
	Pneumonia (COVID-19)	180	24	-	-

1) Segmentation Network Dataset:

The JSRT dataset was released by the Japanese Society of Radiological Technology (JSRT). Total 247 chest posteroanterior (PA) radiographs were collected from 14 institutions including normal and lung nodule cases. Corresponding segmentation masks were collected from the SCR database. The JSRT/SCR dataset were randomly split into training (80%) and validation (20%). For cross-database validation purpose, we used another public CXR dataset: U.S. National Library of Medicine (USNLM) collected Montgomery Country (MC) dataset. Total 138 chest PA radiographs were collected including normal, TB cases and corresponding lung segmentation masks.

2) Classification Dataset:

The dataset resources for the classification network are described in Table II. Specifically, for normal cases, the JSRT dataset and the NLM dataset from the segmentation validation dataset were included. For comparing COVID-19 from normal and different lung diseases, data were also collected from different sources including additional normal cases. These datasets were selected because they are fully accessible to any research group, and they provide the labels with detailed diagnosis of disease. This enables more specific classification of pneumonia into bacterial and viral pneumonia, which should be classified separately because of their distinct clinical and radiologic differences.

In the collected data from the public dataset, over 80% was pediatric CXR from Guangzhou Women and Children’s Medical Center. Therefore, to avoid the network from learning biased features from age-related characteristics, we excluded pediatric CXR images. This is because we aim to utilize CXR radiography with unbiased age distribution for more accurate evaluation of deep neural networks for COVID-19 classification.

Total dataset was curated into five classes; normal, TB, bacterial pneumoia, viral pneumonia, COVID-19 pneumonia. The numbers of each disease class from the data set are summarized in Table III. Specifically, a total of 180 radiography images of 118 subjects from COVID-19 image data collection were included. Moreover, a total of 322 chest radiography images from different subjects were used, which include 191, 54, and 20 images for normal, bacterial pneumonia, and viral pneumonia (not including COVID-19), respectively. The combined dataset were randomly split into train, validation, and test sets with the ratio of 0.7, 0.1, and 0.2.

TABLE III Disease Class Summary of the Data Set

Dataset	Normal	Bacterial	Tuberculosis	Viral	COVID-19	Total
Training	134	39	41	14	126	354
Validation	19	5	5	2	18	49
Test	38	10	11	4	36	99
Total	191	54	57	20	180	502

3) Dataset for Comparison With COVID-Net:

We prepared a separate dataset to compare our method with existing state-of-the art (SOTA) algorithm called COVID-Net. COVID-19 image data collection was combined with RSNA Pneumonia Detection Challenge dataset as described in for a fair comparison between our method and COVID-Net. The reason we separately train our network with the COVID-Net data set is that RSNA Pneumonia Detection Challenge dataset provide only the information regarding the presence of pneumonia, rather than the detailed diagnosis of disease, so that the labels were divided into only three categories including normal, pneumonia, and COVID-19 as in Table IV. More specifically, there were 8,851 normal and 6,012 pneumonia chest radiography images from 13,645 patients in RSNA Pneumonia Detection Challenge dataset, and these images were combined with COVID-19 image data collection to compose a total dataset. Among these, 100 normal, 100 pneumonia, and 10 COVID-19 images were randomly selected for validation and test set, respecitvely as in. Although we believe our categorization into normal, bacterial, TB, and viral+COVID-19 cases is more correlated with the radiological findings and practically useful in clinical environment, we conducted this additional comparison experiments with the data set in Table IV to demonstrate that our algorithm provides competitive performance compared to COVID-Net in the same experiment set-up.

TABLE IV Dataset for Comparison With COVID-Net

Dataset	Normal	Pneumonia	COVID-19	Total
Training	8651	5812	160	14623
Validation	100	100	10	210
Test	100	100	10	210
Total	8851	6012	180	15043