Supplemental Material

We obtained a CNN previously trained using 22,000 frontal chest radiographs and radiologist-defined bounding boxes localizing regions affected by pneumonia from the 2018 RSNA pneumonia challenge1,2, which we refer to as the *initial CNN* algorithm. We propose a new strategy where we apply transfer-learning using more accurate ground truth labels to fine-tune the *initial CNN* and improve the accuracy of pneumonia localization. To provide stronger definition of ground truth than the external data set, we had subspecialty cardiothoracic radiologists generate pneumonia probability maps on an internally curated chest radiograph data set from patients who had both a chest radiograph and CT on the same day.

## Data Sources for Algorithm Development

We retrospectively obtained 1,466 paired frontal chest radiographs and chest CT scans performed on the same day from 1,163 patients at our institution as part of their routine clinical care from January to April of 2020. We refer to this as our *internal data set*.

In addition, two public sources of data were also obtained to complement our internal, matched cohort for algorithm development, which included the chest radiographs and bounding box annotations from the 2018 RSNA pneumonia challenge1,2 used to train the *initial CNN*. Additionally, 138 frontal chest radiographs and lung masks, which we refer to as the Montgomery data set3, were acquired to develop a background lung mask algorithm for calculation of fractional lung involvement.

## Radiologist Annotations

Pneumonia Probability Maps

Two sets of pneumonia annotations were used to improve the *initial CNN* algorithm. First, frontal chest radiographs from the internal *cohort* were each annotated with visually estimated probability maps. Annotations were performed by fellowship trained cardiothoracic radiologists with an average of 4.6 years (range 2-12 years) post-fellowship experience. A concurrently performed CT was displayed alongside the chest radiograph to provide support to the radiologist annotation. No additional clinical information was available to the radiologist.

Second, bounding box annotations from the external RSNA data set were converted to ellipsoid annotations as previously described4. Noting that the elliptical annotations frequently excluded the medial and lateral portions of the lung bases, we elected to re-annotate 518 randomly-sampled chest radiographs from the external RSNA data set that were previously marked as having pneumonia to include the lung bases by a radiology resident ([blinded]).

Lung Segmentations

To create a lung segmentation CNN, we randomly sampled 99 chest radiographs from the external data set. Annotations of the lungs were performed by the same resident ([blinded]). This data was combined with 138 chest radiographs from the Montgomery data set to build a lung segmentation CNN.

## Algorithm Training

Pneumonia Localization CNN

We employed a transfer learning strategy starting from the weights of the *initial CNN*. Data used for transfer learning was divided into cohorts of roughly 80% for training, while reserving the remaining 20% for validation. Patients were mutually exclusive to the training and validation cohorts.

With multiple data sources, we conducted a hyperparameter search balancing several factors, including the relative influence of external and internal data sets, and class imbalance with smaller foci of pneumonia. Specifically, we weighted the contribution of the larger number of cases in the external training data and the greater reliability of the internally curated training data by varying the number of samples used in each epoch of training as well as the balance of positive and negative cases. An additional pixel-wise weighting factor was used to improve the sensitivity for small foci of pneumonia. Choice of loss function was also varied for each run, and included mean squared error, Kullback-Leibler divergence, soft Dice, and binary focal loss. One-hundred and two candidate CNNs were trained from the weights of the *initial CNN* while varying the parameters mentioned above with a batch size of 12 radiographs optimized using the RMSProp algorithm with an initial fixed learning rate of 0.001 for the first 10 epochs with a 0.1 exponential decay for a total of 60 epochs. The CNN with the best joint AUC and Dice improvements on the validation data from the internal matched cohort was selected as the *updated CNN* and used for allsubsequent analyses.

Lung Segmentation CNN

An additional U-net CNN5 was separately trained to segment the left and right lung from chest radiographs. This lung segmentation CNN was trained with a combination of chest radiographs and annotations described above, using Dice loss for 25 epochs that was optimized using the RMSProp algorithm with an initial learning rate of 0.0005. Each lung was then evenly subdivided into upper, middle, and lower regions.

1. Wang X, Peng Y, Lu L, Lu Z, Bagheri M, Summers RM. ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases. *2017 IEEE Conf Comput Vis Pattern Recognition(CVPR)*. Published online 2017:2097-2106.

2. Shih G, Wu CC, Halabi SS, et al. Augmenting the National Institutes of Health Chest Radiograph Dataset with Expert Annotations of Possible Pneumonia. *Radiol Artif Intell*. 2019;1(1):e180041. doi:10.1148/ryai.2019180041

3. Jaeger S, Candemir S, Antani S, Wáng Y-XJ, Lu P-X, Thoma G. Two public chest X-ray datasets for computer-aided screening of pulmonary diseases. *Quant Imaging Med Surg*. Published online 2014. doi:10.3978/j.issn.2223-4292.2014.11.20

4. Hurt B, Yen A, Kligerman S, Hsiao A. Augmenting Interpretation of Chest Radiographs with Deep Learning Probability Maps. *J Thorac Imaging*. Published online 2020. doi:10.1097/RTI.0000000000000505

5. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Vol 9351. ; 2015:234-241. doi:10.1007/978-3-319-24574-4\_28