Supplemental Material

Hyperparameter Optimization

A variety of aspects for training a convolutional neural network were considered as hyperparameters; they were learning rate, optimizer, oversampling technique, and data augmentation as seen in Table S.1. The random hyperparameter optimization was employed given a predefined range for hyperparameters for each models (frontal CNN, lateral CNN, dual CNN) separately. Augmentation and all the other hyperparameters were also optimized during fine-tuning using the RUMC-real dataset with PFT labels.

Strong augmentation consists of brightness and contrast (factors were chosen uniformly from the range of 0.4 to 1.6), rotation (degrees range from -15 to 15), translation, and horizontal flip (probability=0.5). Weak augmentation consists of the rotation, horizontal flip, and translation (0.05 for vertical and 0.1 for horizontal maximum absolute fraction), and the degree of augmentations was reduced compared to strong augmentation. Moderate augmentation consists of the same augmentation techniques (and degree ranges) with the weak augmentation, but it additionally includes random resizing. During resizing, the aspect ratio was preserved, and the corresponding TLV reference standard was adjusted accordingly during training. If oversampling hyperparameter was chosen, the chest x-rays with low and high TLV labels, which were underrepresented in the dataset, were oversampled during training. The selected hyperparameters are shown in Table S.2

Table S.1: The ranges of the hyperparameters that were defined for the optimization procedure.

| Hyperparameter       | Range                      |
|----------------------|----------------------------|
| Learning rate        | $[10^{-5}, 0.001]$         |
| Optimizer            | [Adam, SGD]                |
| Data augmentation    | [strong, moderate, weak]   |
| Oversampling         | [Yes, No]                 |
Table S.2: The final hyperparameters that were selected for the best models on each dataset.

| Dataset                | Optimizer       | Data augmentation | Oversampling | Learning rate |
|------------------------|-----------------|-------------------|--------------|---------------|
| COPDGene-sim           | Adam            | weak              | no           | 0.00018       |
| RUMC-sim               | Adam            | weak              | yes          | 2.78x10⁻⁵     |
| RUMC-real (CT labels) | Adam & Adam     | weak & strong     | no & no      | 0.0001 & 0.0001|
| RUMC-real (PFT labels)| Adam            | strong            | yes          | 2.6x10⁻⁵      |

The selected hyperparameters of the best models in each experiment. The best model for the RUMC-real (CT labels) is an ensemble model, therefore frontal & lateral model hyperparameters are shown, respectively.

Evaluation Metrics

Mean absolute error (MAE) and mean absolute percentage error (MAPE) were used to evaluate the accuracy of the total lung volume prediction with respect to the reference standard. MAE is formulated as follows:

\[
    \text{MAE} = \frac{1}{N} \sum_{t=1}^{N} |y_t - \hat{y}_t|
\]

where \(N\) denotes the number of images, \(y_t\) represents the reference standard TLV, and \(\hat{y}_t\) represents the predicted TLV.

MAPE is formulated as follows:

\[
    \text{MAPE} = \frac{1}{N} \sum_{t=1}^{N} \left| \frac{y_t - \hat{y}_t}{y_t} \right|
\]