Data

We used a combined dataset of **9,397** cases from **3 publicly available datasets** corresponding **1,997 healthy** and **7,400 disease** images. From the VinBigData ChestXray dataset, we used all cases corresponding to ***"Lung Opacity"*** (2,483) and randomly sampled 1,500 cases from the healthy class; all data from the Brixia Score COVID-19 dataset corresponding to the CR modality (2,861) and all cases (2,553) from the MIDRC mRALE Mastermind Challenge. The combined dataset was split into train (**7,517**) and validation (**1,880**) and stratified on the target class.

Image preprocessing

Raw CXR images were pre-processed by cropping them to a square size of 512 x 512 pixels. Then, image intensities were normalized by the mean and standard deviation of the training set. We performed data augmentation during training by flipping, rotating, with a given probability. We also used color augmentation.

Model and Training

We used an inherently interpretable model (denseBagNet) which model the local evidence for the presence of disease as part of its architecture. The BagNet is an implicitly patch-based model based on bag-of-local features that aggregates local evidence from interpretable heatmaps to make predictions. It takes a two-dimensional input which is implicitly split into many small, overlapping patches (size q=33x33 pixels corresponding to the size of the model's effective receptive field), which are independently processed in parallel to compute the local evidence for the presence of lunch opacity. The patchwise predicted local evidence values are combined into a single class evidence map corresponding to a downsampled version of the input image, which then is aggregated using average pooling and passed through a softmax function to output a probability distribution.

The model support screening not only with the final prediction but also shows the model's internal decision-making process through the provided inherent class evidence map which highlights the contribution of small local regions to the final prediction. The low-resolution evidence map is then upsampled to the full image resolution. In contrast to post-hoc saliency map based methods, the class evidence map provided by the dense BagNet is a transparent part of the decision-making process and faithfully captures the local evidence without any posthoc processing. In addition, the model is trained at image level without patch-level annotations, and produces fine localization.

The best model was saved on the basis of the best **weighted log loss score** calculated on the 14 annotated "practical" cases provided for the callibration stage.