DualChexNet: A View-Specific Approach to Chest Pathology Classification

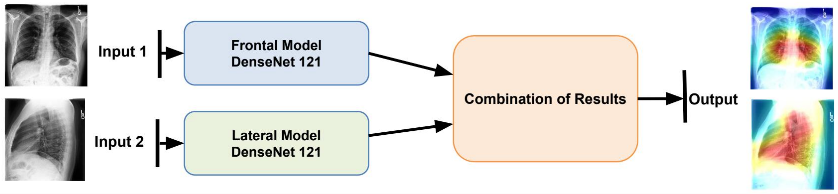
Michael Yan, Ying Chang, Yu Xin Ang
Stanford University

Motivation

Automating chest radiograph interpretation can provide many benefits, ranging from better workflow prioritization to increased screening accessibility in third-world countries without trained radiologists. However, current deep learning models for chest x-ray interpretation either do not make use of different x-ray views or train a single model for all view types [1]. The purpose of this project is to investigate whether combining output from Convolutional Neural Networks specifically trained on different x-ray view types (frontal and lateral) would perform better than a single network on all the views.

Model Architecture

We trained two DenseNet models separately, one for frontal images and one for lateral images, with multi-class cross-entropy loss. Thereafter, we incorporated results from the two models for our final output using two methods: (1) using the average probabilities of frontal and lateral (2) using domain knowledge of the diseases from research.



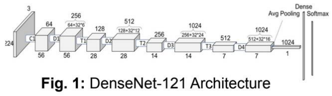


Fig. 1: DenseNet-121 Architecture

DenseNet-121 Hyperparameters

- *Batch Size*: 4
- *Epochs*: 10
- *Learning Rate:* 0.0001 (with decay)
- *Optimizer*: Adam
- Pretrained on ImageNet

Dataset

Dataset Specifications:

- Stanford CheXpert Dataset collected from Stanford Hospital chest radiographic exams.
- 223,414 chest radiographs of 64,540 patients
- 14 classes of common chest pathologies
- 32,387 out of 223,414 radiographs are lateral view images (14.4% of total)

Data Split:

- Our development and test set contained only patients with *both* frontal and lateral images (22,418 out of 64,540).
- We performed a **70/20/10** split for train/dev/test. We include the patients with only frontal or only lateral images in the training set.

	Training	Validation	Test	Total
Frontal	166,573	16,003	8,451	191,027
Lateral	22,592	6,387	3,408	32,387
Total	189,165	22,390	11,859	223,414

References

[1] Irvin et al., (2019): *CheXpert: A Large Chest Radiograph Dataset with Uncertainty Labels and ExpertComparison.CoRRabs/1901.07031*
[2] Pranav et al., (2017): *CheXNet:Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning.CoRRabs/1711.05225*
[3] Rubin et al., (2018): Large Scale Automated Reading of Frontal and Lateral Chest X-Rays using Dual Convolutional Neural Net-works.CoRRabs/1804.07839.

Results

The frontal column shows the AUROC values of the frontal model only. The next two columns show the results from utilizing inputs from the two DenseNets into a final layer. The average column takes the average of the frontal probabilities and lateral probabilities output by the model so as to use both views in its prediction. From our research on pathologies, we find the most appropriate X-ray views for each pathology, and use the appropriate method for each pathology, which thus gives us our research column. The ROC curves of our highest performing pathologies are shown below.

Table 3: AUROC Values for Test Set

	Frontal	Average	Research	Rubin DualNet
No Finding	0.861	0.621	**0.861**	0.758
Enlarged Cardiomediastinum	0.614	**0.883**	0.614	-
Cardiomegaly	0.877	0.734	**0.877**	0.840
Lung Opacity	0.729	0.706	**0.729**	-
Lung Lesion	0.701	0.894	**0.894**	-
Edema	0.887	0.787	**0.887**	-
Consolidation	0.779	0.723	**0.779**	0.632
Pneumonia	0.721	**0.785**	0.721	0.625
Atelectasis	0.776	0.905	**0.905**	0.766
Pneumothorax	0.891	0.921	**0.921**	0.706
Pleural Effusion	0.903	0.741	**0.903**	0.757
Pleural Other	0.748	0.675	**0.748**	0.687
Fracture	0.646	0.879	**0.879**	-
Support Devices	0.863	**0.876**	0.863	-
Average Test AUROC	0.785	0.789	**0.823**	0.721

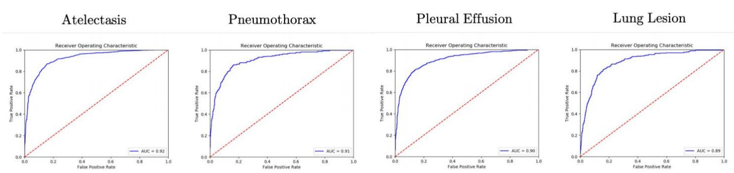


Fig. 2: Operating characteristics for our "Research" model's highest performing pathologies

Discussion

We examine two methods in improving our performance: (1) adding uncertainty labels and (2) using multiple view types in our prediction.

1. Adding Uncertainty

- Uncertainty labels improve the model performance for both Frontal and Lateral view points as shown in the table below.

Table 2: AUROC Values for Test Set

	Frontal	Frontal (w uncertainty)	Lateral	Lateral (w uncertainty)
AUROC	0.773	0.785	0.716	0.761

2. Using Multi-view Approaches

- Average approach incorporating both viewpoints performs better than the Frontal viewpoint only approach.
- Research-based selection approach yields the highest AUROC score.

Finally, our model under the Frontal, Average, and Research-based Selection approaches all have higher AUROC scores than *Rubin et al. (2018)*, which used combined viewpoints as well.

Future Work

- Utilize the DualNet architecture to fully capitalize on the multiple view types during the training process
- Obtain radiologist approval for our research based selection