Chest Pathology Classification using a Multimodel, View-Specific Approach

Shorya Sharma
ss118@iitbbs.ac.in

Abstract: Automating chest radiograph interpretation using deep learning models are highly beneficial as it increases screening accessibility in countries with a lack of radiologists. Current deep learning models do not incorporate both uncertainty labels from the training dataset and different chest X-ray views. In our improved model architecture, we train two DenseNet-121 models on frontal and lateral views separately. Thereafter, we incorporate both viewpoints by averaging the output probabilities for frontal and lateral models, and draw on existing radiology research on viewpoint conventions to determine usage of either frontal or average outputs. We find that the research-based selection approach performs best, yielding an AUROC of 0.823, compared to 0.785 for the frontal model, and 0.789 for the average model.

1 Introduction

Automating chest radiograph interpretation provides 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. The purpose of this paper is to investigate whether combining outputs from Convolutional Neural Networks specifically trained on different view types (frontal and lateral) would perform better than a single network on all the views. The outputs of each model will be combined using a research-guided approach using the optimal viewpoint for that specific pathology class for the final diagnostic prediction.

2 Related Work

2.1 Early Approaches

One of the most well-known papers of classifying diseases based on chest X-rays is that of Rajpurkar et al. (2017). The authors utilized the 121-layer DenseNet CNN architecture to create, CheXNet, a model that could detect pneumonia from chest X-rays on a level that exceeds practicing radiologists. This was extended to give predictions for 14 other diseases. Using the F1 metric, it was found that CheXNet exceeds average radiologist performance.

2.2 Incorporating Labels from Radiology Reports

In a recent paper by Irvin et al. (2019), the model takes as input a single-view chest radiograph and outputs the probability of each of the 14 observations. When more than one view is available, models

output the maximum probability of the observations across the views. The paper also investigated different approaches towards incorporating labels (label extraction from radiology reports) into the training process, using 1 for positive, 0 for negative and u for uncertain. However, they did not make use of multiple views (AP/PA/lateral). In our paper, we hope to use different views for better performance. For example, cardiomegaly can only be accurately assessed in PA images, as the AP view will exaggerate the heart silhouette due to magnification.

2.3 Incorporating Multi-View Images

Rubin et al. (2018) trained separate CNNs based on view positions. They constructed a DualNet architecture whereby both frontal and lateral CXR inputs are processed by separate CNNs and their outputs are combined into a fully connected layer to make a final classification. They used a novel dataset, MIMIC-CXR, and used labelling from radiology reports using NegBio. In our paper, we use a similar architecture which utilises both views but will apply it to the CheXpert dataset with different NLP labels.

3 Data

3.1 Dataset

This project utilized the Stanford CheXpert Dataset collected from Stanford Hospital chest radiographic examinations, performed between October 2002 and July 2017 in both inpatient and outpatient centers, along with their associated radiology reports.

The CheXpert dataset consists of 223,414 chest radiographs of 64,540 patients with 14 classes of labels for common chest radiographic observations, listed as follows: *No Finding, Enlarged Cardiomediastinum, Cardiomegaly, Lung Lesion, Lung Opacity, Edema, Consolidation, Pneumonia, Atelectasis, Pleural Effusion, Pleural Other, Fracture,* and *Support Devices*. Within our dataset, 32,387 out of 223,414 (14.4% of total) radiographs are lateral view images.

3.2 Data Split

Table 1: Dataset Split

	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

In contrast to prior works on the CheXpert dataset, we trained separate models by category of radiograph viewpoints since we want to evaluate whether datasets with both frontal and lateral views will improve the performance of the network. As such, our development and test set contains only patients with both frontal and lateral images. Furthermore, to ensure that a patient's X-ray images do not appear in both training and dev/test set, we grouped the images by patient and sorted them into train/dev/test. Of the patients with both frontal and lateral images (22,418 out of 64,540), we randomly split them by 70/20/10

for train/dev/test. The rest of the patients with only frontal or only lateral images were included in the training set. Overall, the dataset split is balanced in terms of distribution of diseases.

4 Model Architecture and Training Procedure

4.1 Baseline Model Architecture

This project centers on an open source Pytorch replication of Stanford University's CheXNet project which was built using the DenseNet121 architecture. The model feeds size 256×256 pixel images into the network and uses an Adam optimizer with the binary cross entropy loss function.

4.2 Improved Model Architecture

We trained two DenseNet121 models separately, one for frontal view positions and one for lateral view positions, with multi-class cross-entropy loss to support the usage of uncertainty labels.

From these two output sets, we treat the data by two separate methods: (i) Averaging Frontal and Lateral (ii) Research-based Selection. In our first method (i), we average the output probabilities for Frontal and Lateral viewpoints. In our second method (ii), we draw on existing radiology research on viewpoint conventions to determine usage of either frontal or average outputs.

4.3 Training Procedure and Hyperparameters

We used a batch size of 4 to train the two DenseNet models on the CheXpert dataset. We used learning rate decay, starting at an initial learning rate of 0.0001. Furthermore, we trained over 10 epochs and used Adam optimization. The DenseNet models were pretrained on ImageNet.

Even though uncertainty labels exist in the data, in the test phase, we calculated AUROC on a predicted output dataset which predicts only positive or negative for each pathology. To do this, we apply a filter that only includes classes that have certainty labels (either negative or positive, excluding uncertainty classes). We then apply a Softmax to scale the 0 and 2 labels to a [0,1] range.

4.4 Domain Research

We provide a closer explanation of our aforementioned research-based approach here. From expert interviews and a broad survey of the medical literature, frontal images are the predominant viewpoint used for chest x ray diagnoses. Lateral images are typically used as a secondary input given special cases unique to characteristics of the pathology. By this criteria, ten out of fourteen classes use the Frontal viewpoint probability outputs. Notably, studies have shown that the addition of the lateral chest radiograph to the frontal view did not improve the sensitivity or specificity of diagnosis in pneumonia (Lynch T et al. 2004). The four classes for which we used an average probability output of the frontal and lateral viewpoints are the following: (i) Lung Lesion, (ii) Atelectasis, (iii)Pneumothorax, and (iv) Fracture.

(i) For lung lesions, multiple studies have noted the importance of lateral views. The relevance of the lateral view is emphasized by Chotas et al., who found that frontal view represent only 73.6% of the total lung volume and 57% of the total lung area that is not obscured by the superimposed anatomical structures(Chotas et al. 1994).

(ii) One of the prominent symptoms of Atelactasis is loss of lung volume, which a lateral viewpoint is helpful in illustrating. In tutorials of medical imaging diagnosis for Atelactasis, lateral imaging is widely cited (Chandrasekhar, 2015).

(iii) Pneumothorax is closely associated with displaced pleural line. We incorporate the lateral view for Pneumorthorax as studies have cited that the lateral viewpoint provided helpful information to supplement the frontier viewpoint with visibility into the pleural line (Glazer HS et al. 1989).

(iv) According to Khoriati et al. (2013), "Lateral chest radiograph is considered the gold standard for making the diagnosis, because fracture and displacement or dislocation occurs in the sagittal plane."

5 Findings

Our main finding is that incorporating uncertainty labels and multi-views improves model performance. The research-based selection approach does the best, with the highest AUROC of 0.823, compared to 0.789 for the average model, and 0.785 for the solely frontal views model. These AUROC are all comparable to that of Rubin et al. (2018) DualNet's model, which has an AUROC of 0.721.

5.1 Analysis

We examine the effectiveness of adding uncertainty labels and using multi-view approach in improving model performance.

Firstly, incorporating uncertainty labels improves the performance of our model. Table 2 presents the per class AUROC scores calculated on the test set for models trained by Frontal with Certain Labels (Positive / Negative), Frontal with Uncertainty Labels (Positive / Uncertain / Negative), Lateral with Certain Labels, Lateral with Uncertainty Labels. For both Frontal and Lateral view types, the sets with uncertainty Labels had a higher average AUROC score compared to their corresponding Certain Labels-only view types. We therefore observe that uncertainty labels improve model performance for both frontal and lateral view points.

Table 2: AUROC Values for Test Set

	Frontal	Frontal (w uncertainty)	Lateral	Lateral (w uncertainty)
No Finding	0.872	0.861	0.821	0.842
Enlarged Cardiomediastinum	0.612	0.614	0.545	0.568
Cardiomegaly	0.851	0.877	0.799	0.850
Lung Opacity	0.719	0.729	0.694	0.712
Lung Lesion	0.747	0.701	0.620	0.685
Edema	0.850	0.887	0.802	0.856
Consolidation	0.723	0.779	0.711	0.759
Pneumonia	0.737	0.721	0.639	0.670
Atelectasis	0.691	0.776	0.705	0.760
Pneumothorax	0.846	0.891	0.773	0.856
Pleural Effusion	0.870	0.903	0.883	0.911
Pleural Other	0.778	0.748	0.626	0.679
Fracture	0.650	0.646	0.578	0.648
Support Devices	0.868	0.863	0.833	0.857
Average Test AUROC	0.773	**0.785**	0.716	**0.761**

Secondly, using the multi-view approach improves AUROC score. Table 2 presents the per class AUROC scores calculated on the test set for models trained by Frontal with Uncertainty Labels and Lateral with Uncertainty Labels. From these two test output sets, we further treat the data by two separate methods: (i) Averaging Frontal and Lateral (ii) Research-based Selection, which were elaborated upon in section 4.2. As shown by table 3, we thus have three sets of data outputs for performance review: (1) Frontal only (2) Average of frontal and lateral (3) Research-based selection. We find that our Average approach which incorporates both viewpoints yields an average test AUROC of 0.789 across the 14 pathologies, higher than that of the frontal viewpoint only approach of 0.785. However, our research-based selection approach yielded the highest AUROC score of 0.823 (see Table 3).

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

Finally, our model under the Frontal, Average, and Research-based Selection approaches all attained higher AUROC scores than the *Rubin et al. (2018)* paper, which used combined viewpoints as well. The AUROCs shown for *Rubin et al. (2018)* were from their DualNet architecture. **Figure 1** below includes ROC curves for the highest performing classes in our research-guided selection class.

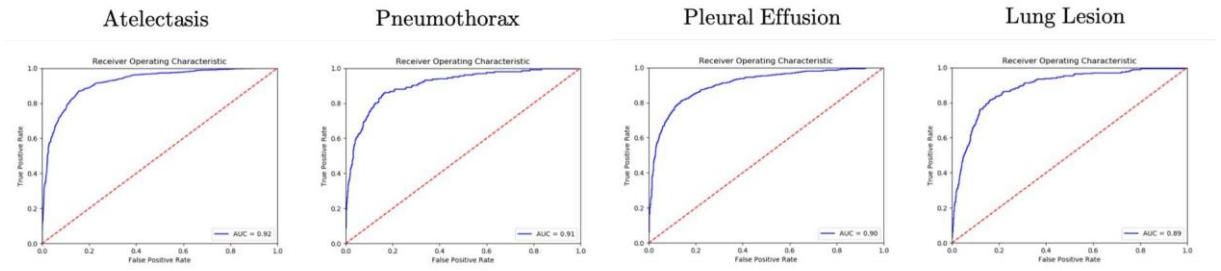

Figure 1: The highest performing classes in the research category.

5.2 Limitations

The model did not have access to other patient information such as gender, age, or patient history as inputs, which may help improve the accuracy of the diagnosis. The model also did not separate the PA and AP view-types of the frontal view. We focused on frontal and lateral viewpoint comparisons, but perhaps a three-viewpoint approach of PA, AP and lateral will allow for even better performance for some of the classes. No statistical tests were performed to evalaute model performance.

6 Conclusions

In our paper, we train our model using uncertainty labels and incorporate both frontal and lateral chest X-ray views. We find that adding uncertainty labels per pathology (in addition to positive / negative) improves model performance. Furthermore, both multi-view approaches increase AUROC performance. The research-based selection approach does best, yielding an AUROC of 0.823, compared to 0.785 for the frontal model and 0.789 for the average model.

6.1 Future Work

To improve accuracy even further, we hope to utilize the DualNet architecture to fully capitalise on the multiple view types during the training process. It would be good to connect both baseline models with a few fully-connected layers to optimise the weights on the output probabilities from the frontal and lateral models for final diagnosis. Moreover, we hope to obtain radiologist approval for our research-based selection.

References

[1] H. G. Chotas & C. E. Ravin (1994): *Chest Radiography: Estimated Lung Volume and Projected Area Obscured by the Heart, Mediastinum, and Diaphragm*. Radiology 193(2), p. 403404, doi:10.1148/radiology.193.2.7972752.

[2] H.S. Glazer, D.J. Anderson, B.S. Wilson, P.L. Molina & S.S. Sagel (1989): *Pneumothorax: Appearance on Lateral Chest Radiographs*. Radiology 173(3), p. 707711, doi:10.1148/radiology.173.3.2813774.

[3] Jeremy Irvin, Pranav Rajpurkar, Michael Ko, Yifan Yu, Silviana Ciurea-Ilcus, Chris Chute, Henrik Marklund, Behzad Haghgoo, Robyn L. Ball, Katie Shpanskaya, Jayne Seekins, David A. Mong, Safwan S. Halabi, Jesse K. Sandberg, Ricky Jones, David B. Larson, Curtis P. Langlotz, Bhavik N. Patel, Matthew P. Lungren & Andrew Y. Ng (2019): *CheXpert: A Large Chest Radiograph Dataset with Uncertainty Labels and Expert Comparison*. CoRR abs/1901.07031. Available at `http://arxiv.org/abs/1901.07031`.

[4] Abraham M. Ittyachen, Anuroopa Vijayan & Megha Isac (2017): *The Forgotten View: Chest X-ray - Lateral view*. Respiratory Medicine Case Reports 22, p. 257259, doi:10.1016/j.rmcr.2017.09.009.

[5] Al-Achraf Khoriati, Ramyah Rajakulasingam & Rakhee Shah (2013): *Sternal Fractures and Their Management*. Journal of Emergencies, Trauma, and Shock 6(2), p. 113, doi:10.4103/0974-2700.110763.

[6] A R Oconnor & W E Morgan (2005): *Radiological review of pneumothorax*. BMJ 330(7506), p. 14931497, doi:10.1136/bmj.330.7506.1493.

[7] Pranav Rajpurkar, Jeremy Irvin, Kaylie Zhu, Brandon Yang, Hershel Mehta, Tony Duan, Daisy Ding, Aarti Bagul, Curtis Langlotz, Katie Shpanskaya, Matthew P. Lungren & Andrew Y. Ng (2017): *CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning*. CoRR abs/1711.05225. Available at `http://arxiv.org/abs/1711.05225`.

[8] Jonathan Rubin, Deepan Sanghavi, Claire Zhao, Kathy Lee, Ashequl Qadir & Minnan Xu-Wilson (2018): *Large Scale Automated Reading of Frontal and Lateral Chest X-Rays using Dual Convolutional Neural Networks*. *CoRR* abs/1804.07839. Available at `http://arxiv.org/abs/1804.07839`.

[9] Chandrashekhara Sohoni (2015): *Peripheral Wedge-Shaped Radiographic Lung Opacity in a Young Patient*. *Lung India* 32(2), p. 184, doi:10.4103/0970-2113.152651.