DeepBreath: An Automated Chest X-Ray Diagnostic Tool

Lisa Ishigame (lih11@stanford.edu)[1], Andrea Oviedo (aoviedob@stanford.edu)[1]
[1]Department of Mechanical Engineering
CS 230: Deep Learning

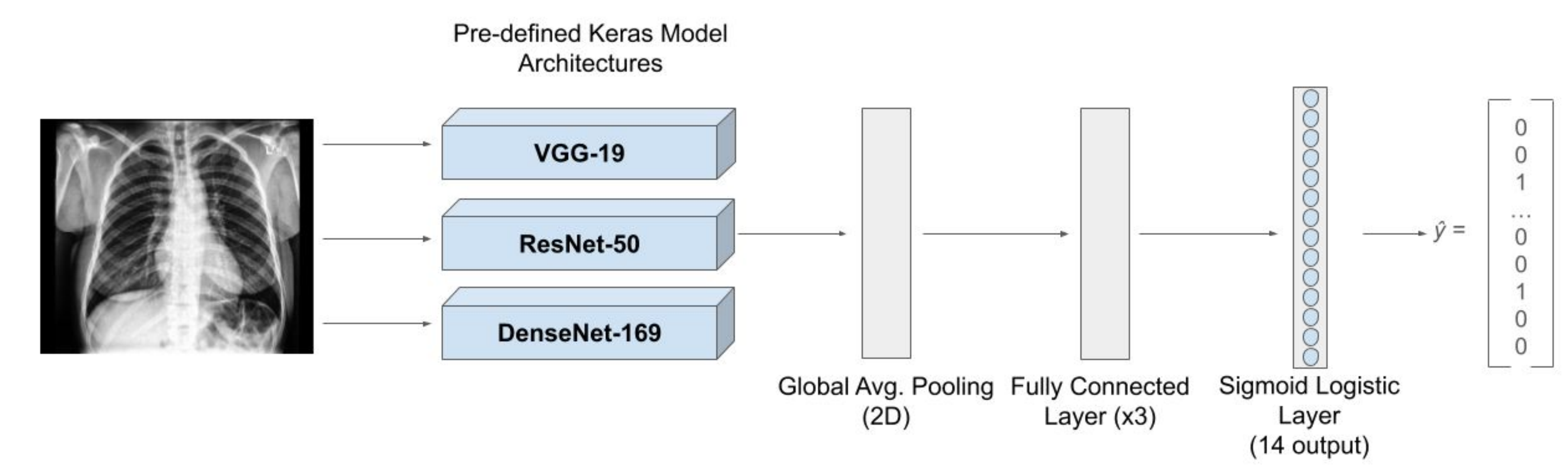
Motivation

- Chest x-rays are one of the most widely used radiological examinations for diagnosing and detecting various thoracic diseases.
- Developing a tool that can help radiologists achieve more accurate diagnoses could have tremendous benefits for under-resourced hospitals and clinics.
- In this project, we present various deep learning architectures that can accurately detect for the presence of 14 different observations from both frontal and lateral chest x-ray scans.
- From the explored architectures, the DenseNet-169 model performed the best, achieving a test accuracy of 86%.

Data & Features

- We used a chest radiograph dataset published by Stanford's Machine Learning group, collected by Stanford Hospital..
- It includes 223,648 chest x-rays of frontal and lateral views of 65,240 patients.
- The dataset is labeled with the presence of 14 common observations: *No Finding, Enlarged Cardiomediastinum, Cardiomegaly, Lung Lesion, Lung Opacity, Edema, Consolidation, Pneumonia, Atelectasis, Pleural Effusion, Pleural Other, Fracture, and Support Devices.*
- We changed all of the uncertain labels to positive labels, randomized the order of the dataset, and divided it into 80%/10%/10% training, validation and test sets.
- In 3 of our models, we considered only the frontal images, and in the best performing model, we considered both the frontal and lateral images.

Models



- Deep model architectures explored: VGG, ResNet, and DenseNet.
- Final layer of all networks were modified to include a global average pooling layer, three densely connected layers, and a final logistics layer of 14 outputs with sigmoid activation for predicting presence of different thoracic conditions.
- Binary cross entropy function was minimized using Adam optimizer.

Results & Discussion

Table 1: Performance assessment of deep models

Model	Test Accuracy	Training Examples (#)	Test Examples (#)
VGG-19 (F)	0.84870	152,983	19,123
ResNet-50 (F)	0.84386	152,983	19,123
DenseNet-169 (F)	0.86099	152,983	19,123
DenseNet-169 (F&L)	0.85976	178,918	22,365

Table 2: Precision, Recall and F1 scores for DenseNet-169 (F)

	precision	recall	f1-score	support
No Finding	0.09	0.06	0.07	1675
Enlarged Cardiomediastinum	0.07	0.00	0.00	2003
Cardiomegaly	0.16	0.08	0.11	3163
Lung Opacity	0.52	0.63	0.57	9966
Lung Lesion	0.00	0.00	0.00	795
Edema	0.31	0.26	0.28	6154
Consolidation	0.12	0.00	0.00	3749
Pneumonia	0.11	0.01	0.01	2083
Atelectasis	0.31	0.14	0.19	6025
Pneumothorax	0.11	0.05	0.07	2031
Pleural Effusion	0.46	0.43	0.44	8681
Pleural Other	0.00	0.00	0.00	458
Fracture	0.01	0.00	0.00	780
Support Devices	0.58	0.57	0.57	10887
micro avg	0.45	0.33	0.38	58450
macro avg	0.20	0.16	0.17	58450
weighted avg	0.36	0.33	0.33	58450
samples avg	0.39	0.31	0.32	58450

- As expected, DenseNet-169 had best performance out of the three explored model architectures.
- F-1 scores for certain conditions, like Lung Opacity and Support Devices, were also high.

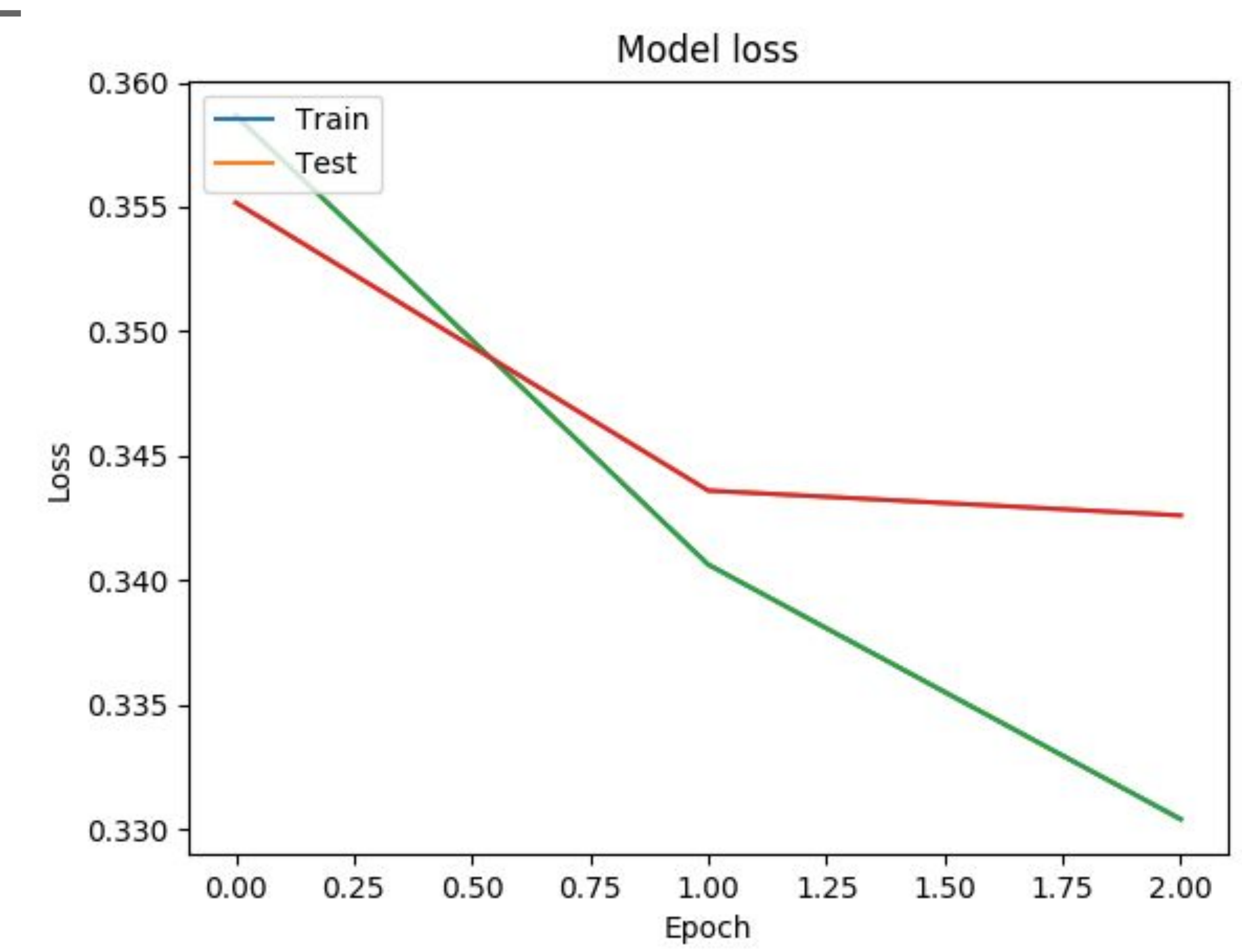


Figure 1: Loss curve for DenseNet-169

Conclusion & Future Work

- Out of the models presented in this project, Dense-169 had the highest accuracy, with 86% accuracy on the test set.
- The same model trained on both frontal and lateral images did not exhibit higher performance, probably because of a smaller dataset of lateral images.
- We are unsure why some categories exhibited lower scores, but we hypothesize that it was due to an unbalanced training dataset in the categories.
- Future work could include data augmentation on categories with fewer data, such as "Enlarged Cardiomediastinum", and lateral images.
- Future work could also include training on images with higher resolution, training for longer epochs, and tuning hyperparameters.

References

- https://github.com/lih11/DeepBreath
- Chexpert: A large dataset of chest x-rays and competition for automated chest x-ray interpretation. https://stanfordmlgroup.github.io/competitions/chexpert/
- Pranav Rajpurkar, Jeremy Irvin, Kaylie Zhu, Brandon Yang, Hershel Mehta, Tony Duan, Daisy Ding, Aarti Bagul, Curtis Langlotz, Katie Shpanskaya, et al. Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. arXiv preprint arXiv:1711.05225, 2017.
- Xiaosong Wang, Yifan Peng, Le Lu, Zhiyong Lu, Mohammadhadi Bagheri, and Ronald M Summers. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 2097–2106, 2017.