Deep Learning Applications in Medical Imaging

Presenter: John Lawless

Data Problem

Can a machine learning algorithm be trained to interpret radiographs?

- A study of the potential for convolutional neural networks to classify findings in x rays
- Understanding possibilities, pitfalls, and limitations to this approach

Image from Fierce Biotech Article:https://www.fiercebiotech.com/medtech/bayer-signs-ai-imaging-platform-pact-blackford-analysis

Purpose of Study

- Diagnostic x rays are rapidly increasing in use
- There is a current and projected worsening physician shortage
- Radiologists are burdened with large demand for interpretation
- Burnout + burden = increased risk of missed/incorrect diagnoses

This month (Dec 2020), Bayer contracted the analytics team Blackford Analysis to develop deep learning applications in radiology

NOT the Purpose of this Study

1. It cannot be overstated that there is ***no replacement for the trained eyes of a radiologist***

2. Radiologists combine patient history, image analysis, and experience

3. An algorithm is limited to differentiating structures on an image

4. There are ethical issues of fully entrusting patient health and safety to an algorithm

Potential for Clinical Application

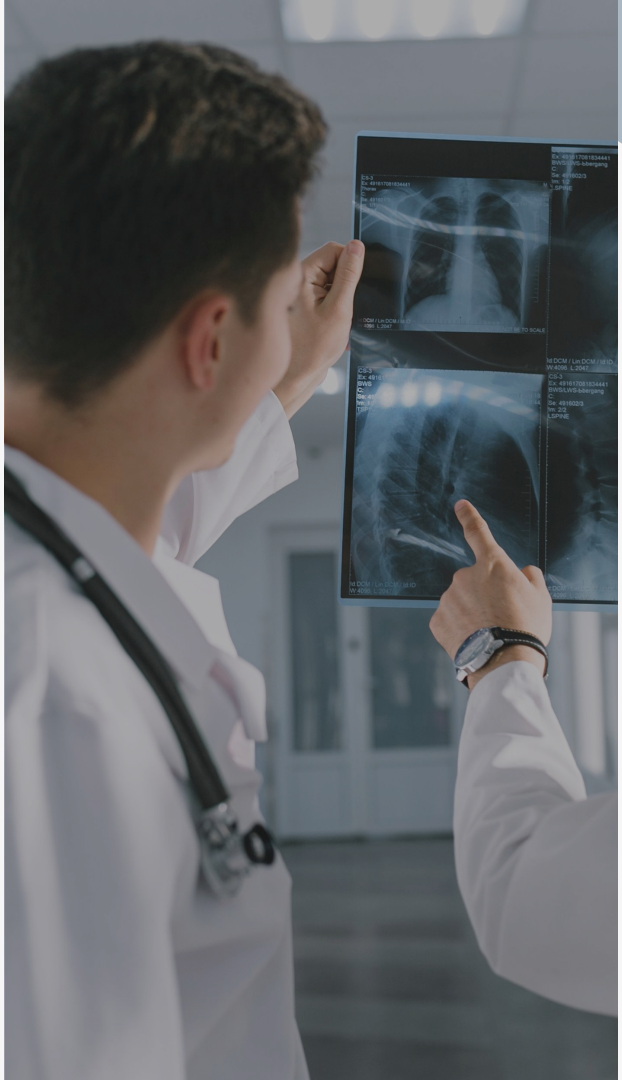
1. Despite limitations, convolutional neural networks can potentially pick up on very subtle changes in an image that human eyes cannot distinguish

2. Combining experienced radiologists with machine learning applications can greatly improve the quality of patient care in medicine

3. Example: a model could read x rays prior to the radiologist interpreting them and “flag” likely cases of disease or positive finding

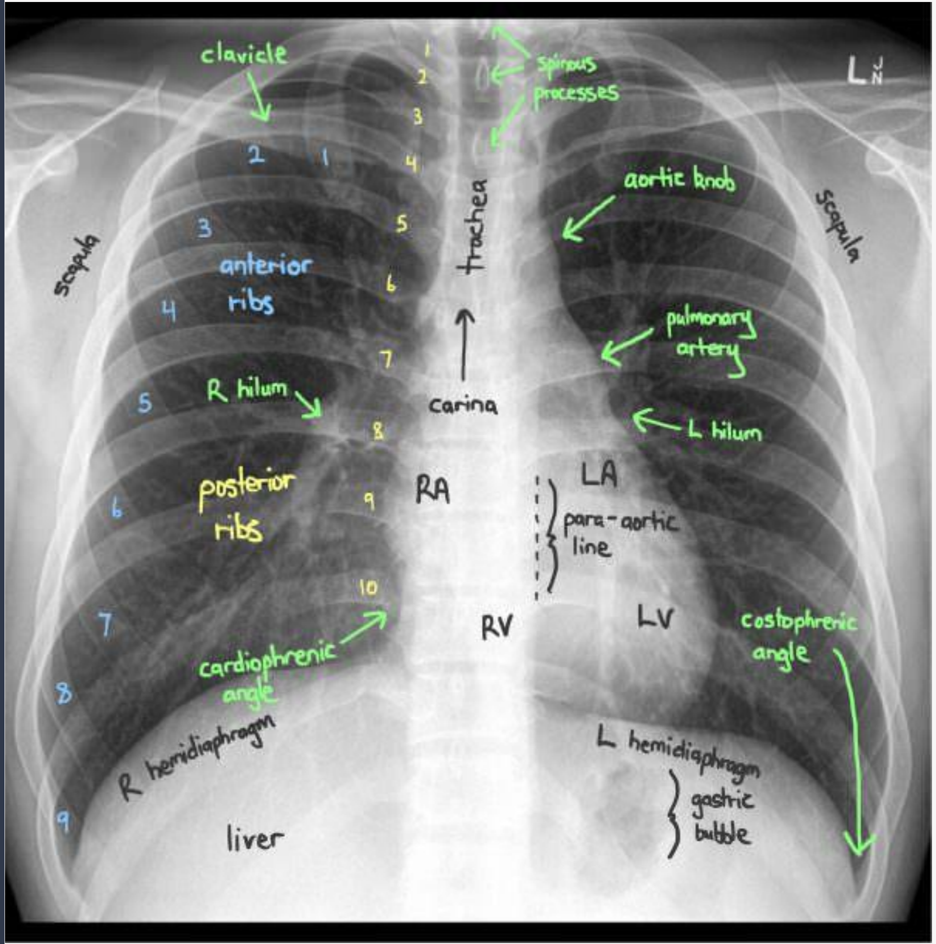
Data Used in this Study

- All X ray images collected from the NIH Clinical Center
- Data consists of 121,120 chest X Rays from over 30,000 patients (45 GB)
- Image findings text mined from radiologist reports (estimated >90% accuracy)
- Image classes are either “No Finding” or among 14 different positive findings - may have more than 1



Analysis of Findings

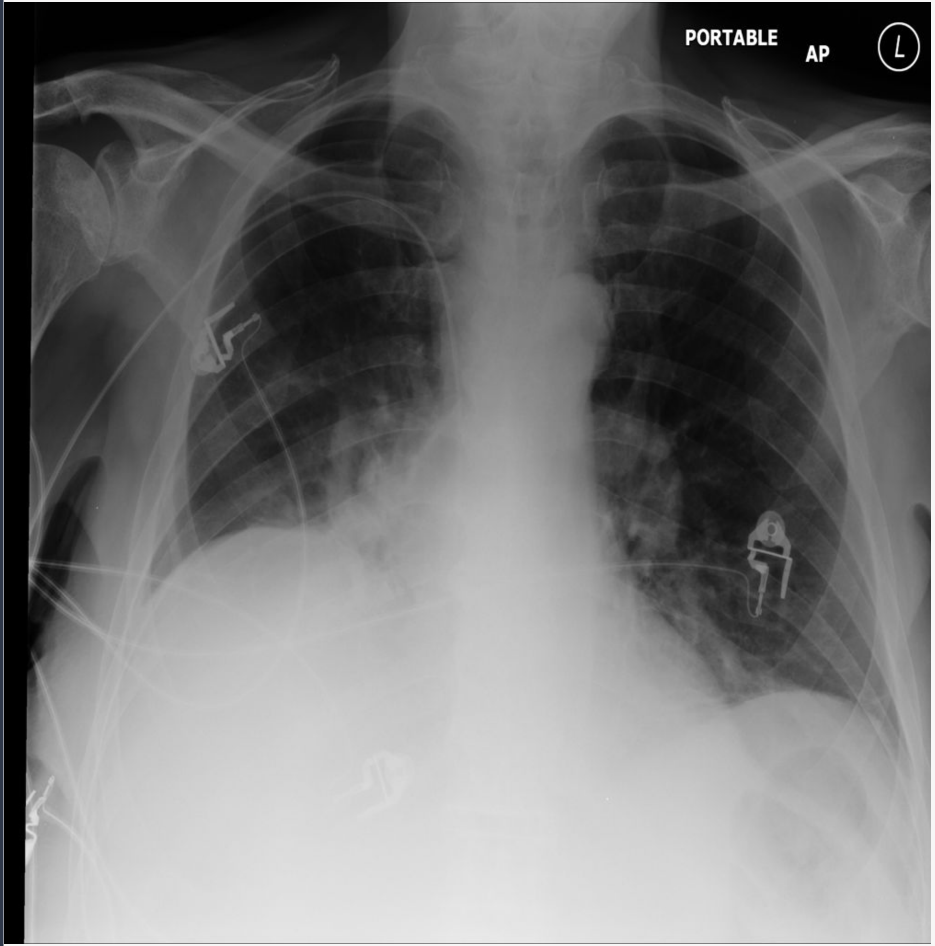
Finding	Total Count	Total Ratio
Atelectasis	11559	10.31%
Cardiomegaly	2776	2.48%
Consolidation	4667	4.16%
Edema	2303	2.05%
Effusion	13317	11.88%
Emphysema	2516	2.24%
Fibrosis	1686	1.50%
Hernia	227	0.2%
Infiltration	19894	17.74%
Mass	5782	5.16%
No Finding	60361	53.84%
Nodule	6331	5.65%
Pleural Thickening	3385	3.02%
Pneumonia	1431	1.28%
Pneumothorax	5302	4.73%



Clinical Background

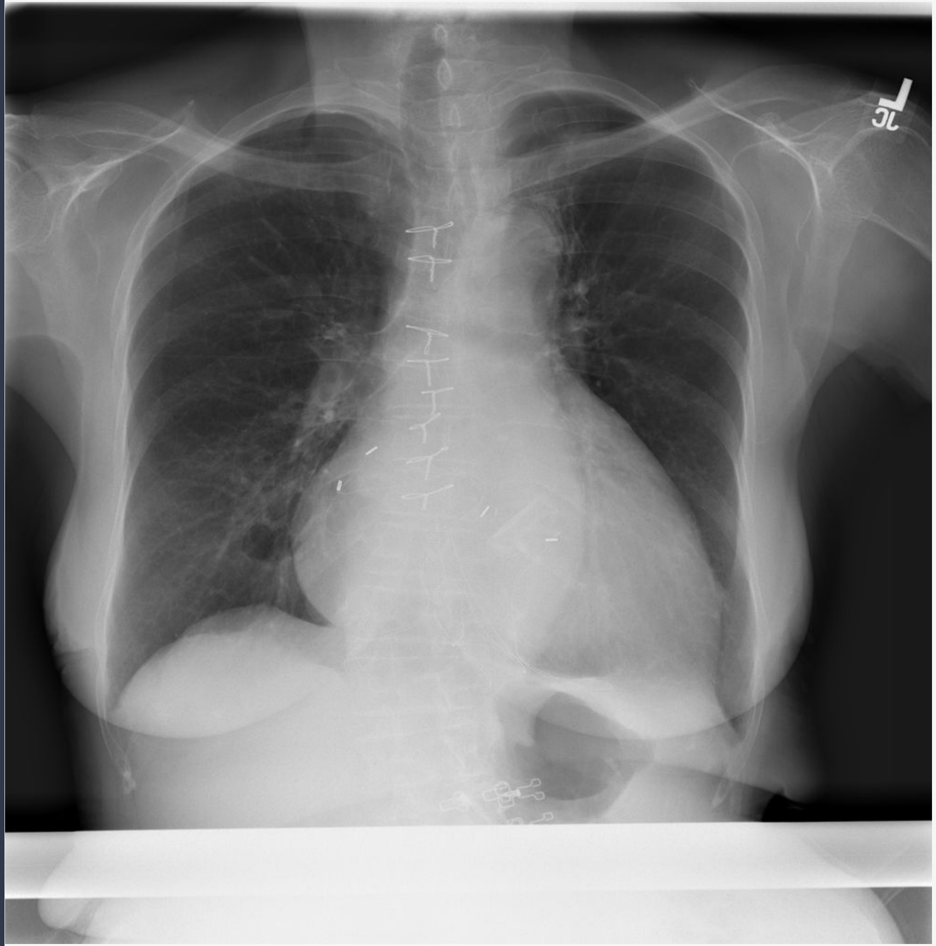
Example of a “Normal” patient

Note how subtle some findings we see will present!



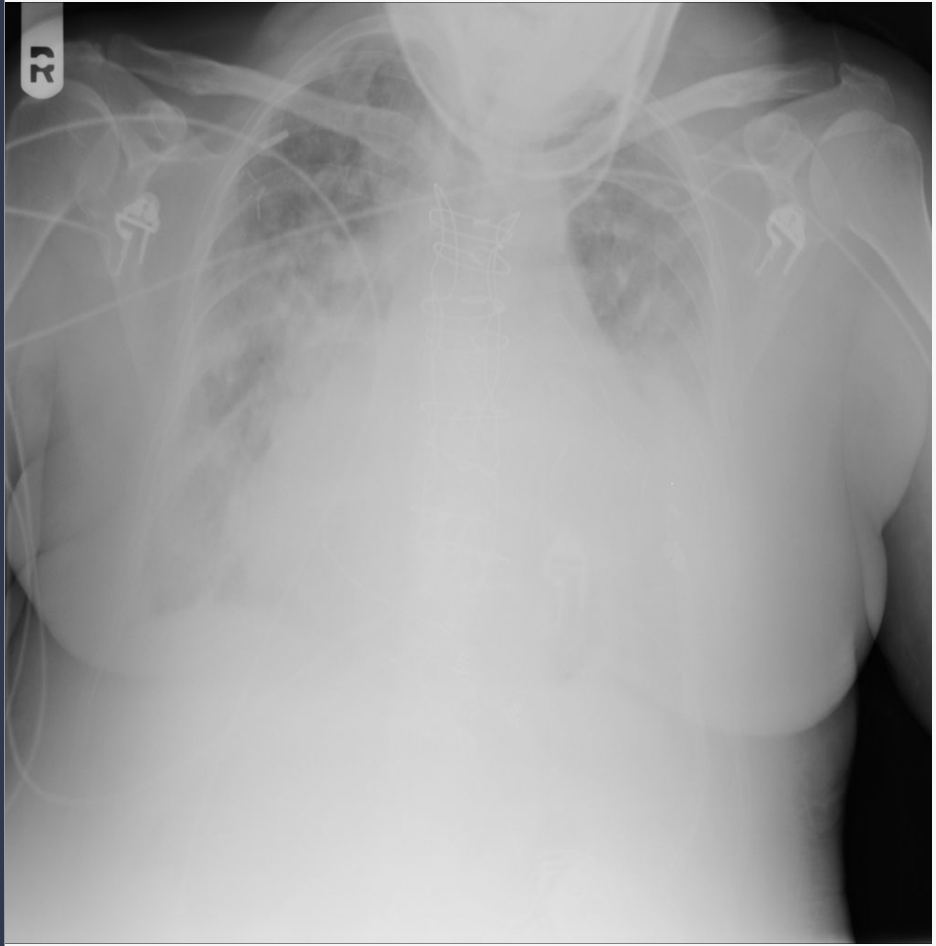
Atelectasis

- Inability to fully inflate lungs
- Common temporarily after anesthesia



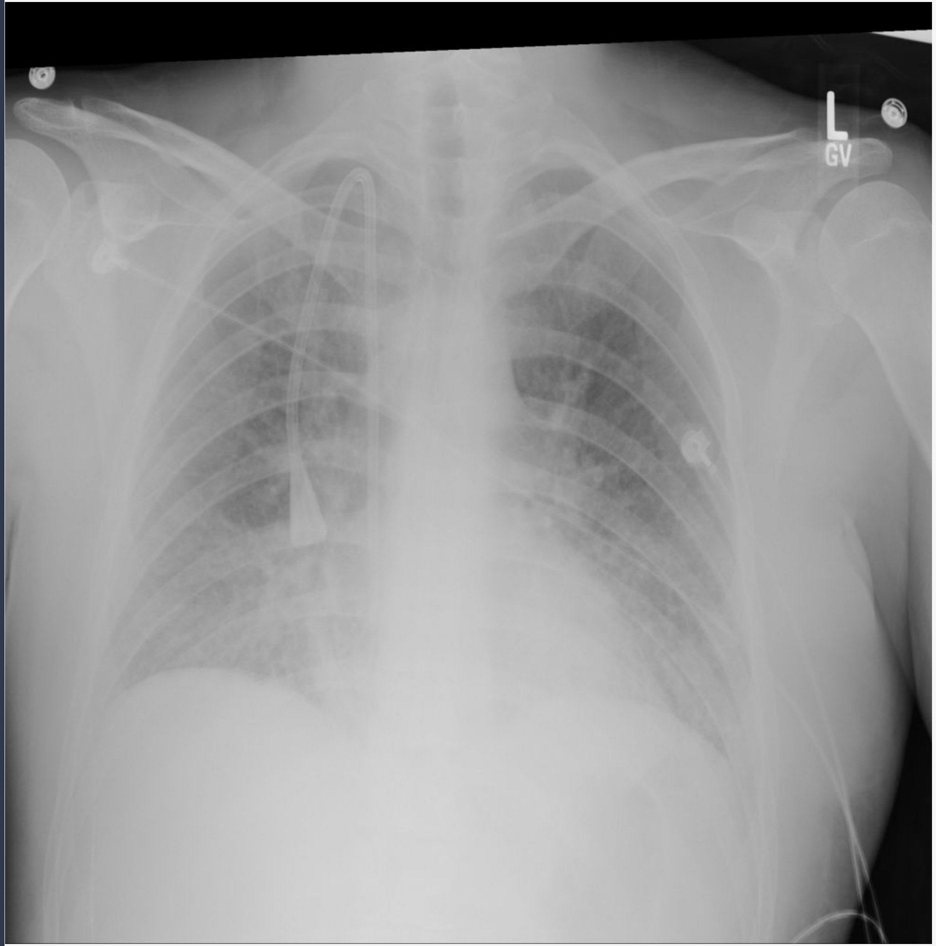
Cardiomegaly

- Enlarged heart
- Some causes: congestive heart failure, hypertension, infection



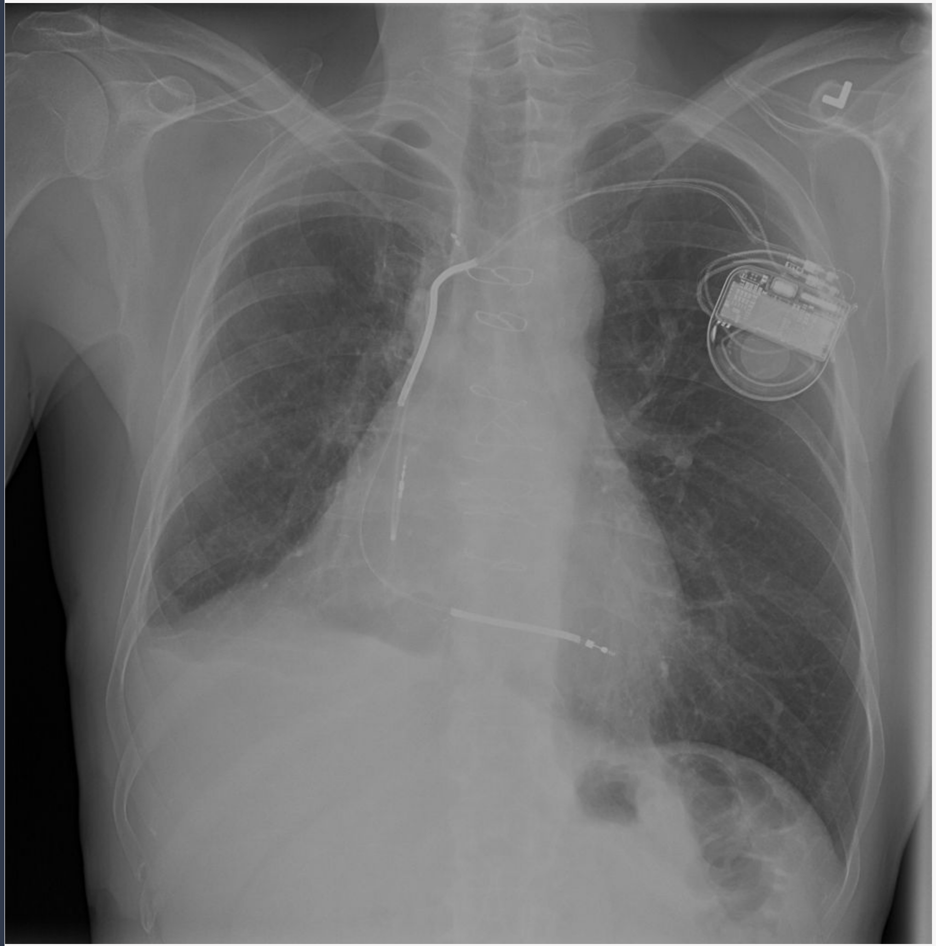
Consolidation

- Very general term: space normally filled with air is instead filled with another substance
- Usually replaced with more specific finding when causes are known
- Note poor image quality may contribute to general label



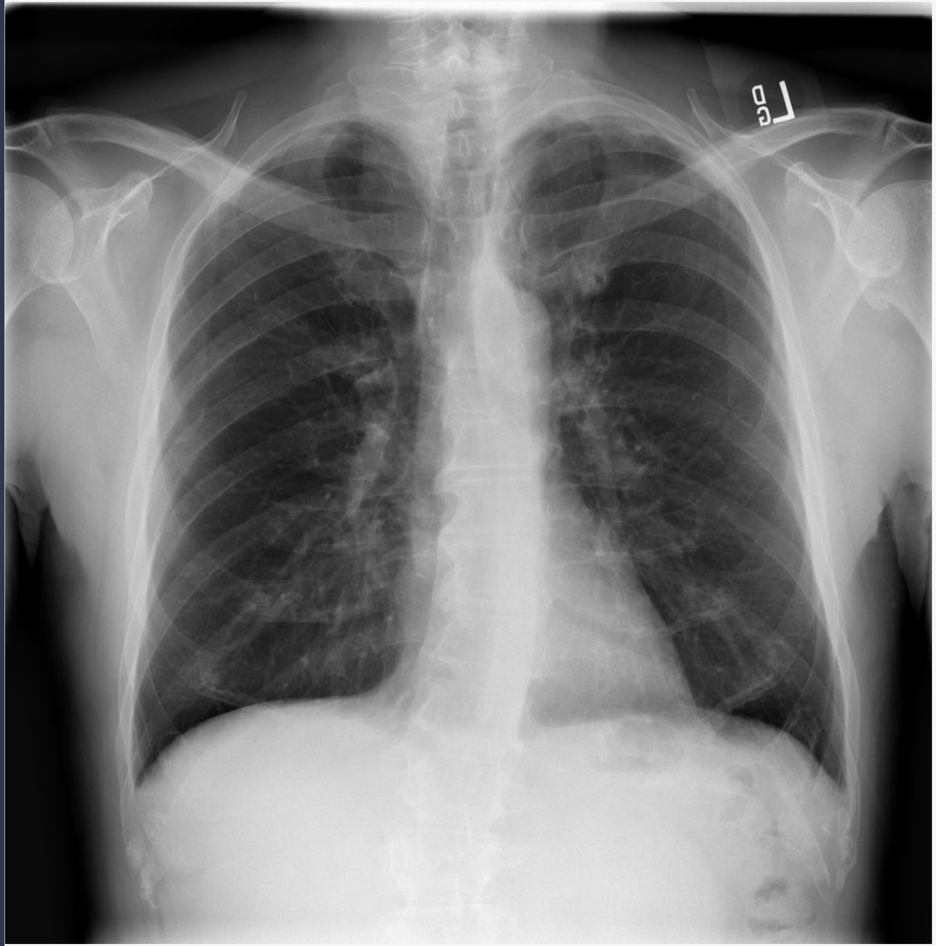
Edema

- Excess fluid in lungs
- Note similarity to previous image!



Effusion

- Excess fluid around lungs
- In pleura - the lining between lungs and chest wall

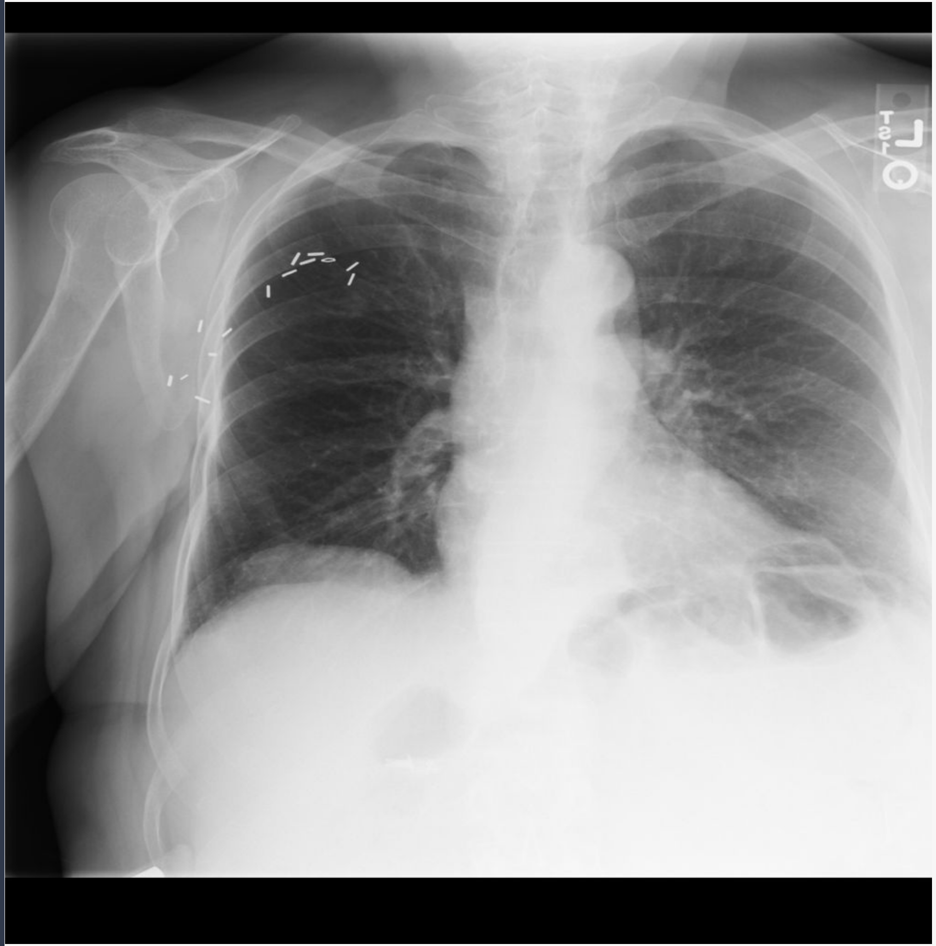


Emphysema

- Weakening/rupturing of air sacs in the lungs
- Most commonly caused by smoking or exposure to heavily polluted air

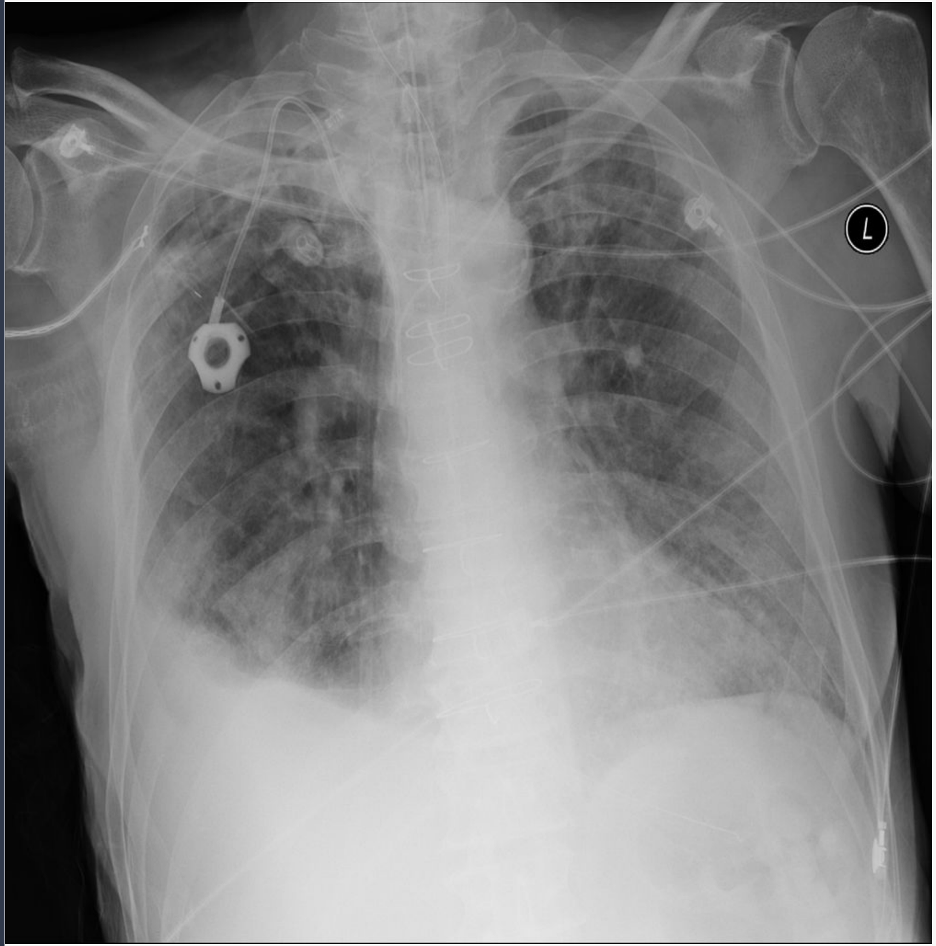
Fibrosis

- Scarring of lung tissue
- Can be caused by severe illness and infection
- Seen rarely in severe COVID 19 cases



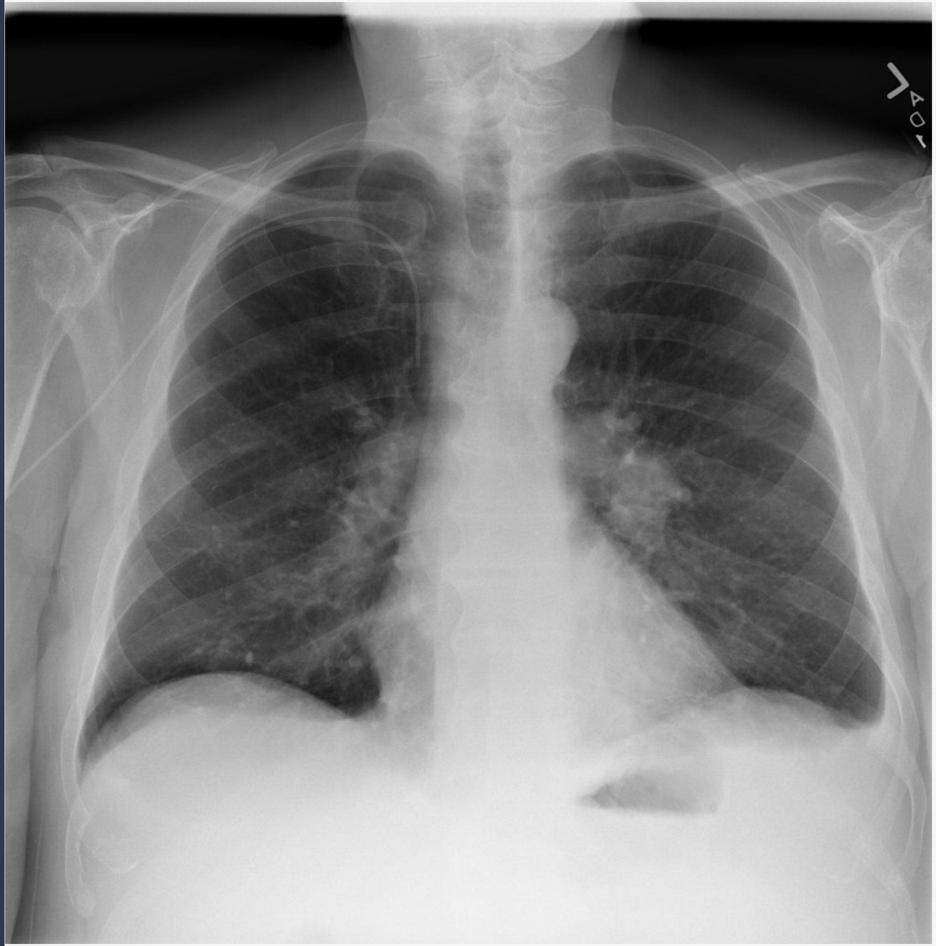
Hernia

- Term for an organ protruding into unusual space
- Chest x rays show hiatal hernia



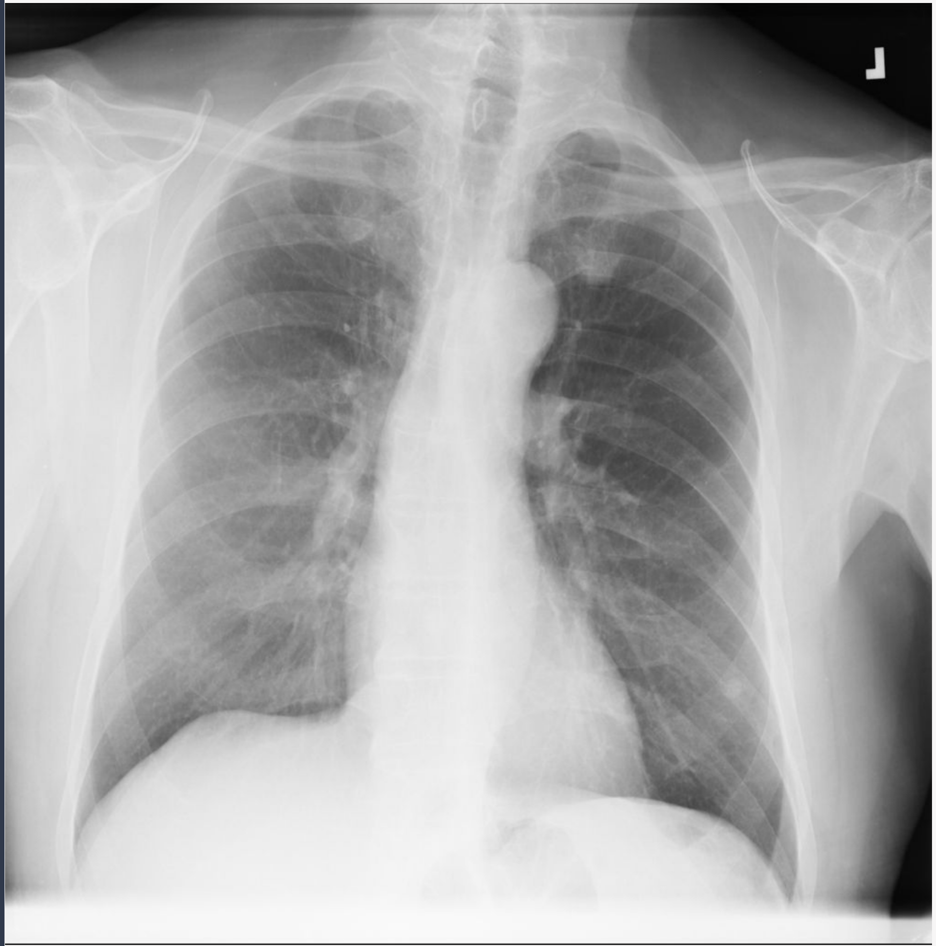
Infiltration

- General term for the presence of abnormal substance in lungs
- VERY similar in appearance and definition to other findings



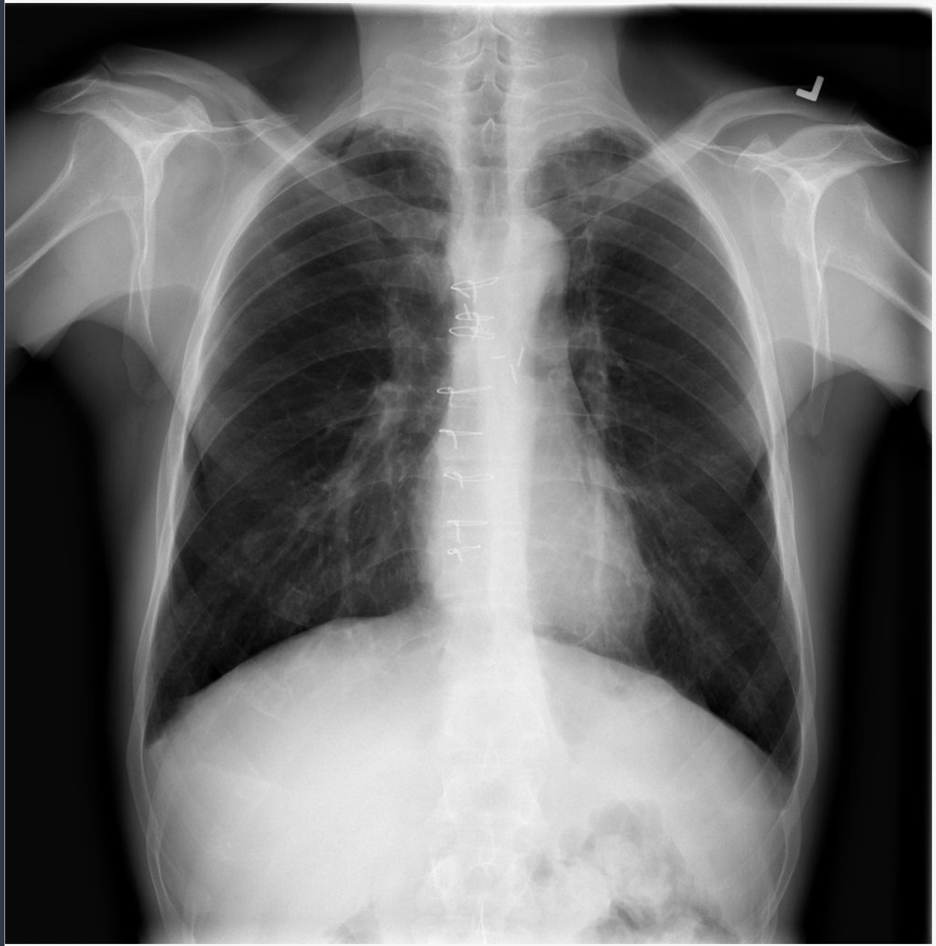
Mass

- Defined as an abnormal growth of over 3cm size
- Indication of cancer
- Difficult to identify and localize



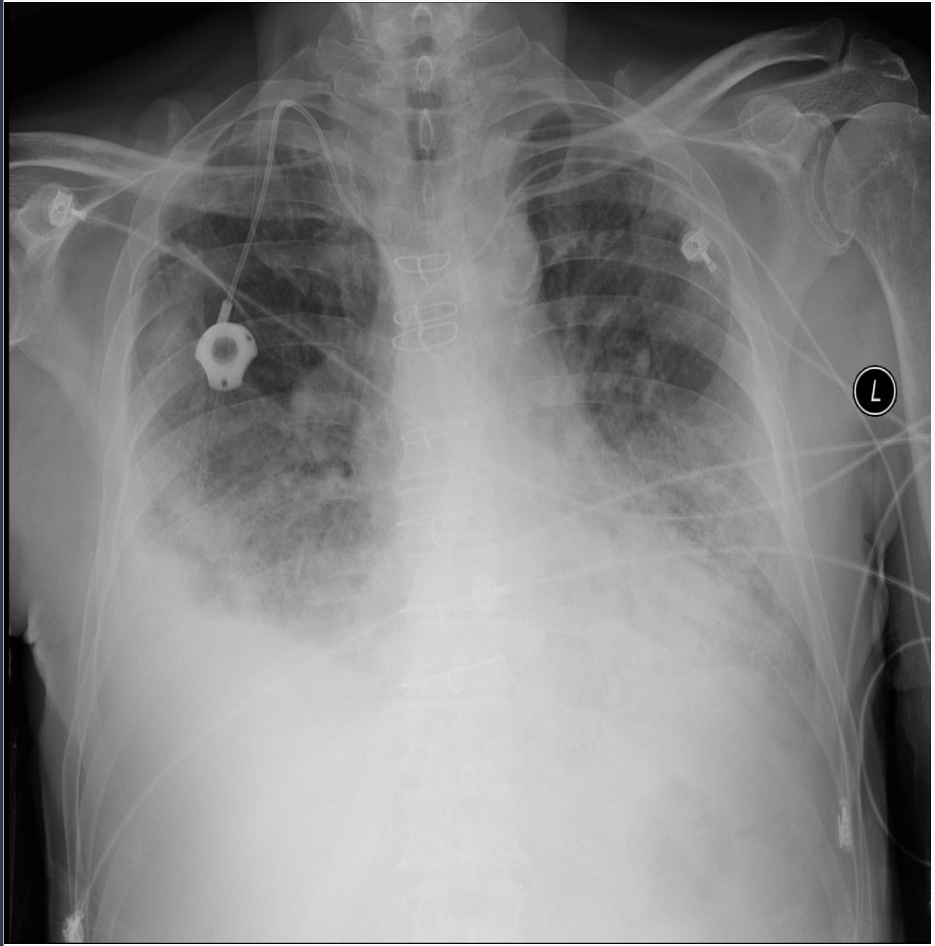
Nodule

- Defined as an abnormal growth of under 3cm size
- Can indicate cancer, most often benign causes
- Extraordinarily difficult to identify and localize



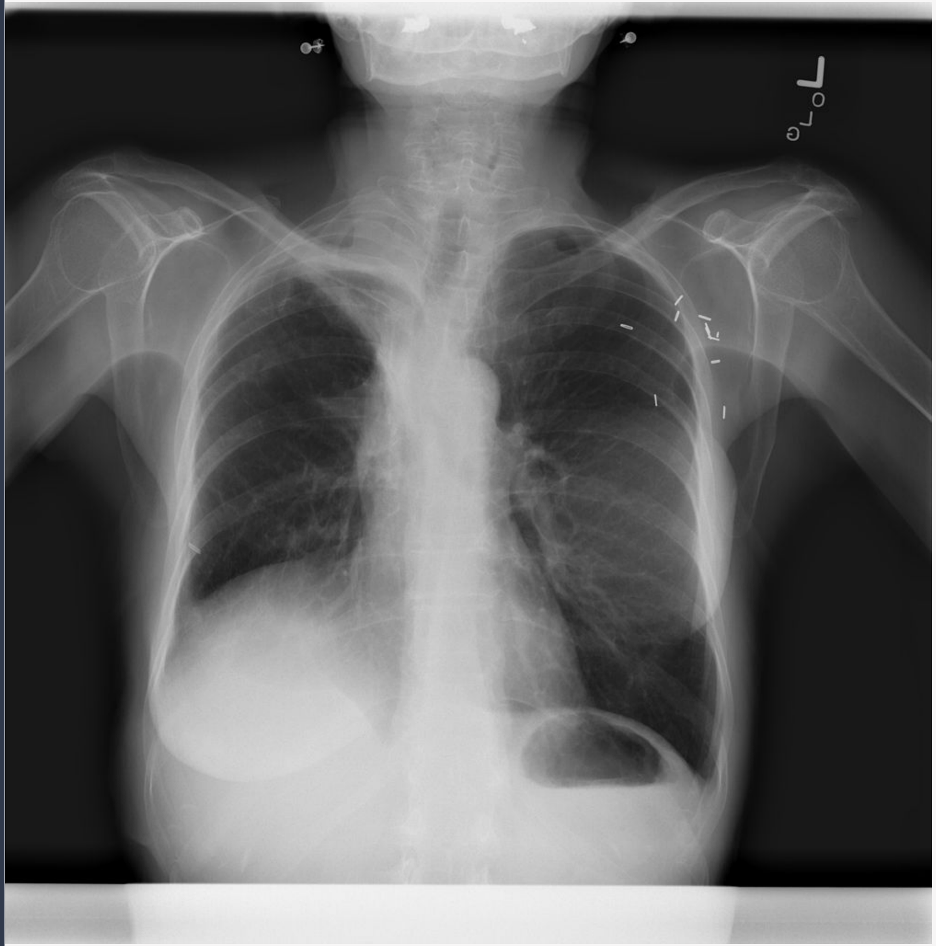
Pleural Thickening

- Pleura normally not visible on an x ray
- Visibility implies thickening
- Can indicate malignant pleural mesothelioma, a rare cancer caused by asbestos exposure



Pneumonia

- Presence of infection that inflames the lungs
- Highly life threatening in the elderly and patients with co morbidities
- A type of infiltration

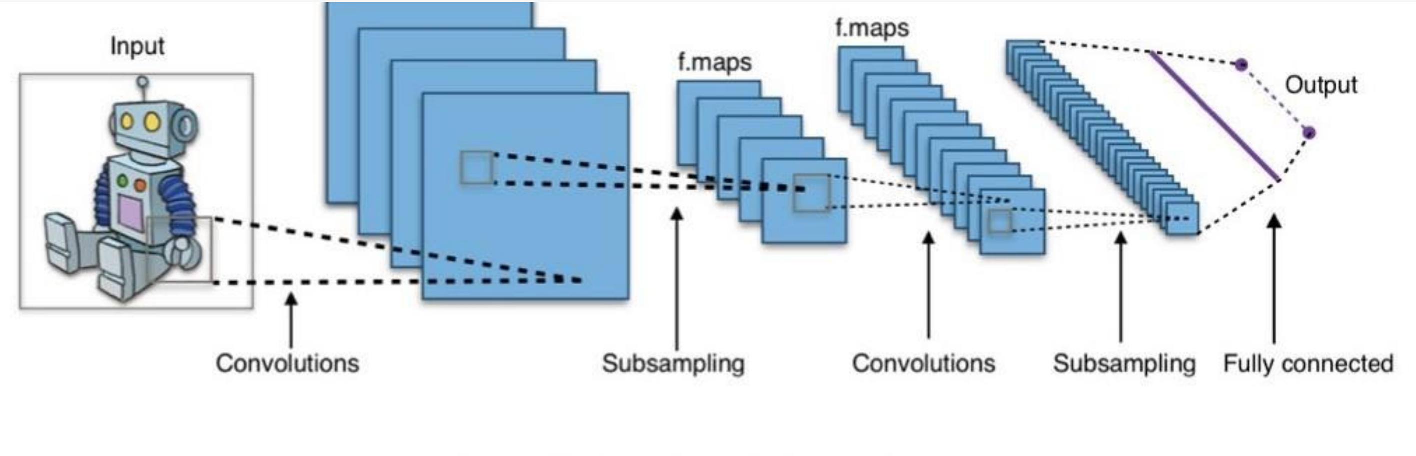


Pneumothorax

- Proper medical term for a collapsed lung
- Caused by pressure outside of lung pressing it closed
- Can resolve on its own, but must be monitored as it can quickly become life threatening

Modeling the Data

- Images were passed into a Convolutional Neural Network built with Tensorflow's Keras API
- These *convolutions* simplify the image data, extracting relevant features to be classified at the network's output layers



Transfer Learning

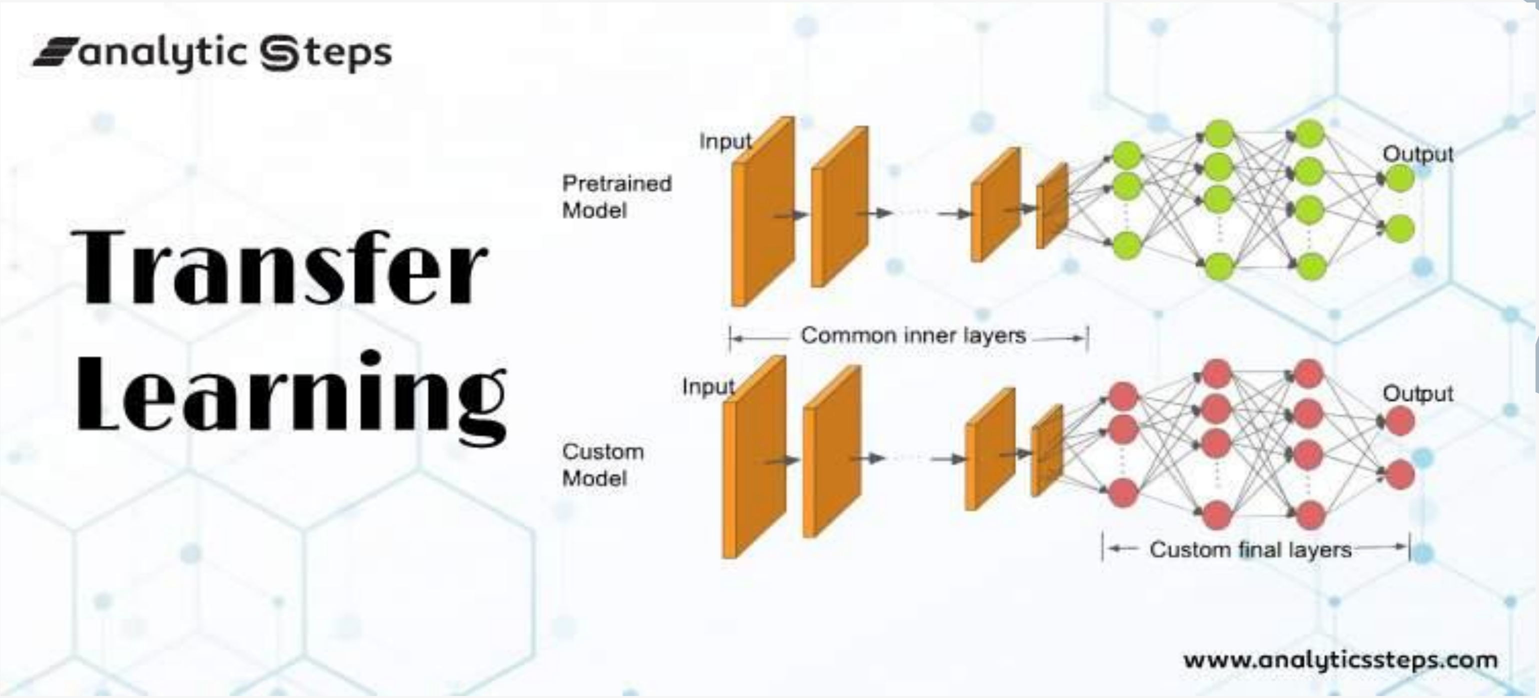
A key factor in this project was utilizing transfer learning

These models use weights that were pre-trained on millions of random images

Classification is achieved by adding layers after the transfer layers

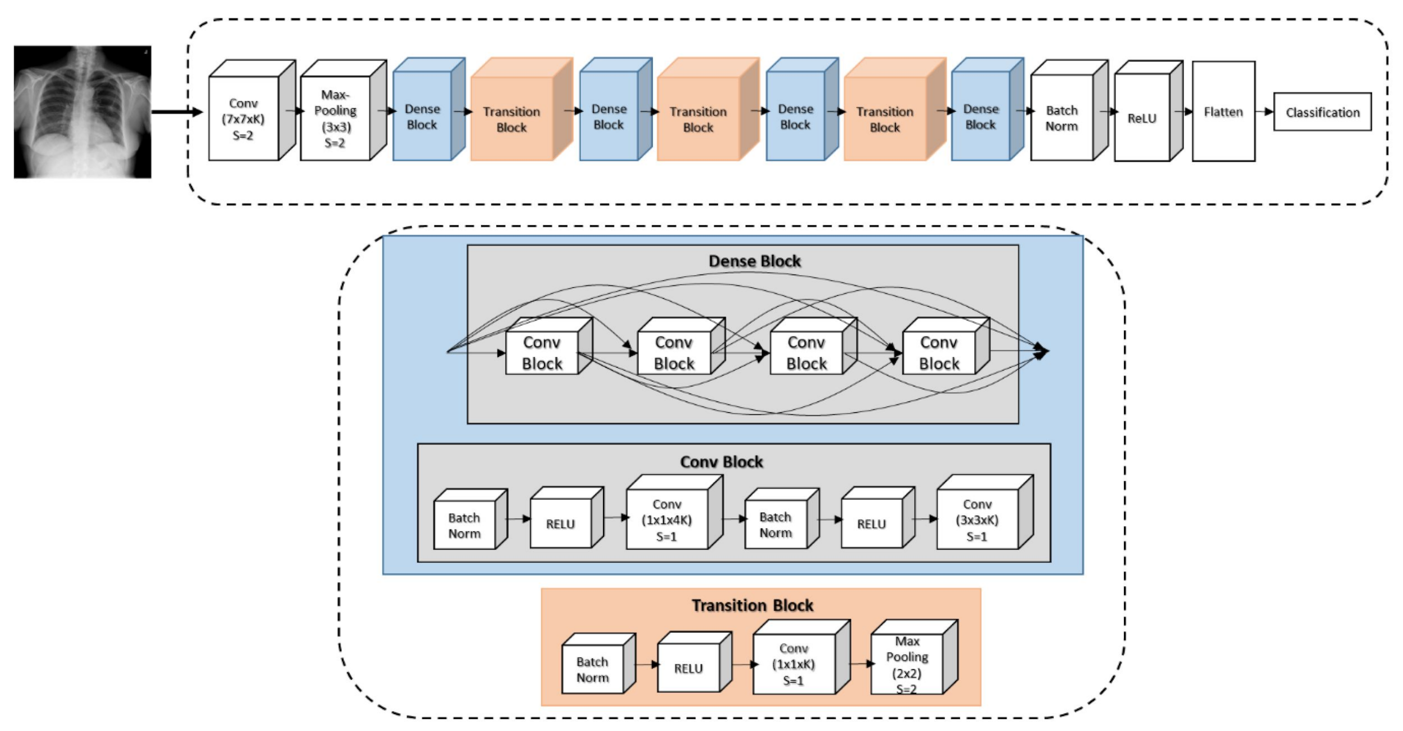
Models were imported from keras, and trained by ImageNet (http://www.image-net.org)

Transfer Learning



Visual Example of DenseNet121

One of the available models, trained on chest x rays as well. It achieves feature extraction with relatively few parameters.



4 Models Created for Comparison

01

Binary: Transfer

Determined “No Finding” vs Positive Finding

02

Binary: Full Train

Determined “No Finding” vs Positive Finding

03

Multilabel: Transfer

Attempted to classify each label (can predict more than 1)

04

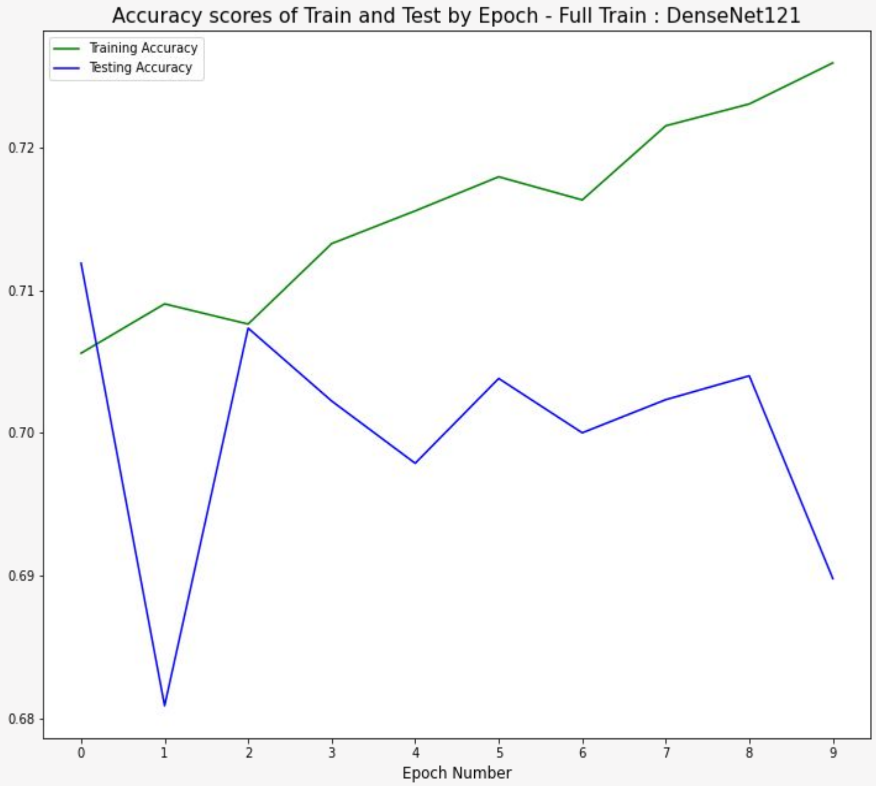
Multilabel: Full Train

Attempted to classify each label (can predict more than 1)

Performance of Base Models

Compared scores on Binary Model's Transfer Scores

Base Model Imported	Training Accuracy	Validation Accuracy
VGG-16	63.27%	62.52%
InceptionResNet	62.42%	61.66%
MobileNet	86.07%	62.71%
Xception	66.8%	66.5%
DenseNet121	67.80%	67.96%



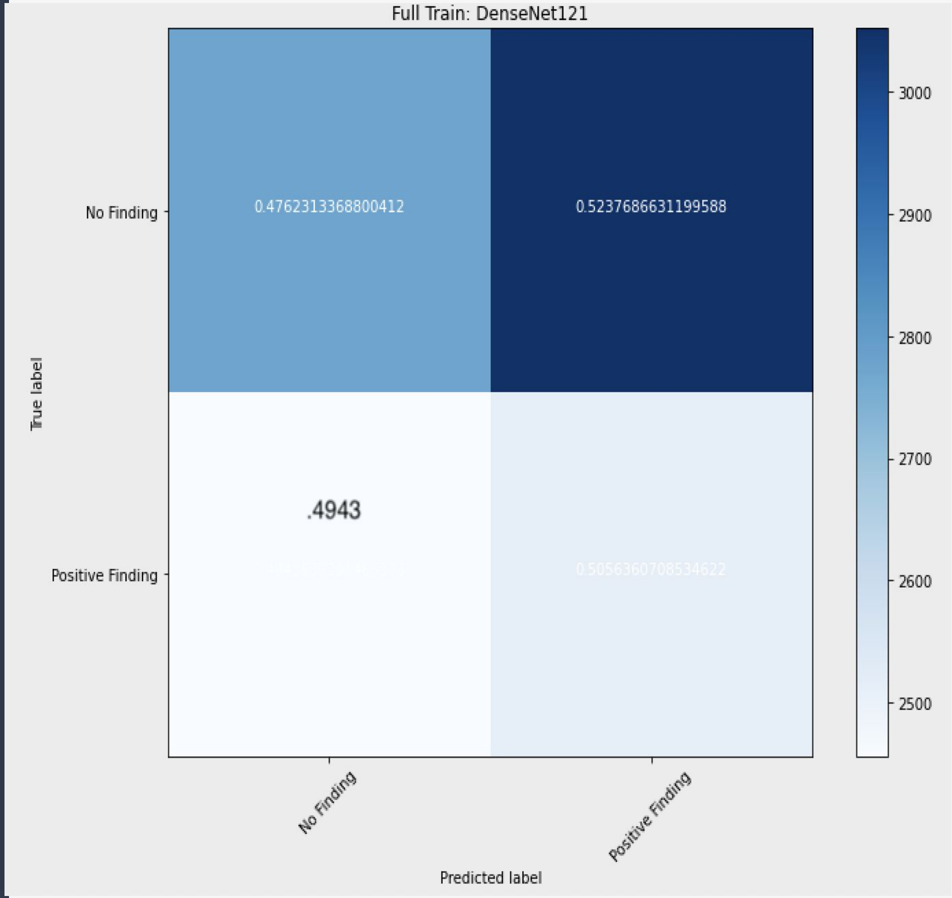
Fully Trained vs Transfer Learning

- Models trained “from scratch,” using model checkpoint saving as training times were often many days long

- Neither multi label classification model performed significantly higher than baseline accuracy

- Binary model trained from scratch achieved 70% test accuracy

Left image on a subset of data for visual purposes



Binary Recall/Specificity

- The binary model currently has a recall rate of 50%
- Specificity is only 47%
- Recall is an important metric if production model will flag positive findings

Note: This model is saved as a part of the repo to utilize in making predictions

Limitations to Data

Factors that potentially contributed to limited accuracy in these models:

- Propagation of uncertainty in class labels
- Algorithms cannot benefit from other relevant information such as patient history
- Heavily imbalanced classes
- High correlation in definition and appearance of many labels
- Restriction to open source computational resources - limited batch sizes, GPU usage, etc
- Necessity of model to identify very subtle changes in images

Next Steps

Balance Classes

Collapse similar classes into one class/reduce number of labels and imbalance

Increase Data

Optimal training can often require hundreds of thousands, or millions, of images

Seek Better Resources

Consider computational resources that allow for better model hyperparameters

Conclusions

Despite limited accuracy scores, this project demonstrates the potential for deep learning applications in medical imaging.

Combining the experience of experienced radiologists and machine learning models could greatly improve the quality of patient care.

CREDITS: This presentation template was created by **Slidesgo,** including icons by **Flaticon**, infographics & images by **Freepik** and illustrations by **Stories**

Please keep this slide for attribution

Thanks

Do you have any questions?