Pneumonia Diagnosis System with Chest X-Ray

APS360: Applied Fundamentals of Deep Learning

Aseem Gill, Dustin Zhang, Kirti Saxena, Ruiling Yang

Motivation

Automating pneumonia detection in chest X-rays

What is **Pneumonia**

- An infection that can cause inflammation in lungs
- In severe cases, could lead to respiratory system failure

Our Motivation

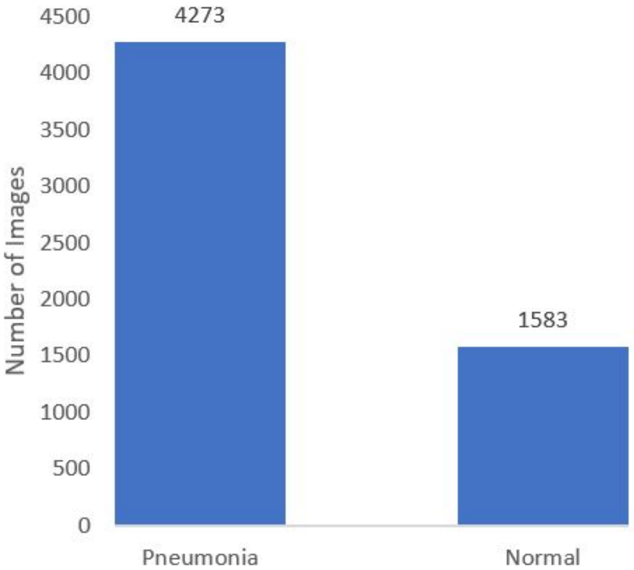
1. Pneumonia has a **500,000 global death rate** among children age < 5 years old (WHO)
2. **Imbalanced medical resources** in underdeveloped areas creates difficulty for such detection
3. Insufficient diagnostic time for every patient, leading to **possible incorrect diagnostic decisions**
   (avg. 3 min/patient with a radiologist)

Pneumonia

Normal

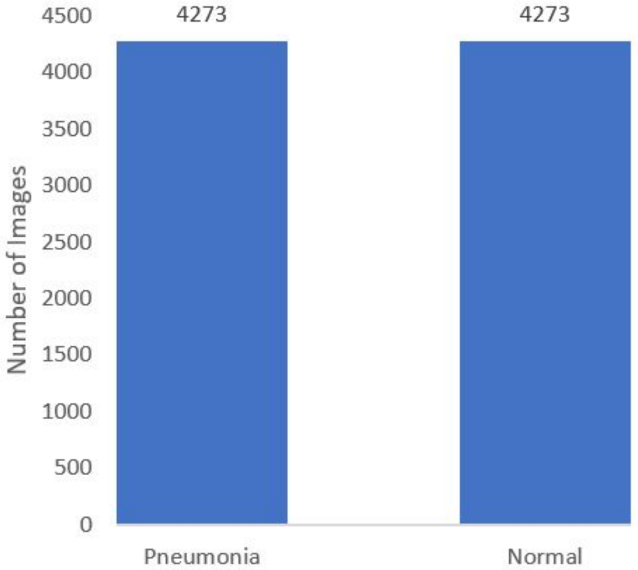
https://www.radiologyinfo.org/en/info/pneumonia

Selecting chest X-ray images for a deep learning neural network for binary classification



Pneumonia Chest X-Ray Images on **Kaggle** (JPEG)

Combine "Normal" images from **NIH Chest X-ray dataset**

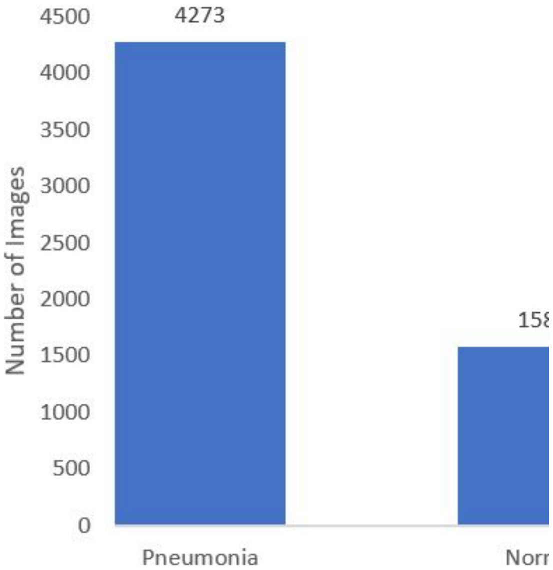


Balanced dataset (JPEG)

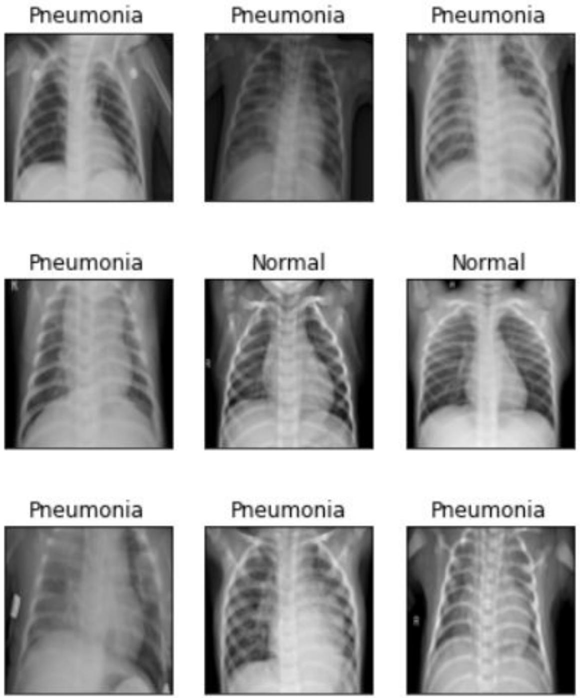
- **8,546** images from over 7,000 patients

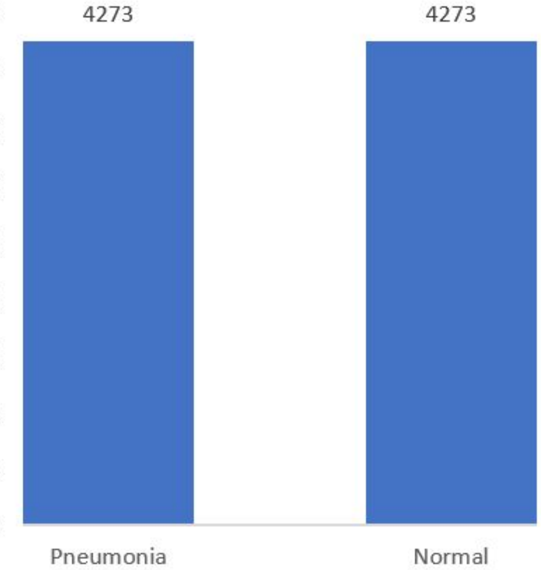
https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia
https://cloud.google.com/healthcare-api/docs/resources/public-datasets/nih-chest

Selecting chest X-ray images for a deep learning neural network for binary classification



Pneumonia Chest X-Ray Im
on **Kaggle** (JPEG)





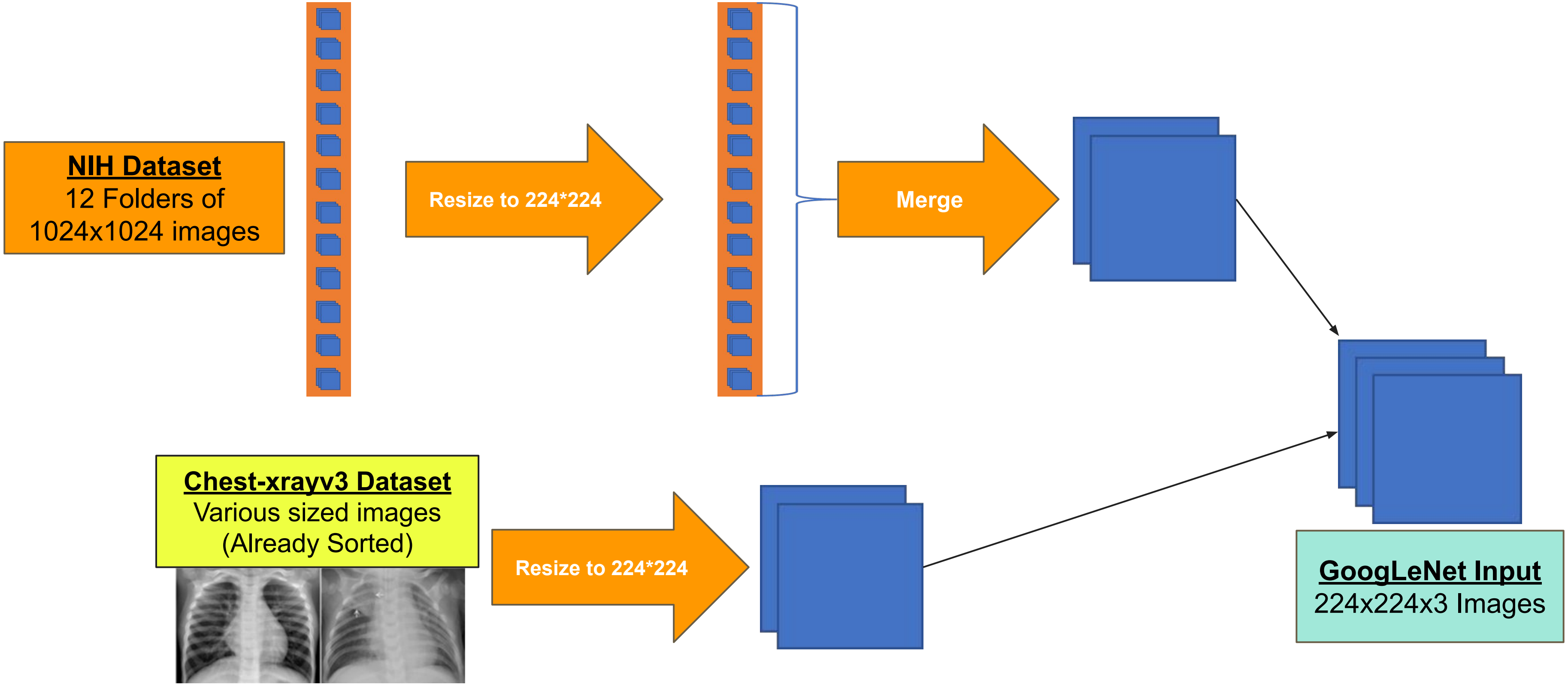
Balanced dataset (JPEG)

- **8,546** images from over 7,000 patients

https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia
https://cloud.google.com/healthcare-api/docs/resources/public-datasets/nih-chest

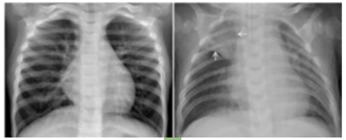
Data Processing

Resizing and merging image dataset



Processing our data distribution for pneumonia detection

Chest-xrayv3 Dataset



Convert labels to list of vectors

NIH Dataset

Image Index	Finding Labels
00000001_000.png	Cardiomegaly

Convert labels to list of vectors

Image Index	Finding Labels
00000001_000.png	[0,1,...0,0]

Filter to 2 classes

Image Index	Finding Labels
00000001_000.png	[0,1]

Remove unusable images

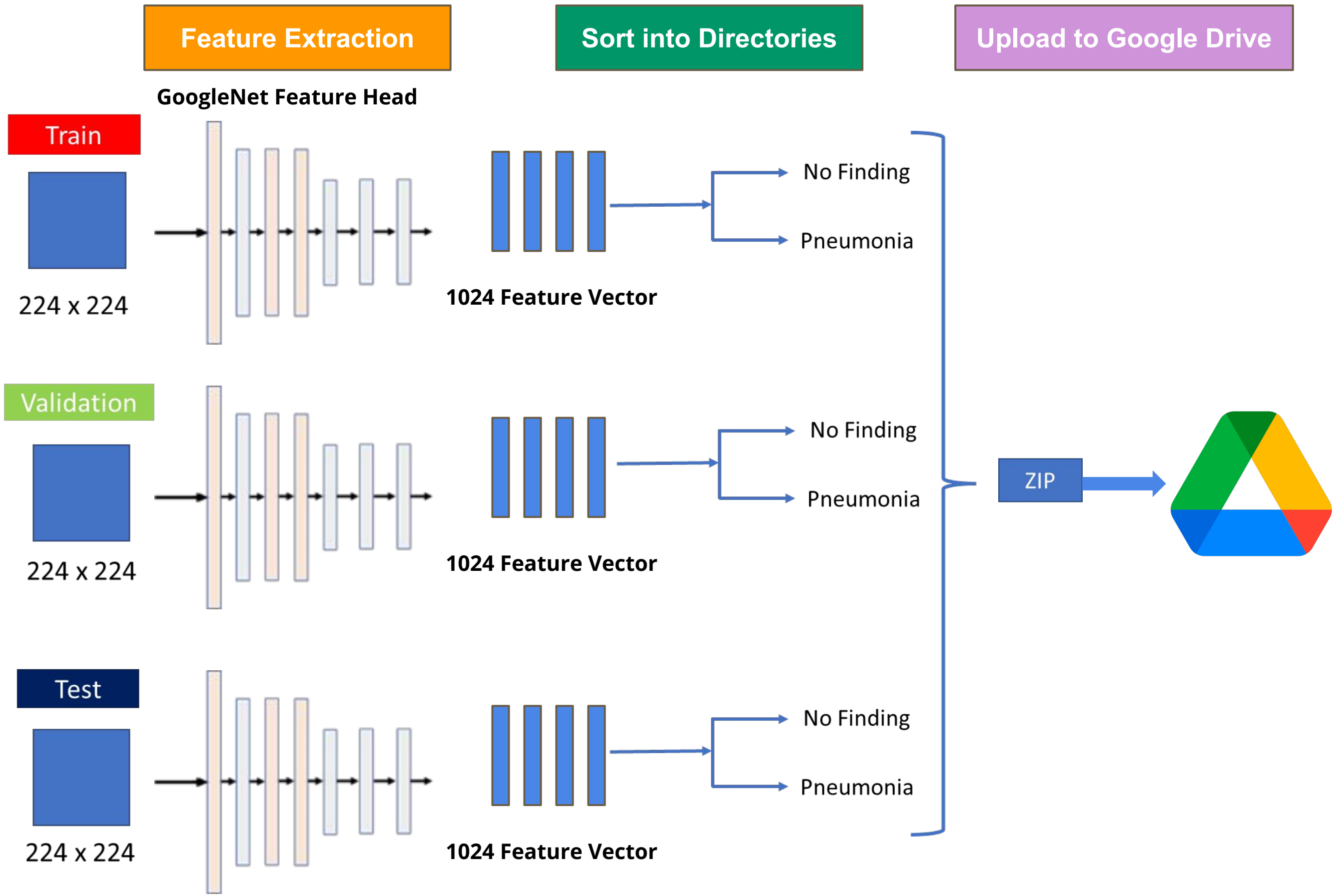
Split and Organize into folders

70% Train 224 x 224

20% Validation 224 x 224

10% Test 224 x 224

Extracting features to speed up training



Advantages

- Performing feature extraction before training has many benefits including:
  - Speeding up training time
  - Reducing the disk space usage on google drive
  - Reducing data loading time

Baseline Model & Primary Model

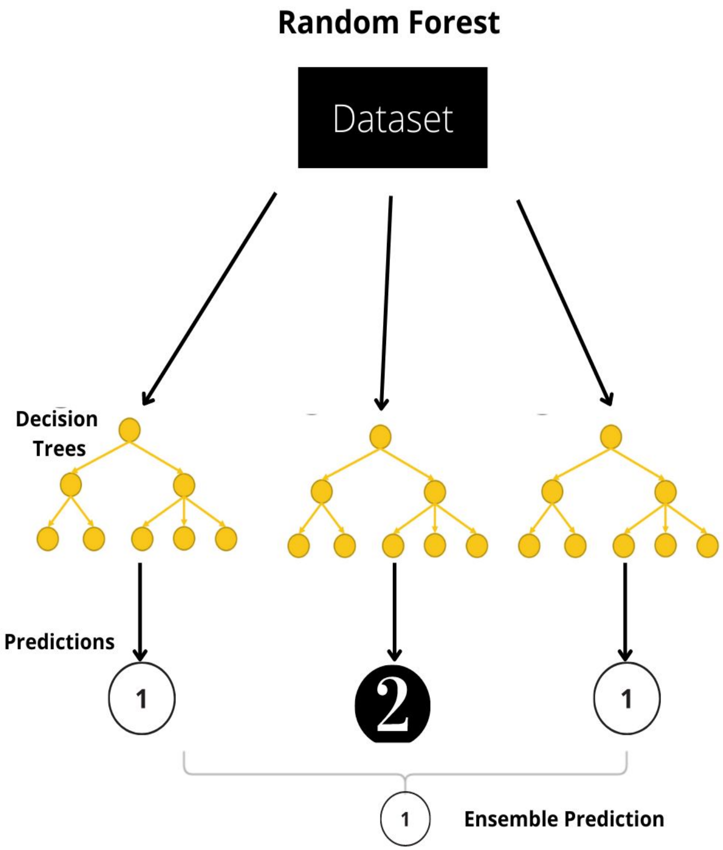
Baseline Model

Baseline Model Selection

- Chose to use a Random Forest Classifier

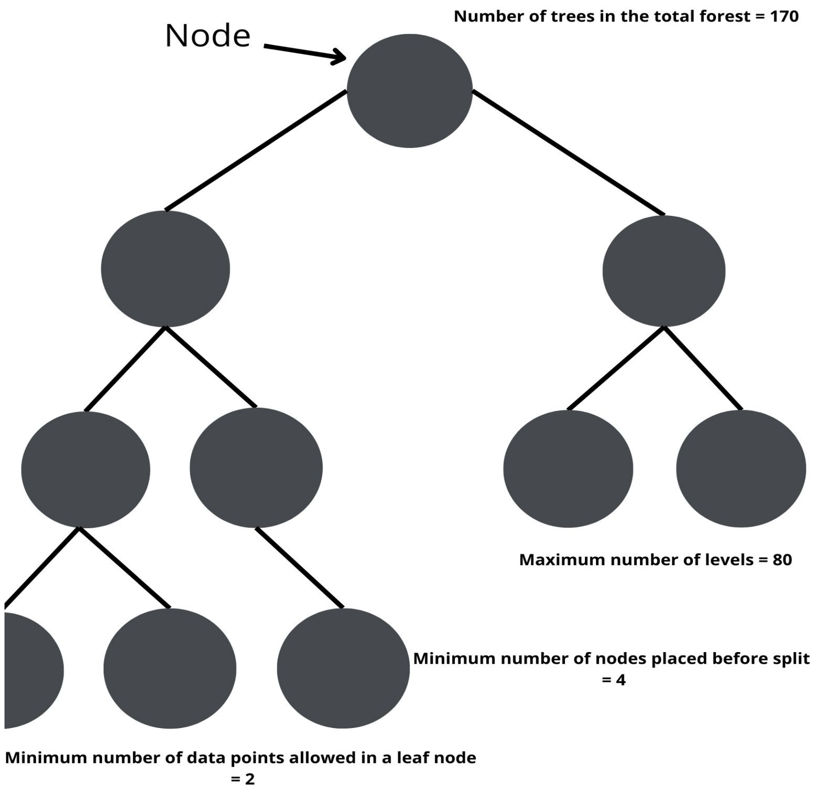
  - Provides higher levels of accuracy
- Decision Trees → Nodes → Predictions

  = Grouped Prediction
- Limitations of the model
  - Cannot extrapolate data



Baseline Model
Model Complexity & Hyperparameters

1. n estimators
   - number of trees in the forest = 170
2. Maximum Depth
   - maximum number of levels in each decision tree = 80
3. Minimum Samples Split
   - minimum number of data points placed in a node before the node is splitted = 4
4. Minimum Samples Leaf
   - minimum number of data points allowed in a leaf node = 2



Primary model: transfer learning provides deep model complexity on a limited dataset

GoogLeNet[1]
Complexity: 22 layers, Inception Blocks

Winning architecture of ILSVRC 2014 competition

1. **6.67%** top-5 error rate

Has only been started to be researched for pneumonia classification since 2020

Was significantly due to the interests of classifying COVID & pneumonia

DenseNet121
Complexity: 58 layers per Dense Block, 4 Dense Blocks

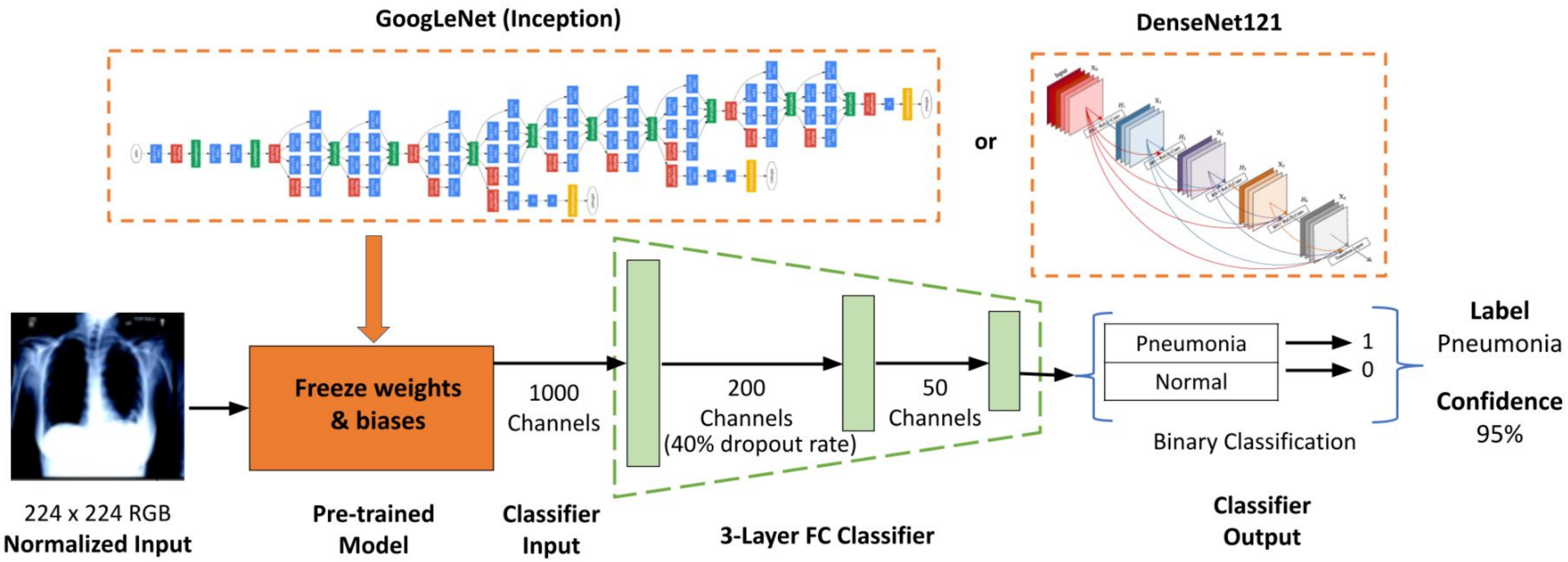
Previous work has been done on similar classification task: **CheXNet**[2]

1. Was built on **DenseNet121** to classify lung diseases in chest X-rays
2. Trained on NIH Chest X-ray Dataset (similar to ours)
3. **76.80%** accuracy on pneumonia detection

Used to benchmark the performance with GoogLeNet

1. Szegedy, Christian, et al. "Going deeper with convolutions." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015.
2. Rajpurkar, Pranav, et al. "Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning." *arXiv preprint arXiv:1711.05225* (2017).

Pretrained models linked with a 3-layer FC classifier



Demonstration

~30 seconds

Results

Baseline model serves to compare the results with the transfer learning models

Baseline Model **Random Forests Classifier ***	Transfer Learning Model **GoogLeNet**	Transfer Learning Model **DenseNet121**
Train Accuracy: 99.37% **Validation Accuracy:** 8.48% **Test Accuracy:** 9.50%	**Train Accuracy:** 99.25% **Validation Accuracy:** 99.00% **Test Accuracy:** 92.40%	**Train Accuracy:** 97.52% **Validation Accuracy:** 97.55% **Test Accuracy:** 59.35%

Four classes of lung conditions classification problem

→

Rescoped to binary classification problem (Normal vs Pneumonia)

DenseNet VS. GoogLeNet

(batch_size = 64, epoch = 5)

Final Training Accuracy	0.9752348420153715
Final Validation Accuracy	0.9755194990037005
Testing Accuracy	0.5935374149659863
Recall	0.15584415584415584

Final Training Accuracy	0.9925891830344444
Final Validation Accuracy	0.9900370054084828
Testing Accuracy	0.9236161616161616
Recall	0.9351851851851851

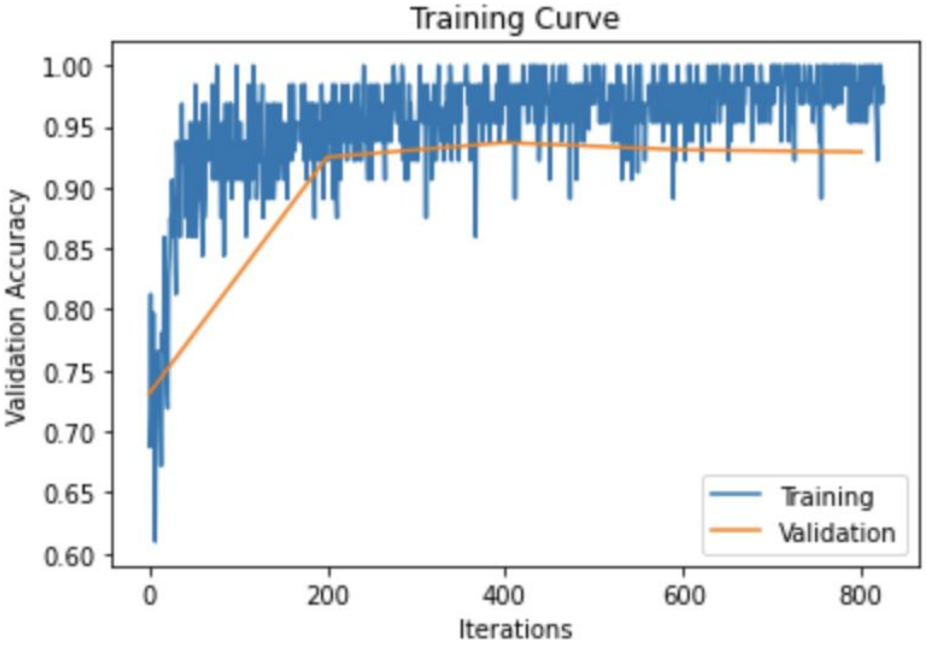
True Positives (TPs): 32	False Positives (FPs): 36
False Negatives (FNs): 184	True Negatives (TNs): 180

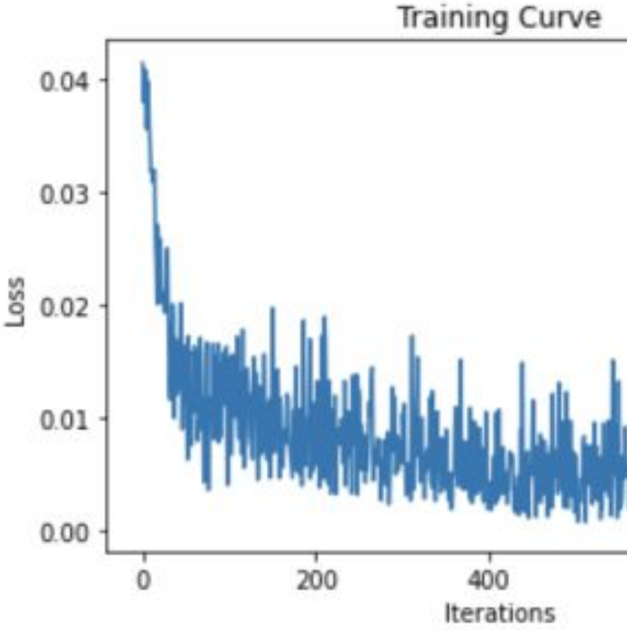
True Positives (TPs): 203	False Positives (FPs): 20
False Negatives (FNs): 13	True Negatives (TNs): 196

Obtaining best results after hyperparameter tuning

Optimizer	Stochastic Gradient Descent
Loss Function	Binary Cross Entropy
Batch Size	64
Learning Rate	0.01 (SGD default)
Momentum	0.9 (SGD default)
Training Epochs	10

Best classification results from GoogLeNet - Qualitative





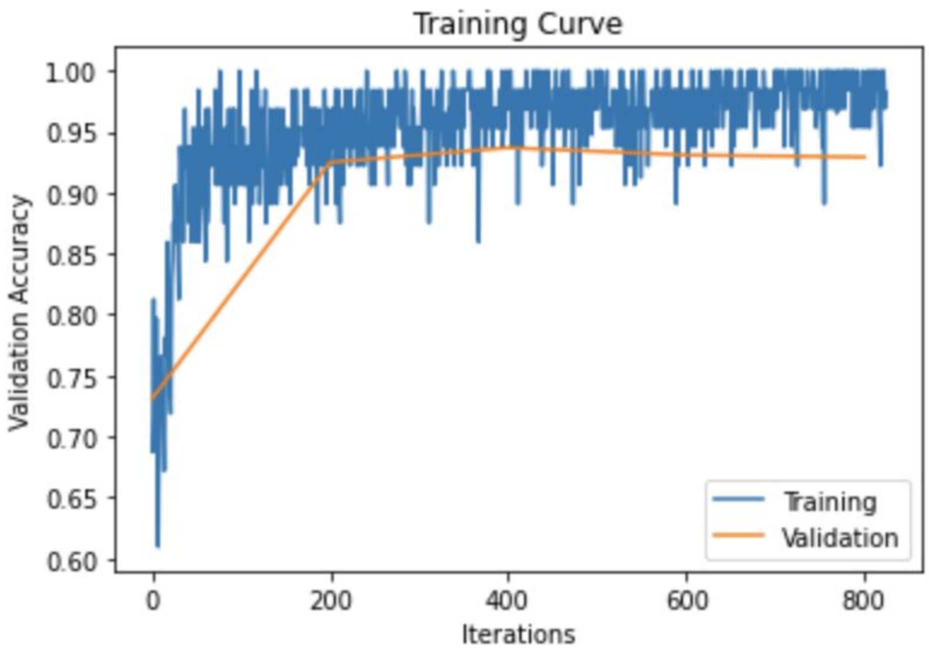
GoogLenet Learning Curves - Quantitative Results

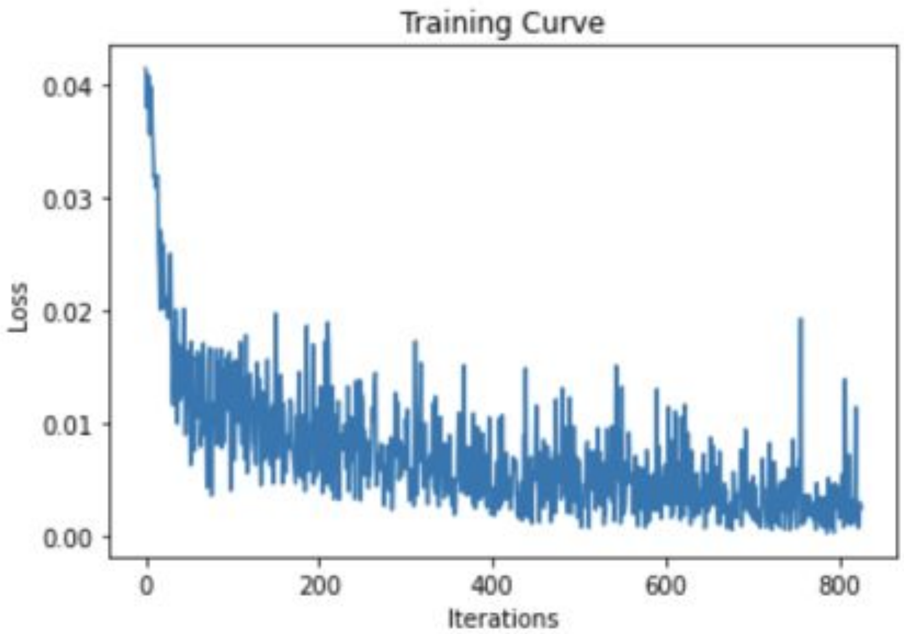
Final Training Accuracy	0.9758041559920296
Final Validation Accuracy	0.9817819527469399
Testing Accuracy	0.9259259259259259
Recall	0.9490740740740740
Precision	0.9070796460974932
F1 Score	0.9276018099966283

True Positives (TPs): 205	**False Positives (FPs):** 21
False Negative (FNs): 11	**True Negatives (TNs):** 195

Discussion

GoogLeNet Model Performance Evaluations





Final Training Accuracy	0.9758041559920296
Final Validation Accuracy	0.9817819527469399
Testing Accuracy	0.9206484641638225

GoogLeNet Model Performance Evaluations

Recall	0.9490740740740740
Precision	0.9070796460974932
F1 Score	0.9276018099966283

Recall
(pneumonia detection rate)
95%

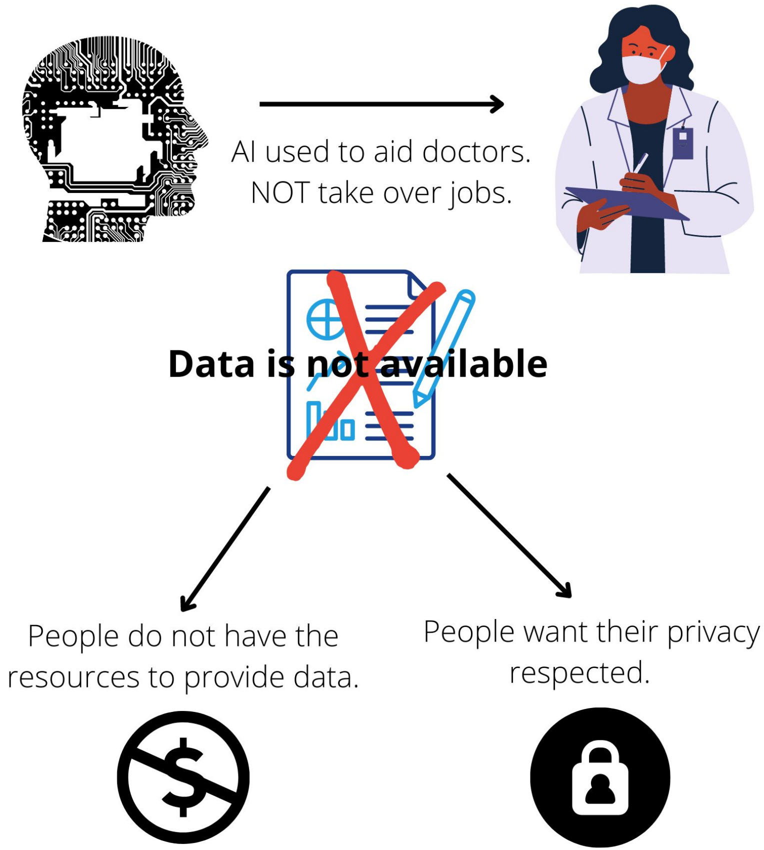
True Positives (TPs): 205	**False Positives (FPs)**: 21
False Negative (FNs): 11	**True Negatives (TNs)**: 195

Final Thoughts

Applications & Ethical Considerations

- People do not want technology overtake the healthcare system
  - Although this may be beneficial in some cases.

- Patient privacy & inconsistent data
  - Marginalized groups who do not wish to share their data or who do not have the resources to obtain scans.



Baseline model

Predictions		Actuals: Pneumonia	Actuals: No Finding
	Pneumonia	**True Positives (TPs)**: 310	**False Positives (FPs)**: 56
	No Finding	**False Negative (FNs)**: 122	**True Negatives (TNs)**: 376

Primary model

		Actuals	
		Pneumonia	No Finding
Predictions	Pneumonia	**True Positives (TPs)**: 410	**False Positives (FPs)**: 42
	No Finding	**False Negative (FNs)**: 22	**True Negatives (TNs)**: 390

Key Takeaways

1. AI-ML field is growing due to diverse people from different academic backgrounds

2. There will be always be questions to be answered, facts that we know and exploration to be done