EMPOWERING PHYSICIANS WITH COMPUTER-AIDED DIAGNOSIS OF LUNG DISEASE

Ash Ang
DSI 23

TABLE OF CONTENTS

01 INTRODUCTION

02 METHODOLOGY

03 EDA & DATA-PROPCESSING

04 MODELLING & EVALUATION

05 LIMITATIONS

06 CONCLUSION

Timely and accurate diagnosis of lung disease is vital to optimize patient's outcome and reduce associated costs

~ 40 M — Cases of asthma or chronic obstructive pulmonary disease (COPD) in the United States

~12 M — Cases of COPD have not been yet been diagnosed

$20.7 B — Estimated annual healthcare costs for asthma alone

Chest X-ray (CXR) is one of the most common and cost-effective diagnostic tools in diagnosing lung diseases

Computer-aided diagnosis can empower physicians to make more timely and accurate diagnosis

Computed-Aided Diagnosis (CAD)

- Some promising work have been reported in the past. E.g. Deep learning work on Tuberculosis (TB) classification

- Achieving clinically relevant CAD in real world medical sites on all data settings of chest X-rays is still very difficult.

Human Diagnosis

- Although CAD has been used in clinical environments for over 40 years, CAD usually does not substitute the doctor or other professional. It can serve as a second opinion.

- A radiologist is still responsible for the final interpretation of a medical image and closing the report.

Objectives:

1. Build a convolutional neural network (CNN) model to perform a multi-label classification of thoracic pathology using CXR images and,

2. Evaluate its performance against transfer learning approaches such as ResNet50 and MobileNetV2 models with metrics such as macro-average recall and F1 score

Understanding multi-label classification

Multi-label classification:
There are two classes or more and every observation belongs to one or multiple classes at the same time (or co-occurrences).

Multi-class vs multi-label classification:
In multi-class problems the classes are mutually exclusive, whereas for multi-label problems each label represents a different classification task, but the tasks are somehow related.

Understanding multi-label classification

Classifying cats, dogs and birds

Image	Labels	Ground-truth binary vector
	[cat, dog]	[1, 1, 0]
	[cat, bird]	[1, 0, 1]
	[bird]	[0, 0, 1]
	[]	[0, 0, 0]

Classifying lung pathology(ies) with CXR images

Atelectasis	Cardiomegaly	Consolidation
1	1	1

Atelectasis	Cardiomegaly	Consolidation
0	0	1

Atelectasis	Cardiomegaly	Consolidation
0	0	0

Complex challenge to build a convolutional neural network (CNN) model to perform a multi-label classification

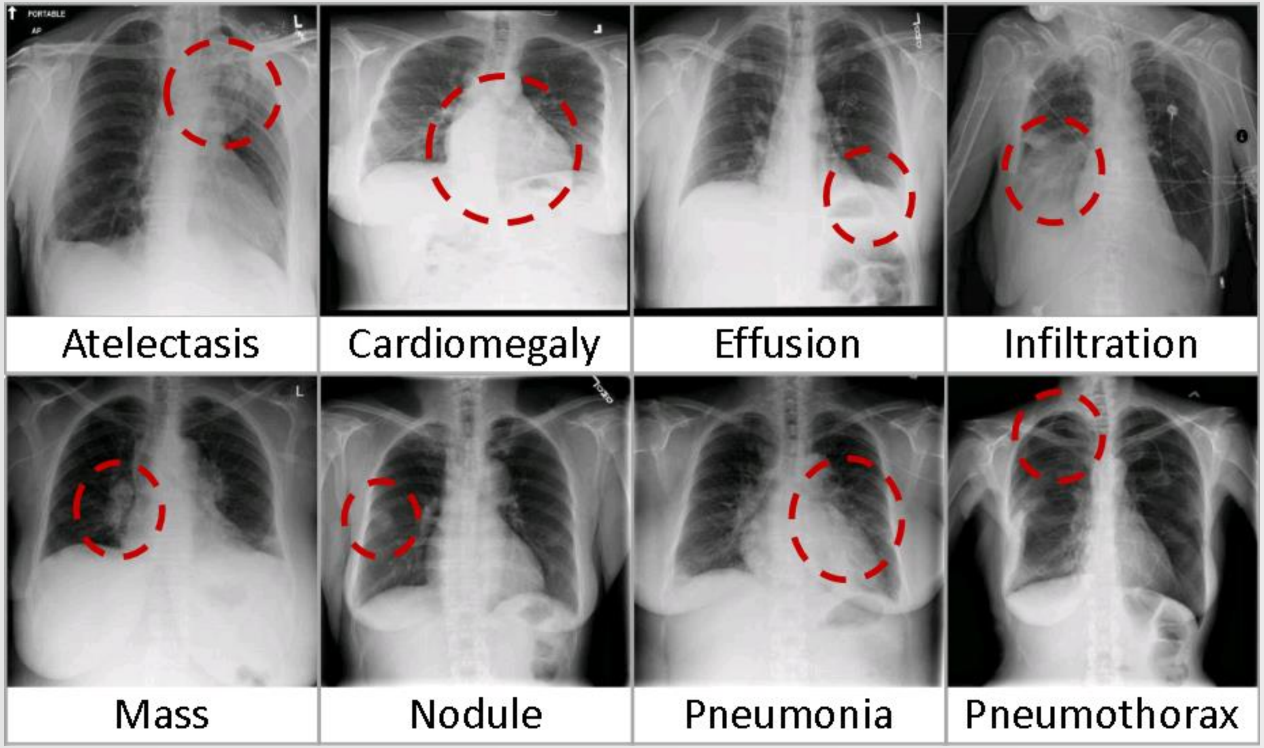
Data source

- Original dataset extracted from clinical Picture Archiving and Communication System (PACS) database at National Institutes of Health Clinical Center in the U.S.
- 112,120 frontal chest x-ray images of 30,085 unique patients in the hospital
- Fourteen disease image labels (where each image can have multi-labels) were mined from the associated radiological reports using natural language processing. Some label examples include 'Atelactasis', 'Cardiomegaly', 'Pneumonia' and 'Effusíon'.

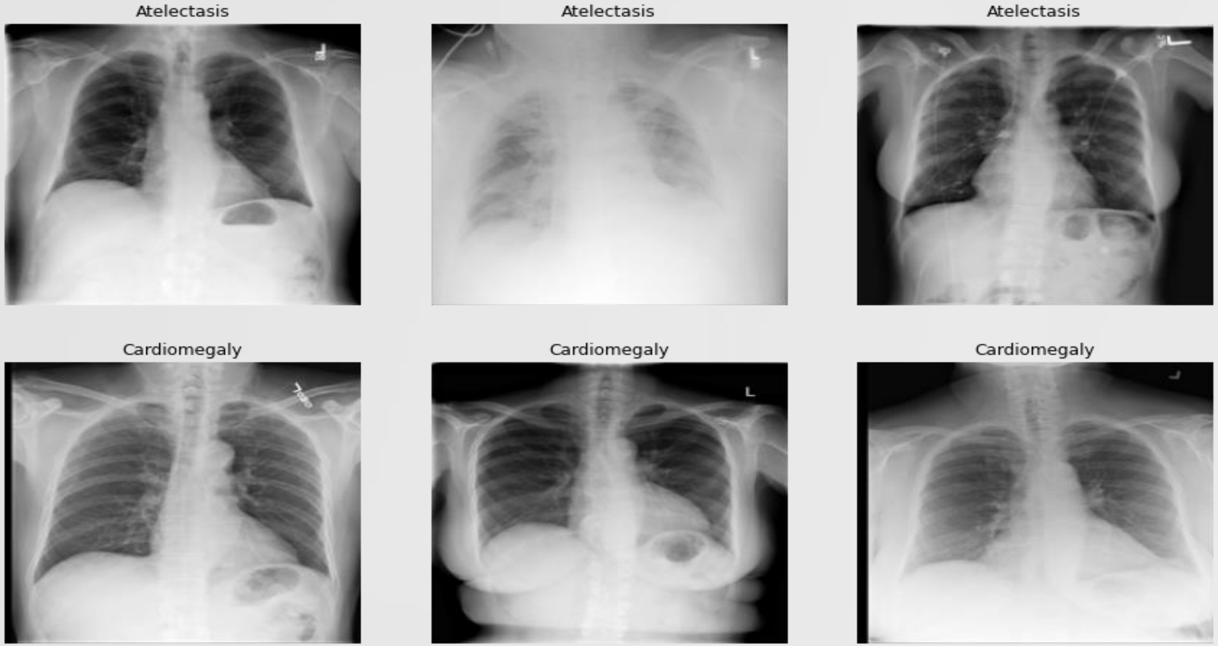
Pre-processing

- Image labels with less than 1,000 counts were removed.
- CXR Images are divided into train/validation and testing by patient level. All studies from the same patient will only appear in either training/validation or testing set.
- Image labels with normal results (i.e. 'No Finding') are removed to reduce sparseness of target matrix.
- Final dataset contains 36,023 samples with 13 labels.

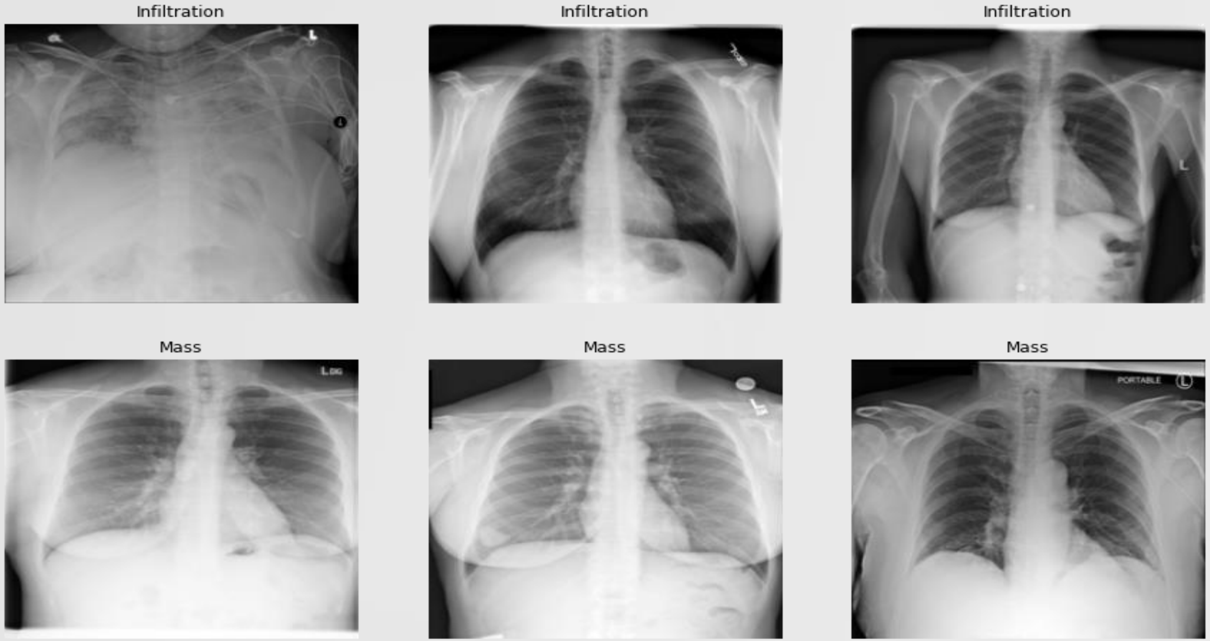
Visualize examples of common thorax diseases and their pattern



An arduous task for the untrained eye to discern and correctly identify single thoracic pathology label, let alone multiple labels

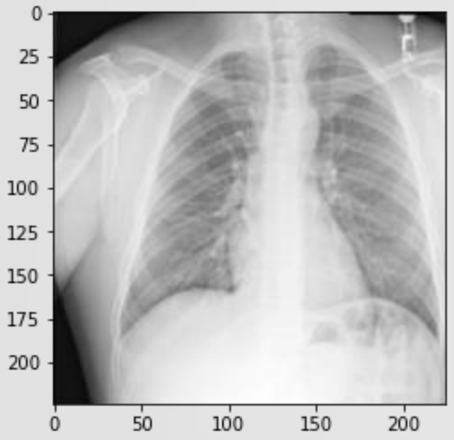


An arduous task for the untrained eye to discern and correctly identify single thoracic pathology label, let alone multiple labels



Data processing and augmentation

01

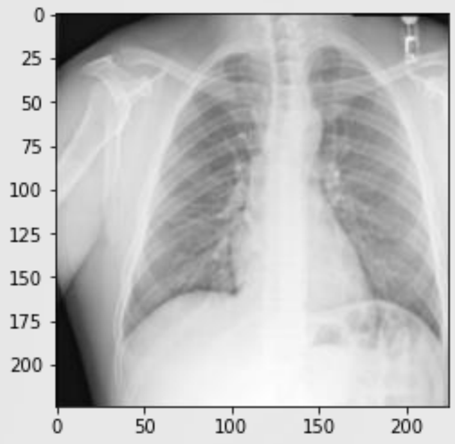


(224, 224, 1)

Color conversion

Convert from gray scale to RGB for subsequent transfer learning

02

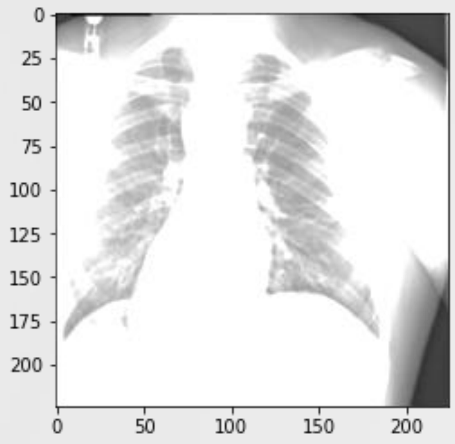


(224, 224, 3)

Data augmentation

Randomize lateral flipping and brightness adjustment to increase diversity

03



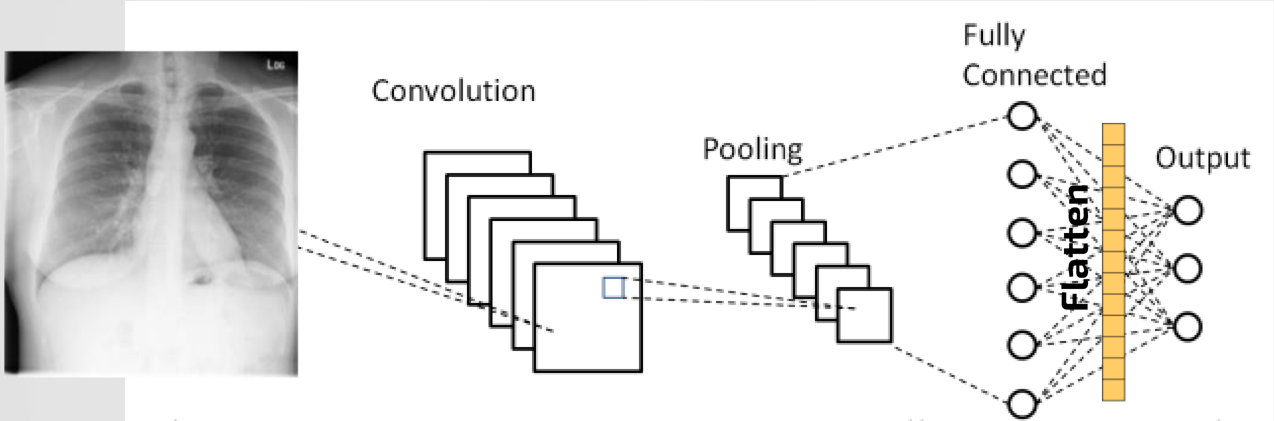
(224, 224, 3)

Designing convolutional neural network architecture

None, or any 1 or more of the pathology labels:

Applied threshold of 0.3

Input CXR Image



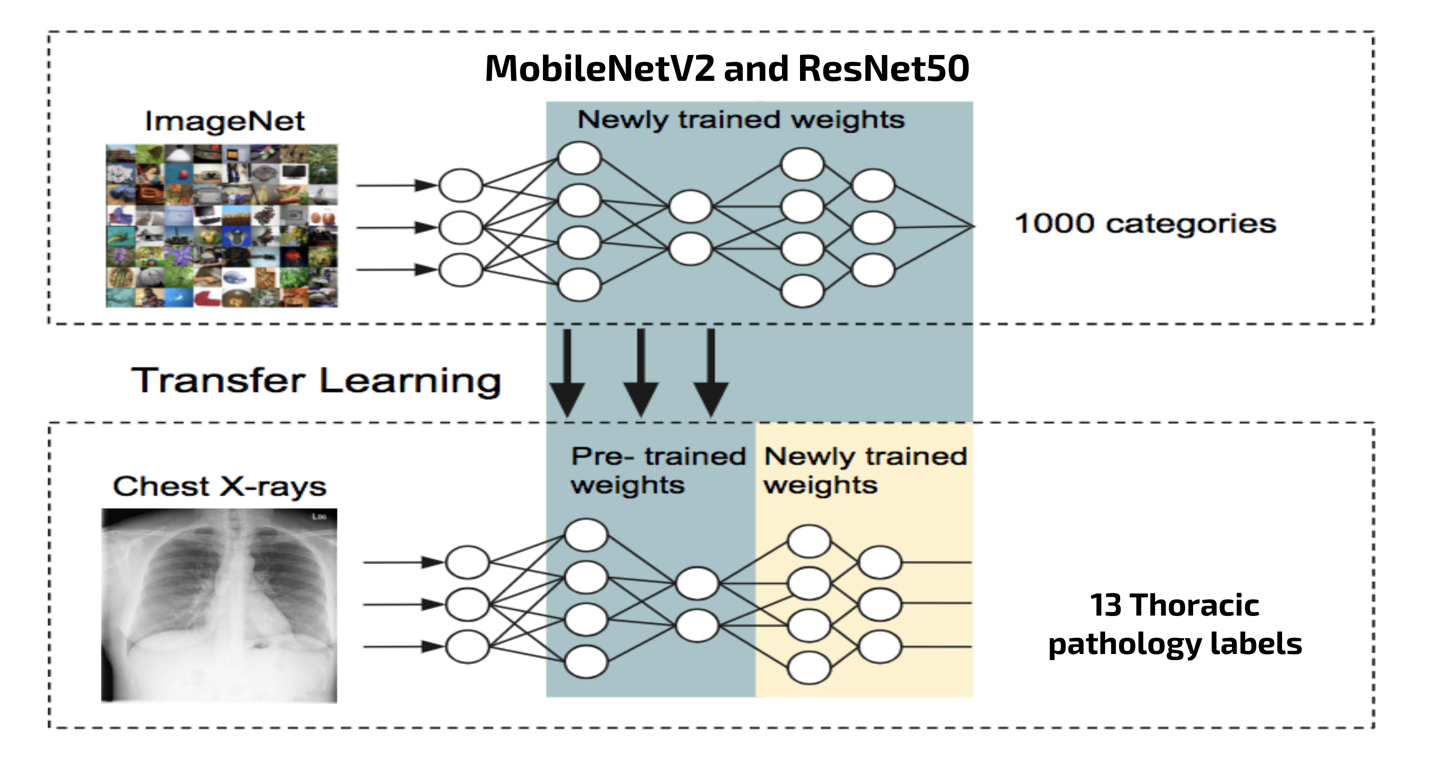
(224, 224, 3)

8 x Convolutional
4 x Batch Normalization
4 x Max Pooling

Sigmoid activation function returns probability [0, 1] for each label

'Atelectasis'
'Cardiomegaly'
'Consolidation'
'Edema'
'Effusion'
'Emphysema'
'Fibrosis'
'Infiltration'
'Mass'
'Nodule'
'Pleural_Thickening'
'Pneumonia'
'Pnuemothorax'

Transfer learning using MobileNetV2 and ResNet50



Model Evaluation

	Macro-average recall	Macro-average F1 score	Train accuracy	Test accuracy
CNN	0.19	0.19	0.893	0.888
MobileNetV2	0.24	0.22	0.918	0.881
ResNet50	0.32	0.31	0.915	0.888

Algorithm's ability to detect positive in any of the thirteen labels would help to alert or draw immediate attention from physicians on any abnormal thorax findings in a patient's CXR study.

Macro-averaging is a single performance indicator obtained by averaging the score of individual classes. Macro-averaging is preferred over micro-averaging in case of imbalanced classes because it weighs each of the classes equally and is not influenced by the number of examples of each class.

ResNet50 Classification Report

	precision	recall	f1-score	support
Atelectasis	0.52	0.38	0.44	1656
Cardiomegaly	0.40	0.45	0.42	342
Consolidation	0.27	0.05	0.08	570
Edema	0.32	0.28	0.29	275
Effusion	0.56	0.68	0.62	1732
Emphysema	0.32	0.36	0.34	285
Fibrosis	0.37	0.04	0.08	250
Infiltration	0.43	0.90	0.58	2756
Mass	0.63	0.21	0.32	806
Nodule	0.36	0.20	0.26	942
Pleural_Thickening	0.23	0.15	0.18	448
Pneumonia	0.00	0.00	0.00	175
Pneumothorax	0.31	0.43	0.36	527
micro avg	0.45	0.49	0.47	10764
macro avg	0.36	0.32	0.31	10764
weighted avg	0.44	0.49	0.42	10764
samples avg	0.46	0.52	0.46	10764

Limitations

- Imbalanced labels often leads to poor classification results, especially when CXR images are hard to distinguish using the available features

- Details and information are lost when CXR images are resized to 224 by 224 pixels.

- While the NLP image labelling accuracy is estimated to be >90%, there are bound to have some erroneous labels.

- Under constraints of processing speed and resources, model parameters did not undergo extensive fine-tuning

Conclusion

Balance labels

- Apply over-sampling technique (e.g. Variational Autoencoders (VAE) or Synthetic Minority Over-sampling Technique (SMOTE))
- Collecting more CXR images of underrepresented labels over time

Model extension

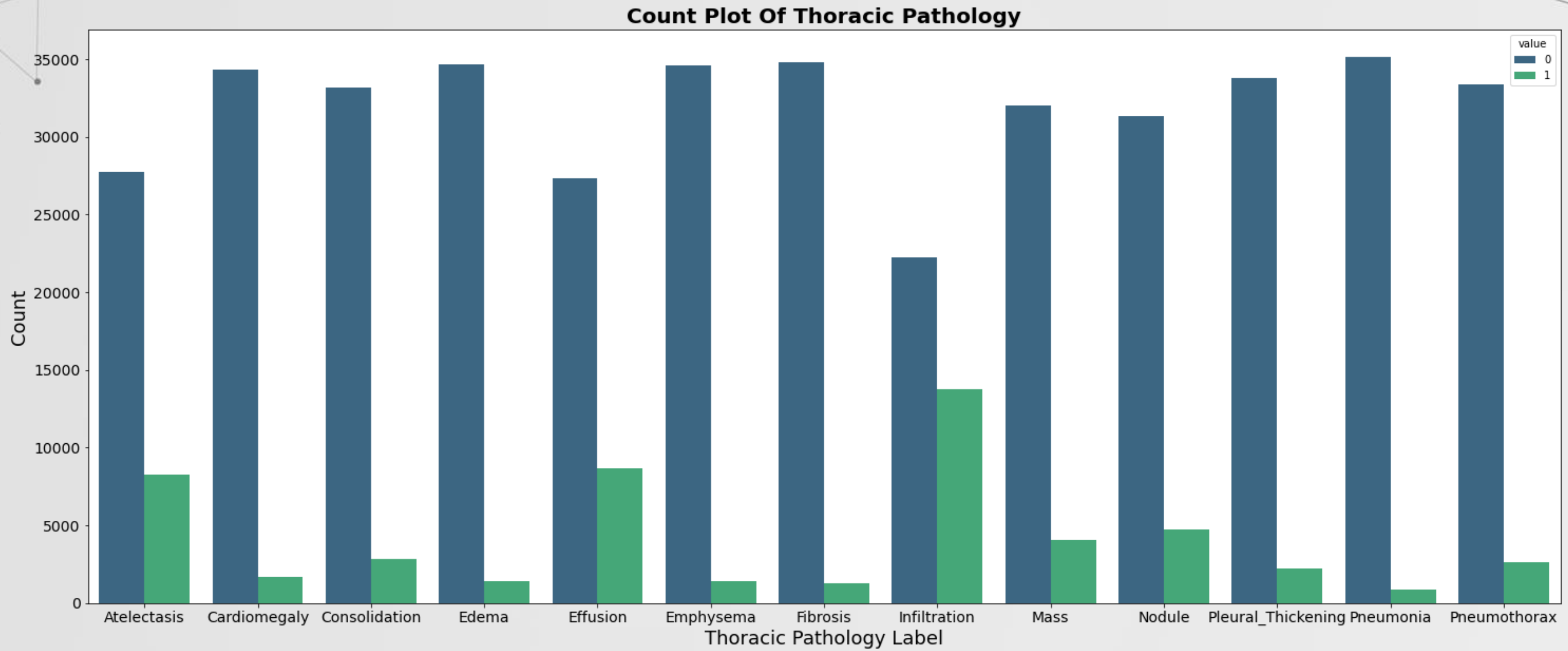
- Other medical specialties such as Cardiology or Orthopedic
- Other imaging diagnostics such as computed tomography or ultrasound scan

Overcome adoption barriers

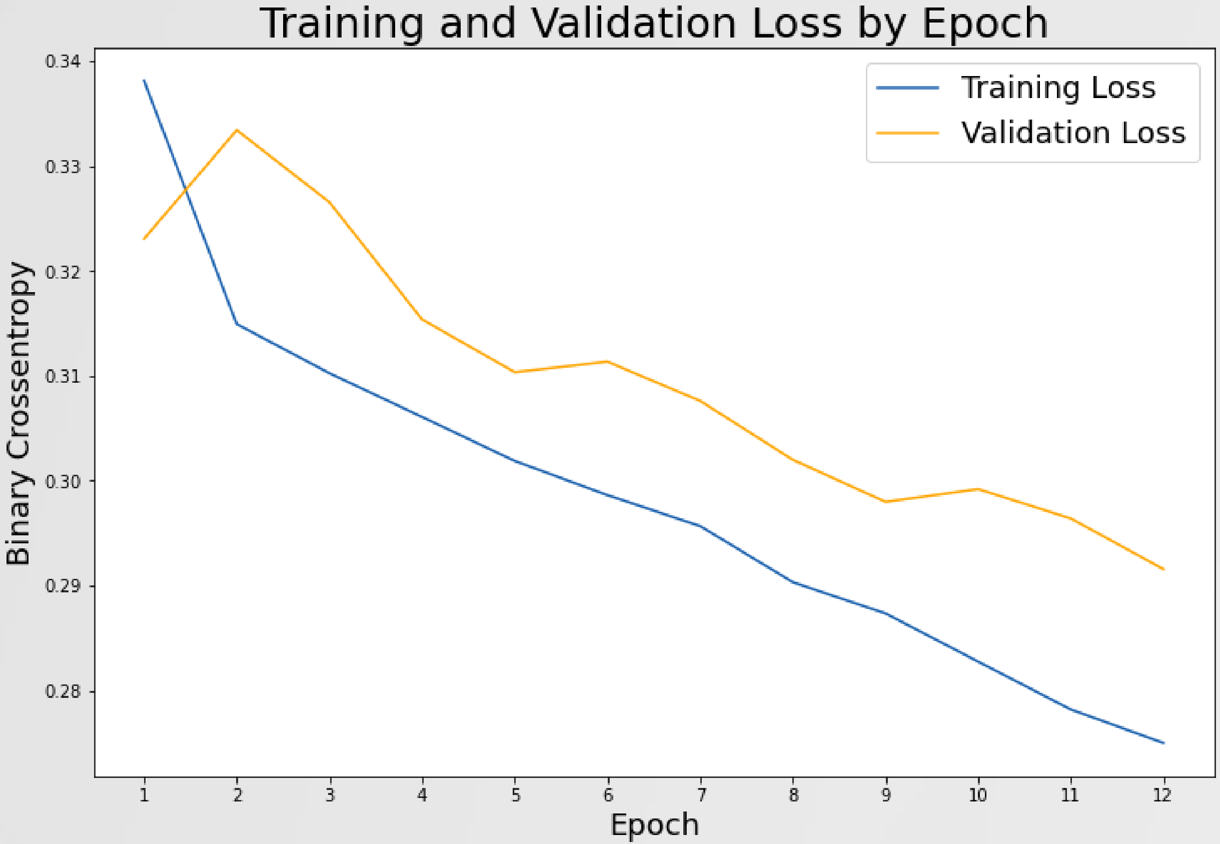
- Address medicolegal issues and overcome other regulatory barriers.
- Design model with patient-safety at the core and test it on real-life cases in a safe and controlled environment.

Thank you,
Thank you very much!

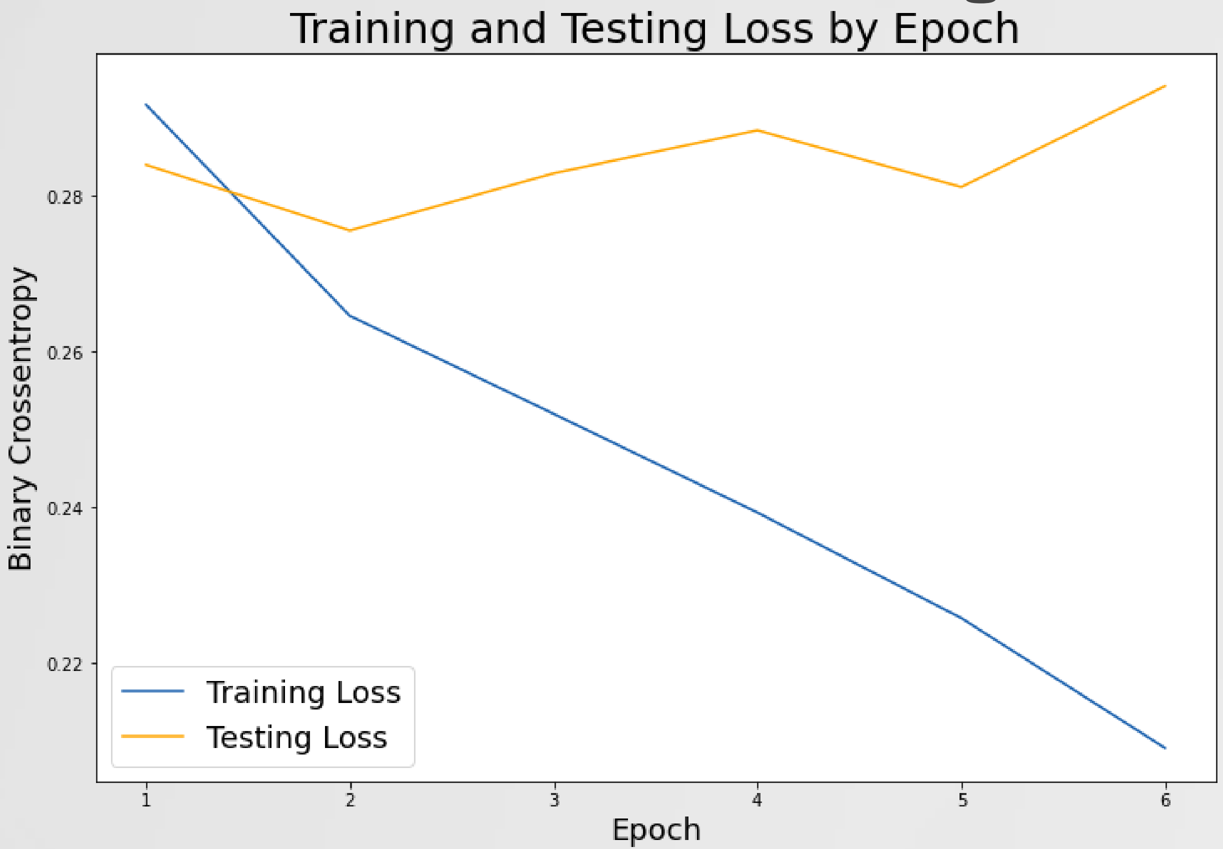
Appendix
Exhibit 1 – Count plot of thoracic pathology labels



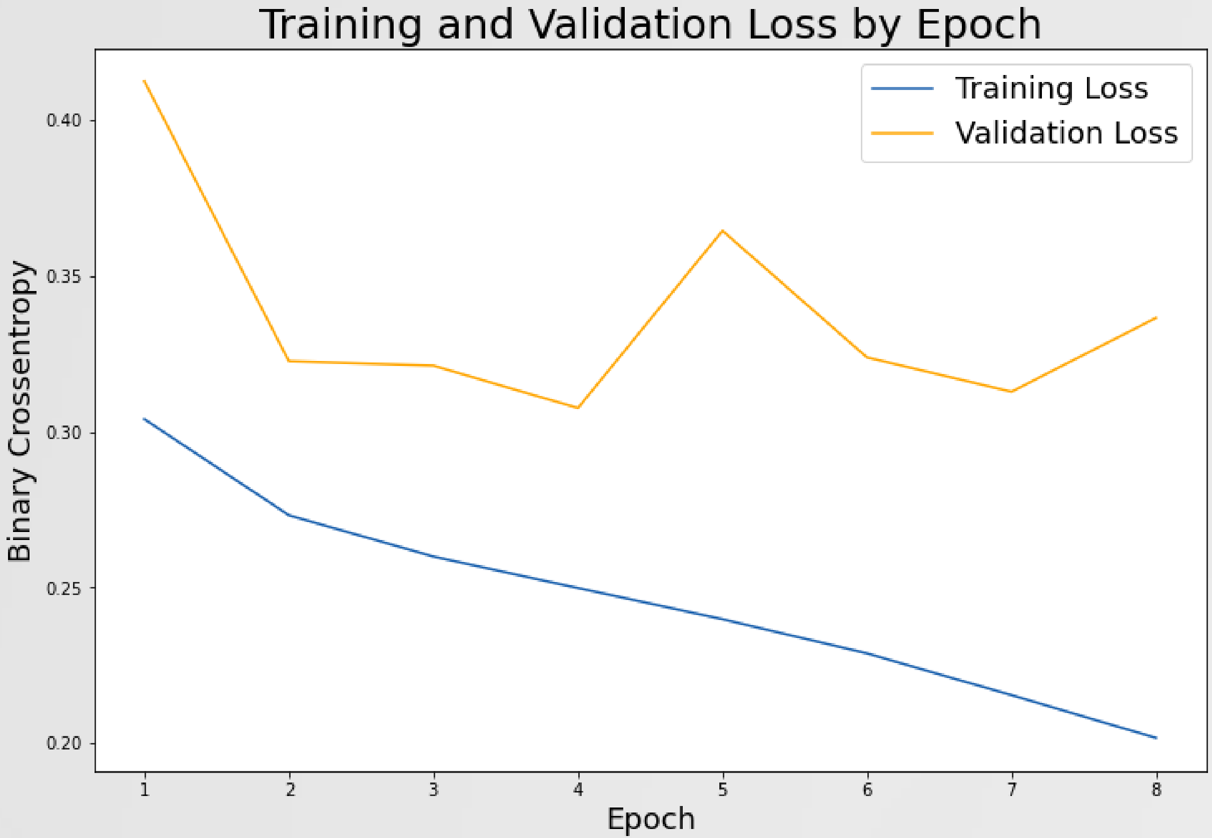
Appendix
Exhibit 2A – CNN Model Learning Curves



Appendix
Exhibit 2B – ResNet50 Model Learning Curves



Appendix
Exhibit 2C – MobileNetV2 Model Learning Curves



Appendix
Exhibit 3A- CNN Model’s Classification Report

	precision	recall	f1-score	support
Atelectasis	0.39	0.47	0.43	1656
Cardiomegaly	0.42	0.14	0.21	342
Consolidation	0.24	0.02	0.03	570
Edema	0.31	0.08	0.12	275
Effusion	0.45	0.61	0.52	1732
Emphysema	0.18	0.01	0.01	285
Fibrosis	0.00	0.00	0.00	250
Infiltration	0.42	0.81	0.56	2756
Mass	0.34	0.18	0.24	806
Nodule	0.25	0.10	0.14	942
Pleural_Thickening	0.38	0.04	0.07	448
Pneumonia	0.00	0.00	0.00	175
Pneumothorax	0.32	0.07	0.12	527
micro avg	0.41	0.41	0.41	10764
macro avg	0.28	0.19	0.19	10764
weighted avg	0.36	0.41	0.34	10764
samples avg	0.39	0.43	0.39	10764

Appendix

Exhibit 3B- MobileNetV2 Model's Classification Report

	precision	recall	f1-score	support
Atelectasis	0.63	0.17	0.27	1656
Cardiomegaly	0.79	0.06	0.10	342
Consolidation	0.12	0.00	0.01	570
Edema	0.29	0.20	0.24	275
Effusion	0.74	0.30	0.42	1732
Emphysema	0.11	0.76	0.20	285
Fibrosis	0.40	0.03	0.06	250
Infiltration	0.50	0.59	0.54	2756
Mass	0.51	0.22	0.31	806
Nodule	0.51	0.12	0.19	942
Pleural_Thickening	0.22	0.13	0.17	448
Pneumonia	0.00	0.00	0.00	175
Pneumothorax	0.23	0.54	0.32	527
micro avg	0.39	0.31	0.35	10764
macro avg	0.39	0.24	0.22	10764
weighted avg	0.50	0.31	0.33	10764
samples avg	0.36	0.34	0.33	10764