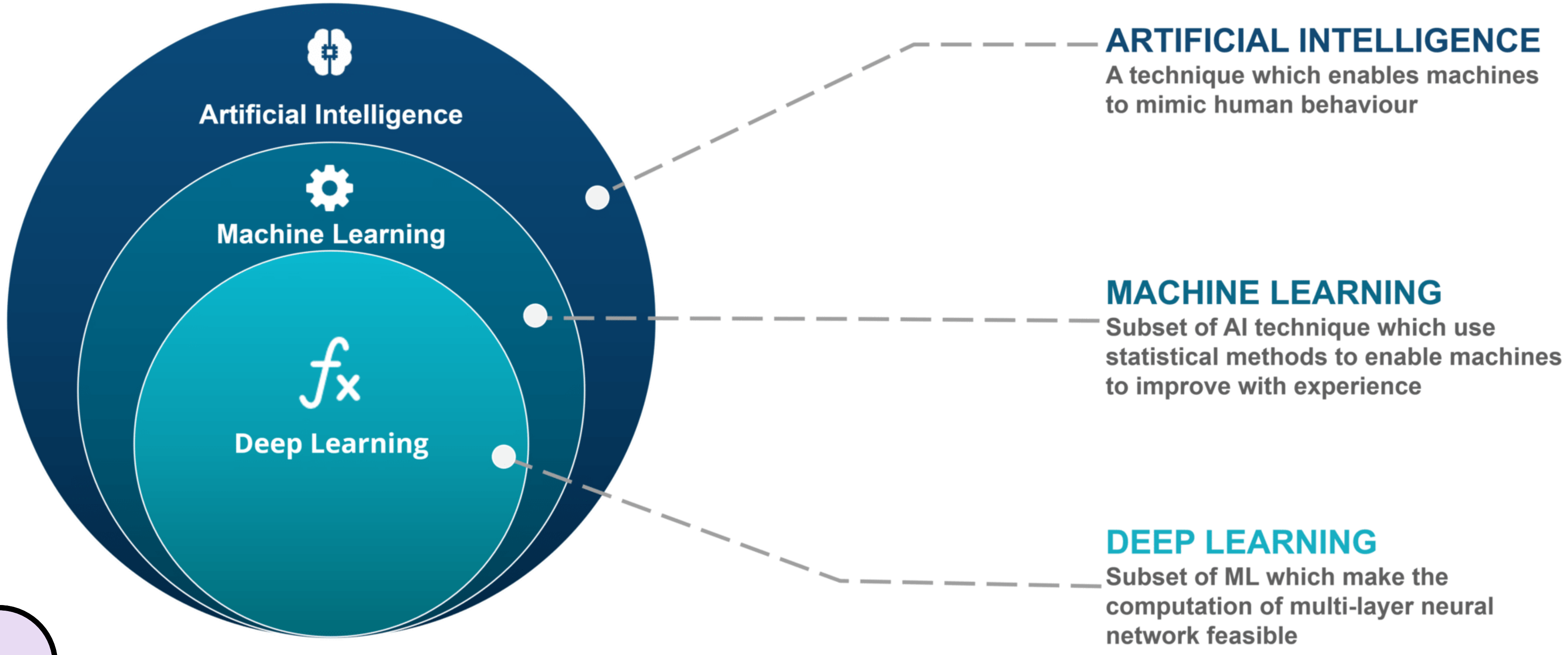
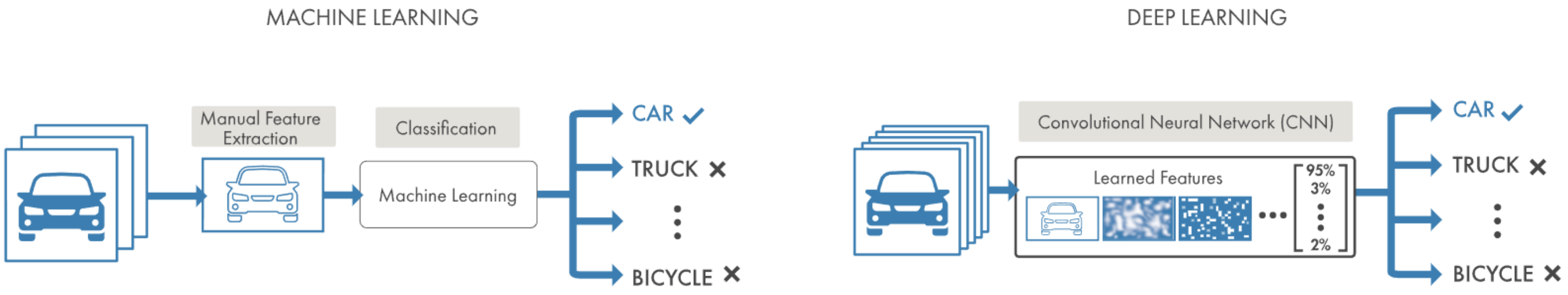
DEEP LEARNING IN HEALTHCARE

PRESENT BY
WING CHAN

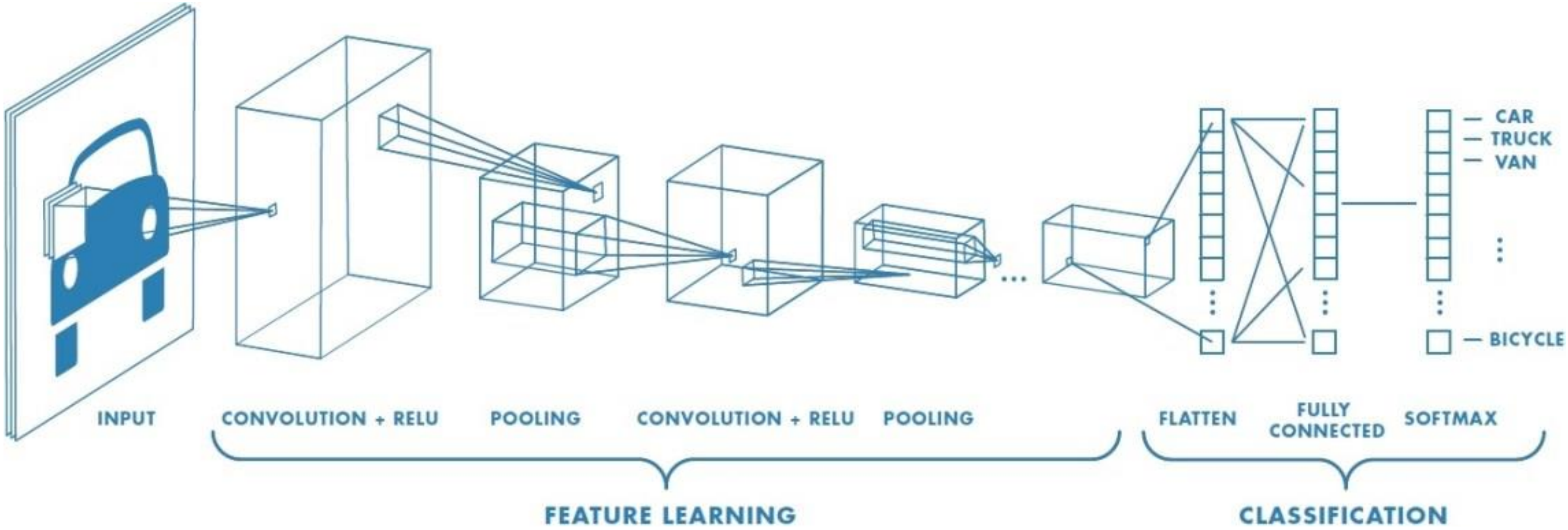

Artificial Intelligence
Machine Learning
Deep Learning
ARTIFICIAL INTELLIGENCE
A technique which enables machines to mimic human behaviour
MACHINE LEARNING
Subset of AI technique which use statistical methods to enable machines to improve with experience
DEEP LEARNING
Subset of ML which make the computation of multi-layer neural network feasible



In machine learning, you manually choose features and a classifier to sort images. In deep learning, feature extraction and modeling steps are automatic.

A successful deep learning application requires a very large amount of labeled data (thousands of images) to train the model, as well as GPUs, or graphics processing units, to rapidly process your data.

https://www.mathworks.com/discovery/deep-learning.html



https://www.mathworks.com/videos/introduction-to-deep-learning-what-are-convolutional-neural-networks--1489512765771.html

Applications of Deep Learning in Healthcare

- Better Imaging & Diagnostic
- Drug Discovery & Research
- Detecting Diseases in Early Stage
- Providing Personalized Treatment
- Clinical Decision Support
- Preventing Medical Insurance Frauds

UNIVERSITY OF ILLINOIS URBANA-CHAMPAIGN

TEAM RESEARCH PAPER

CS 598 – Deep Learning for Healthcare
Instructor: Jimeng Sun

Team: Wing Chan, Xiaohan Tian, Marcus Tan, Joshua Ceaser

https://mediaspace.illinois.edu/media/t/1_j816sn0o

Multi-Label Classification and Visual Explanation of Chest X-ray Images using Neural Networks with Attention Mechanism and Grad-CAM

https://mediaspace.illinois.edu/media/t/1_j816sn0o

Marcus Hwai Yik Tan, Xiaohan Tian, Wing Chan, Joshua Ceaser

University of Illinois, Urbana-Champaign

ABSTRACT

We develop a method to select a subset of images from the full Chest-Xray14 data set to allow for the training of deep learning models with limited computing resources. The resulting subset contains approximately 20% of the original images. To determine if the subset of images is sufficiently large to provide meaningful results, we compute AUROC scores and compare the performance of the DenseNet-121 and ResNet-50 models used in the CheXNet study. Average validation and test AUROC scores as high as 0.7 or greater can be achieved with those models with pre-trained weights. The superior performance of DenseNet-121 over ResNet-50 shown in the CheXNet study is reproduced with the smaller data set. Two variants of an attention mechanism (DenseNet-121-attA and DenseNet-121-attB) are added to the DenseNet-121 model and shown to improve the test AUROC between 0.02 and 0.03 when all models were trained with a learning rate of 0.01 for 8 epochs. Visual explanation of model prediction is provided with heat maps generated using the Grad-CAM method.

1. INTRODUCTION AND BACKGROUND

Chest X-rays are widely used as a diagnostic method to examine a patient's current respiratory system for illnesses and prioritize patient care [3]. Due to the large number of images and limited access to experienced radiologists who can interpret the images accurately, deep learning models have been used on the images to assist in determining the presence of a disease, and if a disease is present, to identify the disease and highlight the diseased region [12].

1.1 Literature Review

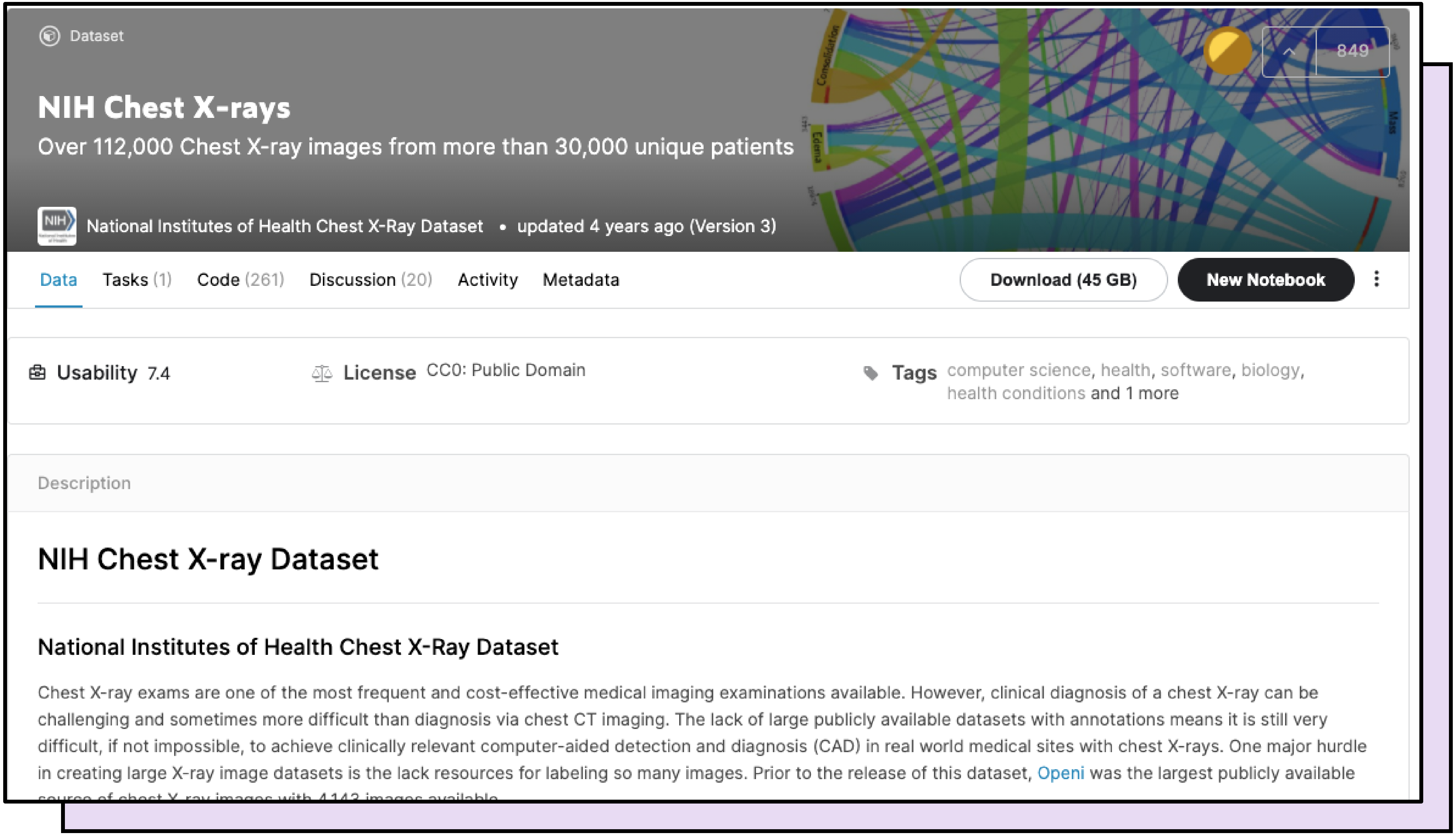
Existing literature predominantly uses three datasets when classifying chest X-rays: ChestX-ray14 [18], CheXpert [9] and JSRT [16]. ChestX-ray14 was originally called ChestX-ray8 as there were 8 disease labels, but was renamed when an additional 6 disease classes were added. The dataset was created by Wang et al. using radiology reports and images from the Picture Archiving and Communication systems of NIH [18]. Reports with keywords corresponding to 8 common thoracic pathologies and the corresponding images were extracted from the system. Natural Language Processing (NLP) techniques were applied on the reports to obtain the disease classifications for the images. AlexNet, GoogLeNet, VGGNet-16 and ResNet-50 models pre-trained on ImageNet images were used. All models had AUROCs between 0.51 to 0.81 for all disease classes and ResNet-50 had the best performance for all classes except one.

Yao et al. proposed an encoder model similar to DenseNet [8] with LSTM decoders exploiting the dependencies between labels. Trained on the ChestX-ray14 dataset, the encoder-only model had higher AUROCs in 13 out of 14 pathologies than the pre-trained models used by Wang et al. [18]. Rajpurka et al. subsequently trained a 121-layer DenseNet to obtain a model called CheXNet that had higher AUROCs than those obtained by Yao et al. [19] and Wang et al. [18]. Liu et al. proposed a segmentation-based model that first trained a model (called Lung Region Generator) on the JSRT dataset to extract the lung regions of the ChestX-ray14 images [11]. Two CNNs were then applied to the entire image and the lung region, which were then combined with a fusion model to predict the pathology labels. Models using the entire image and the lung regions separately outperformed the models by Wang et al. [18] and Yao et al. [19] in all classes except one for the lung-region-only model. Neither model (entire image or lung region only) is consistently better than the other but the fusion model is consistently better than both.

CheXpert is a relatively newer and larger dataset created by Irvin et al. using chest radiographic data from Stanford Hospital [9]. The disease classifications were extracted using a labeler designed by the authors and were claimed to be more accurate than those used to build the ChestX-ray14 dataset. An additional uncertainty label was also introduced. The study trained several CNN models such as ResNet-152, DenseNet-121, Inception-v4 etc. and found DenseNet-121 to have the best performance. The best model outperforms 2 or 3 radiologists in 4 of the disease classes on a test set.

1.2 Data Exploration

The image dataset chosen for this project is the ChestX-ray14 dataset consisting of 112120 frontal view images, where each image is associated with one of fourteen disease classes or a "No Finding" class. This dataset is an augmented version of the original ChestX-ray8, which has a slightly smaller



Dataset

NIH Chest X-rays

Over 112,000 Chest X-ray images from more than 30,000 unique patients

National Institutes of Health Chest X-Ray Dataset • updated 4 years ago (Version 3)

849

Data Tasks (1) Code (261) Discussion (20) Activity Metadata

Download (45 GB) New Notebook

Usability 7.4

License CC0: Public Domain

Tags computer science, health, software, biology, health conditions and 1 more

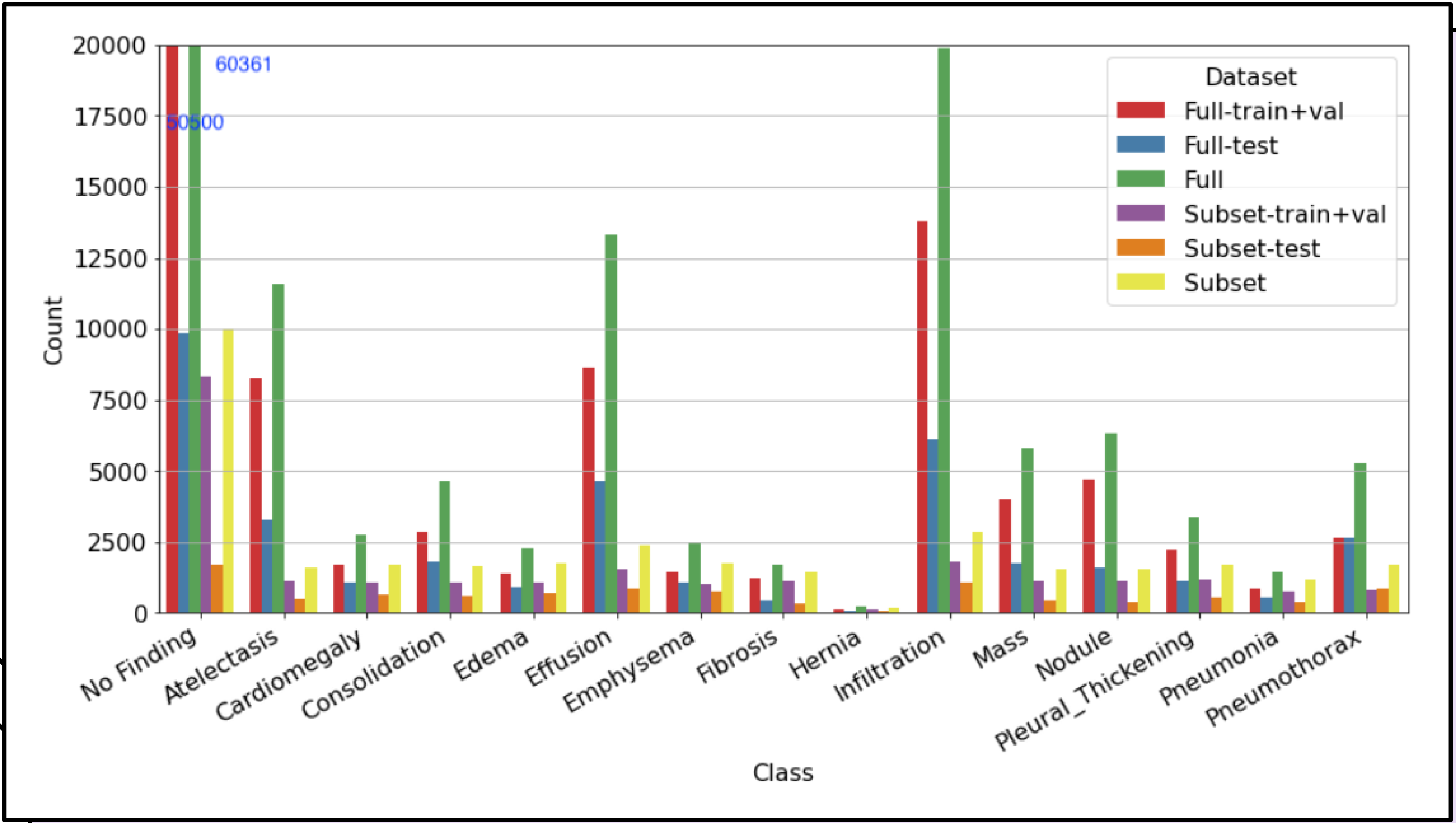
Description

NIH Chest X-ray Dataset

National Institutes of Health Chest X-Ray Dataset

Chest X-ray exams are one of the most frequent and cost-effective medical imaging examinations available. However, clinical diagnosis of a chest X-ray can be challenging and sometimes more difficult than diagnosis via chest CT imaging. The lack of large publicly available datasets with annotations means it is still very difficult, if not impossible, to achieve clinically relevant computer-aided detection and diagnosis (CAD) in real world medical sites with chest X-rays. One major hurdle in creating large X-ray image datasets is the lack resources for labeling so many images. Prior to the release of this dataset, Openi was the largest publicly available

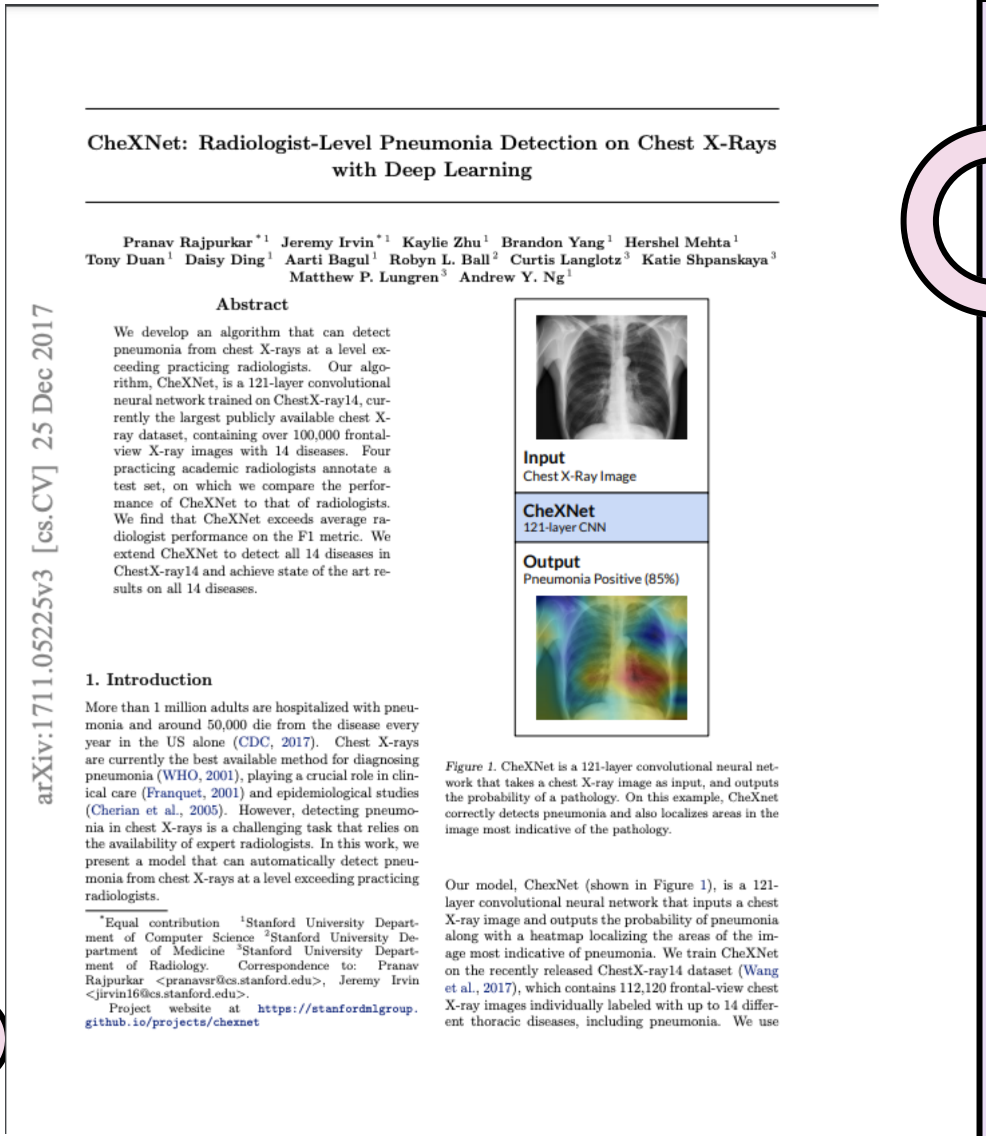
https://www.kaggle.com/nih-chest-xrays/data



Dataset
Full-train+val
Full-test
Full
Subset-train+val
Subset-test
Subset
60361
50500
Count
20000
17500
15000
12500
10000
7500
5000
2500
0
No Finding
Atelectasis
Cardiomegaly
Consolidation
Edema
Effusion
Emphysema
Fibrosis
Hernia
Infiltration
Mass
Nodule
Pleural_Thickening
Pneumonia
Pneumothorax
Class

O B J E C T I V E S

1. Implement **DenseNet-121** and **ResNet-50** models used in the CheXNet paper
2. Use **Attention mechanism** and **Non-image features** to improve DenseNet-121 performance
3. Visualized and interpreted the results using **Heatmaps**



arXiv:1711.05225v3 [cs.CV] 25 Dec 2017

CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning

Pranav Rajpurkar [*1] Jeremy Irvin [*1] Kaylie Zhu [1] Brandon Yang [1] Hershel Mehta [1] Tony Duan [1] Daisy Ding [1] Aarti Bagul [1] Robyn L. Ball [2] Curtis Langlotz [3] Katie Shpanskaya [3] Matthew P. Lungren [3] Andrew Y. Ng [1]

Abstract

We develop an algorithm that can detect pneumonia from chest X-rays at a level exceeding practicing radiologists. Our algorithm, CheXNet, is a 121-layer convolutional neural network trained on ChestX-ray14, currently the largest publicly available chest X-ray dataset, containing over 100,000 frontal-view X-ray images with 14 diseases. Four practicing academic radiologists annotate a test set, on which we compare the performance of CheXNet to that of radiologists. We find that CheXNet exceeds average radiologist performance on the F1 metric. We extend CheXNet to detect all 14 diseases in ChestX-ray14 and achieve state of the art results on all 14 diseases.

1. Introduction

More than 1 million adults are hospitalized with pneumonia and around 50,000 die from the disease every year in the US alone (CDC, 2017). Chest X-rays are currently the best available method for diagnosing pneumonia (WHO, 2001), playing a crucial role in clinical care (Franquet, 2001) and epidemiological studies (Cherian et al., 2005). However, detecting pneumonia in chest X-rays is a challenging task that relies on the availability of expert radiologists. In this work, we present a model that can automatically detect pneumonia from chest X-rays at a level exceeding practicing radiologists.

Input
Chest X-Ray Image
CheXNet
121-layer CNN
Output
Pneumonia Positive (85%)

Figure 1. CheXNet is a 121-layer convolutional neural network that takes a chest X-ray image as input, and outputs the probability of a pathology. On this example, CheXnet correctly detects pneumonia and also localizes areas in the image most indicative of the pathology.

Our model, ChexNet (shown in Figure 1), is a 121-layer convolutional neural network that inputs a chest X-ray image and outputs the probability of pneumonia along with a heatmap localizing the areas of the image most indicative of pneumonia. We train CheXNet on the recently released ChestX-ray14 dataset (Wang et al., 2017), which contains 112,120 frontal-view chest X-ray images individually labeled with up to 14 different thoracic diseases, including pneumonia. We use

[*] Equal contribution [1] Stanford University Department of Computer Science [2] Stanford University Department of Medicine [3] Stanford University Department of Radiology. Correspondence to: Pranav Rajpurkar <pranavsr@cs.stanford.edu>, Jeremy Irvin <jirvin16@cs.stanford.edu>.

Project website at https://stanfordmlgroup.github.io/projects/chexnet

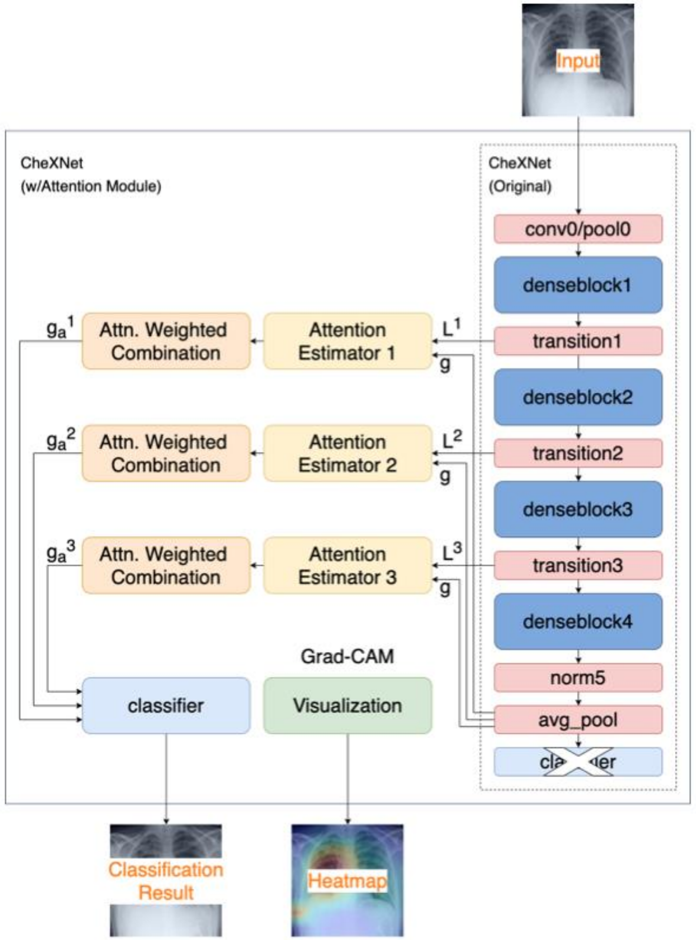
https://arxiv.org/pdf/1711.05225.pdf

Layers	Output Size	DenseNet-121	DenseNet-169	DenseNet-201	DenseNet-264
Convolution	112×112	7×7 conv, stride 2			
Pooling	56×56	3×3 max pool, stride 2			
Dense Block (1)	56×56	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$
Transition Layer (1)	56×56	1×1 conv			
	28×28	2×2 average pool, stride 2			
Dense Block (2)	28×28	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$
Transition Layer (2)	28×28	1×1 conv			
	14×14	2×2 average pool, stride 2			
Dense Block (3)	14×14	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 24$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 32$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 48$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 64$
Transition Layer (3)	14×14	1×1 conv			
	7×7	2×2 average pool, stride 2			
Dense Block (4)	7×7	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 16$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 32$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 32$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 48$
Classification Layer	1×1	7×7 global average pool			
		1000D fully-connected, softmax			

https://amaarora.github.io/2020/08/02/densenets.html

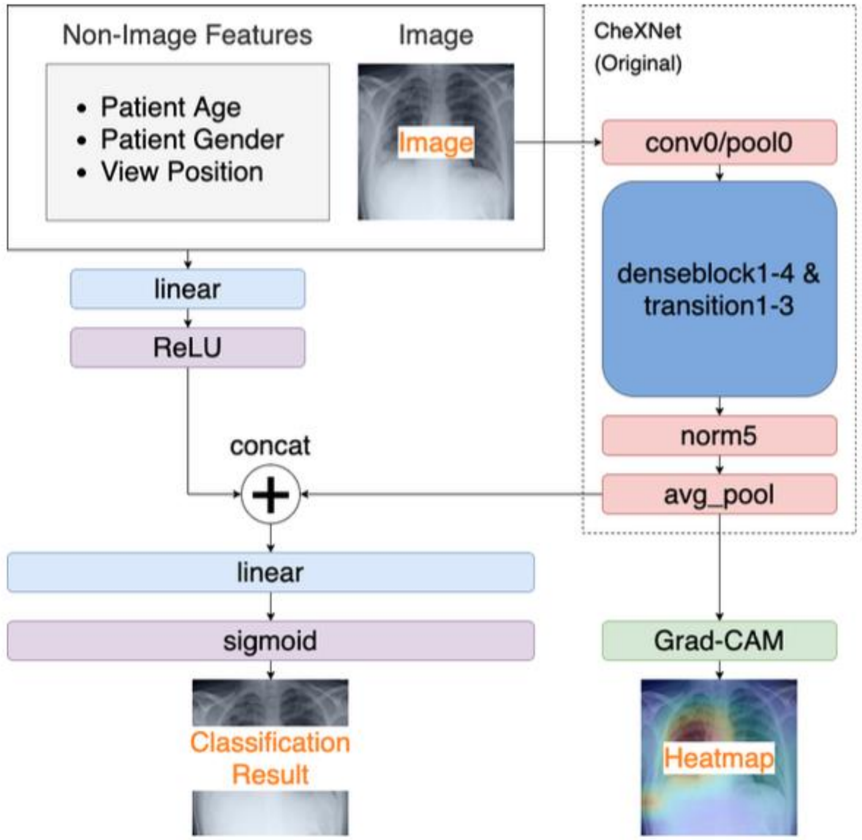
MODELING APPROACH #1

DenseNet-121 with Attention Mechanism



MODELING APPROACH #2

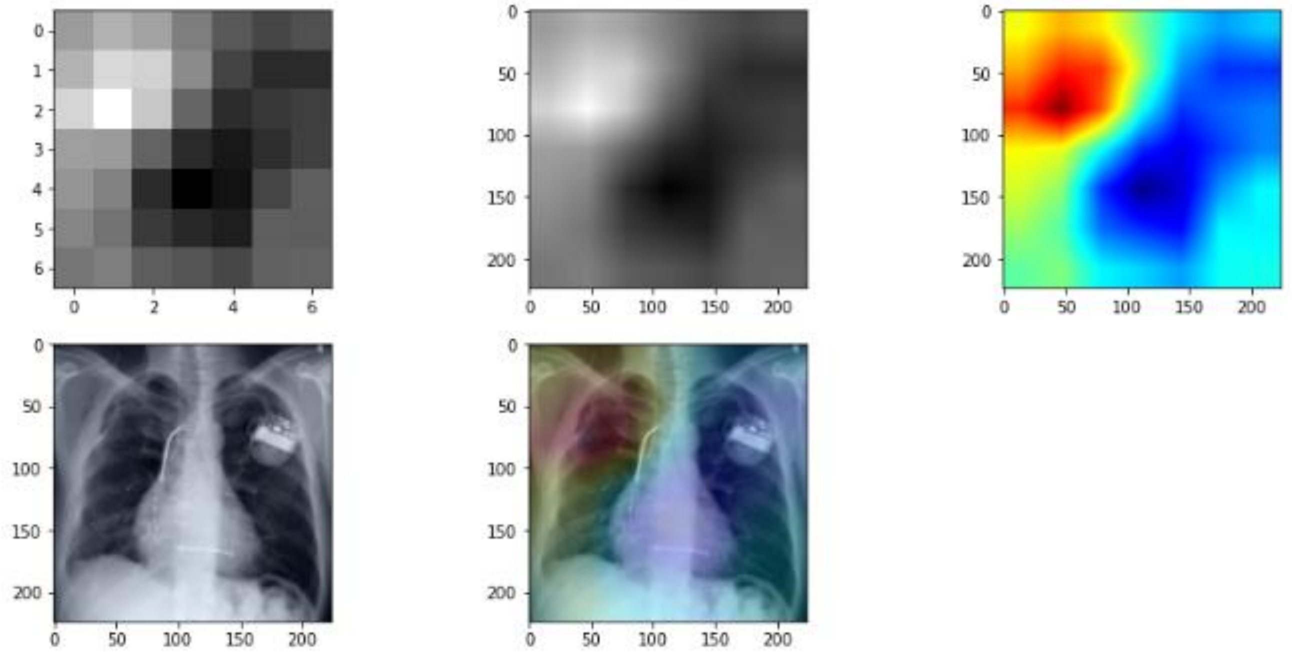
DenseNet-121 with non-image features



Heatmap Generation with Grad-CAM

Grad-CAM is a popular technique for visualizing where a convolutional neural network model is looking. Grad-CAM is class-specific, meaning it can produce a separate visualization for every class present in the image.

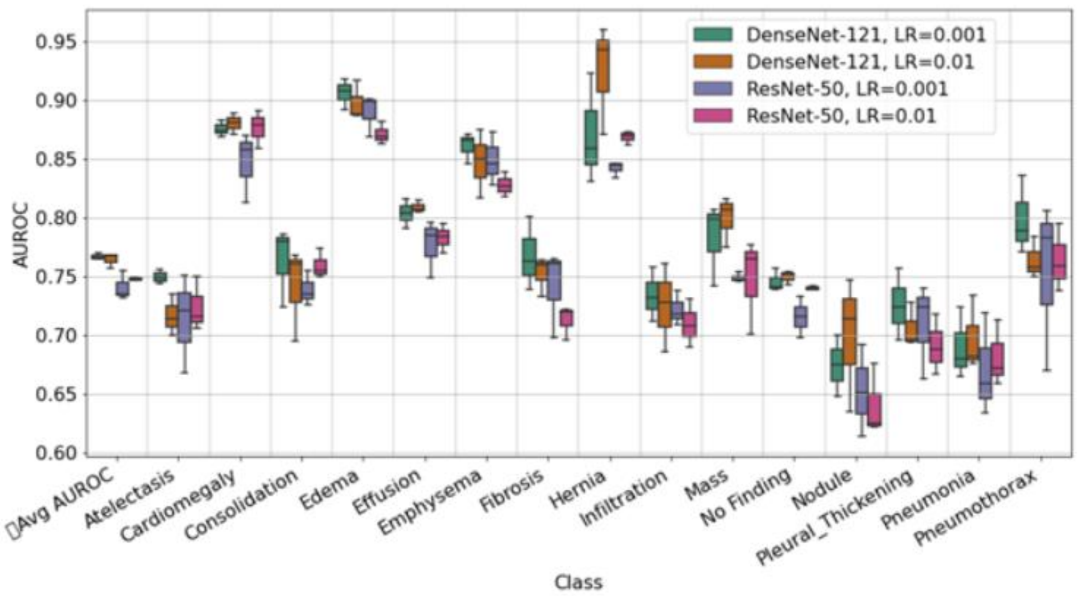
Grad-CAM method developed by Selvaraju et al.
https://arxiv.org/pdf/1610.02391.pdf

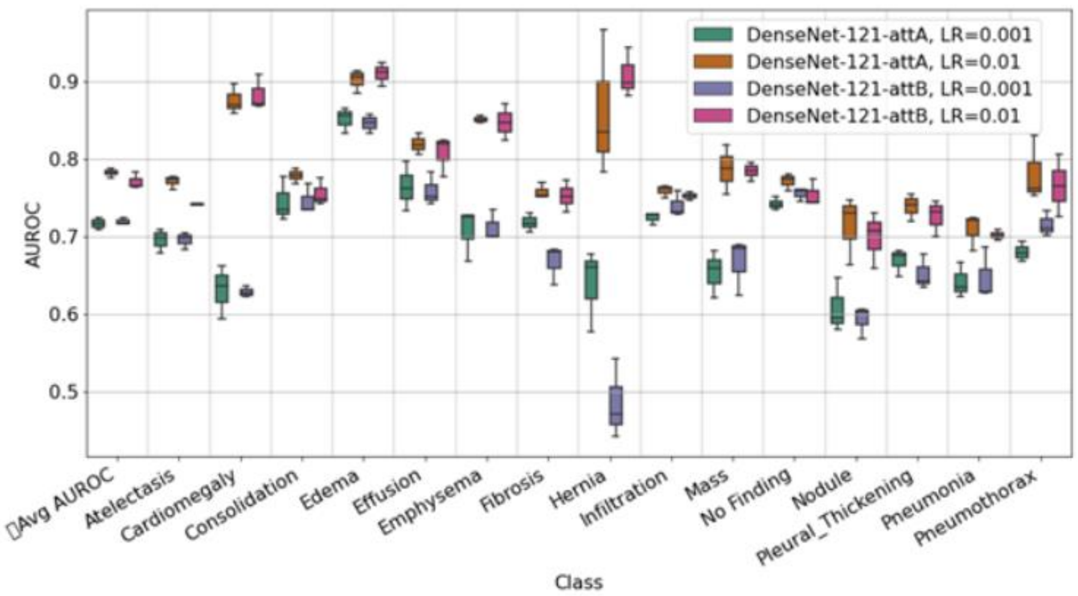


Model Training Environment

- Software Modules
  - Pytorch
  - Pytorch Vision and Pillow Image
  - Numpy
  - Pandas
  - Scikit-Learn
- Hardware Modules
  - PC with Nvidia GeForce RTX
  - AWS EC2 p2.Xlarge instances

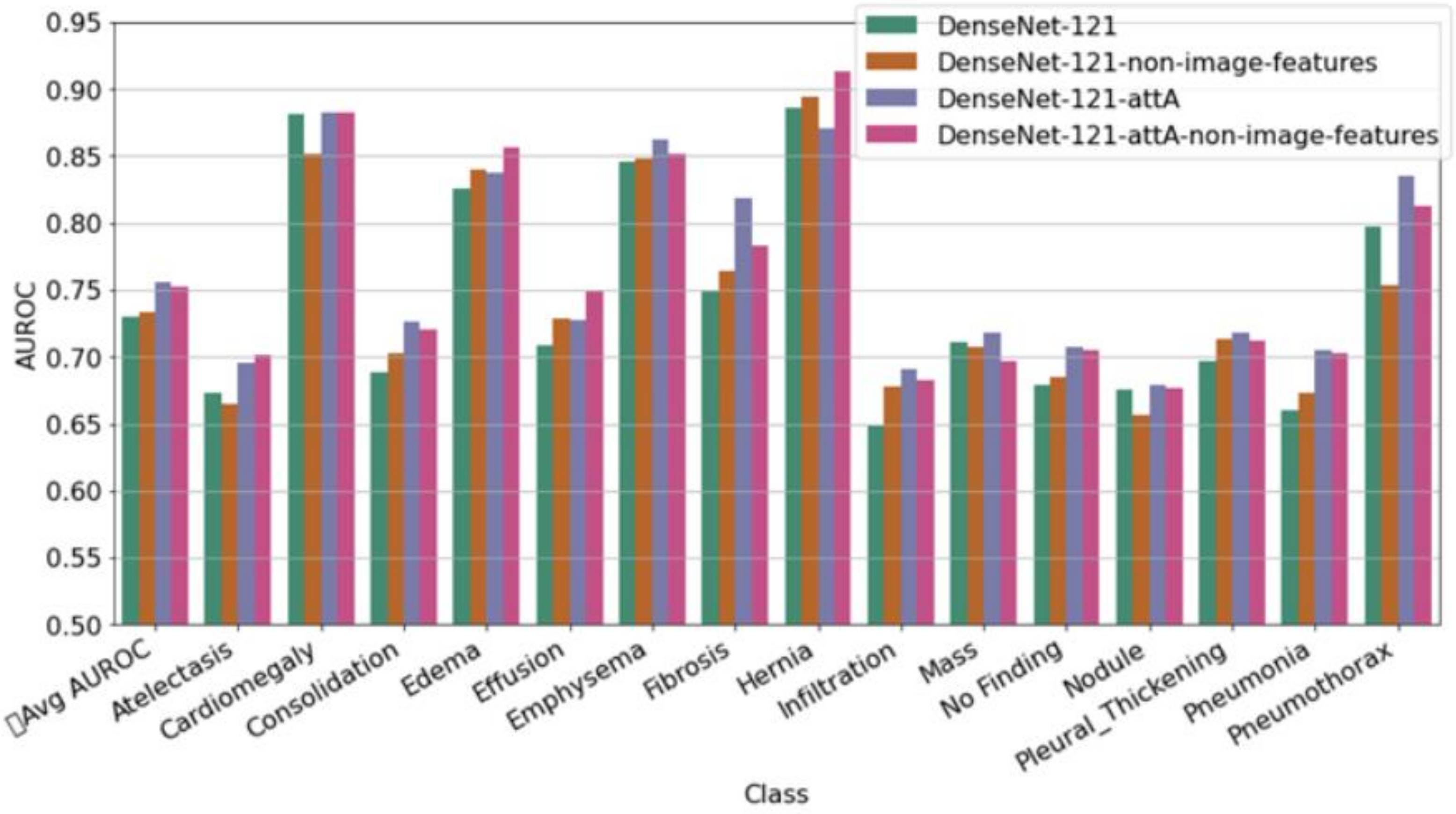
Result: with attention mechanism





1. Baseline models - DenseNet-121 has higher AUROC scores than ResNet-50 like CheXNet study
2. Attention models – have a higher average AUROC scores than baseline models

Result: with non-image features



Result - minor impact of performance by adding non-image features to input

Result: Heat maps