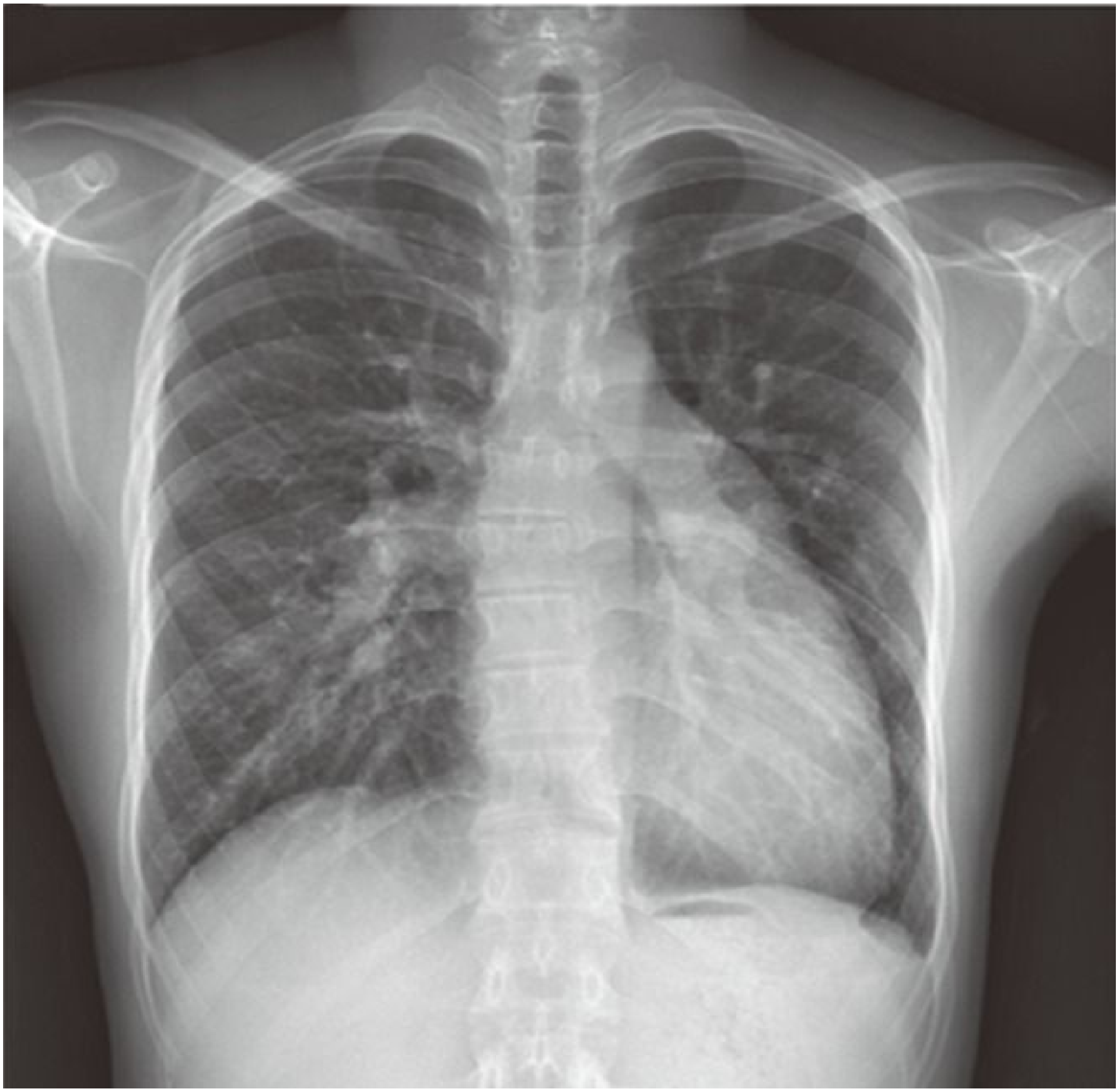


Classification of cardiomegaly using CNN

12.20.2017

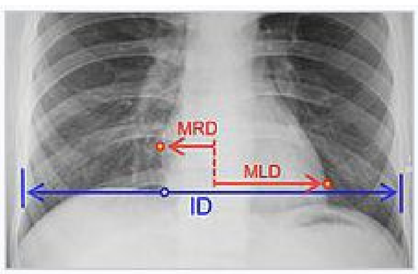
Anoop Singh

Domain Background

From [Wikipedia](#):

Cardiomegaly is a medical condition in which the heart is enlarged. It is more commonly referred to as an enlarged heart. The causes of cardiomegaly may vary. Many times this condition results from high blood pressure (hypertension) or coronary artery disease. An enlarged heart may not pump blood effectively, resulting in congestive heart failure.

X-ray images help see the condition of the lungs and heart. If the heart is enlarged on an X-ray, other tests will usually be needed to find the cause. A useful measurement on X-ray is the cardio-thoracic ratio, which is the transverse diameter of the heart, compared with that of the thoracic cage." These diameters are taken from PA chest x-rays using the widest point of the chest and measuring as far as the lung pleura, not the lateral skin margins. If the cardiac thoracic ratio is greater than 50%, pathology is suspected, assuming the x-ray has been taken correctly. The measurement was first proposed in 1919 to screen military recruits. A newer approach to using these x-rays for evaluating heart health, takes the ratio of heart area to chest area and has been called the two-dimensional cardiothoracic ratio.



Cardiothoracic ratio = $\frac{MRD + MLD}{ID}$

where,[4]

MRD = greatest perpendicular diameter from midline to right heart border

MLD = greatest perpendicular diameter from midline to left heart border

ID = internal diameter of chest at level of right hemidiaphragm

X-ray exams are the first step in diagnosing cardiomegaly in a patient. Once the x-ray's available, a radiologist looks at it and tries to diagnose the disease.

From GlobalDiagnostiX - Context: *According to WHO figures, **more than two thirds of the world's population does not have access** to this essential x-ray imaging equipment. Too often in developing countries, patients die of trivial problems, which, due to a lack of access to diagnosis, take dramatic proportions.*

From Most of the World Doesn't Have Access to X-Rays: *After the 2010 earthquake in Haiti, Mendel and Partners in Health stocked the University Hospital in Mirebalais (UHM) with a CT scanner—the first in a public hospital in Haiti and the first to cost their patients nothing. But the hospital still doesn't have enough money to hire a radiologist to run the machine, Mendel says.*

One solution is telemedicine. UHM uses a picture-archiving and communication system that sends CT scans to a server in Boston, which stores the images and creates an electronic medical record for volunteer radiologists in the U.S. and Canada to read. The volunteers log on twice a week to look over scans, which each take about ten minutes. In 2014, 40 volunteers read approximately 4,000 CT scans.

But telemedicine has limits, especially in an emergency. "Bus accidents happen every day in Nepal," Schwarz explains. "You have literally 25 patients all at once, who are all bleeding. That's challenging enough. You're certainly not waiting for someone in a different country or time zone to tell you what an x-ray shows."

With initiatives like GlobalDiagnostiX, there's hope that low cost x-ray systems will be more readily available in underdeveloped countries as time progresses. Radiologist availability still remains limited and lives are lost while patients wait for diagnosis.

The ImageNet challenge has led to successful advances in the field of computer vision and the ability to use convolutional neural networks for image recognition tasks. Using transfer learning, there have been several instances where a successful ImageNet architecture's used and modified/re-trained to recognize images in a particular field with high accuracy.

Due to the availability of large datasets, progress in computer hardware and an active interest in using machine learning for medical diagnosis, researchers have demonstrated at-par or better performance when compared to medical professionals.

Radiologists will be 'obsolete' in five years led me to gain interest in this particular problem. If we can solve this problem, it can drastically bring down diagnosis time (particularly in developing countries) and potentially make diagnosis cheaper.

When compared to medical professionals, we may get better diagnosis as well due to:

- the ability to train on tens and hundreds of thousands of images, to pick up intricate details
- no overworked radiologists who have to occasionally screen over 100 x-rays a day, potentially leading to errors

Research Citations

- ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases
- CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning

Problem Statement

Develop an algorithm that can detect cardiomegaly from chest X-rays.

This is a binary classification problem: does the image exhibit Cardiomegaly or not?

Inputs for the problem: Images labeled as *Cardiomegaly* or *No Finding*, we'll ignore all the other features.

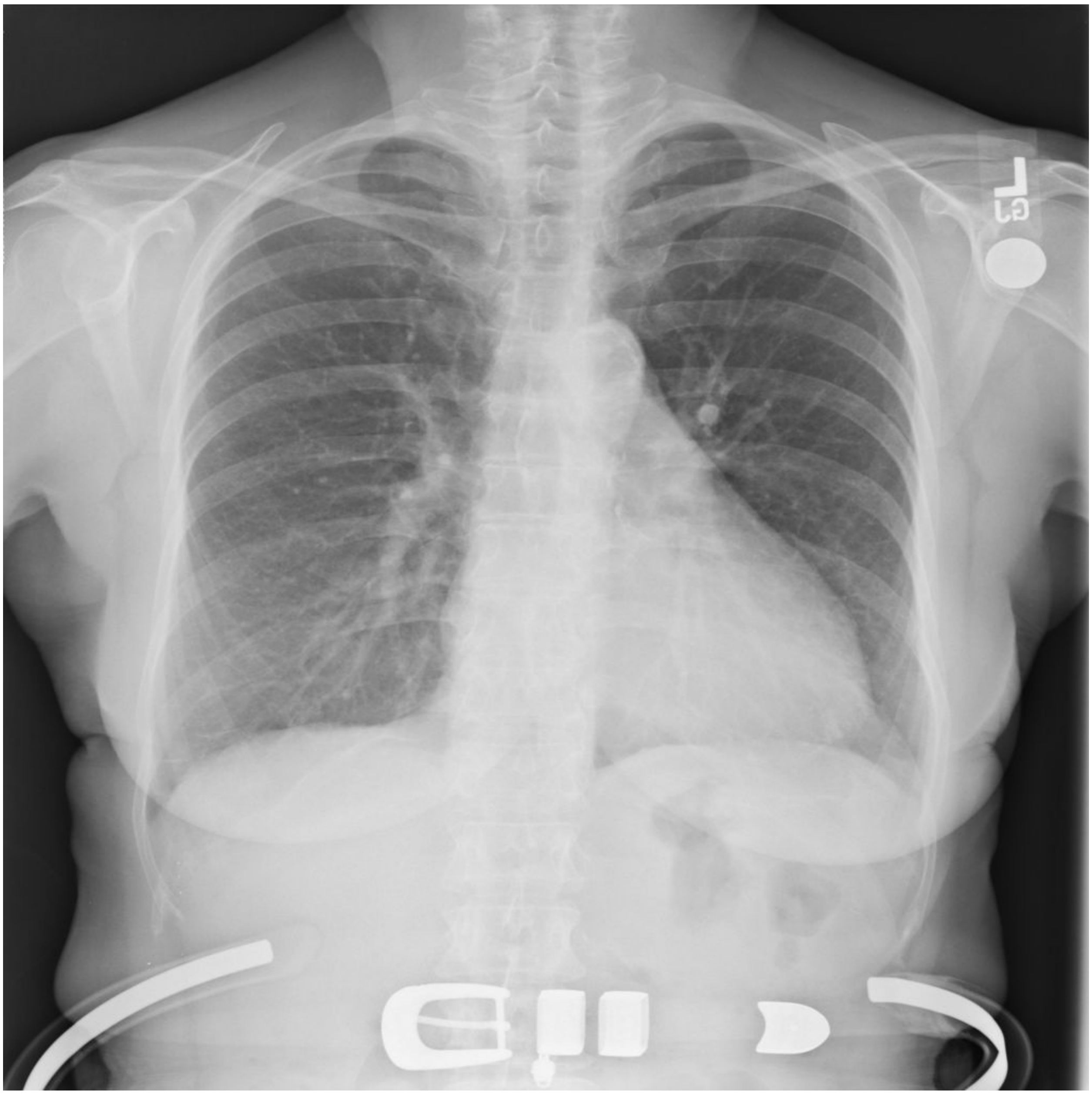
Outputs for the problem: label for the image (*Cardiomegaly* or *No Finding*)

Datasets and Inputs

CXR8 dataset provided by the National institute of Health Clinical Center

It has labels that diagnose a chest x-ray into the following 8 thoracic pathologies: *At-electasis, Cardiomegaly, Effusion, Infiltration, Mass, Nodule, Pneumonia* and *Pneumathorax*. An image may be labeled as one or more of these pathologies. A *No Finding* label indicates that the image was not labeled as any of the 8 pathologies.

Sample image (PNG, 1024 x 1024, 414 KB):



Data:

Image Index,Finding Labels,Follow-up #,Patient ID,Patient Age,Patient Gender,View Position,OriginalImage[Width,Height],OriginalImagePixelSpacing[x,y],

00000001_000.png,Cardiomegaly,0,1,58,M,PA,2682,2749,0.143,0.143

00000001_001.png,Cardiomegaly|Emphysema,1,1,58,M,PA,2894,2729,0.143,0.143

00000002_000.png,No Finding,0,2,81,M,PA,2500,2048,0.171,0.171

..

..

00000008_000.png,Cardiomegaly,0,8,69,F,PA,2048,2500,0.171,0.171 (for sample image)

- How are the images structured?
    - PNG, 1024 x 1024, 400 - 500 KB
- Are there color layers?
    - The images are in grayscale and should not need any color transformation before they can be used
- What are the dimensions (or ranges of dimensions)?
    - 1024 x 1024
- How many examples are there in the dataset that you'll be using?
    - The dataset contains over 100,000 anonymized chest x-ray from more than 30,000 patients and takes 45.14 GB of space when compressed
    - The *Project Design* section talks about making a few changes, restricting us to a subset of the dataset
- How many classes will there be (just two? or more?) and are they balanced?
    - Just two - *Cardiomegaly* and *No Finding*
    - are they balanced? - No, the dataset will be biased towards more images with the *No Finding* label
- How will you split the data into training/validation/testing sets?
    - May start off with a 60 - 20 - 20 split and see how it goes
- Will you do anything to maintain class balances across each subset?
    - Yes, will aim for a consistent ratio of images labeled *Cardiomegaly* to images labeled *No Finding* within each subset

Solution Statement / Project Design

With a subset of the CXR8 dataset, we'll attempt binary classification of images to detect cardiomegaly. We'll try several successful ImageNet CNN architectures as starting points and modify/re-train them to suit our needs (transfer learning).

To limit the scope of the problem and be able to train a solution locally on a MacBook Air, we'll modify the data as follows:

- Consider reducing the image size to 224 x 224 (similar to CheXNet)
- Restrict ourselves to images of *Cardiomegaly* and *No Finding* labels only
- Restrict ourselves to a subset of images - dataset has 2,776 *Cardiomegaly* images and 60,361 *No Finding* images

Without running the numbers, between these three approaches, we should be able to come down from 45 GB to something manageable around 5 GB max.

If training time exceeds a day, AWS or Google Tensor Processing Units will be considered.

In addition to this, we may need to use Keras Image Augmentation API to improve the algorithm's results. Since all images are already in the correct orientation, I doubt if we'll gain much from rotating images in the augmentation step. We may benefit from shifting them a bit.

We'll try our own CNN architecture and get a performance baseline. This'll be compared to the ImageNet architectures mentioned below once re-trained with transfer learning:

- AlexNet
- Inception
- Xception
- VGGNet-16
- VGGNet-19
- ResNet-50

Benchmark Model

We'll come up with our own CNN architecture and measure its AUC. This can later be compared against an ImageNet architecture which's re-trained on our dataset.

Pathology	Wang et al. (2017)	Yao et al. (2017)	CheXNet (ours)
Atelectasis	0.716	0.772	**0.8094**
Cardiomegaly	0.807	0.904	**0.9248**
Effusion	0.784	0.859	**0.8638**
Infiltration	0.609	0.695	**0.7345**
Mass	0.706	0.792	**0.8676**
Nodule	0.671	0.717	**0.7802**
Pneumonia	0.633	0.713	**0.7680**
Pneumothorax	0.806	0.841	**0.8887**
Consolidation	0.708	0.788	**0.7901**
Edema	0.835	0.882	**0.8878**
Emphysema	0.815	0.829	**0.9371**
Fibrosis	0.769	0.767	**0.8047**
Pleural Thickening	0.708	0.765	**0.8062**
Hernia	0.767	0.914	**0.9164**

Table 1. CheXNet outperforms the best published results on all 14 pathologies in the ChestX-ray14 dataset. In detecting Mass, Nodule, Pneumonia, and Emphysema, CheXNet has a margin of >0.05 AUROC over previous state of the art results.

Based on findings from CheXNet, we see AUROC ranging from 0.807 to 0.9248 for *Cardiomegaly*. This can serve as a secondary benchmark.

Evaluation Metrics

I. AUC - Area Under The Curve

Since our data set will have a strong bias towards *No Finding* vs *Cardiomegaly*, AUC is a better metric than accuracy. We'll compare our results with observations from the benchmark model.

Software Requirements

Python, NumPy, Pandas, Scikit-learn, IPython Notebook, Keras, TensorFlow