Building the Standard: Data Curation for All-in-One Detection, Segmentation, and Risk Scoring

Tom Sanford

NIH NATIONAL CANCER INSTITUTE

MIP MOLECULAR IMAGING PROGRAM

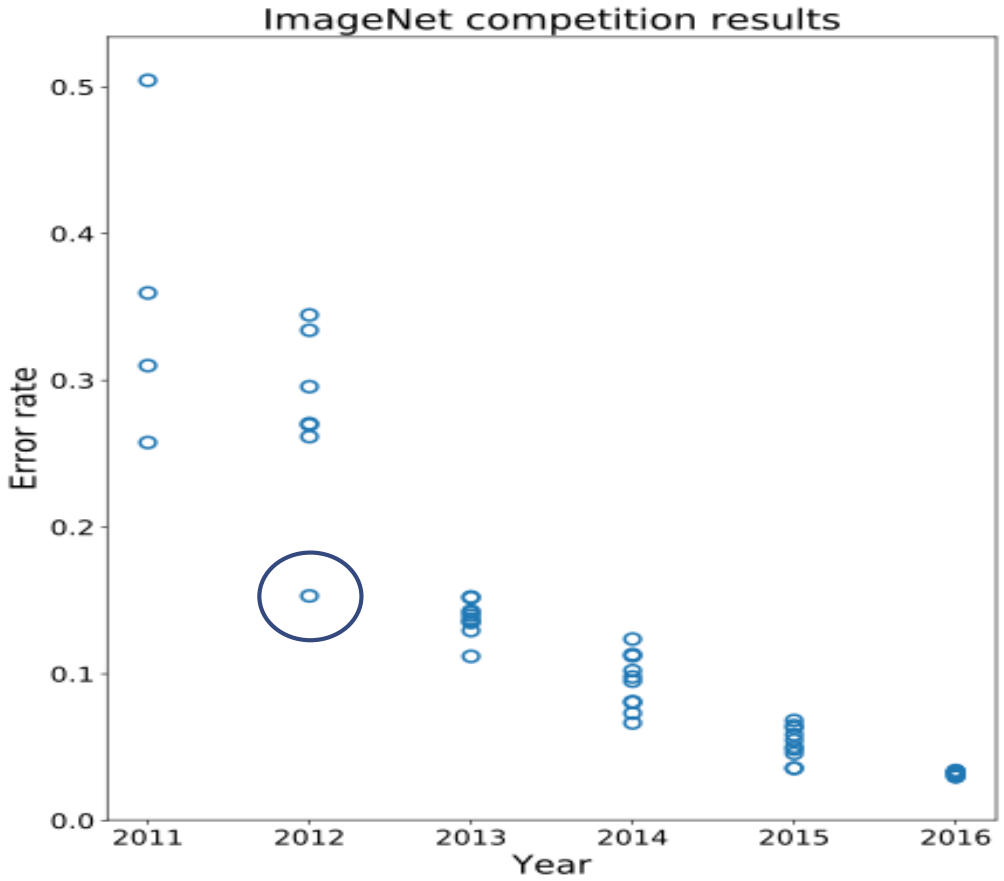
Need for AI in Medical Imaging

Questions Addressed by Urologists for Patients with Prostate Cancer

- Which patients need a biopsy?
- When performing a biopsy, which lesions do we need to target?
- What is the best treatment for patients diagnosed with Prostate Cancer?
- What is the best treatment for various degrees of prostate cancer?
- Which patients need additional treatment after the initial treatment?

Excitement for the Potential of AI in Medical Imaging

Novel
Architecture
+
Well Curated
Dataset

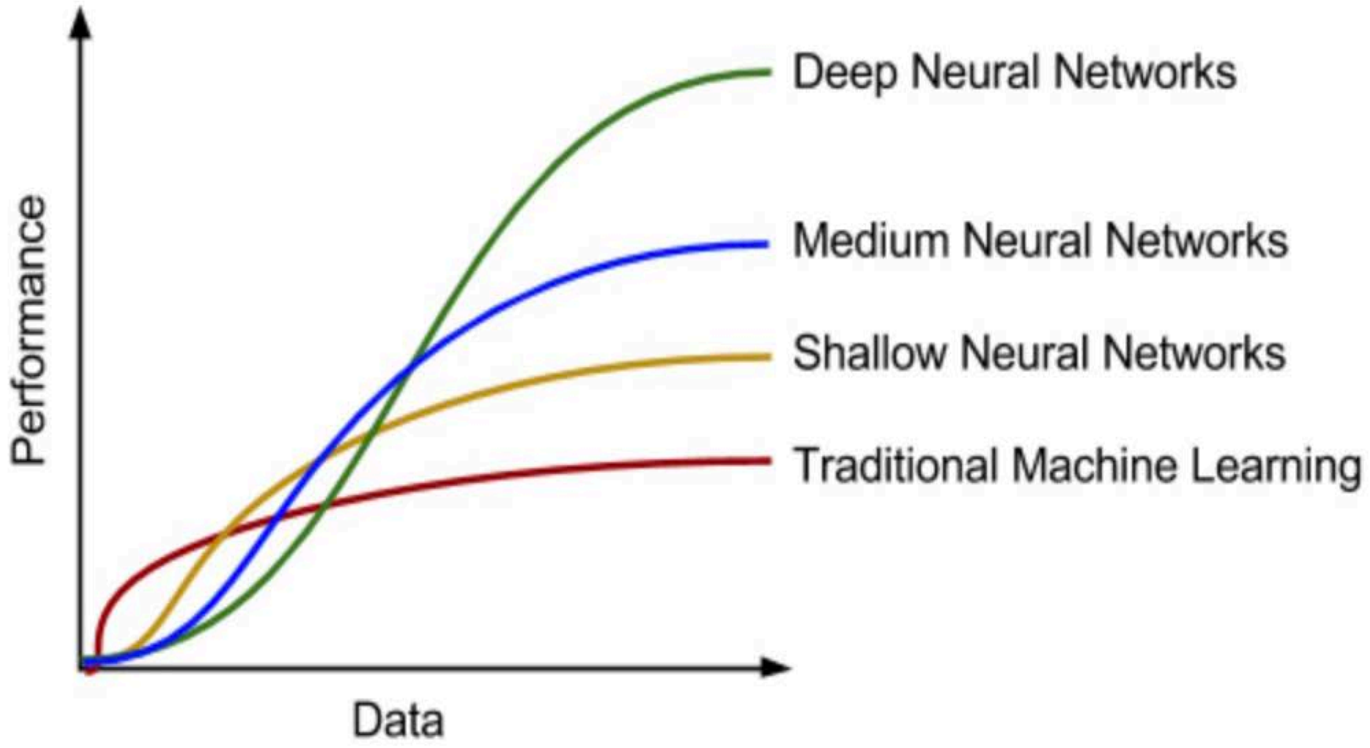


https://en.wikipedia.org/wiki/ImageNet

Medical Images vs Natural World Images: Opportunities for Novel Architectures

RGB images (i.e. Dematology)	Radiology	Pathology
Small features with important consequences (	Greyscale images 3D images represented in 2D slices Multiple phases/series	Multiple resolutions Large file sizes

Deep Neural Networks Require Data



Challenges to Building a Dataset for Medical Imaging

- Access
  - Patients should consent!
  - Institutional Review Board required to review protocol, data storage policies.
- Obtaining Raw Data
  - EMR access restricted
  - Limited network bandwith for transferring files
  - Files with many formats
- Labeling

Obtaining Labeled Data

Assuming one annotation takes 10 minutes, creating a labeled dataset the size of ImageNet would:

- take radiologists 2.4 million hours
- Cost $4.8 billion (over 10% annual NIH budget)

Creating labels in radiology

- It depends on the task

Our Task

- Develop an automated methodology for prostate lesion detection, classification, and grading

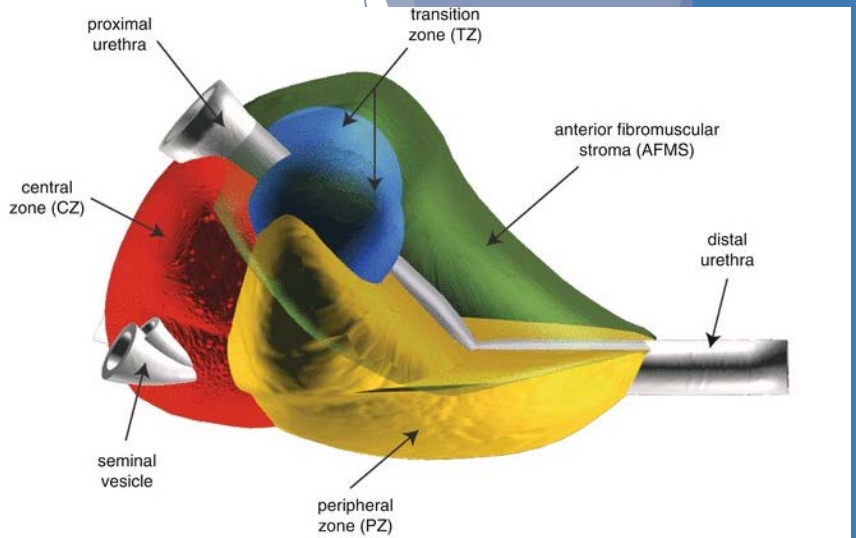
Introduction to the Prostate

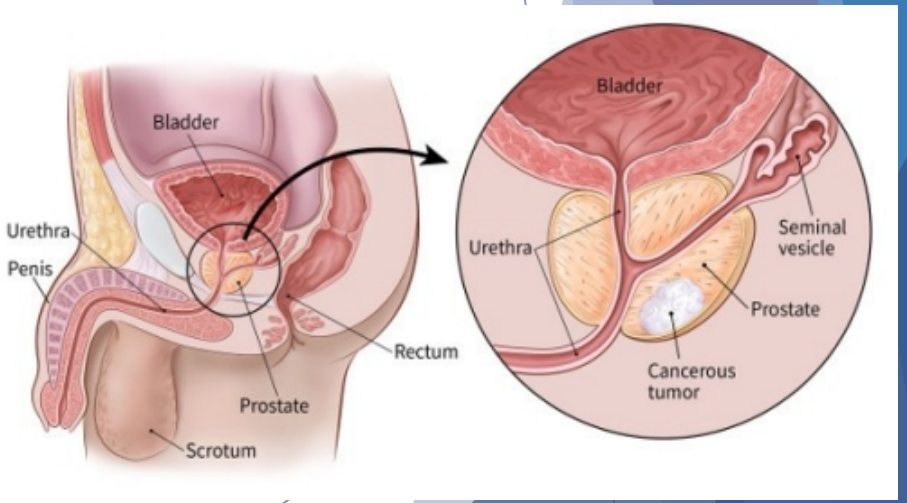
- Organ in close proximity to confounding & confusing anatomy (bladder, urethra, rectum, & seminal vesicle)

- Multiple variable & indistinct zones challenging
  - Peripheral Zone
  - Transition Zone
  - Central Zone
  - Anterior Fibromuscular Stroma

- Diseases of the prostate
  - Benign Growth (Benign Prostatic Hyperplasia)

  - Prostate Cancer:
    - Most common non-skin cancer in men
    - One of leading causes of death
    - Over-treated, so needs stratification by MRI or Fusion guided biopsy
      - Huge opportunity for AI impact



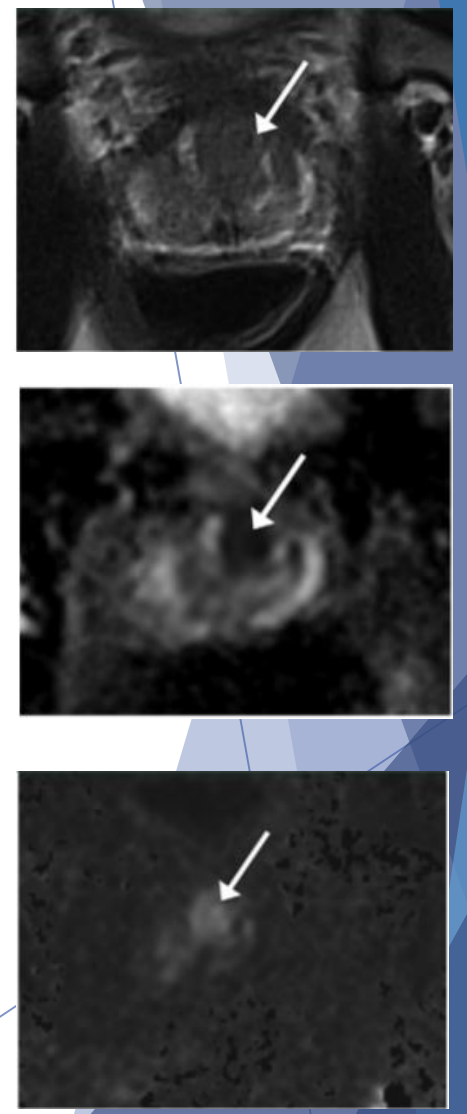


https://www.cancer.org/cancer/prostate-cancer/about/what-is-prostate-cance
https://radiologykey.com/anatomy-of-the-prostate/

Our Task: Prostate Cancer Detection on MRI

Overview of prostate cancer:

- Prostate cancer is the most common non-cutaneous cancer in men with and estimated 175,000 men diagnosed in the US each year
- Prostate cancer is best seen on MRI with multiple phases including D2, diffusion, and dynamic contrast enhanced (DCE).
- Visualization of prostate cancer has improved diagnosis.
- However, there is significant inter-observer variation
  - PIRADS 5 lesion (considered high risk for having cancer) has cancer detection rates of 40-80%.



Sonn et al. European Urology. 2017. S2405-4569(17)30266-3

Steps in Development:

1. Automated Segmentation of the Prostate
2. Classification of Prostate Lesions into Risk Categories
3. Detection of Lesions
4. Generation of an automated report

Creating Labeled Data

Required:

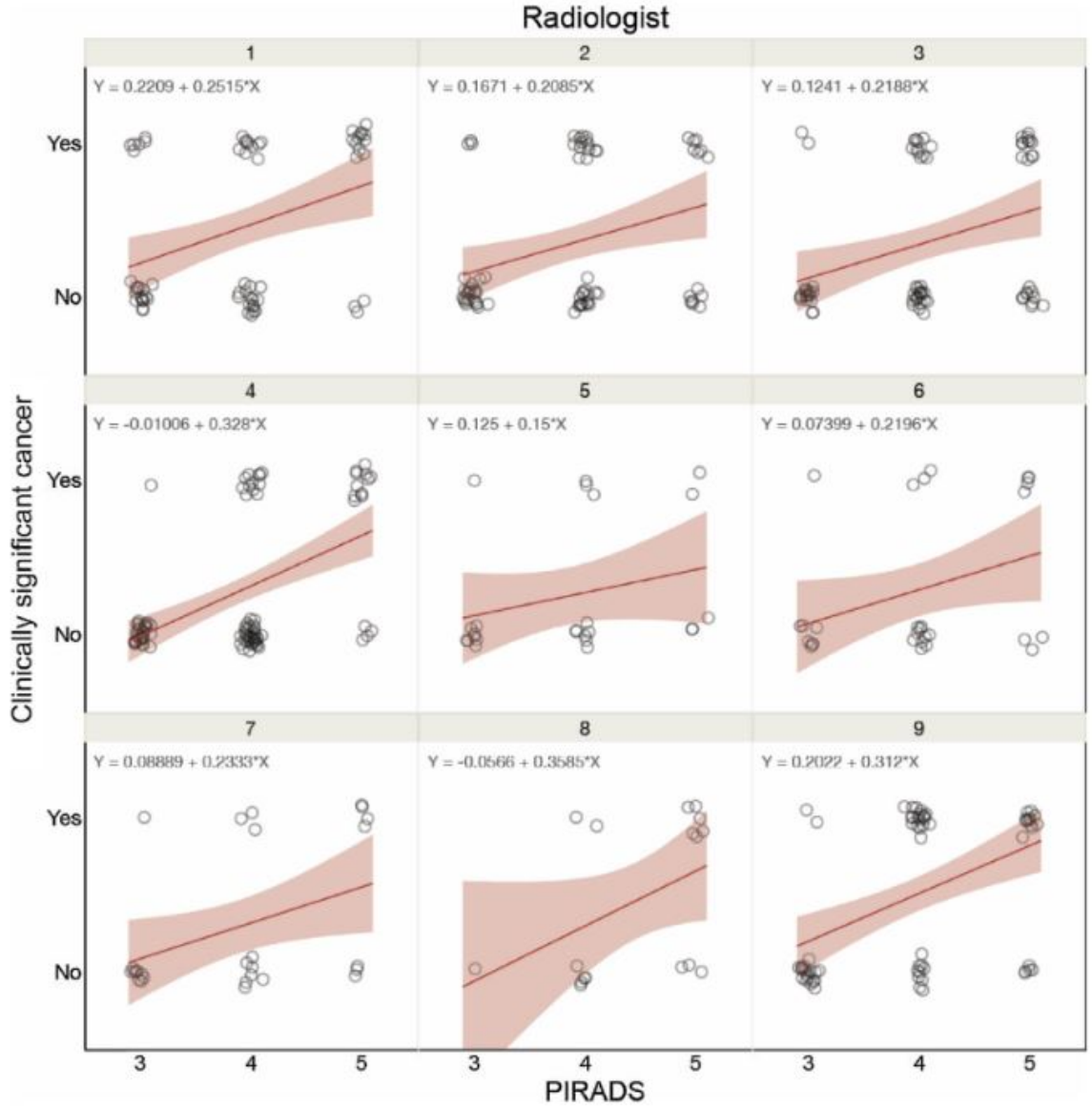
- DICOM (or nifti) files
  - T2
  - ADC
  - bighB value

- Segmentation
  - Whole Prostate Segmentation
  - Transition zone segmentation
  - Tumor Segmentation

Who performs the labeling?

- Options:
  - Expert Radiologists
  - Non-expert radiologists
  - Non-expert non-radiologists physician
  - Non-expert non-radiologist student/fellow
  - Crowd source?

High Variability Among Radiologists



PIRADS 5 associated with significant cancer detection rates of 40-80%, depending on the radiologist.

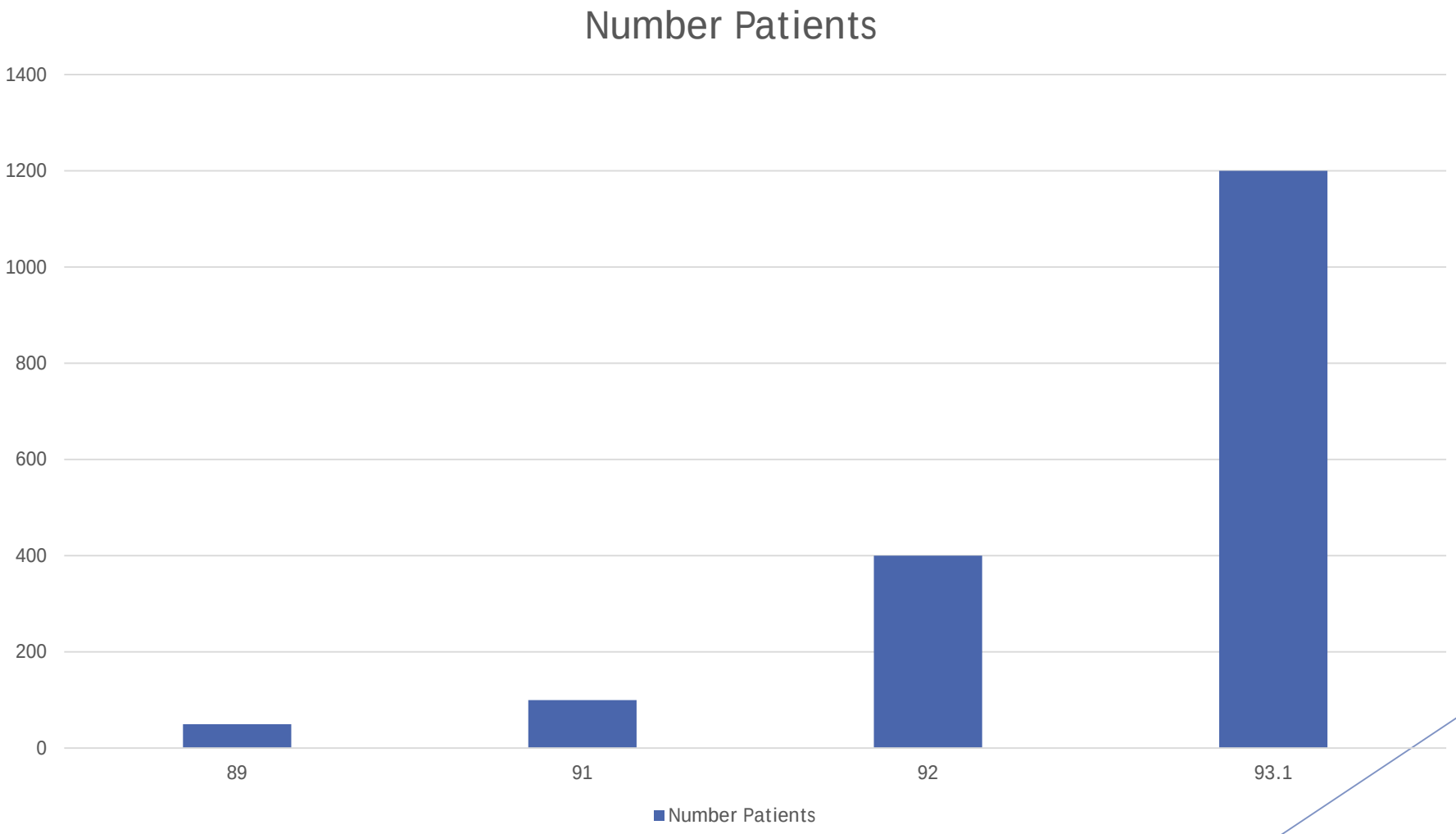
Rates of cancer detection higher in studies from institutions with expert readers

Sonn et al. European Urology. 2017. S2405-4569(17)30266-3

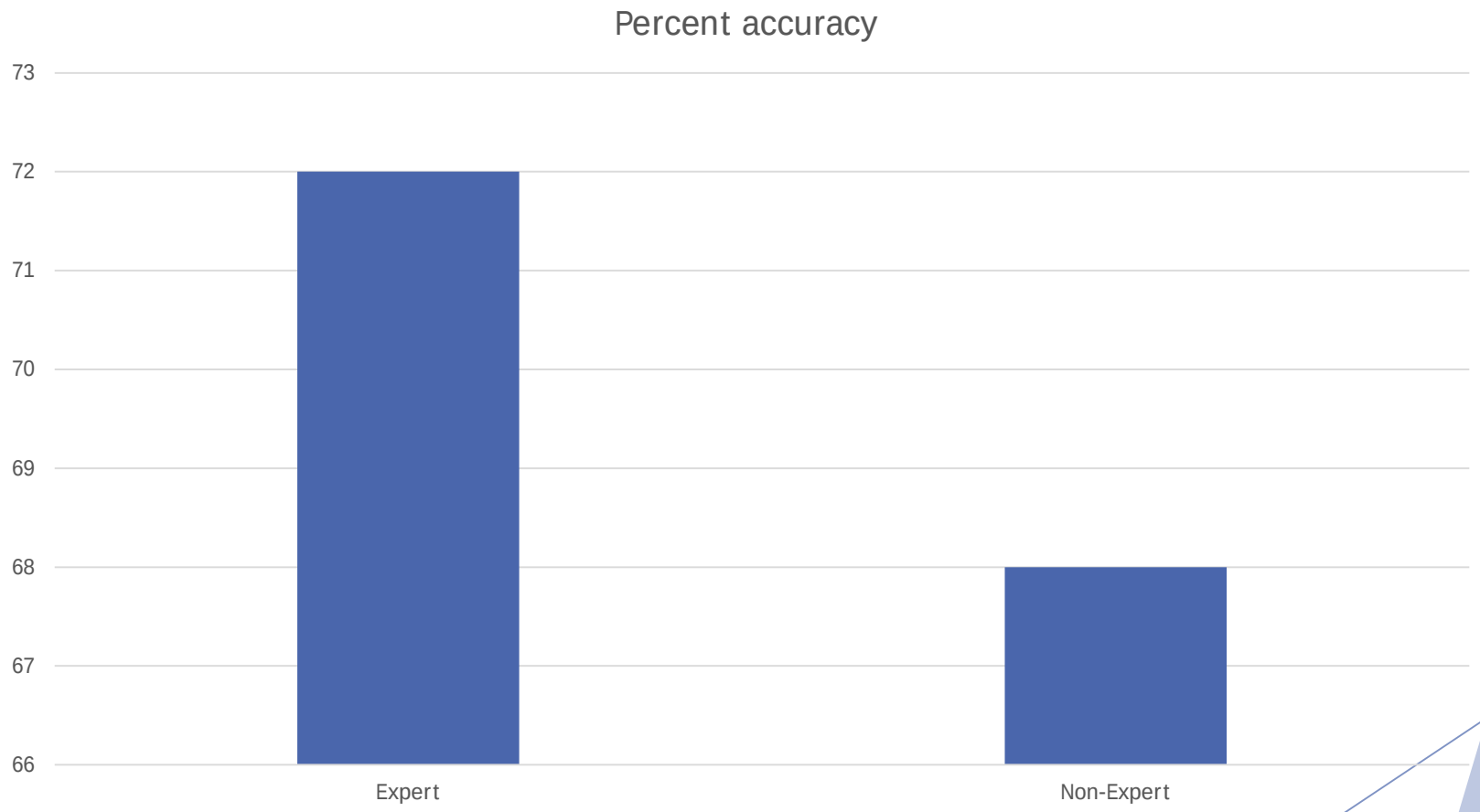
Our Approach: Replicate the expert

- >300 publications in prostate imaging
- Clinically read >10,000 Prostate MRIs over 10 years
- Member of multiple committees for prostate imaging guidelines
- Known, published, rates of detection for each PIRADS classification

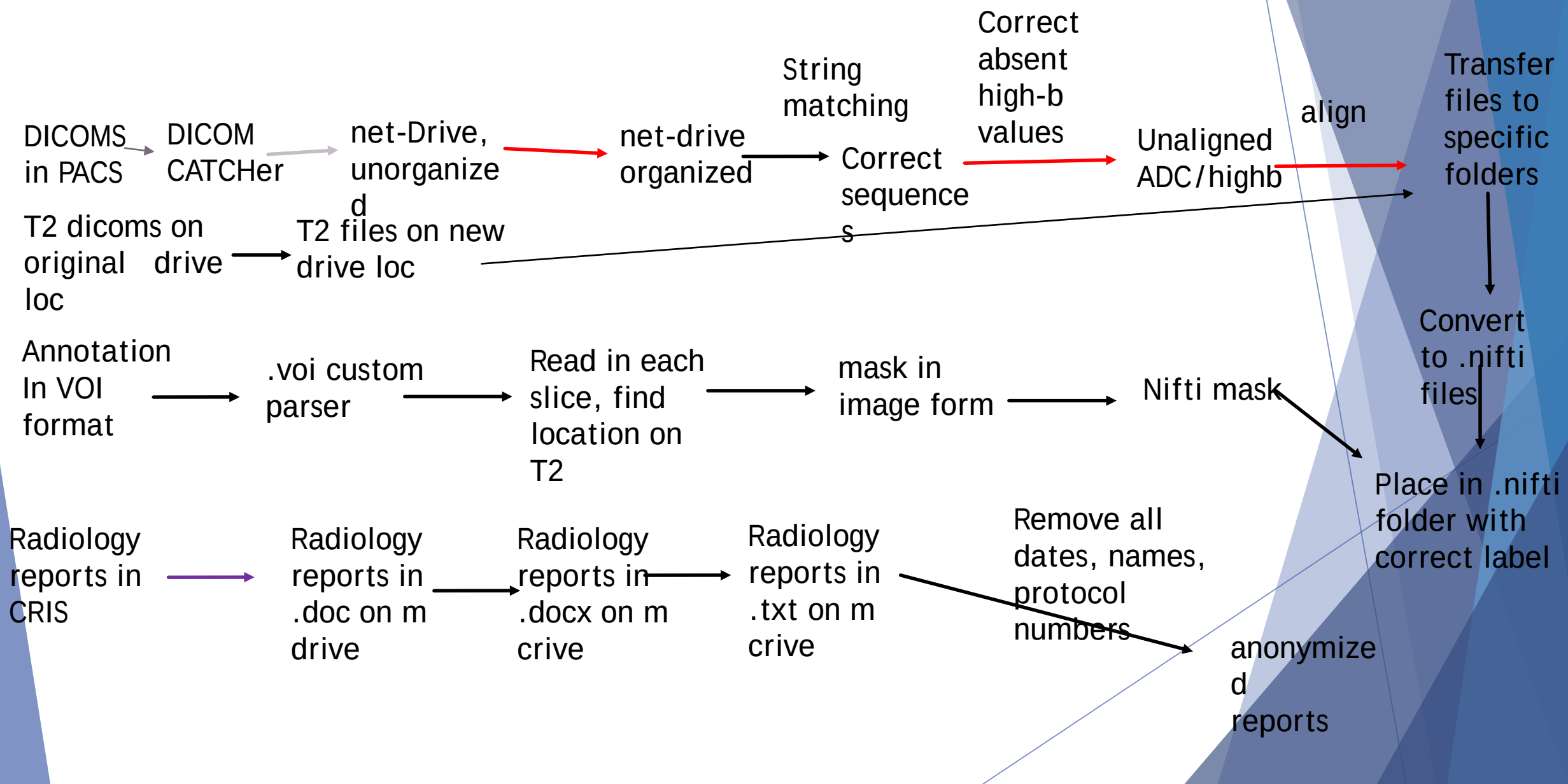
Results: Whole prostate Segmentation



Effect of the Expert

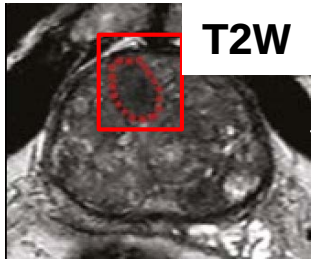


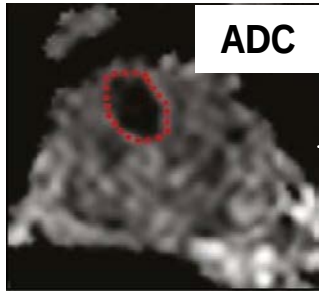
Creating a database for AI: an overview

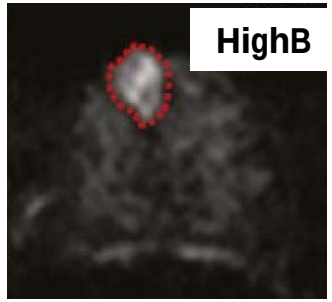


Data Overview

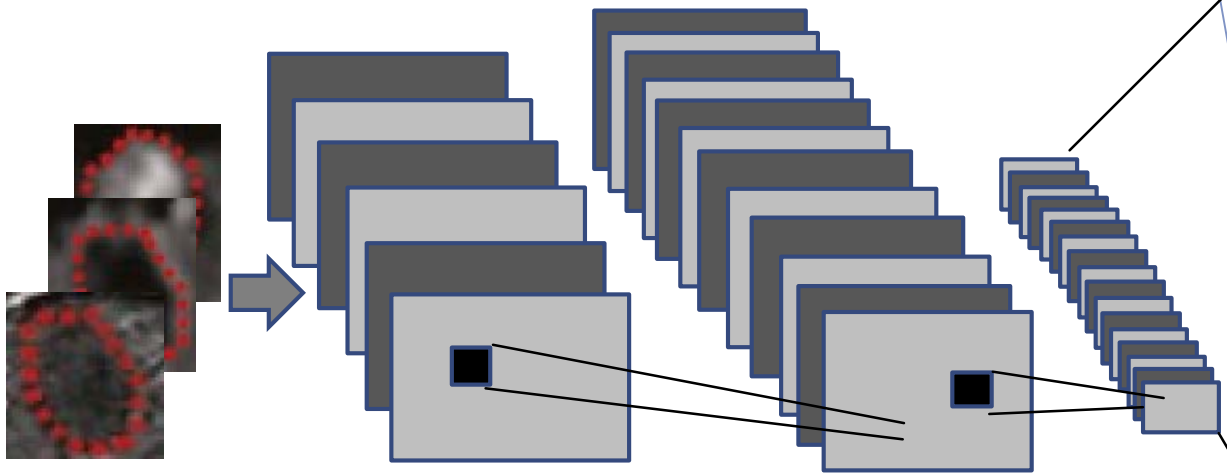
`	ProstateEx	Multi-Center	Consecutive	Surgery	Totals
Status	Complete	Complete	Complete	Complete	
Confidentiality level	Public	Anonymous	Anonymous	Anonymous	
Total Patients	99	465	422	265	1251 (1218 unique)
Whole prostate segmentation	99	465	422	265	1251 (1218 unique)
TZ prostate segmentation	99	465	422	265	1251 (1218 unique)
Urethra segmentation	99	0	150	265	514
Tumor segmentation	280	0	1083	702	2065
Total segmentation files	774	935	2478	2075	**6262**







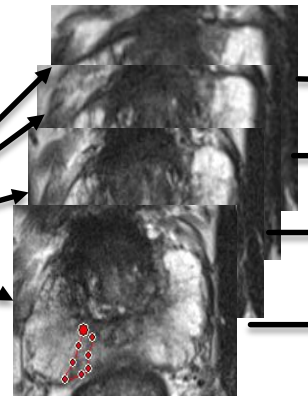
Training Convolution Neural Network



PIRADS 4/5
vs
PIRADS 2/3

Applying the Neural Network Over All Slices

Trained Model



Prediction Slice 1

Prediction Slice 2

Prediction Slice 3

Prediction Slice 4

$$\frac{1}{n}\sum_{i=0}^{n} P$$

$< 0.5 = PIRADS\ 2/3$
$\geq 0.5 = PIRADS\ 4/5$

88% accuracy

Progress

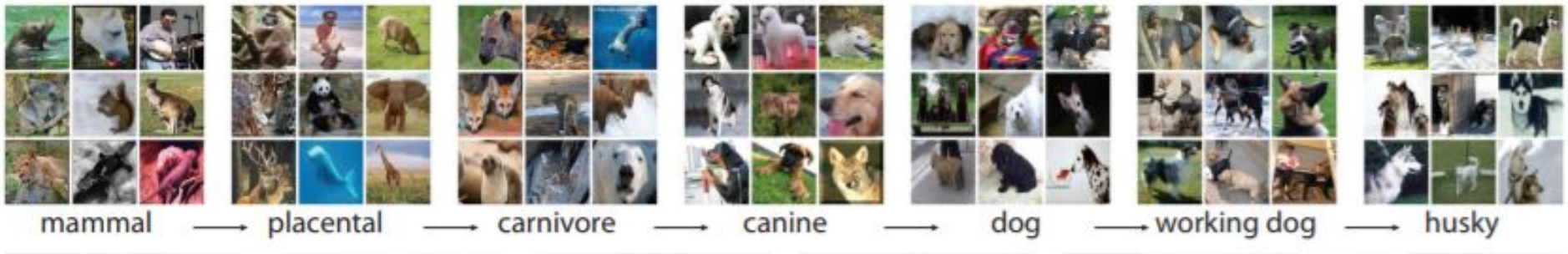
1. Automated Segmentation of the Prostate ✔
2. Classification of Prostate Lesions into Risk Categories ✔
3. Detection of Lesions
4. Generation of an automated report

Take home messages

- Dataset must be able to answer important question
- Expert level annotations
- More data IS better
- Beware of shortcuts.

Future Directions: Developing an Ontology

Imagenet



Medicine

Medical Imaging → Radiology → MRI → Multiparametric → Tumor Detection

Deng et al CVPR 2009