Machine Learning in Cancer Imaging

Dr. Jieqiong WANG
Department of Neurological Sciences
Feb. 23, 2024

University of Nebraska Medical Center
Nebraska Medicine

1

About me

Current:

Assistant professor, Dept. of Neurological Sciences

Postdoc:

University of Pennsylvania
The Hong Kong Polytechnic University
The Chinese University of Hong Kong

Research interests:

Machine learning, Medical imaging processing, Cancer, Neuroscience

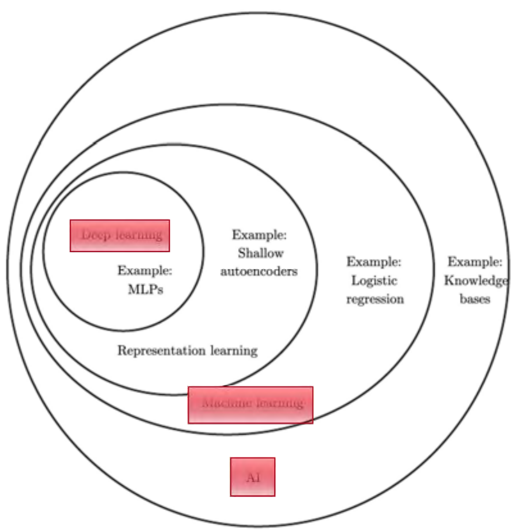
2

Disclosure

I have no financial disclosure or conflicts of interest with the presented materials in this presentation

3

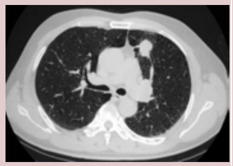
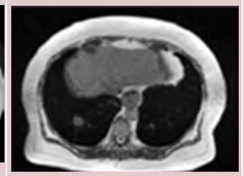
Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL)



Ref: Goodfellow et al., Deep learning, 2016

4

Radiomics (Medical Imaging)

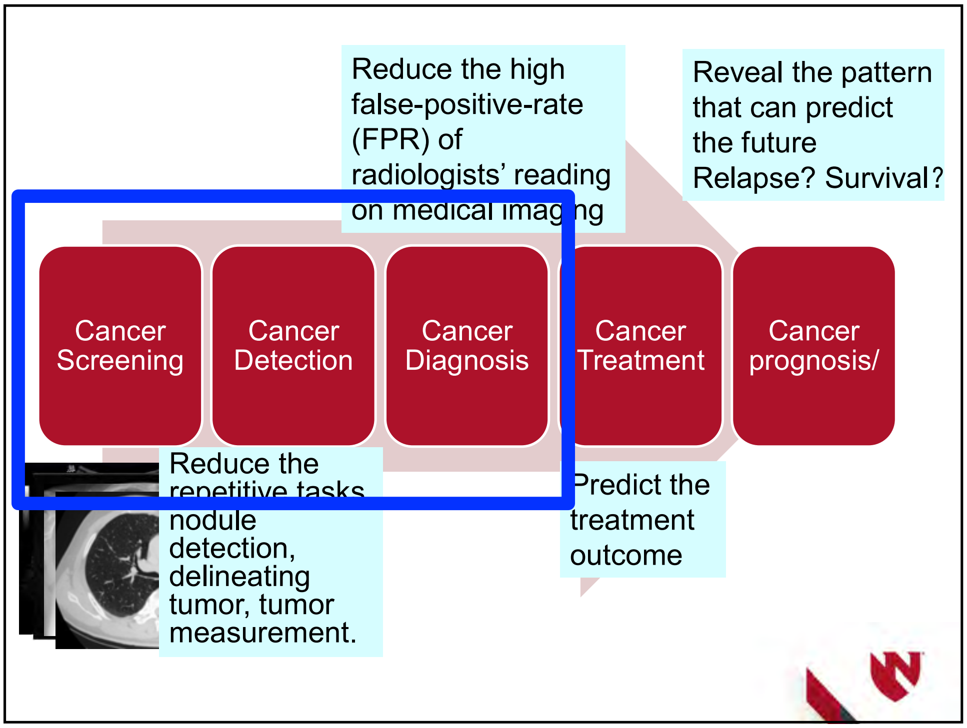
X-ray	CT	MRI	Ultrasound
			

5

ML in cancer imaging (a patient's cancer journey)

Cancer Screening → Cancer Detection → Cancer Diagnosis → Cancer Treatment → Cancer prognosis

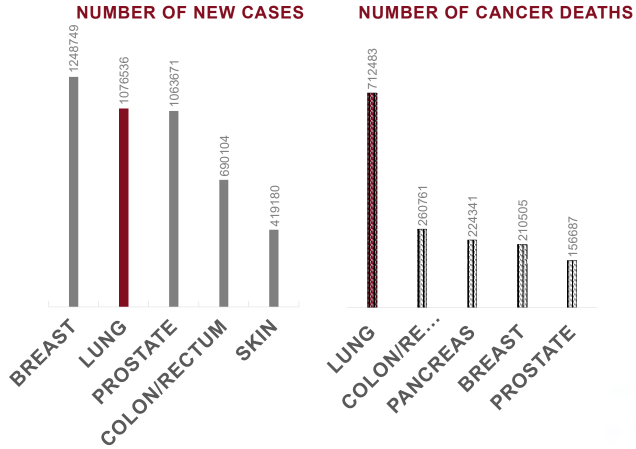
6



7

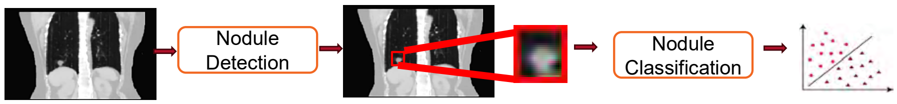
Lung Cancer

Top 5 cancers in the United States from 2016 to 2020



8

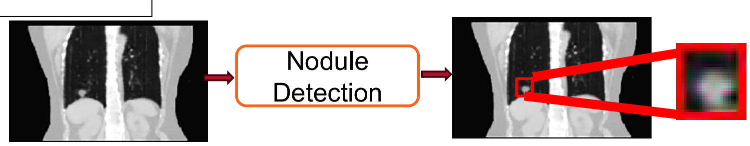
Computer-aid-diagnosis (CAD) Pipeline



Challenge:
Limited annotated dataset
Harmonization (multi-site, multi-scanner)

9

peline (Nodule Detection)



Public datasets:

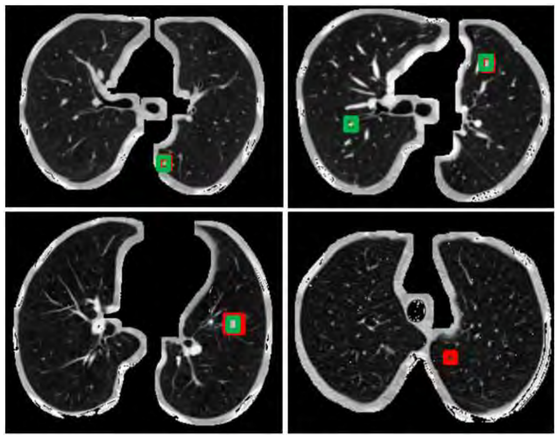
LIDC-IDRI dataset: 1,318 CT scans derived from 1,010 patients
LUNA 16 dataset: 888 cases of CT scans with smaller nodule sizes.

Annotation: Radiologists identified the nodule location

10

CAD Pipeline (Nodule Detection)

The representative nodule detection results on LUNA16 datasets using YOLO algorithm (deep learning model).



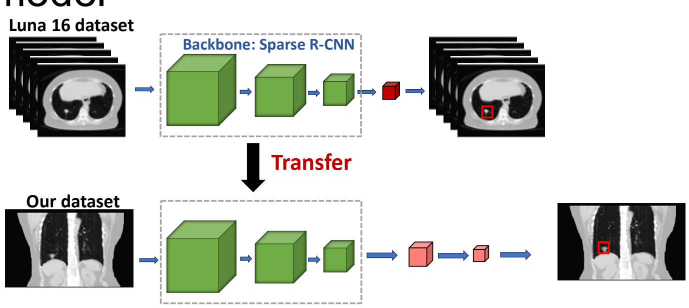
Red boxes: the ground truth.
Green boxes: YOLO results

11

Challenge: Harmonization

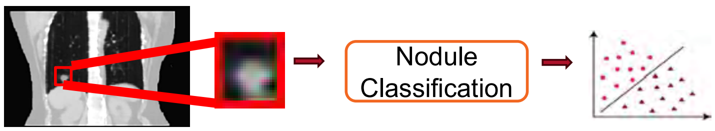
LUNA 16 vs. Our dataset

Transfer learning: annotate a few data from our dataset to fine-tune the deep learning model



12

CAD Pipeline (Nodule Diagnosis)



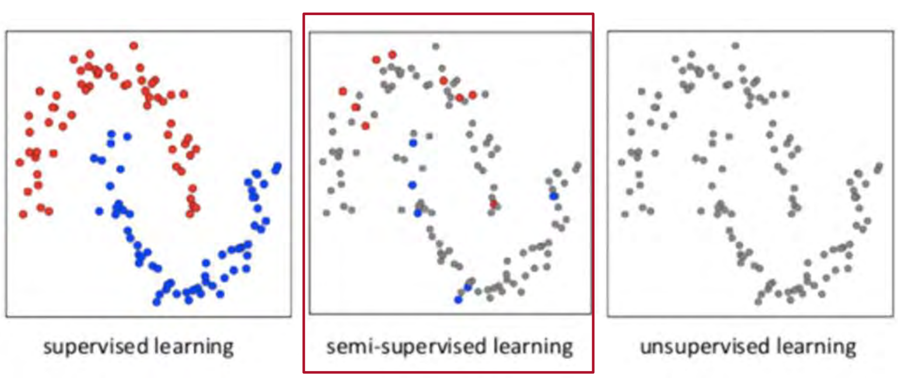
Public datasets: **LIDC-IDRI dataset & LUNA 16 dataset, etc.**

Annotation: Radiologists' reading on CT data. HIGH FPR!!!!

Use biopsy-confirmed results as the annotation

13

Machine Learning Models

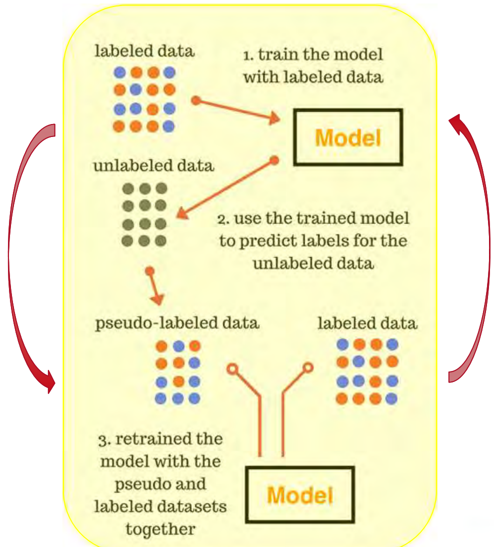


Limited labelled data: biopsy-confirmed label
Large unlabeled data: radiologists' reading on CT or no diagnosis

14

Semi-supervised learning

Pseudo-labeling



Ref: https://blog.roboflow.com/what-is-semi-supervised-learning/

15

Three Important Subareas

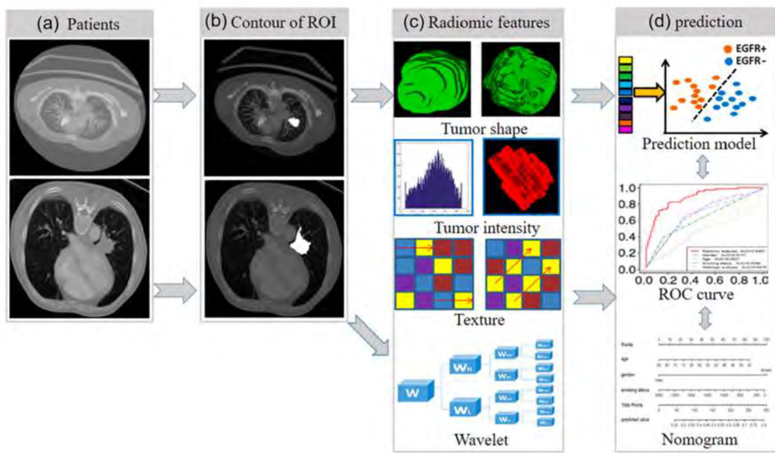
(1) Radiogenomics

(2) Imaging synthesis

(3) Automatic segmentation/detection

16

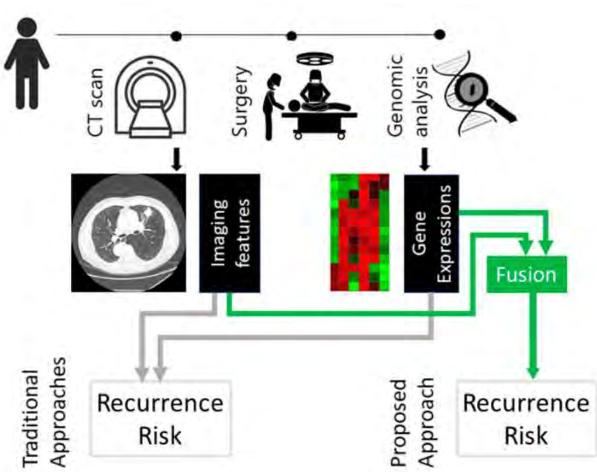
ML: Imaging → Gene mutation



Ref: Zhang et al. Quantitative Biomarkers for Prediction of Epidermal Growth Factor Receptor Mutation in Non- Small Cell Lung Cancer, 2018

17

ML: Imaging + Gene →Recurrence



Ref: Subramanian et al. Multimodal Fusion of Imaging and Genomics for Lung Cancer Recurrence Prediction, 2020

18

Imaging Synthesis

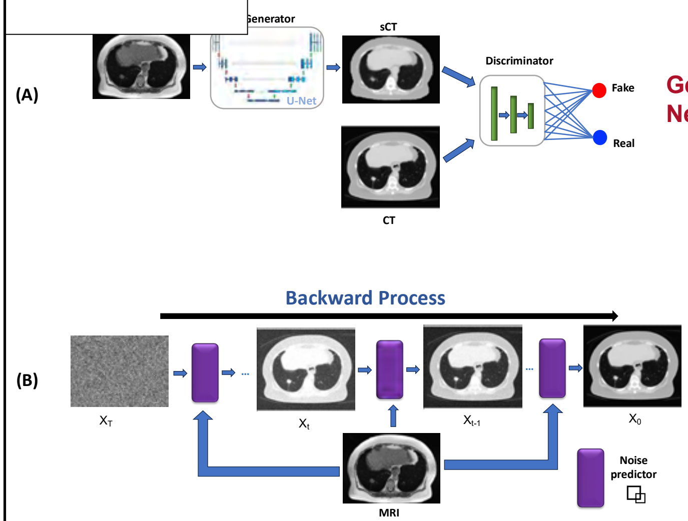
19

Why?

Create more annotated data
Save money: MRI, low-resolution
Synthesized MRI: some patients cannot take MRI
Synthesized CT: reduce harm to patients

20

e CT from MRI



Generative Adversarial Network (GAN)

Diffusion Model

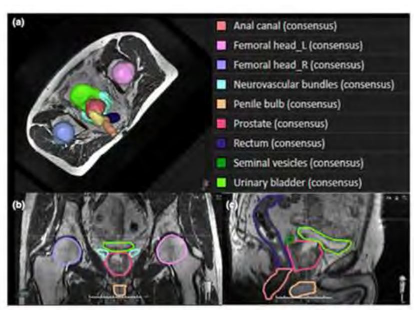
21

Imaging synthesis of pelvis dataset

Datasets: Gold Atlas project – Pelvic area
19 patients, 5 experts delineated 9 organs for each patient

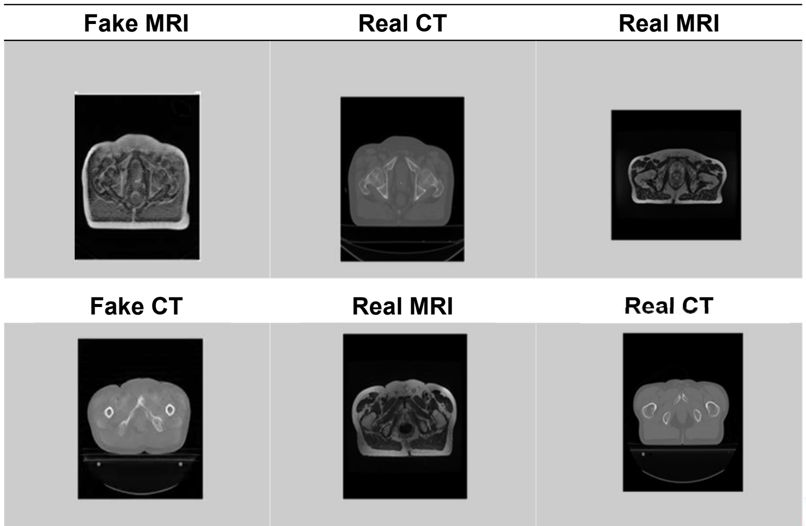
TABLE II. Included structures and their delineation definitions.

Structure	Definition
Urinary bladder	Outer contour inferiorly from its base, superiorly to the dome
Rectum	From above the anal canal (the most caudal slice with perirectal fat) to the turn into the sigmoid colon, including the rectal contents
Anal canal	From anal verge in cranial direction until one identifies mesorectal fat surrounding the anal canal (often most visible in sagittal view). Outer surface of internal anal sphincter (IAS, usual length between 2 and 5 cm). The distinction between external and internal sphincter is often easiest to identify by following the muscular layers from the rectum to the anus
Penile bulb	Hyper intense midline structure on T2-weighted MR images.[12] The most proximal portion of the penis located immediately caudal to the prostate[13] bounded by the crura, corpora spongiosum, and the levator ani muscle
Neurovascular bundles	The neurovascular area bounded by the posterior prostate wall, levator ani and rectum from the superior to the lower border of prostate
Femoral heads (R + L)	Right and left femoral heads circumference
Prostate	From prostate apex to base excluding seminal vesicles
Seminal vesicles	The seminal vesicles outer contour



22

Image synthesis of pelvis dataset



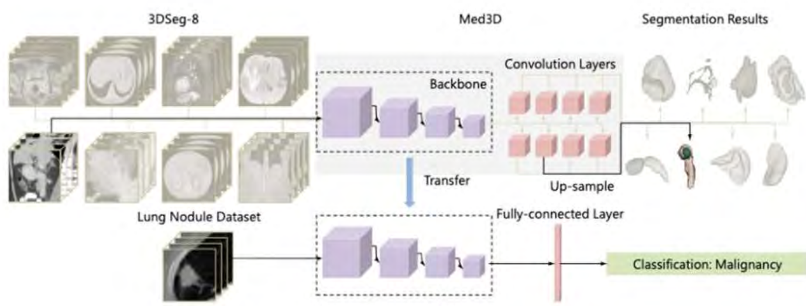
23

Automatic segmentation

24

Automatic segmentation in medical imaging

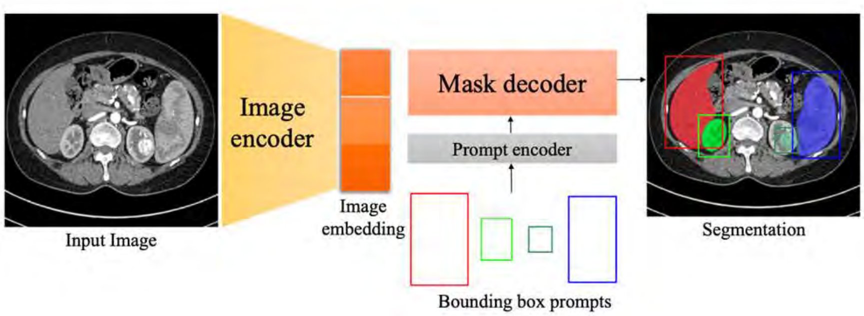
Limited dataset and limited number of data
Generalization problem: from adults to children
Transfer learning



Ref: MED3D: TRANSFER LEARNING FOR 3D MEDICAL IMAGE ANALYSIS, 2019

25

MedSAM



26

Acknowledgements

Funding:

NIH/NCI P30CA036727 Suppl Grant

NIH/NIGMS CoNDA Team Science Administrative Supplement

NIH/NIGMS CoNDA Pilot grant

BCC American Cancer Society IRG Grant

NIH/NIMH NNTC Data Coordinating Center

FIRST Award of Nebraska EPSCoR Program

NIH/NIAAA P50AA030407-5126, ACORN Pilot Core grant

Wang Lab Members

Ziyang Xu, Lie Li, Mingjun Li, Julie Mammen, Xian Hu

Collaborators

Drs.Raymond Bergan, Edward Peters, Apar Ganti, Todd Wyatt, Babu Guda, Shibiao Wan

27

THANK YOU!

28

Cancer classification/prediction

Uniqueness of patients in NE: farmers, exposure to chemistry, …

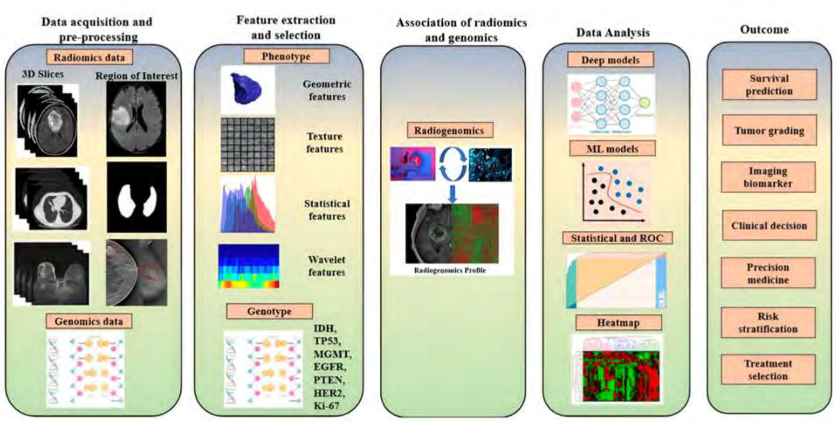
Health disparity

29

Radiogenomics (Imaging + Genomics)

Association

AI/ML



Ref: Saxena et al. Role of Artificial Intelligence in Radiogenomics for Cancers in the Era of Precision Medicine

30

Association radiomics and genomics

Heritability estimation

Polygenic risk scores (PRS) effect on radiomics features (imaging phenotypes)

GWAS to identify SNPs beyond imaging phenotypes